UTS UNIVERSITY OF TECHNOLOGY SYDNEY

Using Data Analytics and Artificial Intelligence-based techniques to promote water-conscious behaviour through personalised recommendations

by Md Shamsur Rahim

Thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

under the supervision of Professor Michael Blumenstein & Professor Damien Giurco

University of Technology Sydney
Faculty of Engineering and Information Technology

June 2021

CERTIFICATE OF ORIGINAL AUTHORSHIP

I, Md Shamsur Rahim declare that this thesis, is submitted in fulfilment of the requirements for the award of Doctor of Philosophy in Information Technology, in the School of Computer Science at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by the Australian Government Research Training Program.

Signature: Production Note: Signature removed prior to publication.

Date: 18/06/2021

Abstract

Water scarcity has become an imminent problem throughout the world due to population growth, prolonged drought, climate change, poor water management, lack of infrastructures and so on. In addition, many metropolitan water utilities are suffering from low water pressure during the hours of peak demand. Several approaches have been taken to overcome these challenges. Rolling out digital or intelligent or smart water meters was one of these approaches. Recent advances in machine learning and data analytics technologies provide excellent opportunities to utilise the vast amount of data generated from these meters.

This study aims to generate new insights from the water end-use data to provide a better understanding of water consumption behaviour and investigate the scope of personalised recommendations for the water industry that will contribute to promoting water conscious behaviours during water restrictions. To achieve this goal, the contribution of this research can be divided into five categories.

First, a systematic literature review was conducted to identify the state-of-the-art and research gaps. The findings from the literature review suggested that highly personalised recommendations through recommender systems (RSs) could play a vital role in promoting water-conscious behaviours. However, before developing an RS for the water industry, some preliminary research issues needed to be addressed. These issues were: advanced household profiling, clustering consumers based on new insights from water consumption patterns, and finally, the integration of RSs in behaviour change interventions.

Second, to achieve personalisation at the household level and identify water consumption habits/behaviour patterns, an advanced household profiling approach is proposed. The proposed approach uses the water end-use data as input and produces two outputs: the time of use and weighted probability of use (TUWPU) data set and the extracted/engineered features (EF) data set. The TUWPU data set provides the probability of an event occurring at a particular time by calculating the weighted probability distribution based on the end-use data using a proposed algorithm. On the other hand, the EF data set is generated by extracting features from the end-use data. The proposed technique is able to identify and address changes in water consumption behaviour which is essential for promoting and sustaining long-term water conservation behaviours.

Next, to understand the water consumption patterns of different groups of households, a clustering approach is presented. Extensive experiments were conducted to gain a better understanding of consumption patterns and determine the optimal number of clusters in terms of clustering quality and performance. Two different data sets which were generated from the aforementioned profiling approach were used in this experiment. For the EF data set, the

optimal number of clusters was determined using the elbow method, and several clustering techniques were applied to determine the most suitable technique. However, for the TUWPU data set, the most suitable profiling interval and linkage method was identified in terms of cluster quality metrics. Moreover, water consumption patterns are also discussed based on the findings from the approach.

Later, a generic model is proposed to integrate RSs in behaviour change intervention studies with an application in water conservation. Although the applications of RSs in behaviour change interventions are increasingly common in recent years, there is no model that can be generalised. Since RSs and behaviour change interventions are two different disciplines, a generic model can reduce the time and effort required to design such a study for students, PhD researchers, and practitioners. The proposed model addresses this research gap. Based on the proposed model, a prototype of an RS was developed, and attitudes towards the prototype of the personalised recommendations were evaluated through a survey questionnaire.

Finally, based on the findings from the research, future research directions are suggested, and recommendations are made that have implications for intelligent water conservation programs.

Acknowledgements

I would like to thank my primary supervisor Professor Michael Blumenstein and co-supervisor Professor Damien Giurco for their continuous support, patience, trust, and feedback. Thank you for giving me the opportunity to work with you and for believing in me and in my work. No words are enough to express my gratitude.

I would also like to thank the partner investigators Professor Rodney Stewart and Dr Khoi Nguyen from Griffith University for their valuable input, support, feedback, and encouragement.

Dr Tanvir Ahmed, Dr Mohammad Saqib, Dr Nabin Sharma, and Md Rafiqul Islam, thank you for your suggestions which made this journey easier.

My sincere gratitude to my teachers from American International University-Bangladesh especially Mr Mashiour Rahman, Dr Mohammad Saiedur Rahaman, Mr A.G.M. Zaman, and Mr Al Imran for helping me to complete my study during a difficult time and giving me life-changing advice. Also, I would like to express my gratitude to Dr Tabin Hasan, Dr Saddam Hossain and Dr Dip Nandi for motivating me to pursue research.

I was lucky to work and collaborate with some outstanding colleagues who always motivated me to do something better. Thank you AZM Ehtesham Chowdhury, Shovra Das, Hasib Hasan, Dr Andrew Levula, and A B M Mehedi Hasan. Also, thanks to Michele Mooney for proofreading this thesis.

I also would like to thank my friends, Irfan Alam Nabil, Amna Islam Badhan, Hridoy Ahmed, Tahmina Akhter, and Umayer Reza for helping me, inspiring me and for their constant belief in me.

Finally, and most importantly, I would like to thank my family for their sacrifice, patience, and continuous encouragement. Special thanks to my wife Ananya Rahman for supporting me in every step. I am indebted to my mother and brother as I was not able to fully shoulder my responsibilities due to my study. Thank you for your patience and for believing in me.

List of Publications

This thesis comprises a series of published and to-be-published articles together with an exegesis. The list of the publications is as follows:

I. **Rahim, M.S.,** Nguyen, K.A., Stewart, R.A., Giurco, D. & Blumenstein, M. 2020, 'Machine Learning and Data Analytic Techniques in Digital Water Metering: A Review', Water, vol. 12, no. 1.

II. **Rahim, M.S.**, Nguyen, K.A., Stewart, R.A., Giurco, D. & Blumenstein, M. 2021, 'Advanced household profiling using digital water meters', Journal of Environmental Management, vol. 288, p. 112377.

III. **Rahim, M.S.**, Nguyen, K.A., Stewart, R.A., Ahmed, T., Giurco, D. & Blumenstein, M. 2021 (in-press), 'A Clustering Solution for Analyzing Residential Water Consumption Patterns', Knowledge-Based Systems, Elsevier.

IV. **Rahim, M.S.**, Nguyen, K.A., Stewart, R.A., Giurco, D. & Blumenstein, M. ' Recommender systems in Behaviour Change Interventions: A generalized model with application in water conservation' [Draft in preparation. Will be submitted to the journal Computers in Human Behavior, Elsevier]

Other relevant publications developed but not included in this thesis include:

I. **Rahim, M.S.**, Nguyen, K.A., Stewart, R.A., Giurco, D. & Blumenstein, M. 2019, 'Predicting Household Water Consumption Events: Towards a Personalised Recommender System to Encourage Water-conscious Behaviour', pp. 1-8. [peer reviewed conference paper]

II. Nguyen, K.A., Stewart, R., Zhang, H., Jones, C., Siriwardene, N., Brown, A., Radion, A., Crook, J., Stevens, M., Smith, N., Giurco, D., Blumenstein, M. & **Rahim, M.S.** 2019, 'Developing a next generation machine learning system for enhanced urban water management: Autoflow', paper presented to the Ozwater19, Melbourne, Australia, 7-9 May 2019. [peer reviewed conference paper]

Table of Contents

Co-authors' statements

Publication I

I, Md Shamsur Rahim, contributed 75% to the paper entitled:

Rahim, M.S., Nguyen, K.A., Stewart, R.A., Giurco, D. & Blumenstein, M. 2020, 'Machine Learning and Data Analytic Techniques in Digital Water Metering: A Review', Water, vol. 12, no. 1.

Signature of Candidate: Production Note: Signature removed prior to publication. Date: 03/06/2021

I, as a co-author, endorse that this level of contribution by the candidate indicated above is appropriate.

Co-Author 1. Khoi Nguyen

Date: 15/06/2021

Signature: Production Note: Signature removed prior to publication.

Co-Author 2. Rodney Stewart

Date: 03/06/2021

Signature: Production Note: Signature removed prior to publication.

Co-Author 3. Damien Giurco

Date: 15/6/21

Signature: Production Note: Signature removed prior to publication.

Co-Author 4. Michael Blumenstein

Date: 14/06/21

Signature: Production Note: Signature removed prior to publication.

Publication II

I, Md Shamsur Rahim, contributed 75% to the paper entitled:

Rahim, M.S., Nguyen, K.A., Stewart, R.A., Giurco, D. & Blumenstein, M. 2021, 'Advanced household profiling using digital water meters', Journal of Environmental Management, vol. 288, p. 112377.

Signature of Candidate: Production Note: Signature removed prior to publication. Date: 03/06/2021

I, as a co-author, endorse that this level of contribution by the candidate indicated above is appropriate.

Co-Author 1. Khoi Nguyen

Date: 15/06/2021

Signature: Production Note: Signature removed prior to publication.

Co-Author 2. Rodney Stewart

Date: 03/06/2021

Signature: Production Note: Signature removed prior to publication.

Co-Author 3. Damien Giurco

Date: 15/6/21

Signature: Production Note: Signature removed prior to publication.

Co-Author 4. Michael Blumenstein

Date: 14/06/21

Signature: Production Note: Signature removed prior to publication.

Publication III

I, Md Shamsur Rahim, contributed 70% to the paper entitled:

Rahim, M.S., Nguyen, K.A., Stewart, R.A., Ahmed, T., Giurco, D. & Blumenstein, M. 2021 (under revision), 'A Clustering Solution for Analyzing Residential Water Consumption Patterns', Knowledge-Based Systems, Elsevier.

Signature of Candidate: Production Note: Signature removed prior to publication. Date: 03/06/2021

I, as a co-author, endorse that this level of contribution by the candidate indicated above is appropriate.

Co-Author 1. Khoi Nguyen

Date: 15/06/2021

Signature: Production Note: Signature removed prior to publication.

Co-Author 2. Rodney Stewart

Date: 03/06/2021

Signature: Production Note: Signature removed prior to publication.

Co-Author 3. Tanvir Ahmed

Signature: Production Note: Signature removed prior to publication. Date: 15/06/2021

Co-Author 4. Damien Giurco

Date: 15/6/21

Signature: Production Note: Signature removed prior to publication.

Co-Author 5. Michael Blumenstein

Date: 14/06/21

Signature: Production Note: Signature removed prior to publication.

Publication IV

I, Md Shamsur Rahim, contributed 75% to the paper entitled:

Rahim, M.S., Nguyen, K.A., Stewart, R.A., Giurco, D. & Blumenstein, M. ' Recommender systems in Behaviour Change Interventions: A generalized model with application in water conservation' [Draft in preparation. Will be submitted to the journal Computers in Human Behavior, Elsevier].

Signature of Candidate: Production Note: Signature removed prior to publication. Date: 03/06/2021

I, as a co-author, endorse that this level of contribution by the candidate indicated above is appropriate.

Co-Author 1. Khoi Nguyen

Date: 15/06/2021

Signature: Production Note: Signature removed prior to publication.

Co-Author 2. Rodney Stewart

Date: 03/06/2021

Signature: Production Note: Signature removed prior to publication.

Co-Author 3. Damien Giurco

Date: 15/6/21

Signature: Production Note: Signature removed prior to publication.

Co-Author 4. Michael Blumenstein

Date: 14/06/21

Signature: Production Note: Signature removed prior to publication.

1. Introduction

1.1 Research background

1.1.1 Water scarcity and demand management

In a report published by the World Economic Forum (*The Global Risks Report* 2019), water scarcity has been identified as one of the largest global risks. Even though water covers 70 per cent of the earth's surface, water scarcity refers to the scarcity of potable water which is only 0.014% of all water on earth. As we cannot live without water and there is no alternative to this resource like other natural resources, serious attention needs to be paid to overcome this risk; otherwise, a shortage of this resource can have dire consequences. As an example of a community that suffered from a catastrophic water shortage due to prolonged drought, South African officials coined the term Day Zero to represent the exact time in the future when municipal water would be turned off in Cape Town due to very low reservoir levels (LaVanchy, Kerwin & Adamson 2019). Fortunately, Cape Town escaped this fate by adopting dramatic water conservation, from 235 L per person per day (L/pp/dd) to 50 (L/pp/dd).

As in Cape Town, ensuring there is enough water for everyone during prolonged drought is one of the major challenges for water utilities in urban areas. Managing demand during peak hours is another challenge for the urban utilities (Nguyen et al. 2016). These challenges indicate the necessity of a smart technology-based, up-to-date water distribution infrastructure that also supports water demand management (WDM) (Nguyen et al. 2018). The aim of WDM is to promote water conservation measures and ensure an efficient water supply. WDM can be divided into five categories: (1) engineering (i.e., replacing inefficient water appliances); (2) economic (e.g., introducing block water tariffs); (3) enforcement (i.e., water restrictions); (4) encouragement (i.e., incentives for water conservation); and (5) education (i.e., encouraging consumers to use water conservation behaviour, such as limiting shower time) (Liu, Giurco & Mukheibir 2016) (Rahim et al. 2020). However, for effective WDM identification and implementation, reliable and preferably real-time information is essential (Sahin, Stewart & Porter 2015) as it helps customers to get instant access to their water consumption data and identify abnormal water usage (e.g., leaks). Digital water meters (DWM) with data analytics (DA) and machine learning (ML) techniques can play significant roles in identifying and implementing WDM strategies by providing reliable and real-time information (Gurung et al. 2016; Randall & Koech 2019; Reynaud, Pons & Pesado 2018; Stewart et al. 2018; Tanverakul & Lee 2015). For this reason, the importance of DWMs has increased in Australia as well as in other countries in recent years.

1.1.2 Digital water meters and residential water end-use

A digital or smart water meter is a standard meter that is linked to a device to provide electronic readings and display water consumption (Hauber-Davidson & Idris 2006). It enables automatic meter reading that negates the need for manual meter reading.. Modern meters can replay their information to the web or a central server using wireless modems, dial up links, LANs or ripple technology (Hauber-Davidson & Idris 2006). A significant number of studies have been conducted to utilise the data collected from DWMs as they have been identified as a potential solution to the water scarcity challenge due to their contribution of data to water conservation and peak demand management (Liu & Mukheibir 2018; Nguyen et al. 2018). Existing studies on DWM analytics are divided into two categories: (1) non-residential; and (2) residential. Non-residential studies are focused on commercial buildings (i.e., small to large businesses, educational institutions, and hospitals) and residential studies include domestic households (*Water Conservation Report* 2018). As the residential sector consumes 65-80% of water in an urban environment, this study focuses on the residential sector only.

Raw digital water meter data collected from residential households can be used to identify the water end-use categories (e.g. shower, tap, gardening etc.) using software like Autoflow (Nguyen et al. 2015). The categorisations of end-use from water flow data provide feedback through a web-enabled portal interface to the customers to promote water conservation and disaggregated water demand forecasting (Nguyen et al. 2016; Nguyen, Stewart & Zhang 2014). The end-use categorisation process performed in Autoflow can be divided into two stages. At the first stage, singe and combined water use events are identified using the hidden Markov model. At the last step, the rest of the events are classified into end-uses using the hidden Markov model, dynamic time wrapping algorithm, self-organising map, deep learning, and time-of-day probability. In this thesis, Autoflow was used to determine the water end-use categories. It is the state-of-the-art machine learning system that has demonstrated classification accuracy of up to 95%. Besides the higher accuracy rate, another advantage of Autoflow is that it offers a user-friendly, cloud-based interface that can be accessed by customers any time from anywhere. However, one disadvantage of Autoflow is the lower classification accuracy (<80%) for bathtub and irrigation end-use events.

1.1.3 Digital Water Metering: past, present, and future

To critically analyse the previous relevant studies on this topic and to identify the current state-of-the-art to determine future research directions, a systematic literature review was carried out. Based on the quality and selection criteria, 105 articles were selected and then categorised into five themes: (1) water-use feedback; (2) water event categorisation; (3) water demand forecasting; (4) behaviour analysis; and (5) socioeconomic analysis. Although paper I (Rahim et al. 2020) discusses the themes of existing literature in detail, this section presents the review of the latest research and the key findings of this study. The contents of the paper I will be further explained in chapter 4.

Under the water-use feedback category, most of the recent studies either provided paper or online-based feedback. To explore the effects of feedback on water consumption, Cespedes Restrepo & Morales-Pinzon (2020) utilised printed reports that consisted of household's consumption, comparison with their neighbours and savings tips in Pereira and Dosquebradas, Colombia. The study concluded that households with higher water consumption that received more information demonstrated higher water savings. In another study, hourly water use data were used to generated bi-monthly paper-based reports for the target groups from 16,900 single-family households by Jessoe et al. (2021). Findings from this study reported that the paper-based reports led to 4-5% immediate reduction in water consumption; however, only for temporary as interventions effects disappeared after five months from the end of the interventions. As paper-based reports do not offer a high level of user engagement, it could make it ineffective for the long-term conservation. Besides, paper-based studies have many limitations (e.g., inability to deliver real-time feedback, absence of advanced data analytics techniques in feedback generation and lower delivery frequency), as discussed in paper I, which made them the second-most effective feedback medium (Rahim et al. 2020).

On the other hand, studies under the online-based subcategory provide feedback through online portals because of the rapid progress in information technology. In recent years (2020-2021), notable studies have been conducted by Cominola et al. (2021), Cahn, Katz & Ghermandi (2020) and Daminato et al. (2021) on the topic of online-based feedback for water conservation. A gamification-based water consumption feedback delivered online based on smart water meters was performed by Cominola et al. (2021) to investigate the persistence of water conservation in the long run. Findings from this research stated that even after two years from the beginning of the program, water consumption behaviours persisted among the participated households and resulted in 8% reduction of volumetric water consumption. This finding provides empirical evidence that durable water conservation behaviours can be achieved through digital water meters-based water consumption feedback and online user engagement. In another study, the effect of daily water consumption and real-time feedback through an online portal was performed by Daminato et al. (2021) in the Canary Islands. Finding from this research suggested that access to digital water metering technology contributed to the reduction of water consumption by 2% on average.

On the contrary, a prototype study to analyse customers' preferences for online feedback applications conducted by Cahn, Katz & Ghermandi (2020) reported that participants were not interested in learning their daily water consumption but rather interested in getting alerts for a leak for abnormal usage. Though this finding contradicts the previous findings on online-based feedback studies, we believe that as it was a prototype study, the participants were yet to understand the usefulness of detailed online-based water consumption feedback. In summary, although online-based feedback studies reported the highest level of water conservation, they do not offer highly personalised recommendations which could boost the water savings further.

Water event categorisation related studies investigate the methods to classify water end-use data into use events (e.g., indoor or outdoor, shower, toilets etc.) and identify leaks. In recent years, studies were performed to classify water use data into indoor and outdoor use (Meyer et al. 2021), and determine the best classification model (Gourmelon et al. 2021). To investigate the relationship among event duration, volume, and intensity to categorise water use into indoor or outdoor, Meyer et al. (2021) developed three classification models (support vector machine, random forest, and ellipsoidal decision surface model) using end-use events data from 200,00 households in Australia and South Africa. The random forest model outperformed the rest of the models with an accuracy of 97.6%. In another study, a comprehensive comparison of machine learning techniques was performed by Gourmelon et al.(2021) to identify water end-use events from simulated data. The findings from the study concluded that supervised approaches demonstrated up to 99% accuracy in classifying water end-use events and thus outperformed unsupervised approaches. Future studies can be conducted to replicate such high accuracy on actual water consumption data. In terms of leak detection, recent state of the art residential digital meters are capable of picking up low flow rate (< 3L/hr). Therefore, to improve the accuracy of the leak detection, future studies can be conducted using the state of the art meters.

Water demand forecasting related studies aim to deploy a broad range of approaches for predicting water demand from digital water meters data. From the latest research, it can be observed that there is a move towards ensemble methods to obtain better performance. To predict monthly water demand in residential households by considering the influence of climate, a method based on the combination of decomposition technique and gradient boost regression was proposed by Carvalho & de Assis de Souza Filho (2021). The results from the research indicated that by including temperature and precipitation signals with the filtered water demand time series data, it was possible to predict demand accurately for at least four months in advance. In another study, 12 statistical techniques were investigated by Lee & Derrible (2020) to determine the suitable data, variable and modelling technique for prediction accuracy. The results from this study concluded that gradient boosting regression (GBR) performed best among all other techniques to predict daily water consumption. Though artificial neural network (ANN) based methods used in this study outperformed by GBR, these are more advanced ANN methods that can outperform GBR. Therefore, future studies can be conducted to address this limitation in daily water demand prediction.

To understand the behaviours and dynamics of consumers based on water consumption data, studies under the behaviour analysis theme can be categorised into three categories: (1) customer segmentation; (2) demand profiling; and (3) habit detection and profiling. The latest studies are found to be more focused on customer segmentation to understand the behaviours and dynamics of consumers. In one study, customer segmentation was used to explore how customers' responses to drought evolved over time by Bolorinos, Ajami & Rajagopal (2020). Findings from this research reported that informed and engaged customers are not always the most committed ones as conservation and rebound were found to be more likely among households with high income and education. In addition, the study did not find any demographic characteristics that is significantly associated with conservation. In another study, characteristics of household water consumption patterns were presented by Abu-Bakar, Williams & Hallett (2021) to improve the accuracy of demand forecasting. Using k-means clustering technique, four household clusters were identified from 10,000 households based on medium resolution digital water meter data. One common limitation in both customer segmentation studies is that they have not considered disaggregated water consumption data. The inclusion of disaggregated water consumption can reveal further interesting, valuable insights. After performing critical analysis on the recent research and based on the literature review performed in Paper I, this thesis identified several key findings.

One of the key findings from the study is identifying a positive relationship between the level of personalisation and the effects on water conservation. That is, when a water conservation program is more personalised, it contributes more to water savings. Most of the current water conservation studies and programs are limited to provide generic feedback. However, personalised messages, the customer cohort and smart metering may be the future of an improved water conservation program (Moglia, Cook & Tapsuwan 2018). Therefore, it is necessary to investigate how personalised messages and customer cohort can be achieved using digital water metering data.

Another key finding is the limitations in the existing user profiling and clustering approach. Most of the current approaches consider total water consumption data only. These approaches do not consider disaggregated water consumption data (i.e., water end-use), varying water consumption depending on the type of day (i.e., weekday, weekend), and recent changes in behaviour. Therefore, there is a research gap to address these limitations from a profiling and clustering perspective. In addition, profiling is one of the common techniques to achieve personalisation. However, to deliver highly personalised messages for water conservation, a new profiling approach needs to be considered.

In summary, after carefully analysing the literature, it has become clear that highly personalised RS may be the future of water conservation. Therefore, further investigation on this topic is required to draw a conclusion based on evidence.

1.2 Research aim

The purpose of this thesis is to advance the application of data analytics and artificial intelligence-based techniques on residential water end-use data to promote water-conscious behaviour through personalised recommendations. This thesis considers the integration of RS in behaviour change interventions to provide personalised recommendations. However, there are research gaps that need to be addressed before implementing such a system such as advanced household profiling, clustering household, and a design behaviour change intervention program using RS as the application of RS in water conservation is a new concept. This thesis brings the application of RSs in water conservation closer by addressing these research gaps. This thesis applies data analytics and artificial intelligence-based techniques on residential water end-use data to achieve personalisation.

1.3 Research questions

Research Question 1:

a. What are the applications of data analytics and machine learning techniques on DWM data in the existing studies from a water conservation and demand management perspective?
b. What are the limitations of the existing studies identified in Research Question 1 in terms of adopting and improving data analytics and machine learning techniques?
c. How can these limitations be addressed?

Research Question 2:

a. How can we achieve personalisation at the household level based on water end-use data?
b. How can we better understand the behaviours/habits of the consumers?
c. How can recent changes in behaviour be tracked and reflected in profiling?

Research Question 3:

a. How can the water consumption behaviour of different groups of households be understood?
b. What is the optimal number of clusters based on household profiling data?
c. What is the most suitable profiling interval in terms of cluster quality metrics?

Research Question 4:

a. How can recommendation systems be integrated in behaviour change intervention studies?
b. How can water-conscious behaviours using a recommender system be promoted in behaviour change interventions?

1.4 Thesis structure

This thesis is structured as a thesis by compilation. As shown in Figure 4, this thesis has two parts: the exegesis and four paper blocks. The exegesis provides an introduction, discusses the research background, explains the research questions, provides an overview of the publications and their contributions, and finally discusses the research findings with conclusions and recommendations. The second part consists of four blocks of published and to-be-published articles on which this thesis by compilation is based.

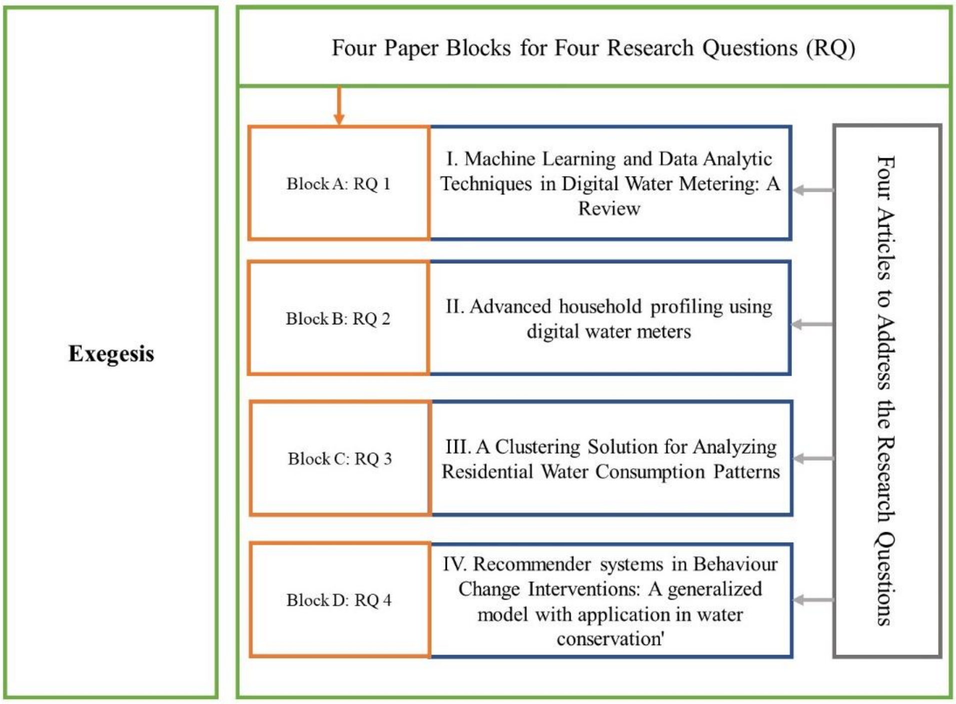


Figure 1. Thesis structure.

As shown in Figure 4, each block consists of one paper that addresses a research question. Block A addresses RQ1 through article one by critically analysing the relevant literature. Block B addresses RQ2 through article two by introducing an advanced profile approach for households based on water end-use data. Block C addresses RQ3 through article three that introduces a comprehensive clustering approach to analyse residential water consumption patterns. Finally, Block D addresses RQ4 by proposing a generic model to integrate RSs in behaviour change interventions and applies this model to develop a prototype of an RS for promoting water-conscious behaviour.

2. Overview of the publications

This section presents an overview of the publications and highlights their links to the research questions. I was the primary author of all the papers and the key contributions were delivered by me. Table 3 summarises my contributions in these four papers and briefly describes how these papers address the research questions.

Table 1. My contributions to the publications

Block	Paper	My Contribution
A	I	As the primary author, I wrote all parts of the paper. I improved the quality of the paper after addressing the feedback from my supervisors (Blumenstein, Giurco), the co-authors (Nguyen, Stewart) and the unknown reviewers. This paper solely contributes to Block A and answers the first set of research questions.
B	II	As the primary author, I contributed to all parts of the paper. I improved the quality of the paper after addressing the feedback from my supervisors (Blumenstein, Giurco), the co-authors (Nguyen, Stewart) and the unknown reviewers. This paper solely contributes to Block B and answers the second set of research questions.
C	III	As the primary author, I wrote all parts of the paper. I improved the quality of the paper after addressing the feedback from my supervisors (Blumenstein, Giurco), the co-authors (Nguyen, Stewart, Ahmed) and the unknown reviewers. This paper solely contributes to Block C and answers the third set of research questions.
D	IV	As the primary author, I wrote all parts of the paper. I improved the quality of the paper after addressing the feedback from my supervisors (Blumenstein, Giurco) and the co-authors (Nguyen, Stewart). This paper solely contributes to Block D and answers the fourth set of research questions.

Block A: Paper I

Research Question 1:

a. What are the applications of data analytics and machine learning techniques on DWM data in existing studies from a water conservation and demand management perspective?
b. What are the limitations of the existing studies identified in Question 1 in terms of adopting and improving data analytics and machine learning techniques?
c. How can these limitations be addressed?

I. Machine Learning and Data Analytic Techniques in Digital Water Metering: A Review

This paper provides answers to the first set of research questions after critically analysing the relevant studies on residential digital water metering data analytics. This paper aims to identify the research gaps in relation to the application of data analytics and artificial intelligence-based techniques on residential digital water metering data and to investigate the impact of highly personalised water consumption feedback on water conservation. To achieve this goal, a systematic literature review was conducted following the 15 steps proposed by Pickering and Byrne (Pickering & Byrne 2014). After analysing the relevant articles, the existing literature on residential digital water metering was categorised into five themes: (1) water-use feedback; (2) water event categorisation; (3) water demand forecasting; (4) behaviour analysis; and (5) socioeconomic analysis. Although not all articles directly applied data analytics or machine learning techniques on DWM data, these studies are also included in this paper as they provide useful insights regarding the usage of DWM data.

Digital water metering enables customised and detailed water-use feedback to consumers. Therefore, researchers investigated different feedback techniques regarding water use. These investigations can be divided into three categories: visual display-based feedback, paper-based feedback, and online feedback.

Visual display-based feedback studies investigate the use of a visual display to provide feedback to customers about their water consumption. Alarming visual display monitors were found effective in shower water and energy conservation by Willis et al. (2010). However, in a later study, it was found that these visual display monitors were effective in the short-term not the long-term, as shower water consumption reverted to the previous level.

The studies which fell into the paper-based sub-category investigate the application of detailed water consumption feedback using paper-based reports. Several studies were conducted by Liu et al. (Liu, Giurco & Mukheibir 2015; Liu, Giurco & Mukheibir 2016; Liu et al. 2016) to investigate the impact of paper-based reports on water conservation. The findings from their studies suggest that paper-based reports contribute to a 7.6% reduction in water consumption. The report used in these studies included water consumption by category, a comparison with similar households, custom insight and water saving tips. In another study, paper-based feedback was used to determine the effectiveness of different communication strategies on a water conservation program. Although paper-based reports have a positive impact on water conservation, they lack real-time content delivery and higher-level personalisation.

Online-based studies provide online feedback to the customer about their water consumption. Online portals have a higher influence than paper-based reports (50% and 30%, respectively) in terms of behaviour change and water-use infrastructure (30% and 10%, respectively) (Liu, Giurco & Mukheibir 2017). Therefore, future studies should investigate this type of feedback. However, the lack of user engagement is a main challenge for an online-based feedback system (Liu, Giurco & Mukheibir 2017; Schultz et al. 2014).

Water event categorisation studies aim to identify the events associated with consumption for better water management. These studies can be divided into two: (1) leak; and (2) water-end use classification. Leak detection helps in reducing the amount of water which is wasted. However, once leaks are fixed, there is no further contribution to water conservation. On the other hand, water-end use classification provides classifications for water events such as shower, irrigation etc. (Nguyen, Stewart & Zhang 2014). It gives the consumers an understanding regarding their consumption (Nguyen et al. 2015). However, the only limitation of these studies is that they do not provide personalised feedback on water conservation.

Water demand forecasting studies predict the water demand for a particular hour, day, week, month or year. Depending on the forecasting horizon, there are three types of forecasting studies: (1) long term (e.g. next 20 years)- infrastructure and capital planning; (2) medium term (e.g. next month or year)- pricing planning; and (3) short term (e.g. hourly intervals)- demand management planning (Emmanuel A. Donkor 2014). Short-term demand forecasting has gained increasing attention in recent years among researchers. The findings from water demand forecasting studies are beneficial to both utilities and government (Nguyen et al. 2016). However, they do not have a direct impact on water conservation.

The studies under socioeconomic analysis discuss the impact of demographic factors in water consumption and conservation. These studies help to identify the factors which contribute to water consumption and conservation. The findings from these studies can lead to effective water management policy such as a targeted community awareness program or education. For instance, it has been observed that a higher price leads to lower consumption and vice versa (Luby, Polasky & Swackhamer 2018). Other socio-demographic determinants that influence water consumption include household size, income, age and education (Bich-Ngoc & Teller 2018).

Lastly, studies on behaviour analysis try to understand the behaviour dynamics of the consumers from the water consumption data. These studies employed several techniques such as profiling (Cominola et al. 2016), habit detection (Cardell-Oliver 2013) and pattern recognition (McKenna, Fusco & Eck 2014). The findings from these studies are useful for planning a water conservation program. However, the common limitations of these studies include considering only overall water consumption data, the absence of user profiling, and segmentation or pattern/habit detection based on disaggregated water end-use event data.

After critically analysing the relevant studies, this paper then highlights the research gaps and provides future research directions on the applications of data analytics and machine learning techniques on DWM data. One of the key findings of this paper is identifying the positive relationship between the level of personalisation and effects of water conservation which has implications for the potential contributions of highly personalised recommendations to promote water-conscious behaviour. Another key finding is the limitations of the existing profiling and clustering approach based on water consumption data. Once these research gaps are identified in paper I, later papers (paper II, paper III, and paper IV) propose data analytics and artificial intelligence-based approaches or techniques.

Block B: Paper II

Research Question 2:

a. How can we achieve personalisation at the household level based on water end-use data?
b. How can we better understand the behaviour/habit of the consumers?
c. How can recent changes in behaviour be tracked and reflected in profiling?

II. Advanced household profiling using digital water meters

The aim of paper II is to provide answers to the second set of research questions. As paper I finds that highly personalised recommendations can be effective in promoting water-conscious behaviour, this paper proposes a comprehensive profiling approach to achieve personalisation that offers the opportunity to accommodate customers' needs and preferences (Eke et al. 2019).

A good number of profiling techniques have been proposed based on the residential DWM data. For instance, demand profiles were proposed to forecast water demand accurately; user profiles were used for habit detection. However, after critically analysing the studies in this category, it is observed that most of the studies do not consider disaggregated water consumption data; instead, they used total consumption data. In addition, other limitations of the existing studies include not considering water consumption depending on the type of day (i.e., weekday, weekend) so they are unable to accommodate recent changes in behaviour. In addition, all the existing studies considered hourly water consumption, however, other frequency intervals (e.g., 15 minutes, 30 minutes) remained unexplored which could provide more insights. Moreover, this profiling approach is not designed to be used to make personalised recommendation or RSs to promote water conservation and demand management. Lastly, to make effective recommendations, the profiling approach needs to identify and track habits and changes in behaviours which is absent in the literature.

In this paper, I propose an advanced household profiling approach that overcomes the limitations of the existing studies. The proposed profiling approach is based on disaggregated water consumption data, performs separate profiling depending on the types of the day (i.e., weekday, weekend), considers different profiling intervals (i.e., 15 minutes, 30 minutes, and 60 minutes), and can accommodate recent changes in behaviours.

The workflow of the study consists of three distinct phases: (1) data collection and preparation; (2) profiling and feature extraction; and (3) analysis and evaluation. In the first phase, raw digital water metering data were collection from 306 single stand-alone households in Melbourne, Australia for over 10 months. Later, the raw data were analysed by Autoflow (Nguyen, Stewart & Zhang 2014) software that provided nine types of water end-use with over 90% accuracy. As part of the data preparation, only events that occurred between 06h00 and 23h59 were considered as the number of events and volume consumed from 00h00 to 05h59 were negligible in comparison. Lastly, type of day (i.e., weekend or weekday) was extracted from the given data to perform profiling separately.

In the second phase, I introduce a profiling algorithm to create the profiles where the algorithm consists of three steps. The final output from the algorithm is a matrix where each row represents the probability distribution of one event and each column represents a time. In addition, to enhance the profiles, I extract several features for each event.

Finally, in the last phase, after analysing the results, I conclude that the proposed profiling approach is suitable to be used in making personalised recommendations. It provides a better understanding of the behaviour of the consumers at 15-minute intervals compared to 30- and 60-minute intervals. In addition, from the probability distribution of each event, it is possible to identify and track changes in behaviour and habits for any water consumption event. The profiling approach also has implications for demand profiling as it offers a better understanding of the customers' demand. The findings from this profiling approach have implications for customers, utilities and policy makers which are discussed in paper II.

After proposing the profiling approach in paper II, I observe that there could be many households with the same consumption patterns. Therefore, instead of considering these separately, households with similar consumption patterns can be grouped together. To achieve this goal, a clustering solution can be provided for the households based on their profiles. This clustering would provide a better understanding of the groups of households which would be beneficial for utilities and policymakers to design future water conservation programs and campaigns. In addition, clustering would enable a social comparison to be performed which is an effective technique in water conservation and can be used to improve the performance of a RS to promote water-conscious behaviour. Therefore, paper III focuses on proposing a clustering solution.

Block C: Paper III

Research Question 3:

a. How to understand the water consumption behaviour of different groups of households?
b. What is the optimal number of clusters based on household profiling data?
c. What is the most suitable profiling interval in terms of cluster quality metrics?

III. 'A Clustering Solution for Analyzing Residential Water Consumption Patterns'

This paper provides answers to the third set of research questions by proposing a clustering solution to analyse residential water consumption patterns based on advanced household profiling.

Clustering is one of the many data analytics and machine learning techniques that can be used to obtain a better understanding of water consumers' behaviour. Several studies have been performed using clustering based on water-consumption data generated from DWMs. These previous studies can be divided into three categories: (i) customer segmentation, (ii) habit detection and profiling, and (iii) demand profiling. After critically analysing these studies, I observed various limitations in the existing studies, such as they do not consider disaggregated consumption data or recent changes in behaviour and that the derived features or patterns are not based on the type of day. However, to fully utilise the data collected from DWMs, these research gaps should be explored. In addition, recent studies (Rahim et al. 2019b; Rahim et al. 2020) have identified the potential of a personalised RS to promote water-conscious behaviour and proposed an advanced customer-profiling approach (Rahim et al. ; Rahim et al. 2021). Therefore, a suitable clustering study is required that can be used to implement a personalised RS for water conservation and demand management.

In this paper, I introduce a clustering solution to overcome the limitations of the existing clustering studies to provide a better understanding of household water consumption patterns which would be suitable for an RS to promote water-conscious behaviour (Rahim et al. 2019b). To accomplish this task, I considered a new and advanced profiling approach that offers in-depth details on consumer water-consumption patterns which I proposed in paper II. This profiling approach is based on disaggregated water-consumption data. It accommodates changes in behaviour with time and separately considers the consumption patterns on weekdays and weekends and includes the derived features (Rahim et al. 2021). My hypothesis was there would be similarity between many households in terms of their water consumption profile. In this scenario, identifying groups of consumers with similar consumption patterns can improve the performance of finding neighbours in collaborative filtering (Guo, Zhang & Yorke-Smith 2015). In addition, it can be useful for performing social comparisons, as social comparisons are effective in promoting water conservation (Schultz 2019). In addition, clustering the data gathered from this new profiling approach can reveal previously unseen water-consumption patterns. Therefore, I chose to investigate the application of clustering techniques for the purpose of exploring water-consumption patterns based on advanced customer profiling.

In this paper, I performed clustering on two data sets: an *engineered features* (EF) data set and a data set including *times of use and weighted probabilities of use* (TUWPU) of water by consumers. The EF data set was created after performing feature engineering on residential water-consumption data. Based on the existing literature, statistical techniques were used to create 13 new features. Later, with the help of the elbow method (Han, Kamber & Pei 2012), the optimal number of clusters was determined, and several clustering techniques (k-means, affinity propagation, mean shift, spectral, BIRCH and hierarchical agglomerative clustering) were applied to identify the most suitable in terms of running time and clustering quality metrics. The K-means clustering algorithm aims to select the centres of the clusters or centroids that minimises the inertia. This algorithm is easy to implement and scalable. However, it performs poorly when the samples are in irregular shapes as it assumes that clusters are convex and isotropic. Affinity propagation is clustering algorithm based on message passing concept between two data points. However, it is not scalable for large amount of data points in terms of run time and memory which makes it suitable for small to medium sized datasets. Mean shift is a centroid-based algorithm that identify blobs from a smooth density of samples. Like affinity propagation, this algorithm is also not highly scalable which is a major drawback to analyse vast amount of data. Spectral cluster is another technique that clusters the data by constructing a similarity graph and then projecting data onto a lower-dimensional space. Although this algorithm suitable for a small number of clusters, it is not recommended for higher number of clusters which is a major drawback. On the other hand, BIRCH is another clustering algorithm that performs hierarchical clustering on large data sets. One key advantage of BIRCH is to cluster incoming data incrementally and dynamically within given restrictions. Lastly, hierarchical agglomerative clustering performs clustering in a bottom-up fashion, where each data point starts as its own clusters and merged with nearest cluster recursively until there is only one cluster. Hierarchical agglomerative clustering is scalable for higher number of data points and clusters. For the EF data set, five clusters were found to be optimal, and the characteristics of these clusters were discussed. In addition, k-means was identified as the most suitable clustering technique after considering performance and clustering quality metrics.

For the TUWPU data set, I performed hierarchical agglomerative clustering as representing data items in a hierarchy is beneficial to data summarisation and visualisation (Han, Kamber & Pei 2012). Also, different linkage methods were employed to identify the best one, as clustering quality may vary depending on the linkage methods. After performing hierarchical agglomerative clustering, I observed that the optimal number of clusters varied based on the profiling interval, water-consumption event and type of day (i.e., weekday or weekend). However, of the three profiling intervals (i.e., 15 minutes, 30 minutes and 60 minutes), the 15-minute profiling interval outperformed the 30-minute and 60-minute profiling intervals to form clusters based on the similarity of consumer behaviour.

After performing clustering on the household profiling and identifying groups of households with similar consumption patterns, the next step is to determine how RS can be implemented in behaviour change interventions to promote water-conscious behaviour. Paper IV proposes a generic model that can be used to design any behaviour change interventions using RS. It also presents the findings of people’s attitude towards the prototype of an RS to promote water-conscious behaviour.

Block D: Paper IV

Research Question 4:

a. How can recommendation systems be integrated in behaviour change interventions studies?
b. How to promote water-conscious behaviours using recommender systems in behaviour change interventions?

IV. Recommender systems in Behaviour Change Interventions: A generalized model with application in water conservation

Paper IV provides answers to the fourth and last set of research questions by proposing a generic model to integrate RSs in behaviour change interventions and applying this to design a prototype of a RS to promote water conservation and demand management.

Behaviour change can be challenging for anyone, whether in the long or short term. To overcome this challenge, behaviour change scientists have proposed many effective methods (Peters, De Bruin & Crutzen 2015). Such methods employ theory-based change processes to influence the psychological variables that result in a behaviour (Peters 2018). Over the past decade, online and computer-tailored intervention methods have become increasingly common due to the wide adoption of the Internet. Although computer-tailored interventions are cost-effective and their effectiveness in changing behaviour has been demonstrated, they are underused due to the following limitations: a lack of engagement, which results in a high dropout rate; minimal or incorrect use of recommendations; and few logins. Recommender systems (RSs) are considered to be a promising approach to overcome the limitations of computer-tailored interventions (Cheung et al. 2019).

An RS is an intelligent system that combines software tools and technologies to recommend a list of items that are most likely of interest to the user (Burke 2007; Resnick & Varian 1997). Today, the application of recommender systems is not only limited to increasing sales; they are widely adopted in various domains like entertainment, content, e-commerce, services and social (Montaner, López & De La Rosa 2003; Ricci, Rokach & Shapira 2015). Generally, the goal of an RS is to assist individuals who lack the experience or competence to select a potential item from an overwhelming number of alternatives offered by a service provider (Ricci, Rokach & Shapira 2015). As RSs can assist an individual in selecting a potential item from many options, therefore, in recent years, many studies investigated the potential of RSs in behaviour change interventions.

Though there are many studies investigating the application of RSs in behaviour change interventions, however, there is no generic model that can be used to investigate the integration of RSs with behaviour change interventions. According to Kok (2018), the major challenges with any behaviour change studies include: "*1) the correct identification of the change objectives (and thereby the evaluation outcomes), 2) the selection and application of appropriate behaviour change methods in an intervention, and 3) adequate implementation of the intervention*". The inclusion of RSs in this context for behaviour change makes it even more challenging. After analysing the literature, I observe that no study has been performed on how RSs can be integrated successfully in behaviour changes.

To address these research gaps, I proposed a generic model that integrates RSs in behaviour change interventions. The model is developed based on the behaviour, RSs, and user engagement theories. Depending on the theories covered or the nature of the tasks, the proposed model has four sections: (1) behaviour theories, (2) intervention design using RSs, (3) user engagement theories, and (4) deployment and evaluation. Later, I designed the prototype of an RS based on the proposed model for promoting water-conscious behaviour among residential customers. To verify the acceptance and potential impact of the RS, I surveyed 300 participants. After examining the survey responses, I concluded that the attitude towards the RSs for promoting water-conscious was positive among the participants which indicates the early acceptance and potential success of the RSs.

3. Discussion of research findings

3.1 Conclusions and Recommendations

Water scarcity is one of the largest global risks affecting many countries around the globe. Many urban water utilities are facing the challenge of ensuring a sufficient water supply during prolonged drought. Avoiding low pressure during the hours of peak demand is another challenge that is also troubling metropolitan water utilities. To overcome these challenges, digital water meters are considered a potential solution. However, the scope to fully utilise the data collected from digital water meters using data analytics and artificial intelligence-based techniques is currently unrealised. This research investigates the applications of data analytics and artificial intelligence-based techniques to promote water-conscious behaviour through personalized recommendations. First, a systematic literature review was conducted to determine the state-of-the-art in digital water metering data analytics. Second, to achieve personalisation at the household level, an advanced household profiling approach was proposed. Third, to understand the water consumption patterns of different groups of households, a clustering approach was proposed. Lastly, a generic model was proposed to integrate recommender systems in behaviour change interventions; a prototype of a recommender system was developed based on the proposed model; and people's attitudes towards the personalised recommendations were identified through a survey questionnaire. The main contributions of this thesis are as follows:

- The systematic literature review on the applications and limitations of machine learning and data analytics techniques for digital water meters identified significant research gaps in the adoption of machine learning and data analytics approaches and concluded that highly personalised recommender systems could improve the current water conservation situation.
- Effective personalised recommendations are challenged by the available resolution of water consumption data. To overcome the limitations of performing user or household profiling based on aggregated water consumption data, I proposed a profiling algorithm and performed features engineering to create user profiles based on disaggregated water consumption data and conclude that this is an effective approach for comprehensive user profiling. The significance is that it can provide better understanding of water consumption patterns for each water end-use (i.e., shower, bathtub, washing machine etc.) that can be used to provide personalised feedback and recommendations.
- Whilst previous studies in user profiling were not able to identify and track changes in water consumption behaviour and habits, I proposed an advanced

profiling algorithm by computing the probability distribution (PD) of each water end-use over three different periods and merging these three PDs after assigning different weights. The significance is that the proposed profiling algorithm can identify and track changes in behaviour and habits and accommodate recent changes (i.e., latest 18 days) in profiling. Furthermore, it provides a better understanding of household water consumption patterns which is beneficial for customers, water utilities and policy makers.

- Existing clustering studies using water consumption data have many limitations (e.g., not considering disaggregated consumption data or recent changes in behaviour, using derived features, patterns are not based on the type of day etc.). In this thesis, I proposed a detailed clustering solution to analyse water consumption to overcome these limitations via clustering techniques to reveal previously unseen, useful water consumption patterns.
- To generate the best insights from clustering, it is crucial to determine the optimal number of clusters and the most suitable clustering technique. For the EF data set, to determine the optimal number of clusters, I used the elbow method and then applied a silhouette plot for further confirmation. After analysing the results, I conclude that the optimal number of clusters for the EF data set is five. In addition, different clustering techniques (e.g., k-means, affinity propagation, mean shift, spectral, BIRCH and hierarchical agglomerative clustering) are explored to determine the most suitable clustering technique. Based on the performance and clustering quality metrics, I conclude that k-means performed better in clustering quality, and it is also scalable.
- The TUWPU data set provides a weighted probability of an event happening at a particular time at three different intervals (i.e., 15 minutes, 30 minutes and 60 minutes) for five water end-use categories (shower, bathtub, washing machine, dishwasher, gardening) for different day types (i.e., weekday, weekend). The application of clustering techniques is still unexplored on this data set which could reveal useful insights. In my thesis, I applied hierarchical agglomerative clustering on the TUWPU data to address this research gap as data items in a hierarchy are beneficial to data summarisation and visualisation. After performing extensive experiments, I identified the optimal number of clusters and the most suitable profiling interval and linkage method in terms of cluster quality metrics.

- Identifying the traits for each cluster is beneficial for water utility companies to better understand their customers and devise effective water-conservation awareness programs. In my thesis, I observed previously unseen, new, insightful traits about the groups of households that can be used to design future water conservation and demand management programs. For instance, for the EF data set, Cluster 1 can be differentiated from other clusters by its lower mean of shower volume and lower mean of maximum shower flow. In contrast, Cluster 5 has the highest value of mean shower volume and maximum shower flow. Another relevant finding is that the mean of daily flushes, the volume of tap water used, and the number of weekly clothes-washer loads for Cluster 4 is the highest compared to other clusters. Even though socio-demographic data are not available for this study, based on previous studies, it can be assumed that households with elderly people and infants might fall in this cluster.

- The population's adoption of hand-hygiene practices is crucial to prevent infection from, for example, the COVID-19 virus. I observed that clustering insights from tap-water usage can be used to determine how hand-hygiene practices have been adopted by the population. For instance, the water consumption patterns of tap-water usage from before and during the pandemic can be compared to determine the acceptance of hand hygiene among different groups of consumers with the inclusion of socio-demographic data and can be useful to policymakers and health experts in deciding their next actions.
- Behaviour change methods employ theory-based change processes to influence the psychological variables that result in a behaviour. Due to the wide adoption of technology and the Internet, online and computer-tailored behaviour change interventions become increasingly common. Though computer-tailored interventions are cost-effective, and they demonstrated effectiveness in changing behaviour, they are underused due to a lack of engagement, minimal or incorrect recommendations and few logins. RSs are considered a promising approach to overcome these limitations. However, there is no generic approach that can be followed to design a behaviour change intervention using RSs. To address this research gap, I proposed a generic model based on behaviour, RSs, and user engagement theories. The significance is that the proposed model can be used to design any behaviour change interventions using RS that would improve the effectiveness of the interventions.
- To address the absence of highly personalised RS to promote water-conscious behaviour, I developed a prototype of an RS for water based on the proposed model. The overall design of the RS is based on behaviour change methods and UX. First, I identified that the RS has six elements: (1) recommended items, (2)

profiling, (3) explanation behind the recommendation, (4) incentive, (5) items preferred by neighbours, and (6) user engagement; moreover, they are mapped with different behaviour change methods. For each recommended item, the source of the data and an explanation are proposed for consciousness raising. To connect behaviour with a consequence and to prompt, repeat, and rehearse as part of reinforcement and guided practice methods consecutively, an incentive in the form of reward points is introduced. This prototype has significance for delivering personalised recommendations to promote water-conscious behaviour.

- As attitude is a crucial determinant in behaviour change interventions, the success of the proposed prototype of the RS for water conservation strongly depends on the people's attitude towards it. To understand the people's attitude towards the RS, I conducted an experiment using a survey questionnaire that comprises ten questions. To understand their attitude, questions were asked to determine as to what degree the respondents agreed or disagreed with the following topics of personalised recommendations: beneficial for water conservation, helpful for saving more water, require less effort to find the scope of water conservation, helpful for identifying further scope for water savings, build a positive attitude towards water conservation, and useful or motivational with incentive points. After analysing 300 responses, I concluded that the overall sentiment of the participants towards the prototype RS was highly positive. This finding has significance for the potential success of the RS when deployed in the real world.

The contributions of this thesis pave the way for delivering personalised messages to promote water-conscious behaviour. Based on the findings from the research and limitations, I make the following recommendations:

- The proposed profiling technique is based on water end-use data only; it does not include any socioeconomic, temporal data. Therefore, further research can be conducted to improve the profiling approach further by including socioeconomic, temporal data. In addition, further investigations can be conducted to explore the applicability of this approach in other areas (e.g., hot water usage profiling, water-energy-gas usage profiling etc.).
- The clustering solution introduced in this thesis is designed solely based on DWM data. However, the inclusion of socio-demographic data can offer more relevant and valuable insights. To this end, further investigation of clustering should be conducted on DWM data with socio-demographic data. In addition, future studies

should also focus on feature engineering, as the inclusion of new data should provide more opportunities to create new features.

- Using the clustering solution introduced in this thesis reveals the consumption patterns of different groups for tap water usage. Based on such patterns, future studies can be conducted to understand the characteristics of different groups of consumers in terms of adopting hand hygiene practices because the understanding of adopting hand hygiene practices is crucial for decision making and deciding the next actions for policymakers and health experts.
- As the applications of RSs in behaviour change interventions are becoming increasingly common, future studies can apply the proposed generic model introduced in this thesis to integrate RSs in behaviour change interventions. The proposed model would benefit students, PhD researchers and partitioners by reducing the time and effort required to design such a study. Further investigations can be performed to improve the proposed model.
- This research introduces a prototype of an RS to promote water-conscious behaviour and validates people's attitude towards it through a questionnaire survey. However, to fully unlock the potential and measure the effectiveness of the prototype of the RS created in this research, the real-world application should be investigated and reported.
- To improve the current water conservation situation, personalised recommendations can be the most effective approach. Therefore, future research should be carried out on this topic to investigate further.

4. Publications

I. **Rahim, M.S.,** Nguyen, K.A., Stewart, R.A., Giurco, D. & Blumenstein, M. 2020, 'Machine Learning and Data Analytic Techniques in Digital Water Metering: A Review', Water, vol. 12, no. 1.

II. **Rahim, M.S.**, Nguyen, K.A., Stewart, R.A., Giurco, D. & Blumenstein, M. 2021, 'Advanced household profiling using digital water meters', Journal of Environmental Management, vol. 288, p. 112377.

III. **Rahim, M.S.**, Nguyen, K.A., Stewart, R.A., Ahmed, T., Giurco, D. & Blumenstein, M. 2021 (submitted), 'A Clustering Solution for Analyzing Residential Water Consumption Patterns', Knowledge-Based Systems, Elsevier.

IV. **Rahim, M.S.**, Nguyen, K.A., Stewart, R.A., Giurco, D. & Blumenstein, M. ' Recommender systems in Behaviour Change Interventions: A generalized model with application in water conservation' [To be published. Will be submitted to the journal Computers in Human Behavior, Elsevier]

4.1 Paper I (Block A)

Machine Learning and Data Analytic Techniques in Digital Water Metering: A Review

Md Shamsur Rahim [1,*], Khoi Anh Nguyen [2], Rodney Anthony Stewart [2], Damien Giurco [3] and Michael Blumenstein [1]

[1] Centre for Artificial Intelligence, School of Computer Science, University of Technology Sydney, Sydney, NSW, Australia; mdshamsur.rahim@student.uts.edu.au, michael.blumenstein@uts.edu.au

[2] School of Engineering and Built Environment, Griffith University, Brisbane, QLD, Australia; k.nguyen@griffith.edu.au, r.stewart@griffith.edu.au

[3] Institute for Sustainable Futures, University of Technology Sydney, Sydney, NSW, Australia; damien.giurco@uts.edu.au

* Correspondence: mdshamsur.rahim@student.uts.edu.au;

Machine Learning and Data Analytic Techniques in Digital Water Metering: A Review

Abstract: Digital or intelligent water meters are being rolled out globally as a crucial component for improving urban water management. This is because of their ability to frequently send water consumption information electronically and later utilise the information to generate insights or provide feedback to water consumers. Recent advances in machine learning (ML) and data analytic (DA) technologies have provided the opportunity to more effectively utilise the vast amount of data generated by these meters. Many studies have been performed on this topic to promote water conservation through analysing the data generated by digital meters and providing feedback to consumers and water utilities. The purpose of this review was to inform scholars and practitioners about the contributions and limitations of ML and DA techniques through critically analysing the relevant literature. We categorised studies into five main themes: (1) water demand forecasting; (2) socioeconomic analysis; (3) behaviour analysis; (4) water event categorisation; and (5) water-use feedback. The review identified significant research gaps in terms of the adoption of advanced ML and DA techniques, which could potentially lead to water savings and more efficient demand management. We concluded that further investigations are required into highly personalised feedback systems, such as recommender systems, to promote water-conscious behaviour. In addition, advanced data management solutions, effective user profiles, and the clustering of consumers based on their profiles require more attention to promote water-conscious behaviours.

Keywords: data analytics; digital metering data; machine learning; personalisation; recommender system; residential water; smart metering data; water conservation

---

1. Introduction

In a recent report published by the World Economic Forum, water scarcity was identified as one of the largest global risks because only 0.014% of all water is both fresh and easily accessible [1]. Four factors can be contribute to water scarcity: (1) uneven geographic distribution of water sources; (2) urbanization with rapid growth in population and economy; (3) poor water resource management; and (4) prolonged drought [2,3]. For these reasons, many metropolitan water utilities are facing challenges, such as ensuring the water supply during water shortages caused by prolonged drought and avoiding low water pressure during the hours of peak demand [4]. These challenges have paved the way for a smart technology-based, updated water distribution infrastructure that supports a safe, reliable, and sustainable water supply to consumers [5], including through supporting water demand management (WDM). WDM aims to ensure a more efficient water supply and promote water conservation measures. WDM has five categories: (1) engineering (i.e., upgrading to more water-efficient appliances); (2) economic (e.g., imposing block water tariffs); (3) enforcement (i.e., imposing water restrictions); (4) encouragement (i.e., rewards or rebate schemes for water conservation); and (5) education (i.e., encouraging water conservation in consumers, such as taking shorter showers) [6]. However, reliable and preferably real-time information is required for identifying and implementing effective WDM strategies [7]. Digital water meters (DWMs), accompanied by machine learning (ML) and data analytic (DA) techniques, can play significant roles by providing reliable and real-time information for identifying and implementing successful WDM strategies [8-12]. This is why the importance of DWMs has continued to escalate in Australia as well as other countries over the past few years [13]. A healthy number of investigations have been undertaken to generate useful insights through applying ML and DA techniques to data collected from DWMs. However, further investigations are required to develop new ML and DA techniques to improve the outcomes from current conservation and demand management programs.

To identify novel methods of applying ML and DA techniques, it is important to first identify what has been done thus far, what could have been done, and what the limitations of previous studies are. To achieve these goals, we performed a systematic literature review in the field. Despite previous review studies having been conducted in this field [14-23], none—to the best of our knowledge—have been conducted from the perspective of ML and DA techniques. We believe that by providing a focused investigation of the literature, this review will benefit ML and DA practitioners as well as researchers in the water industry who are working on digital water metering and closely related fields. Furthermore, we discuss the potential solutions and research directions that can be used as a baseline for future research.

After critically and carefully analysing the literature, we identified several noteworthy findings.

- First, the literature indicates that a positive relationship exists between the level of personalisation and effects on water conservation. That is, if a water conservation program is more personalised, then it contributes more to conservation. For instance, water end-use feedback programs result in more water conservation than does demand forecasting because the former is more personalised. Thus, the absence in this field of highly personalised systems such as recommender systems is notable. Recommender systems (RSs) are defined as intelligent systems that combine software tools and technologies to recommend a list of items that are most likely of interest to the user [24-26]. Generally, the goal of an RS is to assist individuals who lack the experience or competence to select a potential item from an overwhelming number of alternatives offered by a service provider [25]. Such RSs in the water sector may help consumers to choose and perform appropriate actions for promoting water-conscious behaviours. From a business-intelligence perspective, the existing studies mostly fall under descriptive or predictive analytics. Therefore, a research gap exists in terms of contributions to the highest level of analytics—prescriptive analytics.
- Second, in cases of water end-use feedback, online/web-based programs perform better than paper-based or visual display-based feedback systems. Therefore, future research related to water end-use feedback should be undertaken that considers online/web-based programs.
- Third, effective factors and measures that will be useful to direct future research (e.g., online feedback, communication strategies, water consumption data, and social comparison) have been identified in terms of promoting water conservation through water end-use feedback.
- Fourth, the literature shows that during the last 10 years, short-term residential water demand forecasting has attracted more attention than medium-term demand forecasting. This is because of the availability of high-frequency data generated by DWMs. However, further studies are required to improve the overall accuracy of predictions by reducing errors.
- Fifth, the absence of a data management solution such as a data warehouse (DW) was noted; in many studies, water consumption data were stored in a plain text file, which is unsuitable for performing analytics with a large amount of data in a real-world scenario.
- Lastly, we observed that behaviour analysis studies are mostly based on total water consumption data. Thus, further research on behaviour analysis using disaggregated water consumption data to extract novel and useful knowledge is essential for promoting effective water conservation.

The remainder of this report is structured as follows: Section II discusses the scope and method of this study, Section III presents the literature review, Section IV discusses the findings of the study, and finally Section V draws the conclusion of the paper.

2. Scope and Method

The purpose of this literature review was to identify research gaps in relation to the application of DA techniques to the DWM data of residential consumers, as well as the importance of highly personalised water consumption feedback. To accomplish this task, we followed the systematic quantitative literature review method proposed by Pickering and Byrne [27]. This method has 15 steps that allow researchers to produce a structured quantitative summary of the field by systematically analysing relevant literature [27]; Figure 1 depicts these 15 stages. We present a summary of the process as follows.

When examining the literature, it soon became clear that research on digital water metering analytics can be divided into two broad categories: nonresidential and residential. Nonresidential studies include those on commercial buildings (i.e., small to large businesses, educational institutions, and hospitals) and residential studies include those on domestic households [28]. However, this study only focused on the residential sector because it consumes the highest percentage of water in an urban environment at 65–80% [28,29]. Therefore, the literature related to nonresidential water metering was outside this study's scope. After defining the topic as digital water metering data analytics, we formulated the research question in step two, which was as follows:

- *How are machine learning and data analytic techniques applied to residential digital water metering data to promote water conservation among residential consumers and manage water demand in an urban environment?*

Based on this research question, we identified keywords using concept mapping to search databases. This literature review employed three research databases, namely Scopus, IEEE Xplorer, and ProQuest Science & Technology. The keywords used to search these three databases included but were not limited to:

- "digital water meter" and data, residential
- "smart water meter" and data, residential
- "intelligent water meter" and data, residential
- "water meter" and feedback and "water conservation"
- "water meter" and machine learning
- "water meter" and artificial intelligence
- "water meter" and data analysis or analytics

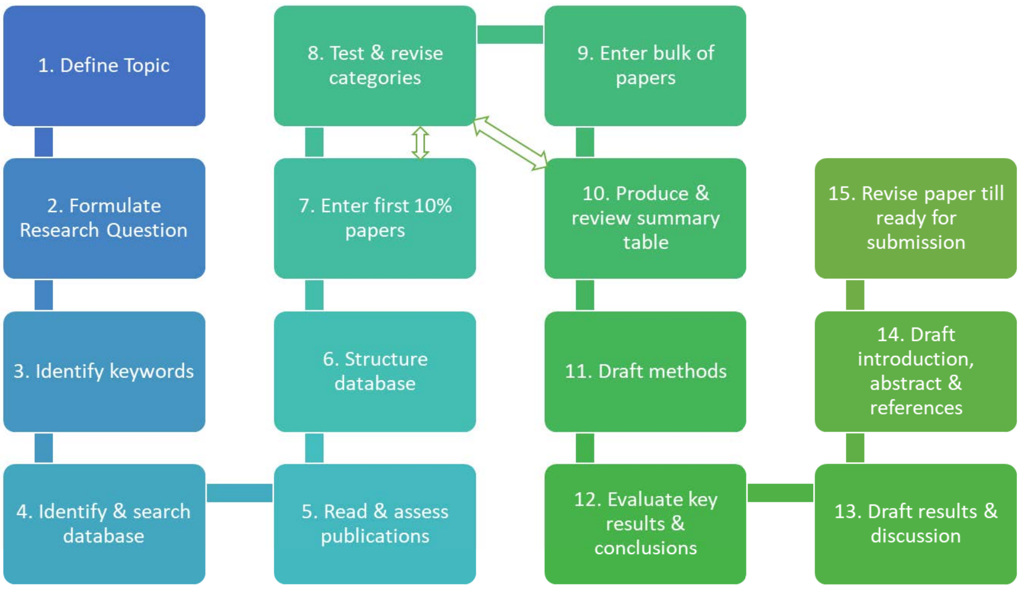


Figure 1. Steps followed to undertake a systematic quantitative literature review, as proposed by Pickering and Byrne [27].

The identified keywords were used to generate queries, which were used to search in the databases. Furthermore, because of rapid progress in artificial intelligence and ML and DA techniques in recent years, studies published before 2010 were excluded from this review. For each query, each of the databases returned several documents that included journal articles, conference papers, books, book chapters, reviews, and reports depending on the database. Table 1 represents a sample of query executions in the Scopus database along with the number of returned documents.

Table 1. Sample queries and their returned number of documents from the Scopus Database.

Query	**Documents returned**
water AND (use OR consumption) AND (classification OR categori*ation OR disaggre*)	899
water AND (use OR consumption) AND (classification OR categori*ation OR disaggre*)	1264
water AND meter* AND data AND (demand OR categorisation OR categorization OR forecast OR predict* OR leak OR usage OR consumption OR insight*) AND NOT (electri* OR soil OR irrigation OR ocean OR dust OR desalin* OR irrigation OR gas OR energy OR waste OR quality OR network OR remote)	354
(digital OR smart OR intelligent OR advance*) AND water AND meter* AND (consumption OR use OR usage) AND (data OR information OR summary OR detail*) AND NOT electr*)	178

Based on titles and abstracts, we downloaded 223 relevant papers identified from the search results. To manage the articles and references, we used the EndNote reference management software package (Clarivate Analytics, Philadelphia, United States). After going through the articles in smaller batches, it soon became clear that the existing literature could be categorised into five main themes. Therefore, this review focused on the five main themes that appear frequently in the literature: (1) water-use feedback; (2) water event categorisation; (3) water demand forecasting; (4) behaviour analysis; and (5) socioeconomic analysis. Water-use feedback provided to consumers compares their water consumption with that of others and provides general tips on water saving that are not highly personalised. Water event categorisation provides consumers with a better understanding of their consumption, as well as a leak-detection facility. Depending on the forecast horizon, forecasting models can be classified into three classes. Long-term forecasting (e.g., the next 20 years) is beneficial for infrastructure and capital planning, whereas medium-term forecasting (e.g., years or months) helps to decide water rates. Conversely, very short-term forecasting (e.g., hourly intervals) is crucial for water utilities to plan how to manage demand in advance. Studies under the behaviour analysis theme have mainly aimed to understand the behaviours—as well as the dynamics—of consumers based on water consumption data. Lastly, socioeconomic studies attempt to determine social, economic, and demographic effects on water conservation. Figure 2 illustrates the comprehensive themes from the literature on the digital water metering analytics of residential households.

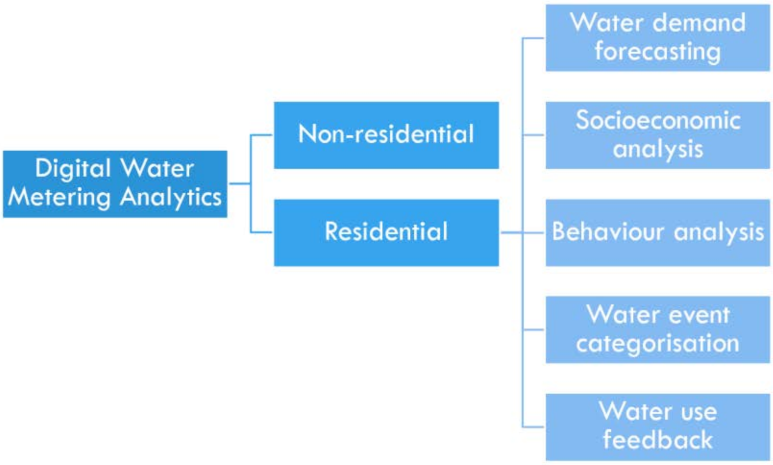


Figure 2. Comprehensive themes of literature on residential digital water metering.

After further macro-level reading (i.e., paragraphs, overall structure/arguments, and illustrations), 118 articles were removed because of a lack of relevance or doubts about quality/authenticity. To ensure quality in this study, most articles published in conference proceedings were ignored; only a small number were selected based on having a high citation rate or particular relevance. Furthermore, some articles were selected using the 'snowballing' technique [30]; thus, the total number of articles expanded to 105. After this step, we produced and reviewed summary tables, drafted methods, and evaluated key results and conclusions. Figure 3 presents the number of papers from each theme for the years 2010 to 2019. From this figure, it is clear that water demand forecasting has received more attention from the research community compared with other categories. The following section presents the findings from our critical evaluation of the studies from these five categories.

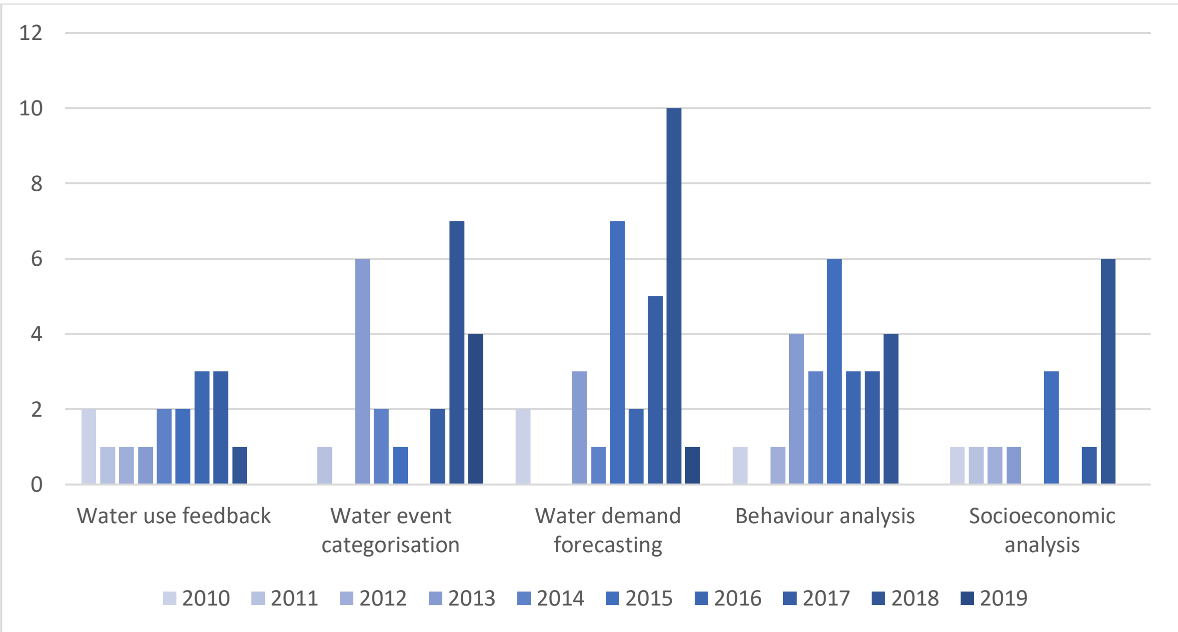


Figure 3. Year-wise distribution of articles for the types of literature.

3. ML and DA techniques in digital water metering

Depending of their nature, different studies have adopted different approaches where ML or DA techniques may not be the primary focus. For this reason, many of these studies have not employed any ML or DA techniques at all. However, these studies are highly relevant because their findings are crucial to understanding the usage of DWM data for promoting water conservation and identifying the room of improvement and challenges in applying ML and DA techniques to improve the situation. Thus, in this section, we present a critical analysis of the studies that involve any relevant analysis or apply ML or DA techniques to the digital water metering data of residential consumers.

3.1 Water-use feedback

Digital water metering provides the opportunity to provide customised and detailed water-use feedback to consumers, which can promote water conservation. Since the deployment of digital meters, researchers have investigated the influence of several feedback techniques regarding water-use. Therefore, the studies under this theme can be divided into three main categories: visual display-based feedback, paper-based feedback, and online feedback. We present a critical analysis of articles in these categories in the following subsections.

3.1.1 Visual display-based feedback

The studies under this category mainly use a visual display to provide feedback to consumers about their water consumption. In 2010, the effects of alarming visual display monitors on shower water and energy conservation were investigated by Willis et al. [31], revealing a positive effect. In 2013, another study investigated the short- and long-term effects of providing a visual display monitor feedback with an alarm feature on conserving shower water [32]. In that study, the authors used a visual display monitor that triggered an alarm if consumers used more water than the value established by the researchers. The findings showed that the integration of alarming visual display monitors reduced water consumption in the form of shower water volume by 27% in the short term. However, after 4 months, water consumption reverted back to pre-retrofit levels. The authors employed descriptive statistics and surveys to generate their results, and finally concluded that informing consumers of their resource consumption via technological devices might not be effective unless changes occurred in their habits or attitudes. Although we agree with the authors' conclusion, we argue that the inclusion of personalised engagement with consumers and feedback on other water end-use events might have improved the outcomes of their study. Furthermore, these studies did not consider other consumption events except showers.

3.1.2 Paper-based feedback

Many studies seeking to unlock the full potential of detailed paper-based water end-use feedback leveraged from digital water data have been undertaken [6,15,33-36]. Although the ultimate goals of these studies are similar in nature, they fall into different categories based on their unique approach.

To understand the effect of paper-based reports on water conservation, Liu et al. conducted several studies [6,15,34]. Their outcomes revealed that during winter, paper-based reports were associated with 7.6% more water savings than for the control group [6]. The reports included water consumption by category, a consumption comparison with similar households, custom insights, and water saving tips. However, during summer, the consumption level of the intervention group was higher than that of the control group. Because the baseline data were only recorded for a few days, this may have influenced the summer result. Furthermore, we argue that these studies may have been improved if more personalised information had been provided along with more frequent consumption feedback.

Depending on the communication strategy, the effectiveness of water conservation programs may vary. Therefore, based on this hypothesis, a study was conducted to identity the effectiveness of four communication strategies adopted from the psychological literature [35]. These strategies were social norms, social identity, personal identity, and the knowledge deficient approach. The analysis demonstrated that the social norms, social identity, and personal identity methods provided superior prospects for water saving. Here, social norms were represented through a comparison of water consumption with users' immediate neighbour, coupled with social approval (i.e., a sad face or happy face if consumption was above or below the mean). Social identity consisted of a short communication regarding water conservation behaviour in the local city area using highly inclusive language. Finally, personal identity communication was similar to social identity except it used personal identity language. The study also reported an interesting finding regarding the ineffectiveness of the knowledge deficient approach (i.e., water saving tips) among high water consumers [35]. Although the study did not employ a personalised feedback system, its findings indicated the possibility of future research adopting clustering techniques to perform comparisons of water consumption with immediate neighbours.

Table 2. Key characteristics of the literature on water-use feedback.

Reference	Medium of feedback	Location of the study	Number of participants	Feedback generation technique(s)	Water savings
[37]	Paper and online	New South Wales, Australia	Paper: 68 households Online: 120 households	Descriptive analytics	Paper: 8.0% Online: 4.2%
[38]	Online	New South Wales, Australia	120 households	Descriptive analytics	4.2%
[33]	Paper	Tokyo, Japan	246 households	Descriptive analytics	-
[6], [34]	Paper	New South Wales, Australia	68 households	Descriptive analytics	7.6%
[35]	Paper	Los Angeles, USA	374 households	Descriptive analytics	-
[39]	Paper and online	San Diego, USA	301 households	Descriptive analytics	-
[40], [31]	Visual display	Gold Coast, Qld, Australia	151 households	Descriptive analytics	27% in shower volume
[41]	Online	Dubuque, IA, USA	303 households	Descriptive analytics	6.6%
[36]	Paper	Sacramento County, CA, USA	100 households	Descriptive analytics	-

In addition, research was undertaken to identify the effectiveness of two water conservation programs, namely the data logger and water-wise house call (WWHC) programs in California's Sacramento County in the United States (US) [36]. The data logger program provided a detailed report of water consumption for each fixture (e.g., shower, irrigation, dishwasher, and clothes washer) and the WWHC provided findings and recommendations to participants via a 1-hour call from a trained water-efficiency professional. The findings suggested that both programs were effective, but the data logger was more effective (in 84% of households) than the WWHC program (62% of households). Although we acknowledge the findings from this study, we argue that the incorporation of near real-time feedback along with personalised recommendations would have improved the effectiveness of the programs.

The effectiveness of water conservation programs may differ depending on water stress levels. The literature has mainly focused on areas where water resources are under stress. However, one study was undertaken by Otaki et al. [33] to identify the most effective type of feedback in promoting water conservation in Tokyo, Japan, where water shortages are not yet a concern. After employing three types of feedback information—actual and mean consumption, water consumption rankings, and emoticons—with three groups of households, the study revealed that the effectiveness of feedback information varies for high and low water consumers. Emoticons played a significant role in reducing the water consumption of high consumers, and water consumption rankings further reduced the water consumption of low consumers. However, we suggest that the exclusion of the summer season from this study may have affected the outcomes. Moreover, reports were sent to consumers with no highly personalised feedback every 2 weeks, which meant that the study missed the opportunities to inform consumers about their consumption in real time and to enable them to take corrective measures, such as stopping a leak.

The evidence suggests that paper-based reports have a positive effect on water conservation. However, we argue that relevant studies have some common limitations. First, paper-based reports cannot be delivered in real/near real time, which results in poor communication about abnormal consumption. Second, ML techniques are absent in feedback generation because they are limited to water consumption feedback and comparisons. The inclusion of highly personalised recommendations for how to conserve water could promote water conservation more effectively. Lastly, we conclude that because of their delivery frequency, paper-based reports cannot contribute to managing peak hour water demands, which is a great challenge for water utilities.

3.1.3 Online feedback

Because of the rapid progress in information technology, several studies have examined the promotion of water conservation through online portals. The literature on this topic can be broadly classified into two categories: experimental studies and conceptual studies.

Among the experimental studies, Liu et al. [38] performed an investigation to understand the long- and short-term effects of an online portal on water conservation. The authors claimed that online feedback had a greater and long-term effect as it reduced the water consumption (by 4.2%) among an intervention group compared with a control group. They also left open the question of how to motivate and increase the interest and engagement of consumers with such systems, as user engagement was low. We acknowledge this problem but consider that integrating advanced ML techniques to incorporate highly personalised feedback and an effective incentive/rebate scheme may have improved the situation. In a similar study, Erickson et al. [41] showed that the inclusion of near real-time water consumption data, along with a social comparison, weekly games, and news and chat delivered via a web portal led to a 6.6% reduction in water consumption. They also mentioned that hourly consumption data (88%) were the most accessed feature, followed by comparisons with neighbours (66%) and game results (48%). This finding indicates that the inclusion of gamification was not particularly appealing to consumers. Gamification can be defined as the "use of video game elements in non-gaming context to improve user experience and user engagement"[42]. However, we noted some limitations in their study. First, the participants were volunteers; the results might differ in the real world because of the level of interest among participants. Second, the study did not provide any detailed water consumption feedback. We consider that the inclusion of such feedback would promote enhanced water conservation. Lastly, highly personalised recommendations for water conservation were lacking.

Although conceptual studies do not measure the effectiveness of the proposed web portals, because of their contribution in generating novel ideas, they were included in this review. Kossieris et al. [43] proposed a web platform to monitor and control water and energy consumption by providing valuable information and feedback in real time. Their study presented the novel idea of collecting several critical pieces of information through a web portal, such as the characteristics of a household, instruments, and geospatial data. Furthermore, it proposed integrating an eLearning module into the web portal, which could reduce the cost of traditional educational programs. Lanzarone et al. [44] presented the concept of another interactive web portal that used hourly consumption data to perform comparisons, as well as a multiple-choice questionnaire, to generate a customer profile based on consumption data, consumer behaviour, and building characteristics. This idea would be an effective replacement for a traditional survey as a method for obtaining demographic data.

It is clear that the effectiveness of online feedback systems suffers from a lack of user engagement. Other than that, they can play a crucial role in delivering and collecting information that would benefit both consumers and water suppliers.

Studies related to water-use feedback systems have shown that online feedback is the most effective and potentially useful medium of the three water-use feedback types for promoting water conservation. We concluded that a visual display-based feedback system is not effective in the long run and the cost of implementation might be a problem for a wide rollout. Two studies were performed to identify the most effective of the remaining two types: paper-based and online feedback. In the first study [37], the authors analysed the effect of these two feedback types in 68 and 120 households, respectively. They concluded that an online portal had a greater influence than paper-based reports (50% and 30%, respectively) in terms of behaviour change and water-using infrastructure (30% and 10%, respectively). However, because of the lack of user engagement with the online portal, the overall water saving metric was lower for online feedback (4.2%) than for paper-based reports (8.0%). The second study reported the same phenomenon [39]: only 18% of participants accessed the website from the 141 households chosen to receive web-based materials. These studies suggest that users with higher engagement in online portals are able to save more water than receivers of paper-based feedback. Thus, after careful analysis of the literature, we concluded that online feedback has greater potential for water conservation compared with paper-based feedback, which is supported by the findings from this study [16]. Along with these studies, the necessity for further investigation has been discussed to identify the most effective type of feedback in terms of information content, granularity, and frequency of delivery [45]. Table 2 summarises the key characteristics of the literature on water-use feedback.

3.2 Water demand forecasting

Water demand prediction or forecasting is a challenging yet highly desirable task for water utilities in metropolitan areas for avoiding water shortages or low water pressure during peak water usage periods [4]. Many approaches have been proposed for forecasting water demand using the data collected from DWMs. The literature covers a wide range of approaches for predicting water demand from short (hourly, daily, weekly, and monthly) to long (10–15 years) term. Based on the nature of these methods, studies can be placed into four categories: neural network (NN)-based methods, regression tree-based methods, stochastic-based methods, and hybrid methods.

3.2.1 Neural network-based methods

The dominance of NN-based methods is notable in water demand prediction because of the large volume of smart metering data. Authors have employed a range of artificial NNs (ANNs) in their proposed methods [46-50].

Several water end-use demand forecasting models were developed using feed-forward back propagation networks and radial basis function networks by Bennett, Stewart, and Beal [50]. Using household demographic, socioeconomic, and water appliance stock efficiency data, this study can provide short-term water demand forecasting and identify the key determinants. To determine the most suitable machine learning method for water demand forecasting, Mouatadid and Adamowski conducted a study [48]. They deployed NN, support vector regression (SVR), extreme learning machine (ELM), and multiple linear regression (MLR) models to forecast water demand with 1–3 days lead time. The study found that ELM models outperformed other methods in terms of accuracy. In another study, a three-layer feedforward NN was proposed with three neurons in each layer [46]. Eight predictors were used to predict the water demand 6 hours ahead. The study demonstrated that the NN outperformed the regression tree in terms of training errors. However, the proposed approach suffered from high standard deviation errors, making it unfit for further adoption. A hybrid particle swarm optimisation–ANN (PSO-ANN) was deployed in one study to predict water demand [47]. This study introduced a climatic factor-based approach where maximum temperature, rainfall, and solar radiation were identified as robust predictors. In evaluating the performance of the proposed method, the authors showed that the PSO-ANN produced a lower RMSE score than a hybrid backtracking search algorithm ANN (BSA-ANN) in terms of a fitness function. Furthermore, the study presented the correlation coefficient values for stochastic and seasonal data at the validation stage, where $R = 0.972$ and 0.982, respectively. However, the study was based on historical monthly water consumption data from 2006 to 2015 collected by Yarra Valley Water and consisted of water consumption in mega litres (1 mega litre = 1 million litres). Thus, the effectiveness of the proposed approach is yet to be tested for predicting water demand over shorter periods (i.e., hours, days, and weeks) using high-resolution metering data. In another study, an NN was deployed with summary statistics by Walker et al. [49]. Although the method could predict the general trend, it encountered difficulties in forecasting peak demands.

Table 3. Summary of the studies in water demand forecasting.

Forecast horizon	Forecast frequency	Forecasting technique			
		Stochastic	Regression	ANN	Hybrid
Short term	Every second	[51]			
	Hourly	[52]	[53], [54], [55], [46], [56]	[49], [46], [56]	[4], [57]
	Daily	[58]	[48], [59]	[50], [48]	[60], [61], [62]
	Event-based		[63]		
Medium term	Monthly				[64], [65], [47]
	Bimonthly		[66]		
	Quarterly	[52], [67]			
	Yearly	[29]			

Although the application of NNs has improved results significantly, we observed the absence of advanced NNs, such as deep learning (DL) NNs. DL models [68] have several training layers that can outperform previous base results [69]. To improve forecasting accuracy, future studies can be conducted that focus on DL.

3.2.2 Regression-based methods

After NN-based methods, regression-based methods are the next most common type of method in the literature for predicting water demand [46,48,53-55,59,63,66]

To forecast hourly water demand, a support vector regression (SVR) based model was identified as the best performing model from different machine learning methods by Herrera et al. [56]. Besides SVR, the other machine learning methods were Artificial Neural Network (ANN), Random forest, Multivariate adaptive regression splines (MARS) and Projection Pursuit Regression (PPR). Though SVR performed better compared to AAN, however, advanced neural networks such as Long short term memory (LSTM) [70] may improve the outcome. Similarly, to characterise and forecast hourly water demand, a support vector machine (SVM)-based regression model was investigated by Candelieri [53,54]. However, the proposed approach performed better in forecasting at an aggregated level (urban water demand) compared with the individual level (single customer) because of the variability in behaviour at the individual customer level. Furthermore, the studies were based on data collected from 26 automatic meter readings (AMRs) over 3 months, which was another limitation. To identify the key determinants and forecast water demand for shower and indoor water consumption, Makki et al. [59,63] performed regression analysis. They identified the key factors behind residential water consumption, and their proposed models provided better accuracy in predicting short-term water demand based on those key factors. Polebitski and Palmer [66] developed a regression-based water forecasting model at a bimonthly time-step for individual multi-houses. They found that water demand was more elastic to price and income effects during summer months than in winter months for single-family homes. In [55], the authors fitted models based on ensembles of regression trees using the

LSBoost algorithm to predict outdoor residential water consumption. Although the proposed approach had lower absolute error values, its predictability was limited to 1 hour.

Regression-based approaches are effective at identifying key factors that drive water consumption and can predict the demand based on these factors. Still, many of the previous studies have only considered the water consumption data or combined consumption data with demographic data or climatic data. Thus, future studies should consider the consumption data along with climatic and sociodemographic data to enhance forecasting models. In addition, Table 3 shows that most of the regression-based studies were on hourly or daily forecasting. Therefore, the applicability of regression-based methods for predicting monthly, quarterly, or yearly water demand can be investigated.

3.2.3 Stochastic-based methods

After regression and ANN-based techniques, the application of stochastic-based techniques is notable for forecasting water demand. In several studies [29,51,52,58,67], researchers have applied different stochastic approaches.

Abadi et al. [58] proposed a mixture of nonhomogeneous hidden Markov models (MixJNHMM) to cluster and forecast short-term water demand. The aim of this study was to cluster consumption behaviour series into several groups and forecast future behaviours for each group of consumers separately. The proposed method provided 80% accuracy for predicting water consumption behaviours.

To predict water demand at multi-spatial and temporal scales, Rathnayaka et al. applied a stochastic model [52]. It performed better at predicting mid-to-short-term water demand compared with their short-term demand forecasting model. Thus, short-term prediction models require further improvements. Blokker, Vreeburg, and Van Dijk [51] developed a stochastic end-use demand model based on the statistical information of users and end-uses. Although the developed model could forecast water demand at 1-second intervals with high accuracy at the simulation level, it lacked an application on real data.

In addition, an auto-regressive integrated moving average (ARIMA)-based approach was proposed by Yalçintaş et al. [29] for forecasting urban water supply and demand in the Istanbul Metropolitan Municipality, Istanbul, Turkey. Although the model could forecast both water supply and demand, the size of the dataset was too small compared with other studies, because the water demand data were annual and only for 8 years. On the other hand, for forecasting quarterly water demand, Sebri investigated several methods such as seasonal auto-regressive integrated moving average (SARIMA) and ANN models [67]. The study finally concluded that the NN model outperformed all the competing models in terms of forecasting accuracy.

3.2.4 Hybrid-based methods

Hybrid-based methods in this study refer to approaches where more than one technique was employed for forecasting water demand. A healthy amount of research has been undertaken to forecast water demand with the help of hybrid-based methods. Moreover, various studies have applied different hybrid techniques.

In a hybrid approach, Autoflow© [4,71] combined different algorithms (i.e., dynamic harmonic regression, Kalman filter, and fixed interval smooth) to predict short-term water demand. The proposed model was verified for three datasets. In the results of 100 tests, the value of the coefficient of determinism, R^2, was greater than 0.9, and the mean absolute percentage error (MAPE) was less than 5%. However, for further verification of the efficiency of the proposed model, it is essential to perform more testing using additional datasets. Furthermore, the proposed model was limited to forecasting only up to 24 hours ahead. Therefore, its effectiveness for predicting water demand more than 24 hours ahead must be investigated.

An ANN was combined with a time-series model to predict daily water demand by Al-Zahrani and Abo-Monasar [60]. They proposed combining the general regression NN (GRNN) model with a moving average method, and concluded that the hybrid method performed better than the time-series or ANN-based methods. Similarly, a fourier time series process over a SVR was proposed by Brentan et al. [57]. Seo, Kwon, and Choi proposed an ELM integrated with variational model decomposition (VMD) for forecasting short-term water demand [61]. Zubaidi et al. proposed using singular spectrum analysis

(SSA) and a linear autoregressive model [62] for predicting water demand. Furthermore, Duerr et al. [64] evaluated the following spatiotemporal statistical models and ML algorithms to forecast monthly water demand: linear and linear mixed models with month effects, a multiple linear regression model, and time-series models (AR(1) and ARIMA, spatiotemporal Gaussian process models, generalized additive models [GAM], random forests [RF], Bayesian additive regression trees [BART], and gradient boosting machines [GBM]). The study found that time-series models outperformed other models, indicating the temporal dynamics of water consumption. A hybrid regression

After critically analysing the literature, it became clear that ANN- and time series-based methods have performed better than hybrid-based methods. Further studies can be performed to observe the results when advanced ANN-based methods are integrated with time-series approaches for forecasting water demand.

3.3 Water event categorisation

The aim of water event categorisation is to identify the events associated with consumption for more effective water management. Studies that have focused on this topic can be divided into two categories: leaks and water end-use classification. Detecting leaks and notifying consumers about them can prevent vast amounts of water being wasted and help to keep bills at normal levels.

3.3.1 Leaks

Detecting leaks in water distribution is a challenging yet crucial task. This is because in most cases, water lines are situated underground, and furthermore, spikes in water bills may come as a shock to consumers because dampness may not appear on the surface. Real-time leak detection and alerts can lead to prompt repairs of leaks, thereby saving a substantial amount of water. To accomplish this task, many approaches have been employed [72-76]. Among these approaches, clustering techniques is the most popular for leak detection followed by hybrid approach. However, further research can be done to identify very small leaks (i.e., slow dripping taps) that create flows of such a low flow rate (< 3L/hr) that cannot be picked up by the meter (if the water meter does not record higher flow rates). Still, in the long run, leak detection techniques exhibit minimal effectiveness because they cannot promote water conservation once a leak is repaired.

3.3.2 Water end-use classification

In a 2010 study, categorising water end-use was identified as the most critical challenge and one that required urgent attention, because existing approaches to water end-use analysis required much time [77]. Since then, several approaches have been proposed [4,5,78,79] to classify water end-use. These can be placed into three broad categories: decision tree methods; data mining techniques applied to data collected from sensor devices integrated into water appliances; and hybrid approaches.

In previous studies [80,81], decision tree methods were mostly associated with the three physical attributes of an event—volume, duration, and flow rate—to classify water end-use. Pressure-based sensors were combined with water appliances to perform water end-use categorisation using data mining techniques [82,83]. However, both of these approaches lacked accuracy, and thus have not been deployed widely. Lastly, the hybrid category contains studies that combined multiple techniques to solve the problem of categorising water end-use. In this category, Autoflow© [4] was identified as making the most prominent contribution, and pattern recognition and data mining techniques were employed to perform the categorisation. For autonomous water end-use classification, the authors combined a hidden Markov model, dynamic time wrapping algorithm, and ANN with one hidden layer and 20 neurons to predict six categories. In most cases, the proposed approach achieved over 90% accuracy; however, for categorising bathtub and irrigation water end-use, the accuracy was 88.1% and 85.9%, respectively, which reduced the overall accuracy of the proposed approach. The relevant literature clearly indicates that categorising water end-use still requires substantial attention.

Table 4. Summary of the studies on water end-use categorisation.

End-use categorisation	Technique used				
	Clustering	Regression	SVM	Hybrid	Other
Leak	[74], [76]	[76]		[72]	[75], [73]
End-use events	[84]		[85]	[86], [87], [88], [89], [71], [90]	[91], [92], [93], [94], [95]

3.4 Socioeconomic analysis

Several studies have been undertaken to understand the effect of socioeconomic and demographic factors in water consumption and conservation [19,32,96]. Although these studies have not promoted immediate water savings, they have helped to identify the social, economic, and demographic factors behind water consumption patterns, which may lead to effective water management policies such as targeted community awareness programs or education.

In addition, studies performed to identify the determinants of household water consumption are essential for understanding the factors behind increasing or decreasing water usage. According to Bich-Ngoc and Teller [19], the examined factors in the literature can be categorised into six groups: (a) economic; (b) sociodemographic; (c) physical characteristics; (d) technological; (e) climatic; and (6) special factors. Higher income, older tenants, presence of infants, double income family, larger house area, irrigation, holiday, number of generations, and renter-occupied household are some of the drivers behind increased water consumption. By contrast, lower-income, younger tenants, absence of infants, single income family, smaller house, lower educated residents and owner-occupied household are the factors for decreased water usage. Notably, a few factors contradict each other. For instance, in most of the studies, a higher family income led to greater water consumption. However, Beal et al. [96] found that higher family income may lead to lower water usage, which was later supported by the findings of Willis et al. [32]. Similarly, in the case of family size, Willis et al. [32] reported that a larger family size contributed to decreased water consumption. A few studies have discussed the possible effect of pricing on water conservation; for instance, most US cities have a lower price per unit for high consumption than for low consumption, which may lead to water shortages in the future [97]. In other investigations, increasing block tariffs along with real-time information seem to have been effective but this requires confirmation in further research [98,99].

3.5 Behaviour analysis

Studies under the behaviour analysis theme have mainly aimed to understand the behaviours and dynamics of consumers based on water consumption data. Various techniques such as profiling, habit detection, and pattern recognition have been deployed to achieve this goal.

Cases of habit detection and profiling [100,101] have followed a cluster-based approach. Cardell-Oliver [101] identified four types of pattern: continuous-flow days; exceptional peak-use days; programmed patterns with recurrent hours; and normal use patterns. However, Cominola et al. [100] identified three profiles based on the eigenbehaviours of hourly consumption. Cole and Stewart [102] provided a detailed breakdown of hourly use by volume for the peak hour, peak day, and peak month to make accurate estimations of outdoor and indoor components. Later, a habit detection algorithm was proposed by Cardell-Oliver [103], but it could not guarantee the detection of all habits in a time series because of heuristics.

Moreover, a few studies have applied clustering techniques to segment or cluster consumers based on their water consumption data. For instance, Leyli-Abadi et al. [104] proposed a MixJNHMM model to cluster consumption behaviour and forecast future behaviour. Ji et al. [105] clustered residents by family structure, job type, or lifestyle based on water consumption data using a fuzzy clustering algorithm. In their study, users were divided into five clusters; although the size of the dataset was relatively small, the proposed approach provided interesting insights into users' lifestyles.

With regard to predicting water demand, many studies have been conducted to extract demand patterns. McKenna, Fusco, and Eck [78] proposed an approach for classifying demand patterns using Gaussian mixture models and K-means clustering. The authors averaged weekday demands over a full month to deal with noisy values, which was a superior approach to those used in other studies. Gurung et al. [106] developed a water demand profile for modelling water demand using diurnal patterns and clustering techniques; however, they only used 14 weeks of data, which was insufficient for understanding the effect of seasonality on water consumption. Later, Cheifetz et al. [107] identified eight relevant usage profiles from water consumption data using clustering and modelling techniques. However, scalability was a concern with the proposed approach.

A limited number of studies—albeit notable ones—have been conducted in the area of behaviour change. For example, Novak et al. [108] and Fraternali et al. [109] studied behaviour changes through gamified incentive models for simulating water savings, and found reduced consumption and positive user feedback. On the hand, Quesnel and Ajami [110] observed that changes in water consumption are linked to heavy news media coverage of water- and drought-related issues. Jorgensen et al. [111] tested explanations behind the lack of association between water conservation and intentions, and found that intentions, habit strength, and their interactions were not good predictors for future water conservation. Lastly, Fielding et al. [112] reported that intervention groups showed a reduced level of water consumption over a period of 12 months, and after that time, their water consumption returned to pre-intervention levels.

Table 5. Findings from the studies on behaviour analysis.

Reference	Type	Technique	Considered disaggregated events?
[104]	Customer segmentation	Hidden Markov Model	No
[113]	Demand profiling	Self-Organising Map, K-means, Dendrogram	No
[114]	Customer segmentation	hierarchical cluster analysis	No
[107]	Demand profiling	K-means, Fourier Regression Mixture model	No
[115]	Customer segmentation	K-means clustering	No
[103]	habit detection and profiling	Time series analysis	No
[116]	Customer segmentation	Descriptive analytics	No
[100]	habit detection and profiling	K-means clustering	No
[106]	Demand profiling	Diurnal pattern, clustering	No
[105]	Customer segmentation	Fuzzy clustering	No
[117]	Customer segmentation	Self-Organising Map	No
[94]	Demand profiling	Diurnal pattern	No
[118]	Demand profiling	Diurnal pattern	No
[78]	Demand profiling	Gaussian Mixture Model	No
[102]	habit detection and profiling	Descriptive statistics	No
[101]	habit detection and profiling	Signature pattern Clustering	No
[119]	habit detection and profiling	Factor Analysis, Cluster Analysis, Discriminant Analysis	No

Behaviour analysis studies are important for understanding the behaviour dynamics of consumers. Such an understanding is useful for water utilities to derive innovative and effective water conservation programs. Studies to date have been based on total water consumption data. However, behaviour analysis based on disaggregated water consumption may reveal interesting and useful insights that may lead to more effective water conservation. Thus, we have identified a gap in the literature involving behaviour analysis based on disaggregated water events.

4. Findings and Discussion

The aim of this section is to highlight the findings that emerged from the critical analysis of the literature pertaining to the five themes identified for this study. Among several interesting findings, one of the most important is the relationship between personalisation of information and effectiveness of water conservation. We observed that more strongly personalised programs have more effect on water savings. That is, if a water conservation program provides more personalised feedback or information then it contributes towards more water savings. Figure 4 summarises this finding based on the literature showing that water-use feedback programs have the highest levels of personalisation by providing customised water consumption summaries and statistics. Studies related to this theme reported from 4.2% to 27.0% water savings, which is the highest level among all themes.

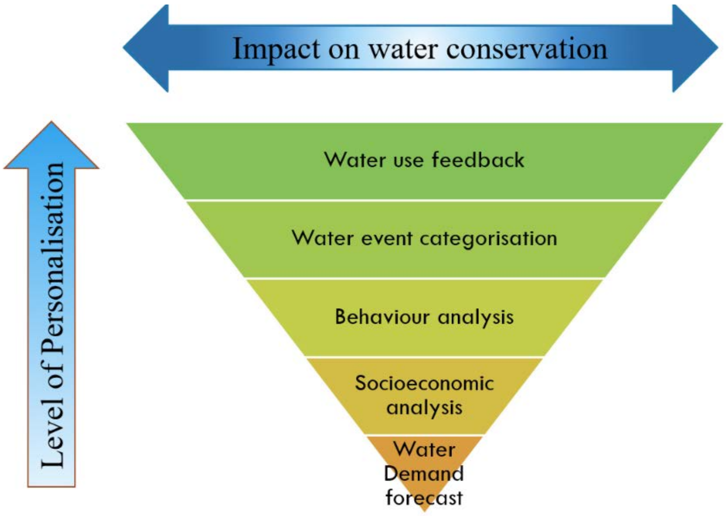


Figure 4. Relationship between personalisation and water conservation.

Water event categorisation helps consumers to detect leaks and provides information on water consumption by category. This type of program may help consumers to save water by facilitating the detection of leaks and reduced consumption based on the categorised water consumption data. However, consumers may not be aware of the potential water saving opportunities provided by this information. In addition, such programs cannot provide highly personalised feedback to promote water conservation. Behaviour analysis and socioeconomic analysis studies have identified the factors driving water consumption. This knowledge may assist in the design of water conservation programs for large populations that involve a lower level of personalisation and have less influence on water savings in the short term. Lastly, water demand forecasting studies have predicted the future water demand for a specific area or municipality but with less personalisation. This type of prediction helps utilities with efficient demand management but has no effect on water conservation by consumers. This confirms that effects on water conservation are strongly related to personalisation. However, the scope remains to adopt highly personalised systems, such as recommender systems, to promote water-conscious behaviour through prescriptive analytics.

From a business-intelligence perspective, prescriptive analytic is the highest level of analytics; they identify and suggest a set of actions to achieve business goals based on descriptive and predictive analytics. Because previous studies mostly fall under descriptive and predictive analytics, the absence of prescriptive analytics for promoting water-conscious behaviour was noted in this study, as shown in Figure 5.

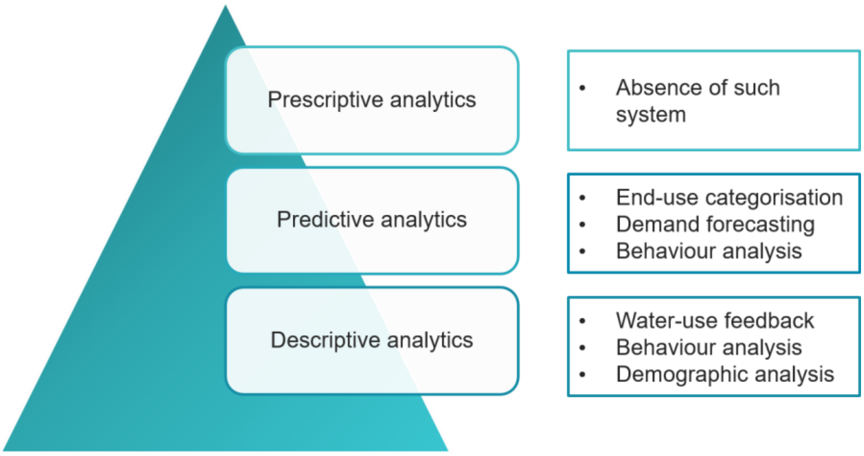


Figure 5. Grouping existing works from a business intelligence perspective.

Table 2 condenses the findings from the literature on water-use feedback. Although it may seem that paper-based feedback performs better than visual display or online feedback systems, one study [37] revealed that online portals have a greater influence than do paper-based reports (50% and 30%, respectively) in terms of behaviour change and water-use infrastructure (30% and 10%, respectively). However, because of the lack of user engagement with the online portal, the overall water saving metric was lower for online feedback (4.2%) than for paper-based reports (8.0%). In addition, [39] reported the same phenomenon: only 18% of participants accessed the website from 141 households chosen to receive web-based materials. These studies suggest that users with higher engagement with online portals are able to save more water than those receiving paper-based feedback. Thus, after carefully analysing the literature, we concluded that online feedback has greater potential for water conservation than does paper-based feedback, as supported by the findings of [16]. These studies indicate the necessity for further investigation to identify the most effective type of feedback in terms of information content, granularity, and frequency of delivery [45].

In the case of water demand forecasting, the literature analysed in the present study shows that over the last decade, short-term water demand for households has received more attention than medium-term demand forecasting. Figure 6 shows that from 2010 to 2019, the cumulative number of studies on short-term demand forecasting reached 16, whereas for medium-term forecasting the number was only six. Clearly, the availability of higher-frequency water consumption data in recent years has resulted in more research on short-term water demand forecasting.

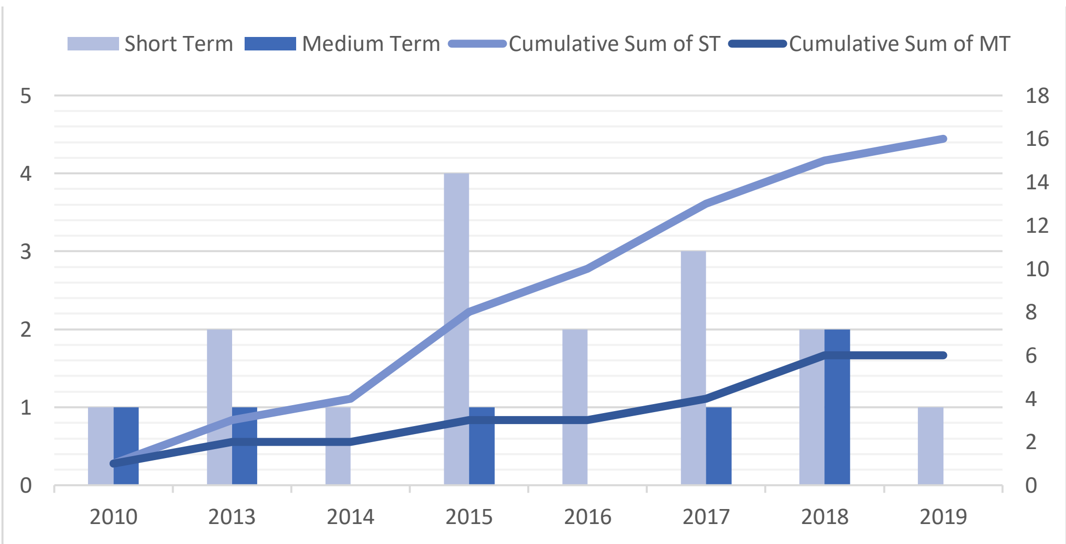


Figure 6. Number of studies and their cumulative sum over the years for short- and medium-term water demand forecasting.

Table 3 summarises studies on water demand forecasting based on the forecast horizon, forecast frequency, and forecasting techniques. In the case of short-term forecast horizons, hourly and daily forecasting frequencies have dominated over shorter or longer frequencies. This is because these two forecast frequencies are essential for water utilities to ensure the supply of water according to forecasted demand. Another finding is that for short-term water demand forecasting, regression and hybrid techniques have been the most commonly applied. However, in recent years, the performance of ANN-based extreme learning machines indicates that further studies are required to improve the overall accuracy of prediction through reducing errors.

Table 4 summarises the techniques applied in the water end-use categorisation literature to detect leaks and disaggregate end-use events. In terms of leak detection, clustering techniques were the most popular, followed by minimum night flow techniques. To ensure greater effectiveness, future research should focus on proactive leak-detection techniques. Hybrid techniques are mostly used in disaggregated end-use event classification, yet the current state-of-the-art accuracy of only around 93% suggests that scope exists for further research on this topic.

Furthermore, a healthy number of studies have reported that socioeconomic factors largely influence water consumption. Although these variables do not promote water conservation directly, identifying these variables is crucial to understand the reason for increments or decrements in water consumption, as well as to determine effective water conservation program and policies. Figure 7 presents the socioeconomic variables that increase or decrease water consumption according to [32,96,120-123]. Higher family income and larger family size are the two determinants that may lead to either increases or decreases in water usage. Therefore, future studies should address this issue in greater detail.

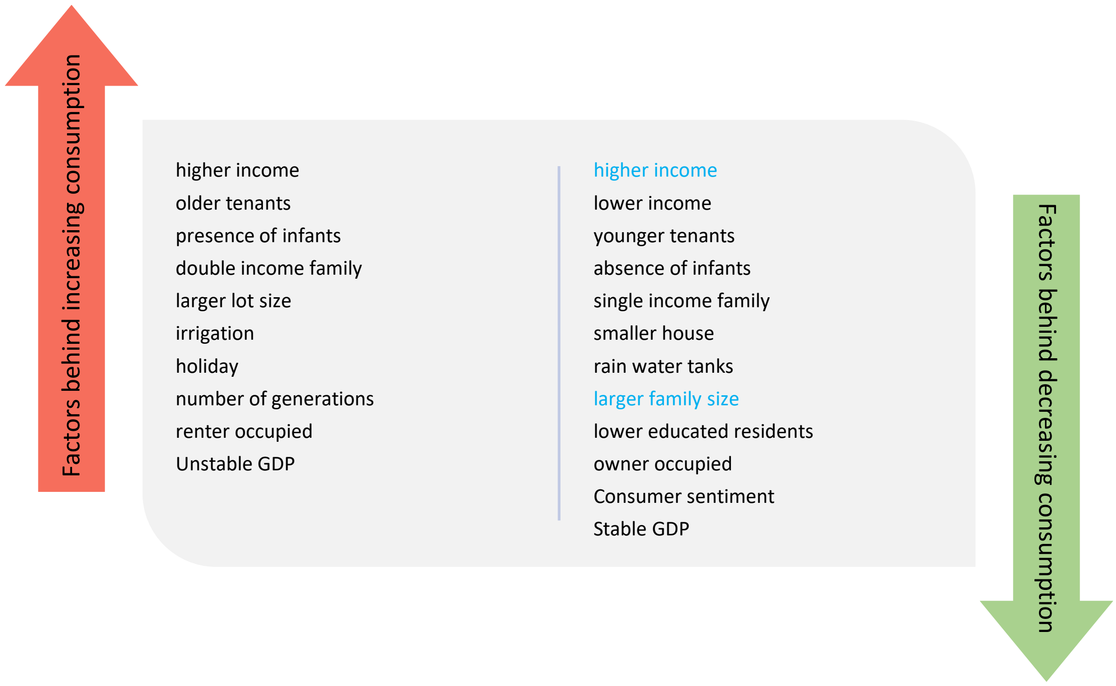


Figure 7. Socioeconomic factors that determine water consumption from mains water.

The literature related to behaviour analysis plays a vital role in understanding consumer behaviour. This understanding is useful for water utilities to develop and deliver innovative and effective water conservation programs. However, we identified a gap in behaviour analysis research based on disaggregated water events: studies to date have been based on total water consumption data, but behaviour analysis based on disaggregated water consumption may reveal noteworthy and useful insights that may lead to more effective water conservation.

Table 6. Summary of studies on behaviour analysis.

Behaviour Analysis	Applied Technique			
	Clustering	Hybrid	Descriptive analytics	Other
Habit detection and profiling	[101], [100]	[119]	[102]	[103]
Customer segmentation	[114], [105], [115]		[116]	[104], [117]
Demand profiling		[106], [107], [113]		[118], [94], [78]

Table 7. Summary of the studies based on sample size, data collection duration and data frequency.

Sample size	Data collection duration						Data sample frequency
	<2 weeks	2 weeks - < 1 month	1-3 months	4-6 months	7-12 months	Over 12 months	
1-100	[51], [36]		[89], [6], [34]		[85]		very short
	[105]		[54], [49], [53]	[78]		[73]	short
					[124]	[72], [66]	medium
101-200		[32], [40], [31], [63], [40], [31]	[106]			[38], [37]	very short
			[4]		[100]	[101]	short
							medium
201-300		[96]	[50]	[86], [59]		[86], [94]	very short
				[33]			short
							medium
301-400				[41]			very short
							short
							medium
400+						[88]	very short
			[58,79]		[103], [95]	[104], [95], [75], [52], [58], [55]	short
						[64]	medium

To understand the nature of the DWM data used in previous studies, we also collected various information such as sample size, data collection duration, and data frequency. Sample size mainly refers to the number of DWMs or households from which data were collected. Data collection duration represents the amount of time for which data were collected. We categorised data frequency into three categories: very short, short, and medium. The very short category includes the studies where the data frequency was 1 second, 5 seconds, 10 seconds, or 15 minutes. Hourly or daily data are considered short frequency data. Lastly, data collected at weekly, bimonthly, monthly, or yearly frequencies fall into the medium frequency category. Table 7 presents a condensed summary of the studies based on sample size, data collection duration, and data frequency. From the table, it is quite clear that in the case of very-short-frequency data, very rarely have studies been performed with more than 300 households/DWMs and data collection durations over 6 months. One of the reasons for this could be the trial rollout of DWMs. However, once DWMs are rolled out fully, this issue should be resolved.

Thus, we conclude that there is further scope to apply novel and advanced ML and DA techniques such as recommender systems to develop highly personalised feedback systems for promoting water-conscious behaviour. Further, to improve customer engagement with the system, integration of gamification may be helpful. To make recommender systems scalable, various clustering techniques need to be investigated. In addition, the introduction of data management solutions for the big data generated from DWMs would be an interesting development. Finally, user profiling or clustering based on disaggregated water consumption data may reveal knowledge that is essential but remains unexplored.

5. Conclusion

Water scarcity and low pressure during peak usage times are the two key challenges facing water utilities in urban areas. Trial rollouts of DWMs in many countries have revealed the potential of DWMs in water conservation, as is evident in the literature. Among the many findings, the review of these studies led us to conclude that a direct relationship exists between the level of personalisation and water conservation. However, in this review, through critically analysing the relevant studies, we identified research gaps and future opportunities regarding DWMs for promoting water conservation from ML and DA perspectives.

The identified research gaps or findings in this area are briefly discussed below:

- **Absence of highly personalised feedback systems**: As mentioned earlier, we observed a direct relationship between the level of personalisation and effects on water conservation. However, the absence in this field of highly personalised systems such as recommender systems [26] is notable. By generating a list of custom-tailored suggestions, such highly personalised system would promote water-conscious behaviour more effectively [125].
- **Absence of advanced ML and DA techniques**: A good number of ML and DA techniques have been applied to the data collected from DWMs. Many of these applied techniques were either basic or a mixture of several techniques. However, the application of advanced ML and DA techniques such as deep learning [68], deep reinforcement learning [126], anomaly detection [127], and recommender systems [26] in this field is rare. For instance, deep learning can be adopted to improve the accuracy of water demand forecasting, anomaly detection based techniques for abnormal water consumption (i.e., leak, theft) detection models, and deep reinforcement learning can be used to determine suitable actions for promoting water conservation. If applied properly, these advanced techniques may improve results.
- **Limitations in customer profiling and clustering**: Existing customer profiling and clustering studies have mostly been based on total consumption. For this reason, it is almost impossible to create customer profiles or perform clustering based on each water consumption event, such as shower, dishwashing, and gardening. Although disaggregated water consumption events are available, the gap in customer profiling and clustering is noticeable.
- **Absence of data management solutions**: Storing data in a plain text file is not suitable for performing analytics with a large amount of data in a real-world scenario. However, we

observed that in many studies, DWM data are stored in such files. This indicates the absence of data management solutions for DWM data.

- **Water demand forecasting and accuracy**: In this study, we found that short-term water demand forecasting has gained more attention in recent years compared with medium- or long-term water demand forecasting. However, we noted that further research scope exists in this area to improve prediction accuracy.
- **Effectiveness of the feedback-delivery medium**: The success of a water conservation program largely depends on the medium of its feedback. Among the various media, we observed that an online or web portal-based medium is the most effective when users were active.
- **Lack of user engagement with online portals:** Although existing works show that online portals are the most effective medium for delivering feedback, the lack of user engagement is still a challenge.
- **Limitations of clustering techniques**: Among the many clustering techniques, we noticed that the application of k-means clustering was very common. However, the k-means clustering technique has some limitations, such as in determining the value of k, the impact of the initial centroid value on the final result, and sensitiveness to the size of the data [128]. Furthermore, computational cost and scalability are challenging issues for any clustering technique. Therefore, besides k-means, other clustering techniques for big data [129] such as CLARANS [130], BIRCH [131], and CURE [132] should be investigated.
- **Factors affecting water consumption and conservation**: We listed the socioeconomic factors appearing in the articles that affect water consumption and conservation. These factors are crucial to consider for future research in this area. However, we noted that two determinants (higher income and family size) can be responsible for both increments and reductions in water consumption.
- **Limitation of DWM data**: While reviewing the literature, it soon became clear that some limitations exist in high-frequency DWM data in terms of the number of participating households, duration of data collection, and frequency of DWM data. In case of high-frequency data (5 second, 10 second, and 1-minute intervals), most of the studies collected data from fewer than 300 households and for less than 1 month in duration. However, high-frequency data can provide more insights compared with weekly, monthly, and yearly data.

Based on the findings and identified research gaps from the literature review, we provide the following recommendations for utilities and governments, as well as future research directions for the research community:

- **Highly personalised feedback and recommender systems**: Recommender systems can play a vital role in promoting water-conscious behaviours through providing highly personalised feedback [133]. Because this area is still unexplored, future research can be conducted on this topic.
- **Deploy advanced ML and DA techniques**: To improve accuracy in disaggregating water events, water demand forecasting, leak detection, customer profiling, and clustering, further research can be performed that deploys advanced ML and DA techniques such as deep learning, reinforcement learning, and anomaly detection.
- **Customer profiling and clustering based on disaggregated data**: Previous customer profiling and clustering studies have mostly been based on hourly total water consumption data. However, customer profiling and segmenting based on high-frequency disaggregated water consumption data may provide more insights. Therefore, future research should address these issues.

- **Research on the data warehouse solution**: Data warehouses are well-known for optimising analytics. However, no studies have been conducted on this topic, and thus further research can be conducted on developing data warehouse solutions. Such solutions would be beneficial for storing and analysing the vast amount of data generated by DWMs.
- **Feedback-delivery medium for future research**: Compared with other feedback-delivery media such as paper and visual displays, online or web portal systems perform better in terms of water conservation. Therefore, future research can be implemented using online or web portal-based feedback delivery.
- **Increasing user engagement in online portals**: Researchers have studied the impact of gamification and reward or rebate programs on user engagement. However, no comparative studies have been conducted to determine which is the most effective. Therefore, further investigation can be conducted to identify the most effective user engagement technique in online portals.
- **Application of clustering techniques**: Because many studies rely heavily on the k-means clustering technique, which has some limitations, further research can be conducted to identify alternative techniques that may improve the results of behaviour analysis, and water end-use categorisation.
- **Dealing with limited DWM data**: Collecting a large volume of DWM data is not an easy task because DWMs are still in the pilot stage, and furthermore, participation in such programs is mostly voluntary. To overcome this limitation, a synthetic data generation technique [134] was proposed, but further research can be conducted on this topic. Furthermore, investigations can be conducted to develop advanced ML techniques that work on a smaller dataset.

Supplementary Materials: The following are available online at www.mdpi.com/xxx/s1, Figure S1: title, Table S1: title, Video S1: title.

Author Contributions: Conceptualization, M.S.R., K.A.N., R.A.S., D.G., and M.B.; Methodology, M.S.R., K.A.N., R.A.S., D.G., and M.B.; Formal Analysis, M.S.R., K.A.N.; Data Curation, M.S.R., K.A.N.; Writing—Original Draft Preparation, M.S.R.; Writing—Review and Editing, M.S.R., K.A.N., R.A.S., D.G., and M.B.; funding acquisition, K.A.N., R.A.S., D.G., and M.B.

Funding: This research was funded by Australian Research Council (ARC), grant number LP160100215 that also includes funding from the following industry partners: Yarra Valley Water, City West Water, Southeast Water, and Aquiba.

Conflicts of Interest: The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

ReferencesUncategorized References

1. *The Global Risks Report 2019*; World Economic Forum: Geneva, Switzerland, 2019.
2. Harou, J.J.; Medellín‐Azuara, J.; Zhu, T.; Tanaka, S.K.; Lund, J.R.; Stine, S.; Olivares, M.A.; Jenkins, M.W. Economic consequences of optimized water management for a prolonged, severe drought in California. *Water Resources Research* **2010**, *46*.
3. Jiang, Y. China's water scarcity. *Journal of environmental management* **2009**, *90*, 3185-3196.
4. Nguyen, K.A.; Sahin, O.; Stewart, R.A.; Zhang, H. Water demand forecasting with AUTOFLOW© using State-Space approach. In Proceedings of 8th International Congress on Environmental Modelling and Software.
5. Nguyen, K.A.; Stewart, R.A.; Zhang, H.; Sahin, O.; Siriwardene, N. Re-engineering traditional urban water management practices with smart metering and informatics. *Environmental Modelling & Software* **2018**, *101*, 256-267, doi:10.1016/j.envsoft.2017.12.015.
6. Liu, A.; Giurco, D.; Mukheibir, P. Urban water conservation through customised water and end-use information. *Journal of Cleaner Production* **2016**, *112*, 3164-3175.

7. Sahin, O.; Stewart, R.A.; Porter, M.G. Water security through scarcity pricing and reverse osmosis: a system dynamics approach. *Journal of Cleaner Production* **2015**, *88*, 160-171, doi:10.1016/j.jclepro.2014.05.009.
8. Stewart, R.A.; Nguyen, K.; Beal, C.; Zhang, H.; Sahin, O.; Bertone, E.; Vieira, A.S.; Castelletti, A.; Cominola, A.; Giuliani, M., et al. Integrated intelligent water-energy metering systems and informatics: Visioning a digital multi-utility service provider. *Environmental Modelling & Software* **2018**, *105*, 94-117, doi:https://doi.org/10.1016/j.envsoft.2018.03.006.
9. Tanverakul, S.A.; Lee, J. Impacts of metering on residential water use in California. *American Water Works Association* **2015**, *107*, E69-E75, doi:10.5942/jawwa.2015.107.0005.
10. Reynaud, A.; Pons, M.; Pesado, C. Household water demand in Andorra: Impact of individual metering and seasonality. *Water (Switzerland)* **2018**, *10*, doi:10.3390/w10030321.
11. Randall, T.; Koech, R. Smart water metering technology for water management in urban areas. *Water e-Journal* **2019**, *4*.
12. Gurung, T.R.; Stewart, R.A.; Beal, C.D.; Sharma, A.K. Smart meter enabled informatics for economically efficient diversified water supply infrastructure planning. *Journal of Cleaner Production* **2016**, *135*, 1023-1033, doi:10.1016/j.jclepro.2016.07.017.
13. Beal, C.D.; Flynn, J. Toward the digital water age: Survey and case studies of Australian water utility smart-metering programs. *Utilities Policy* **2015**, *32*, 29-37.
14. Cominola, A.; Giuliani, M.; Piga, D.; Castelletti, A.; Rizzoli, A.E. Benefits and challenges of using smart meters for advancing residential water demand modeling and management: A review. *Environmental Modelling & Software* **2015**, *72*, 198-214, doi:10.1016/j.envsoft.2015.07.012.
15. Liu, A.; Giurco, D.; Mukheibir, P.; White, S. Detailed water-use feedback: A review and proposed framework for program implementation. *Utilities Policy* **2016**, *43*, 140-150.
16. Liu, A.; Mukheibir, P. Digital metering feedback and changes in water consumption – A review. *Resources, Conservation and Recycling* **2018**, *134*, 136-148, doi:10.1016/j.resconrec.2018.03.010.
17. Boyle, T.; Giurco, D.; Mukheibir, P.; Liu, A.; Moy, C.; White, S.; Stewart, R. Intelligent Metering for Urban Water: A Review. *Water* **2013**, *5*, 1052-1081, doi:10.3390/w5031052.
18. Monks, I.; Stewart, R.A.; Sahin, O.; Keller, R. Revealing Unreported Benefits of Digital Water Metering: Literature Review and Expert Opinions. *Water* **2019**, *11*, 838.
19. Bich-Ngoc, N.; Teller, J. A review of residential water consumption determinants. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2018; Vol. 10964 LNCS, pp 685-696.
20. Miller, C.; Nagy, Z.; Schlueter, A. A review of unsupervised statistical learning and visual analytics techniques applied to performance analysis of non-residential buildings. *Renewable and Sustainable Energy Reviews* **2018**, *81*, 1365-1377, doi:10.1016/j.rser.2017.05.124.
21. Emmanuel A. Donkor, S.M.A.T.A.M.R.S.a.J.A.R., P.E.4. Urban Water Demand Forecasting: Review of Methods and Models. *JOURNAL OF WATER RESOURCES PLANNING AND MANAGEMENT* **2014**, 10.1061/(ASCE)WR.1943-5452, doi:10.1061/(ASCE)WR.1943-5452.
22. Makropoulos, C.; Savić, D. Urban Hydroinformatics: Past, Present and Future. *Water* **2019**, *11*, 1959.
23. Moglia, M.; Cook, S.; Tapsuwan, S. Promoting Water Conservation: Where to from here? *Water* **2018**, *10*, doi:10.3390/w10111510.
24. Burke, R. Hybrid Web Recommender Systems. In *The Adaptive Web*, 2007; 10.1007/978-3-540-72079-9_12pp. 377-408.
25. Ricci, F.; Rokach, L.; Shapira, B. Recommender Systems: Introduction and Challenges. In *Recommender Systems Handbook*, 2015; 10.1007/978-1-4899-7637-6_1pp. 1-34.
26. Resnick, P.; Varian, H.R. Recommender systems. *Communications of the ACM* **1997**, *40*, 56-58.

27. Pickering, C.; Byrne, J. The benefits of publishing systematic quantitative literature reviews for PhD candidates and other early-career researchers. *Higher Education Research & Development* **2014**, *33*, 534-548.
28. *Water Conservation Report*; Sydney Water: Sydney, 2018.
29. Yalçintaş, M.; Bulu, M.; Küçükvar, M.; Samadi, H. A framework for sustainable urban water management through demand and supply forecasting: The case of Istanbul. *Sustainability (Switzerland)* **2015**, *7*, 11050-11067, doi:10.3390/su70811050.
30. Wohlin, C. Guidelines for snowballing in systematic literature studies and a replication in software engineering. In Proceedings of Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering, London, England, United Kingdom; pp. 1-10.
31. Willis, R.M.; Stewarta, R.A.; Panuwatwanich, K.; Jones, S.; Kyriakides, A. Alarming visual display monitors affecting shower end use water and energy conservation in Australian residential households. *Resources, Conservation and Recycling* **2010**, *54*, 1117-1127, doi:10.1016/j.resconrec.2010.03.004.
32. Willis, R.; Stewart, R.; Giurco, D.; Talebpour, M.; Mousavinejad, A. End use water consumption in households: impact of socio-demographic factors and efficient devices. *Journal of Cleaner Production* **2013**, *60*, 107-115, doi:http://dx.doi.org/10.1016/j.jclepro.2011.08.006.
33. Otaki, Y.; Ueda, K.; Sakura, O. Effects of feedback about community water consumption on residential water conservation. *Journal of Cleaner Production* **2017**, *143*, 719-730, doi:10.1016/j.jclepro.2016.12.051.
34. Liu, A.; Giurco, D.; Mukheibir, P. Motivating metrics for household water-use feedback. *Resources, Conservation and Recycling* **2015**, *103*, 29-46, doi:10.1016/j.resconrec.2015.05.008.
35. Seyranian, V.; Sinatra, G.M.; Polikoff, M.S. Comparing communication strategies for reducing residential water consumption. *Journal of Environmental Psychology* **2015**, *41*, 81-90, doi:https://doi.org/10.1016/j.jenvp.2014.11.009.
36. Tom, G.; Tauchus, G.; Williams, J.; Tong, S. The Role of Communicative Feedback in Successful Water Conservation Programs. *Applied Environmental Education & Communication* **2011**, *10*, 80-90, doi:10.1080/1533015X.2011.575632.
37. Liu, A.; Giurco, D.; Mukheibir, P. Advancing household water-use feedback to inform customer behaviour for sustainable urban water. *Water Science and Technology: Water Supply* **2017**, *17*, 198-205.
38. Liu, A.; Giurco, D.; Mukheibir, P.; Mohr, S.; Watkins, G.; White, S. Online water-use feedback: household user interest, savings and implications. *Urban Water Journal* **2017**, *14*, 900-907, doi:10.1080/1573062X.2017.1279194.
39. Schultz, P.W.; Messina, A.; Tronu, G.; Limas, E.F.; Gupta, R.; Estrada, M. Personalized Normative Feedback and the Moderating Role of Personal Norms: A Field Experiment to Reduce Residential Water Consumption. *Environment and Behavior* **2014**, *48*, 686-710, doi:10.1177/0013916514553835.
40. Stewart, R.A.; Willis, R.M.; Panuwatwanich, K.; Sahin, O. Showering behavioural response to alarming visual display monitors: longitudinal mixed method study. *Behaviour & Information Technology* **2013**, *32*, 695-711, doi:10.1080/0144929x.2011.577195.
41. Erickson, T.; Podlaseck, M.; Sahu, S.; Dai, J.D.; Chao, T.; Naphade, M. The dubuque water portal: evaluation of the uptake, use and impact of residential water consumption feedback. In Proceedings of Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Austin, Texas, USA; pp. 675-684.
42. Deterding, S.; Sicart, M.; Nacke, L.; O'Hara, K.; Dixon, D. Gamification. using game-design elements in non-gaming contexts. In Proceedings of CHI'11 extended abstracts on human factors in computing systems; pp. 2425-2428.

43. Kossieris, P.; Kozanis, S.; Hashmi, A.; Katsiri, E.; Vamvakeridou-Lyroudia, L.S.; Farmani, R.; Makropoulos, C.; Savic, D. A web-based platform for water efficient households. In Proceedings of Procedia Engineering; pp. 1128-1135.
44. Lanzarone, G.A.; Zanzi, A. Monitoring gas and water consumption through icts for improved user awareness. *Information Communication and Society* **2010**, *13*, 121-135, doi:10.1080/13691180902992962.
45. Sønderlund Anders, L.; Smith Joanne, R.; Hutton Christopher, J.; Kapelan, Z.; Savic, D. Effectiveness of Smart Meter-Based Consumption Feedback in Curbing Household Water Use: Knowns and Unknowns. *Journal of Water Resources Planning and Management* **2016**, *142*, 04016060, doi:10.1061/(ASCE)WR.1943-5452.0000703.
46. Pesantez, J.E.B., Emily Zechman; Kaza, Nikhil Modeling and Forecasting Short-Term Water Demand Reported by Smart Meters. In Proceedings of 1st International WDSA / CCWI 2018 Joint Conference, Kingston, Ontario, Canada.
47. Zubaidi, S.L.; Dooley, J.; Alkhaddar, R.M.; Abdellatif, M.; Al-Bugharbee, H.; Ortega-Martorell, S. A Novel approach for predicting monthly water demand by combining singular spectrum analysis with neural networks. *Journal of Hydrology* **2018**, *561*, 136-145, doi:10.1016/j.jhydrol.2018.03.047.
48. Mouatadid, S.; Adamowski, J. Using extreme learning machines for short-term urban water demand forecasting. *Urban Water Journal* **2016**, *14*, 630-638, doi:10.1080/1573062x.2016.1236133.
49. Walker, D.; Creaco, E.; Vamvakeridou-Lyroudia, L.; Farmani, R.; Kapelan, Z.; Savić, D. Forecasting Domestic Water Consumption from Smart Meter Readings Using Statistical Methods and Artificial Neural Networks. *Procedia Engineering* **2015**, *119*, 1419-1428, doi:10.1016/j.proeng.2015.08.1002.
50. Bennett, C.; Stewart, R.A.; Beal, C.D. ANN-based residential water end-use demand forecasting model. *Expert Systems with Applications* **2013**, *40*, 1014-1023, doi:10.1016/j.eswa.2012.08.012.
51. Blokker, E.J.M.; Vreeburg, J.H.G.; van Dijk, J.C. Simulating residential water demand with a stochastic end-use model. *Journal of Water Resources Planning and Management* **2010**, *136*, 19-26, doi:10.1061/(ASCE)WR.1943-5452.0000002.
52. Rathnayaka, K.; Malano, H.; Arora, M.; George, B.; Maheepala, S.; Nawarathna, B. Prediction of urban residential end-use water demands by integrating known and unknown water demand drivers at multiple scales II: Model application and validation. *Resources, Conservation and Recycling* **2017**, *118*, 1-12, doi:10.1016/j.resconrec.2016.11.015.
53. Candelieri, A. Clustering and Support Vector Regression for Water Demand Forecasting and Anomaly Detection. *Water* **2017**, *9*, 224, doi:http://dx.doi.org/10.3390/w9030224.
54. Candelieri, A.; Soldi, D.; Archetti, F. Short-term forecasting of hourly water consumption by using automatic metering readers data. In Proceedings of Procedia Engineering; pp. 844-853.
55. Platsko, V.B., Peter van. Identification, Prediction, and Explanation of Outdoor Residential Water Consumption Using Smart Meter Data2018. In Proceedings of 1st International WDSA / CCWI 2018 Joint Conference, Kingston, Ontario, Canada.
56. Herrera, M.; Torgo, L.; Izquierdo, J.; Pérez-García, R. Predictive models for forecasting hourly urban water demand. *Journal of hydrology* **2010**, *387*, 141-150.
57. Brentan, B.M.; Luvizotto Jr, E.; Herrera, M.; Izquierdo, J.; Pérez-García, R. Hybrid regression model for near real-time urban water demand forecasting. *Journal of Computational and Applied Mathematics* **2017**, *309*, 532-541.
58. Abadi, M.L.; Same, A.; Oukhellou, L.; Cheifetz, N.; Mandel, P.; Feliers, C.; Chesneau, O. Predictive Classification of Water Consumption Time Series Using Non-homogeneous Markov Models. In Proceedings of 2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA); pp. 323-331.

59. Makki, A.A.; Stewart, R.A.; Beal, C.D.; Panuwatwanich, K. Novel bottom-up urban water demand forecasting model: Revealing the determinants, drivers and predictors of residential indoor end-use consumption. *Resources, Conservation and Recycling* **2015**, *95*, 15-37, doi:10.1016/j.resconrec.2014.11.009.
60. Al-Zahrani, M.A.; Abo-Monasar, A. Urban residential water demand prediction based on artificial neural networks and time series models. *Water Resources Management* **2015**, *29*, 3651-3662, doi:10.1007/s11269-015-1021-z.
61. Seo, Y.; Kwon, S.; Choi, Y. Short-term water demand forecasting model combining variational mode decomposition and extreme learning machine. *Hydrology* **2018**, *5*, doi:10.3390/hydrology5040054.
62. Zubaidi, S.L.; Kot, P.; Alkhaddar, R.M.; Abdellatif, M.; Al-Bugharbee, H. Short-term water demand prediction in residential complexes: Case study in Columbia city, USA. In Proceedings of Proceedings - International Conference on Developments in eSystems Engineering, DeSE; pp. 31-35.
63. Makki, A.A.; Stewart, R.A.; Panuwatwanich, K.; Beal, C. Revealing the determinants of shower water end use consumption: enabling better targeted urban water conservation strategies. *Journal of Cleaner Production* **2013**, *60*, 129-146, doi:10.1016/j.jclepro.2011.08.007.
64. Duerr, I.; Merrill, H.R.; Wang, C.; Bai, R.; Boyer, M.; Dukes, M.D.; Bliznyuk, N. Forecasting urban household water demand with statistical and machine learning methods using large space-time data: A Comparative study. *Environmental Modelling and Software* **2018**, *102*, 29-38, doi:10.1016/j.envsoft.2018.01.002.
65. Sharvelle, S.; Dozier, A.; Arabi, M.; Reichel, B. A geospatially-enabled web tool for urban water demand forecasting and assessment of alternative urban water management strategies. *Environmental Modelling and Software* **2017**, *97*, 213-228, doi:10.1016/j.envsoft.2017.08.009.
66. Polebitski, A.S.; Palmer, R.N. Seasonal Residential Water Demand Forecasting for Census Tracts. *Journal of Water Resources Planning and Management* **2010**, *136*, 27-36, doi:10.1061/(asce)wr.1943-5452.0000003.
67. Sebri, M. ANN versus SARIMA models in forecasting residential water consumption in Tunisia. *Journal of Water Sanitation and Hygiene for Development* **2013**, *3*, 330-340, doi:10.2166/washdev.2013.031.
68. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *nature* **2015**, *521*, 436-444.
69. Hinton, G.; Deng, L.; Yu, D.; Dahl, G.; Mohamed, A.-r.; Jaitly, N.; Senior, A.; Vanhoucke, V.; Nguyen, P.; Kingsbury, B. Deep neural networks for acoustic modeling in speech recognition. *IEEE Signal processing magazine* **2012**, *29*.
70. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural computation* **1997**, *9*, 1735-1780.
71. Nguyen, K.A.; Stewart, R.; Zhang, H.; Jones, C.; Siriwardene, N.; Brown, A.; Radion, A.; Crook, J.; Stevens, M.; Smith, N., et al. DEVELOPING A NEXT GENERATION MACHINE LEARNING SYSTEM FOR ENHANCED URBAN WATER MANAGEMENT: AUTOFLOW. In Proceedings of Ozwater19, Melbourne, Australia.
72. Bragalli, C.; Neri, M.; Toth, E. Effectiveness of smart meter-based urban water loss assessment in a real network with synchronous and incomplete readings. *Environmental Modelling and Software* **2019**, 10.1016/j.envsoft.2018.10.010, 128-142, doi:10.1016/j.envsoft.2018.10.010.
73. Farah, E.; Shahrour, I. Smart water for leakage detection: Feedback about the use of automated meter reading technology. In Proceedings of 2017 Sensors Networks Smart and Emerging Technologies, SENSET 2017; pp. 1-4.
74. Kermany, E.; Mazzawi, H.; Baras, D.; Naveh, Y.; Michaelis, H. Analysis of advanced meter infrastructure data of water consumption in apartment buildings. In Proceedings of

Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; pp. 1159-1167.
75. Schultz, W.; Javey, S.; Sorokina, A. Smart Water Meters and Data Analytics Decrease Wasted Water Due to Leaks. *Journal - American Water Works Association* **2018**, *110*, E24-E30, doi:10.1002/awwa.1124.
76. Candelieri, A.; Archetti, F.; Messina, E. Analytics for supporting urban water management. *Environmental Engineering and Management Journal* **2013**, *12*, 875-881.
77. Stewart, R.A.; Willis, R.; Giurco, D.; Panuwatwanich, K.; Capati, G. Web-based knowledge management system: linking smart metering to the future of urban water planning. *Australian Planner* **2010**, *47*, 66-74, doi:10.1080/07293681003767769.
78. McKenna, S.A.; Fusco, F.; Eck, B.J. Water Demand Pattern Classification from Smart Meter Data. *Procedia Engineering* **2014**, *70*, 1121-1130, doi:10.1016/j.proeng.2014.02.124.
79. Leyli-Abadi, M.; Samé, A.; Oukhellou, L.; Cheifetz, N.; Mandel, P.; Féliers, C.; Chesneau, O. Predictive classification of water consumption time series using non-homogeneous markov models. In Proceedings of Proceedings - 2017 International Conference on Data Science and Advanced Analytics, DSAA 2017; pp. 323-331.
80. Wizard, T. Trace Wizard water use analysis tool. Users Manual. Aquacaft. Inc: 2003.
81. Kowalski, M.; Marshallsay, D. A system for improved assessment of domestic water use components. In Proceedings of II International Conference Efficient Use and Management of Urban Water Supply.
82. Froehlich, J.E.; Larson, E.; Campbell, T.; Haggerty, C.; Fogarty, J.; Patel, S.N. HydroSense: infrastructure-mediated single-point sensing of whole-home water activity. In Proceedings of Proceedings of the 11th international conference on Ubiquitous computing; pp. 235-244.
83. Froehlich, J.; Larson, E.; Saba, E.; Campbell, T.; Atlas, L.; Fogarty, J.; Patel, S. A longitudinal study of pressure sensing to infer real-world water usage events in the home. In Proceedings of International conference on pervasive computing; pp. 50-69.
84. Srinivasan, V.; Stankovic, J.; Whitehouse, K. WaterSense: Water flow disaggregation using motion sensors. In Proceedings of BuildSys 2011 - Proceedings of the 3rd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings, Held in Conjunction with ACM SenSys 2011; pp. 19-24.
85. Gao, Y.; Hou, D.; Banerjee, N.K.; Banerjee, S. Water fixture identification in smart housing: A domain knowledge based case study. In Proceedings of Proceedings - 2016 15th IEEE International Conference on Machine Learning and Applications, ICMLA 2016; pp. 904-909.
86. Nguyen, K.A.; Stewart, R.A.; Zhang, H. An autonomous and intelligent expert system for residential water end-use classification. *Expert Systems with Applications* **2014**, *41*, 342-356, doi:10.1016/j.eswa.2013.07.049.
87. Pastor-Jabaloyes, L.; Arregui, F.; Cobacho, R. Water End Use Disaggregation Based on Soft Computing Techniques. *Water* **2018**, *10*, doi:10.3390/w10010046.
88. Nguyen, K.A.; Stewart, R.A.; Zhang, H.; Jones, C. Intelligent autonomous system for residential water end use classification: Autoflow. *Applied Soft Computing* **2015**, *31*, 118-131, doi:10.1016/j.asoc.2015.03.007.
89. Fontdecaba, S.; Sánchez-Espigares, J.A.; Marco-Almagro, L.; Tort-Martorell, X.; Cabrespina, F.; Zubelzu, J. An Approach to Disaggregating Total Household Water Consumption into Major End-Uses. *Water Resources Management* **2013**, *27*, 2155-2177, doi:10.1007/s11269-013-0281-8.
90. Nguyen, K.A.; Stewart, R.A.; Zhang, H. An intelligent pattern recognition model to automate the categorisation of residential water end-use events. *Environmental Modelling & Software* **2013**, *47*, 108-127, doi:10.1016/j.envsoft.2013.05.002.
91. Soares, E.S.; Oliveira-Esquerre, K.P.; de Aguiar, A.M.; Botelho, G.L.P.; Kiperstok, A. Development of a model to identify combined use in residential water end use events. In *Computer Aided Chemical Engineering*, 2018; Vol. 44, pp 1951-1956.

92. Pastor-Jabaloyes, L.; Arregui, F.J.; Cobacho, R. A filtering algorithm for high-resolution flow traces to improve water end-use analysis. *Water Supply* **2019**, *19*, 451-462, doi:10.2166/ws.2018.090.
93. Guragai, B.; Hashimoto, T.; Oguma, K.; Takizawa, S. Data logger-based measurement of household water consumption and micro-component analysis of an intermittent water supply system. *Journal of Cleaner Production* **2018**, *197*, 1159-1168, doi:10.1016/j.jclepro.2018.06.198.
94. Beal, C.D.; Stewart, R.A. Identifying Residential Water End Uses Underpinning Peak Day and Peak Hour Demand. *Journal of Water Resources Planning and Management* **2014**, *140*, doi:10.1061/(asce)wr.1943-5452.0000357.
95. Cole, G.; Stewart, R. Smart meter enabled disaggregation of urban peak water demand: precursor to effective urban water planning. *Urban Water Journal* **2013**, *10*, 174-194, doi:http://dx.doi.org/10.1080/1573062X.2012.716446.
96. Beal, C.; Stewart, R.A.; Spinks, A.; Fielding, K. Using smart meters to identify social and technological impacts on residential water consumption. *Water Science & Technology: Water Supply* **2011**, *11*, 527-533, doi:http://dx.doi.org/10.2166/ws.2011.088.
97. Luby, I.H.; Polasky, S.; Swackhamer, D.L. U.S. Urban Water Prices: Cheaper When Drier. *Water Resources Research* **2018**, *54*, 6126-6132, doi:10.1029/2018wr023258.
98. Sahin, O.; Bertone, E.; Beal, C.D. A systems approach for assessing water conservation potential through demand-based water tariffs. *Journal of Cleaner Production* **2017**, *148*, 773-784, doi:10.1016/j.jclepro.2017.02.051.
99. Strong, A.; Goemans, C. The impact of real-time quantity information on residential water demand. *Water Resources and Economics* **2015**, *10*, 1-13, doi:10.1016/j.wre.2015.02.002.
100. Cominola, A.; Moro, A.; Riva, L.; Giuliani, M.; Castelleti, A. Profiling residential water users' routines by eigenbehavior modelling. **2016**.
101. Cardell-Oliver, R. Water use signature patterns for analyzing household consumption using medium resolution meter data. *Water Resources Research* **2013**, *49*, 8589-8599, doi:http://dx.doi.org/10.1002/2013WR014458.
102. Cole, G.; Stewart, R.A. Smart meter enabled disaggregation of urban peak water demand: precursor to effective urban water planning. *Urban Water Journal* **2013**, *10*, 174-194, doi:10.1080/1573062x.2012.716446.
103. Cardell-Oliver, R. A habit detection algorithm (Hda) for discovering recurrent patterns in smart meter time series. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2016; Vol. 9546, pp 109-127.
104. Leyli-Abadi, M.; Same, A.; Oukhellou, L.; Cheifetz, N.; Mandel, P.; Feliers, C.; Chesneau, O. Mixture of Non-homogeneous Hidden Markov Models for Clustering and Prediction of Water Consumption Time Series. In Proceedings of Proceedings of the International Joint Conference on Neural Networks.
105. Ji, Y.; Yong, L.; Jingfeng, Y.; Ke, K.; Yuehua, H.; Wenguang, Q. Analysis of urban residential water consumption based on smart meters and fuzzy clustering. In Proceedings of Proceedings - 15th IEEE International Conference on Computer and Information Technology, CIT 2015, 14th IEEE International Conference on Ubiquitous Computing and Communications, IUCC 2015, 13th IEEE International Conference on Dependable, Autonomic and Secure Computing, DASC 2015 and 13th IEEE International Conference on Pervasive Intelligence and Computing, PICom 2015; pp. 1295-1301.
106. Gurung, T.R.; Stewart, R.A.; Beal, C.D.; Sharma, A.K. Smart meter enabled water end-use demand data: platform for the enhanced infrastructure planning of contemporary urban water supply networks. *Journal of Cleaner Production* **2015**, *87*, 642-654, doi:10.1016/j.jclepro.2014.09.054.

107. Cheifetz, N.; Noumir, Z.; Samé, A.; Sandraz, A.-C.; Féliers, C.; Heim, V. Modeling and clustering water demand patterns from real-world smart meter data. *Drinking Water Engineering and Science* **2017**, *10*, 75-82, doi:10.5194/dwes-10-75-2017.
108. Novak, J.; Melenhorst, M.; Micheel, I.; Pasini, C.; Fraternali, P.; Rizzoli, A.E. Integrating behavioural change and gamified incentive modelling for stimulating water saving. *Environmental Modelling & Software* **2018**, *102*, 120-137, doi:10.1016/j.envsoft.2017.11.038.
109. Fraternali, P.; Baroffio, G.; Pasini, C.; Galli, L.; Micheel, I.; Novak, J.; Rizzoli, A. Integrating Real and Digital Games with Data Analytics for Water Consumption Behavioral Change: A Demo. In Proceedings of Proceedings - 2015 IEEE/ACM 8th International Conference on Utility and Cloud Computing, UCC 2015; pp. 408-409.
110. Quesnel, K.J.; Ajami, N.K. Changes in water consumption linked to heavy news media coverage of extreme climatic events. *Science Advances* **2017**, *3*, doi:10.1126/sciadv.1700784.
111. Jorgensen, B.S.; Martin, J.F.; Pearce, M.W.; Willis, E.M. Aligning theory and measurement in behavioral models of water conservation. *Water Policy* **2015**, *17*, 762-776, doi:10.2166/wp.2014.084.
112. Fielding, K.S.; Spinks, A.; Russell, S.; McCrea, R.; Stewart, R.; Gardner, J. An experimental test of voluntary strategies to promote urban water demand management. *J Environ Manage* **2013**, *114*, 343-351, doi:10.1016/j.jenvman.2012.10.027.
113. Padulano, R.; Del Giudice, G. A Mixed Strategy Based on Self-Organizing Map for Water Demand Pattern Profiling of Large-Size Smart Water Grid Data. *Water Resources Management* **2018**, *32*, 3671-3685, doi:10.1007/s11269-018-2012-7.
114. Vieira, P.; Jorge, C.; Covas, D. Efficiency assessment of household water use. *Urban Water Journal* **2018**, *15*, 407-417, doi:10.1080/1573062x.2018.1508596.
115. Garcia, D.; Puig, V.; Quevedo, J.; Cugueró, M.A. Big Data Analytics and Knowledge Discovery Applied to Automatic Meter Readers. In *Real-time Monitoring and Operational Control of Drinking-Water Systems*, 2017; 10.1007/978-3-319-50751-4_20pp. 401-423.
116. Cardell-Oliver, R.; Wang, J.; Gigney, H. Smart Meter Analytics to Pinpoint Opportunities for Reducing Household Water Use. *Journal of Water Resources Planning and Management* **2016**, *142*, doi:10.1061/(asce)wr.1943-5452.0000634.
117. Laspidou, C.; Papageorgiou, E.; Kokkinos, K.; Sahu, S.; Gupta, A.; Tassiulas, L. Exploring patterns in water consumption by clustering. In Proceedings of Procedia Engineering; pp. 1439-1446.
118. Gurung, T.R.; Stewart, R.A.; Sharma, A.K.; Beal, C.D. Smart meters for enhanced water supply network modelling and infrastructure planning. *Resources, Conservation and Recycling* **2014**, *90*, 34-50, doi:10.1016/j.resconrec.2014.06.005.
119. Solanas, J.L.; Cussó, M.R. Multivariate consumption profiling (MCP) for intelligent meter systems: a methodology to define categories and levels. *Water Science and Technology: Water Supply* **2010**, *10*, 710-720, doi:10.2166/ws.2010.374.
120. Kenney, D.S.; Goemans, C.; Klein, R.; Lowrey, J.; Reidy, K. Residential water demand management: lessons from Aurora, Colorado 1. *JAWRA Journal of the American Water Resources Association* **2008**, *44*, 192-207.
121. Kim, S.; Choi, S.; Koo, J.; Choi, S.; Hyun, I. Trend analysis of domestic water consumption depending upon social, cultural, economic parameters. *Water Science and Technology: Water Supply* **2007**, *7*, 61-68.
122. Rozos, E.; Butler, D.; Makropoulos, C. An integrated system dynamics–cellular automata model for distributed water-infrastructure planning. *Water Science and Technology: Water Supply* **2016**, *16*, 1519-1527.
123. Baki, S.; Rozos, E.; Makropoulos, C. Designing water demand management schemes using a socio-technical modelling approach. *Science of the Total Environment* **2018**, *622*, 1590-1602.

124. Hou, B.; Yang, R.; Zhan, X.; Tian, W.; Li, B.; Xiao, W.; Wang, J.; Zhou, Y.; Zhao, Y. Conceptual framework and computational research of hierarchical residential household water demand. *Water (Switzerland)* **2018**, *10*, doi:10.3390/w10060696.
125. Rahim, M.S.; Nguyen, K.A.; Stewart, R.A.; Giurco, D.; Blumenstein, M. Predicting Household Water Consumption Events: Towards a Personalised Recommender System to Encourage Water-conscious Behaviour. In Proceedings of 2019 International Joint Conference on Neural Networks (IJCNN), 14-19 July 2019; pp. 1-8.
126. Mnih, V.; Kavukcuoglu, K.; Silver, D.; Rusu, A.A.; Veness, J.; Bellemare, M.G.; Graves, A.; Riedmiller, M.; Fidjeland, A.K.; Ostrovski, G. Human-level control through deep reinforcement learning. *Nature* **2015**, *518*, 529.
127. Chandola, V.; Banerjee, A.; Kumar, V. Anomaly detection: A survey. *ACM computing surveys (CSUR)* **2009**, *41*, 15.
128. Rahim, M.S.; Ahmed, T. An initial centroid selection method based on radial and angular coordinates for K-means algorithm. In Proceedings of 2017 20th International Conference of Computer and Information Technology (ICCIT); pp. 1-6.
129. Shirkhorshidi, A.S.; Aghabozorgi, S.; Wah, T.Y.; Herawan, T. Big data clustering: a review. In Proceedings of International conference on computational science and its applications; pp. 707-720.
130. Ng, R.T.; Han, J. CLARANS: A method for clustering objects for spatial data mining. *IEEE Transactions on Knowledge & Data Engineering* **2002**, 1003-1016.
131. Zhang, T.; Ramakrishnan, R.; Livny, M. BIRCH: an efficient data clustering method for very large databases. In Proceedings of ACM Sigmod Record; pp. 103-114.
132. Guha, S.; Rastogi, R.; Shim, K. Cure: an efficient clustering algorithm for large databases. *Information systems* **2001**, *26*, 35-58.
133. Rahim, M.S.; Nguyen, K.A.; Stewart, R.A.; Giurco, D.; Blumenstein, M. Predicting Household Water Consumption Events: Towards a Personalised Recommender System to Encourage Water-conscious Behaviour. 2019.
134. Kofinas, D.T.; Spyropoulou, A.; Laspidou, C.S. A methodology for synthetic household water consumption data generation. *Environmental Modelling & Software* **2018**, *100*, 48-66, doi:https://doi.org/10.1016/j.envsoft.2017.11.021.

4.2 Paper II (Block B)

Advanced Household Profiling Using Digital Water Meters

Md Shamsur Rahim[a]
[a]Centre for Artificial Intelligence, School of Computer Science
University of Technology Sydney, Sydney, Australia
e-mail: mdshamsur.rahim@student.uts.edu.au

Khoi Anh Nguyen[b], Rodney Anthony Stewart[b]
[b]School of Engineering and Built Environment
Griffith University, Gold Coast, Australia
e-mail: k.nguyen@griffith.edu.au; r.stewart@griffith.edu.au

Damien Giurco[*, c]
[c]Institute for Sustainable Futures
University of Technology Sydney, Sydney, Australia
e-mail: damien.giurco@uts.edu.au

Michael Blumenstein[d]
[d]Centre for Artificial Intelligence, School of Computer Science
University of Technology Sydney, Sydney, Australia
e-mail: michael.blumenstein@uts.edu.au

[*] Corresponding author

Advanced Household Profiling Using Digital Water Meters

ABSTRACT

Advanced householder profiling using digital water metering data analytics has been acknowledged as a core strategy for promoting water conservation because of its ability to provide near real-time feedback to customers and instil long-term conservation behaviours. Customer profiling based on household water consumption data collected through digital water meters helps to identify the water consumption patterns and habits of customers. This study employed advanced customer profiling techniques adapted from the machine learning research domain to analyse high-resolution data collected from residential digital water meters. Data analytics techniques were applied on already disaggregated end-use water consumption data (e.g., shower and taps) for creating in-depth customer profiling at various intervals (e.g., 15, 30, and 60 minutes). The developed user profiling approach has some learning functionality as it can ascertain and accommodate changing behaviours of residential customers. The developed advanced user profiling technique was shown to be beneficial since it identified residential customer behaviours that were previously unseen. Furthermore, the technique can identify and address novel changes in behaviours, which is an important feature for promoting and sustaining long-term water conservation behaviours. The research has implications for researchers in data analytics and water demand management, and also for practitioners and government policy advisors seeking to conserve valuable potable-water resources.

KEYWORDS

User profiling, digital water meter, water conservation, water consumption data, behaviour change, Recommender System.

1. INTRODUCTION

In recent years, ensuring water supply during periods of shortage caused by drought and avoiding low pressure during the hours of peak demand have been two of the challenges troubling many metropolitan water utilities (Nguyen et al., 2016). Rolling out Digital Water Meters (DWMs) can be considered a potential solution for overcoming these challenges because the data they collect can contribute to water conservation and peak demand management. In an experiment to measure the effects of feedback information on water consumption, researchers found that when households received more information with higher water consumption, they demonstrated higher water saving (Cespedes Restrepo and Morales-Pinzon, 2020). Another study reported that detailed water consumption feedback can contribute mean savings of 5.5% across 25 studies (Liu and Mukheibir, 2018). In addition, a short-term water demand forecasting model was proposed by Nguyen et al. (2016) based on the data collected from DWMs which stated significant impact on peak demand management. A healthy number of investigations have been undertaken to generate insights from DWM data. These studies can be divided into five categories: (1) water-use feedback; (2) water event categorisation; (3) water demand forecasting; (4) behaviour analysis; and (5) socioeconomic analysis (Rahim et al., 2020). However, further scope for improvement exists for the current situation through incorporating highly personalised systems, because there is a direct relationship between the level of personalisation and effects on water conservation. Personalisation can be achieved through comprehensive user profiling, which provides the opportunity to meet users needs and preferences (Eke et al., 2019).

A user or customer profile is defined as a "summary of the user's interest, characteristics, behaviours, and preferences," whereas user profiling is the "system of collecting, organizing and inferring the user profile information" (Eke et al., 2019). User personalisation through user profiling is now a widely adopted technique in various domains, such as artificial intelligence, data science and information science (Gauch et al., 2007). Recommender systems (RSs) can be considered one of the notable applications of user personalisation. Recommender systems are intelligent systems that recommend a list of items that are most likely of interest to the user (Burke, 2007). At present, RSs are deployed in various application domains as their application is no longer limited to selling more products or recommending news. The most common RS applications are employed in five domains: Entertainment, Content, E-commerce, Services, and Social (Montaner et al., 2003; Ricci et al., 2015). However, the water industry is yet to adopt RSs, although great potential exists for RSs to promote water-conscious behaviour (Rahim et al., 2019). Comprehensive user profiling based on water consumption data collected from DWMs can provide valuable information regarding users' water use patterns along with the ability to track any changes in water use behaviours. For this reason, in-depth user profiling is a must for the success of such RSs. Although a healthy number of user profiling techniques have been proposed in relevant studies (e.g., demand profiling, habit detection), they are still not suitable for RSs because of their limitations.

Many user profiling techniques have been adopted based on water consumption data collected from residential DWMs. For instance, hourly water consumption data have been used to build demand profiles, enabling water demand to be forecasted accurately. Gaussian mixture models (GMM's) based approach was proposed by McKenna et al. (2014) to represent demand patterns and then classify the demand patterns. In another approach, Padulano and Del Giudice (2018) introduced a two phases procedure: clustering and classification to detect water consumption patterns. In addition, the probability of a particular event occurring at a particular hour was used in water event categorisation (Nguyen et al., 2014). Furthermore, a habit detection algorithm (Cardell-Oliver, 2016) was proposed where the user profile was based on five constraints, which were expressed using five parameters. Moreover, behaviour analysis-related studies have mainly aimed to understand the behaviours and dynamics of consumers based on water consumption data. Although different user profiles are created to reach different goals, they have some common limitations. First, in many studies, weekends and weekdays are not considered separately for profiling, despite the consumption patterns varying. Second, all related studies have only considered hourly consumption; however, other frequency intervals such as 15 and 30 minutes may provide more insights. Third, almost all investigations have considered total consumption data rather than individual water event consumption data. Lastly, existing studies have been unable to accommodate recent changes in behaviours because total or average consumption is used for profiling. These limitations of relevant studies indicate the necessity for a new comprehensive user profiling approach to make effective recommendations for promoting water conscious behaviours.

In this study, we introduced a new comprehensive user profiling approach that overcomes the limitations of prior relevant studies. In numerous studies, weekday and weekend water consumption have not been considered separately. Hence, in our proposed approach we performed profiling separately for weekdays and weekends. In terms of the profiling interval, our hypothesis was that more frequent interval profiling provides a greater understanding of the behaviours and habits of customers. Therefore, along with hourly profiling, we created two other user profiling frames at 15- and 30-minute intervals. In addition, instead of profiling based on total water consumption data, the user profiles were created based on disaggregated end-use water consumption data (e.g. showers and taps). Furthermore, to accommodate recent changes in behaviours, profiles were created by giving higher priority to recent data. Finally, an algorithm was introduced that performs user profiling for each household. To enhance the user profiles further, other information that characterises behaviours such as shower duration, volume of water used in the shower, and washing machine usage was collected. After comparing customer profiling at various intervals (e.g., 15, 30, and 60 minutes), we concluded that more frequent intervals (i.e., 15 minutes) of profiling provided a better understanding of user behaviours that was previously unseen. Moreover, through tracking recent changes in behaviours, it became possible to identify and track habits and changes in behaviours.

The findings from our experiments implied that the proposed approach addresses two aspects of the promotion of water-conscious behaviours. First, it addresses water conservation through the collection, organisation, and inference of water usage and savings scopes (i.e., the average volume of water used in the shower, average shower duration, and number of times performing laundry). Second, the approach provides the opportunity to manage water demand by characterising the patterns in water consumption behaviours and habits by providing the likely time a particular event will occur at various intervals. Thus, this research introduces a comprehensive user profiling approach that accommodates recent changes in behaviours and has the potential for promoting water-conscious behaviours. The major contributions of this study are as follows:

- It has proposed a comprehensive user profiling approach that addresses the limitations (e.g. not considering disaggregated water consumption data, profiling at shorter intervals, tracking and reflecting recent changes in behaviour, profiling based on the type of day) of the most recent state-of-art studies.
- It has introduced an advanced profiling algorithm to create user profiles based on dis-aggregated water consumption data.
- It has identified the most suitable profiling interval among three profiling intervals.
- It has highlighted the key benefits from such profiling approach for consumers, utilities, and policy makers.

The remainder of the paper is organised as follows. Section 2 presents a critical analysis of the related works. Section 3 examines the methodology followed in this study. Section 4 discusses the findings of the study and finally, section 5 draws the conclusion of the paper.

2. RELATED WORKS

To the best of the authors' knowledge, no studies have been conducted on in-depth user profiling with the purpose of promoting water-conscious behaviours through an RS, as the concept of RS in the water industry is relatively new (Rahim et al., 2019). However, some studies have involved some sort of user profiling for solving other problems. These studies can be categorised into three: (1) behaviour analysis; (2) socioeconomic analysis; and (3) water end-use categorisation. Although the user profiles developed in these studies were used to solve specific problems, they have some limitations that must be overcome to develop comprehensive user profiling.

Numerous studies have been performed on behaviour analysis based on simple user profiling to understand the behaviours and dynamics of consumers from water metering data. These studies can be categorised into three: (1) habit detection and profiling; (2) demand profiling; and (3) customer segmentation. In case of habit detection and profiling, four types of pattern were identified by Cardell-Oliver (2013): continuous-flow days; exceptional peak-use days; programmed patterns with recurrent hours; and normal use patterns. By contrast, three profiles were identified by Cominola et al. (2016) after segmenting water consumers based on their eigen behaviours. In another study, a detailed breakdown of hourly water use by volume for different times (i.e., peak hour, day, and month) was performed by Cole and Stewart (2013) to provide an accurate estimation of indoor and outdoor consumption. In addition, a habit detection algorithm was introduced by Cardell-Oliver (2016) based on time series data from water meters; however, because of heuristics, it could not guarantee the detection of all habits. Demand profiling has been performed in many studies for predicting water demand. For instance, McKenna et al. (2014) utilised hourly consumption data for classifying demand patterns. Later, a demand profile was proposed (Gurung et al., 2015) based on diurnal patterns of efficiency-rated appliances for modelling water demand. Lastly, eight relevant usage profiles from water consumption data were identified using clustering and modelling techniques on water consumption data by Cheifetz et al. (2017). However, these studies were based on hourly aggregated water consumption data rather than disaggregated data. Only Nguyen et al. (2016) considered hourly disaggregated data. Customer segmentation or clustering techniques have been applied in a few studies as part of their behaviour analysis. For forecasting future behaviour and clustering consumption behaviour, a hybrid model was proposed by Leyli-Abadi et al. (2018). In another study, residents were divided into five clusters by family structure, job, or lifestyle using a fuzzy clustering algorithm based on water consumption data. In a recent study, customer segmentation based on eigen-behaviour analysis was adapted and used to identify three main water end-use profile clusters (i.e., showering, clothes washing, and irrigation) (Cominola et al., 2019). Furthermore, they observed the existence of time-of-use and intensity-of-use within each class. However, in this study, the authors did not consider recent changes in the behaviours.

In socioeconomic studies, consumers' socioeconomic and demographic factors have been studied to understand their effect on water conservation (Beal et al., 2011; Bich-Ngoc and Teller, 2018; Willis et al., 2013). Such studies have used consumers' socioeconomic and demographic data to identify the determinants of water consumption (Beal et al., 2011; Bich-Ngoc and Teller, 2018; Willis et al., 2013). Although these studies have helped to determine the factors behind water consumption, further information is required for users' profiles to promote water conservation.

In the case of water end-use categorisation, the use of user profiles is very rare. The probability of time-of-use for each event over a 24-h period based on total collected data was used by (Nguyen et al., 2014) to categorise water end-use. However, recent changes in behaviours were also omitted from this study.

After examining relevant studies, we identified the following research gaps that need to be overcome to create comprehensive user profiling:

- Consideration of disaggregated water consumption data: Disaggregated water consumption data provide more information about water consumption data (i.e., for which purpose the water is used, such as showering or irrigation). Because existing studies have mostly considered total consumption data instead of disaggregated data, they have missed the opportunity to utilise disaggregated water consumption data to generate useful insights. Therefore, future studies on profiling should consider disaggregated consumption data.
- Profiling at reduced intervals: Existing profiles are mostly on an hourly interval. However, profiling at reduced intervals may provide more specific information and a greater understanding of consumers' habits and behaviours. Thus, future studies should consider user profiling frames at a reduced interval (e.g., 15 or 30 minutes).
- Tracking and reflecting recent changes in behaviour: Studies based on profiling have been unable to reflect recent changes in behaviour in users' profiles. However, if changes go unnoticed and are not considered in the profiles, then the profiles will become less effective. Hence, the approach must be adapted to track and reflect recent changes in water consumption behaviour.
- Profiling based on the type of day: Water consumption patterns vary depending on the type of day (i.e., weekday or weekend). Therefore, user profiling should be performed based on the type of day to improve the effectiveness.

3. MATERIAL AND METHODS

The proposed approach in this study for creating comprehensive user profiling can be divided into three consecutive steps: (1) data set collection and preparation; (2) profile creation; and (3) feature extraction. In the first step, a raw data set was collected from 306 single stand-alone households in Melbourne, Australia, and in the second step, a profiling algorithm was incorporated to create profiles. Finally, in the last step, statistical modelling was adopted for extracting features to enhance the profiles. Figure 1 represents the workflow of the study. The following subsections discuss the steps in details.

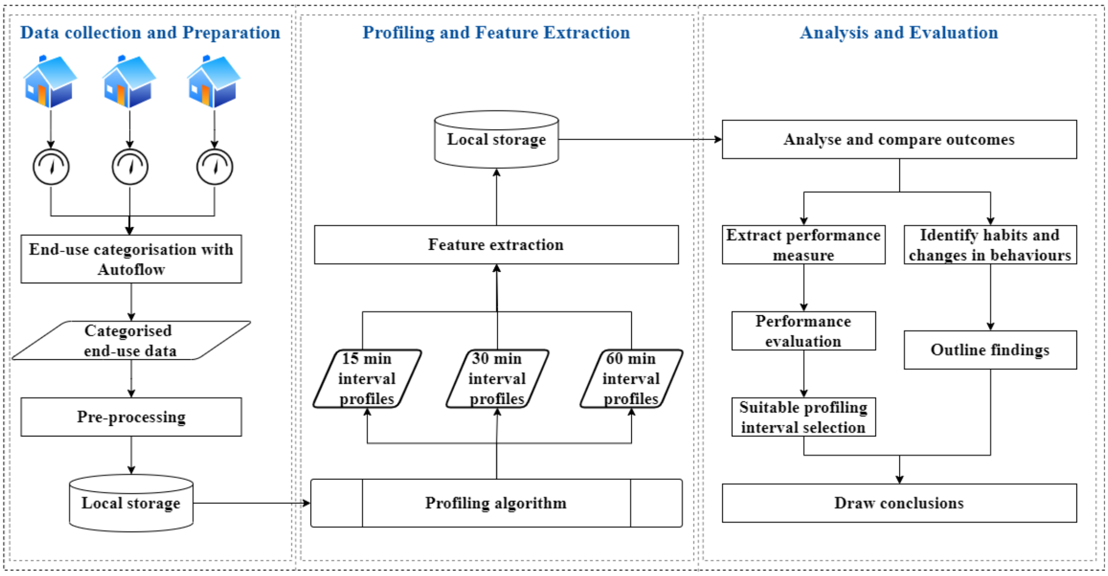


Figure 1. Workflow of the study.

3.1. Data collection and preparation

The raw data for this study originated from 306 single stand-alone households in Melbourne, Australia. The data set consists of recordings of high-resolution data at 5-second intervals that were collected for 10 months (February–December 2010). The collected raw data were then analysed with Autoflow (Nguyen et al., 2014), an intelligent metering system. Nine types of water end-use were classified by Autoflow with 90% accuracy. Autoflow provided the classifications of water end-use events: taps, dishwashers, leaks, evaporating coolers, washing machines, showers, toilets, irrigation, and bathtubs. These formed the primary data set for this research. Figure S1 depicts the steps in the primary data collection process.

After collecting the primary data set, the next step was to perform the initial pre-processing. For profiling, we considered only those events that occurred between 06h00 and 23h59 because the number of events and volume consumed from 00h00 to 05h59 are negligible in comparison. After this step, the pre-processed data were stored in a database for further processing. Table 1 presents a summary of the data set.

Table 1. Summary of the data set after initial pre-processing

Item	Description
Number of households	306
Data collection duration	11-02-2010 to 11-10-2010 (dd-mm-year)
Hours considered	06h00-23h59
Data collection interval	5 seconds
Number of end-uses	9
Number of events	56 310 576
Total shower events	181 358
Total bathtub events	18 274
Total washing machine events	30 996
Total dishwasher events	34 524
Total irrigation events	9 975
Total toilet events	1 173 902
Total tap events	4 152 386
Total leak events	50 709 161

From Table 1, it is obvious that the number of leak events dominates all other events because one long continuous leak event is a set of many tiny discrete events recorded every 5 seconds. Similarly, the number of tap events is the second highest but in each event on average 0.87 litres of water were used. Because these two events occurred frequently but an insignificant amount of water was used, these two events were not considered for profiling to reduce computational time and maintain scalability. In addition, toilet (flush) and evaporating cooler events were excluded from profiling because toilet events are spontaneous and evaporating coolers are responsible for less than 1% of total water consumption. For these reasons, we considered high water consumption events (shower, bathtub, irrigation, washing machine, and dishwasher) (Rahim et al., 2019), because when combined, these events account for nearly 70% of the total water consumption (Stewart et al., 2010).

Many studies have overlooked considering water consumption patterns on weekends and weekdays separately. However, this information is particularly crucial because water consumption patterns may differ depending on the type of day. Hence, as the last step of pre-processing, we extracted the type of day (i.e., weekend or weekday) from the given date. Figure S2 illustrates the consumption patterns of 200 sample households for 15 days, which clearly depict the difference in consumption patterns, especially for showers. During weekdays, 8:00–8:30 am is the peak time for taking showers. Conversely, on weekends, 11:00 am is the peak time for taking showers.

After the pre-processing step, the final data set consisted of 12 attributes. Table 2 presents a description of these attributes.

Table 2. Description of the attributes in the final data set

Attribute	Description	Example
Site	Unique identifier for each household in the data set.	Site001, Site002.
Start date	Start date of an event	11-Feb-2010
Start time	Start time of an event	08:38:15
End date	End date of an event	11-Feb-2010
End time	End time of an event	08:42:10
Category	Water end-use category	Shower, dishwasher
Duration	Amount of time an event took place.	0:03:55
Volume	Quantity (in litre) of water used in an event	36.81
Max flow	Maximum flow rate (Litre per minute) recorded for an event	10.83
Mode flow	Mode of flow rate (Litre per minute) recorded for an event	9.33
Cyclic event	Multiple intakes of water during one single event	C1 (Cycle 1 for washing machine), D5 (Cycle 5 for dishwasher)
Type of day	If the day is Saturday or Sunday then weekend, otherwise weekdays.	Weekend, weekday

The data set used in this study has some limitations. First of all, it does not have any socio-demographic data. Second, it does not have any weather data. Though the inclusion of these data would be interesting, however, we believe the absence of these data will not undermine the current profiling study.

3.2. Profile creation

Profile creation is the most important step in the proposed approach which is based on a profiling algorithm. The proposed algorithm is designed in such a way that it can address the limitations of existing studies. The proposed algorithm consists of three steps. First, a data structure is initialised. Next, the probability distribution (PD) is computed using probability mass function (PMF). Finally, profiles are created by concatenating multiple vectors created in the previous steps. Algorithm 1 describes the algorithm with a detailed explanation for each step.

Algorithm 1. Consumption profile creation algorithm

ConsumptionProfile(s, sd, ed, $intrvl$, td,w):creates water consumption profile for household s, based on the data where the start date is sd and end date is ed, $intrvl$ is the value of interval for profiling, td is the type of the day (i.e, weekday, weekend) and w is weight for different periods of time.

Here:
$s \in \boldsymbol{S}$, where **S**= $\{Site001, Site002 \ldots Site306\}$
$intrvl \in \boldsymbol{I}$, where **I**= $\{15\ minutes, 30\ minutes, 60\ minutes\}$
$e \in \boldsymbol{E}$, where **E**= $\{Shower, Bathtub, Clotheswasher, Dishwasher, Irrigation\}$
$t \in \boldsymbol{T}$, where **T** is the set of times at *intrvl* from 06h00 to 23h59.

The major steps of the algorithm are as follows:
1. Initialise data structure
2. Probability Distribution computation
3. Profile creation

We now examine the steps in detail.

1. Initialise data structure.
   To calculate and store the distribution of an event e over a given time t, for household s, an n-dimensional vector $v_{et}{}^{1} = (v_{et_1},\ v_{et_2}, \ldots, v_{et_{n-1}}, v_{et_n})$is taken and it is initialised to 0. Note that, the value of dimension depends on the value of *intrvl*. For instance, when the value of *intrvl*= 30 minutes, then dimension will be total hours from 06h00 to 23h59 $\times 2 = 18 \times 2 = 36$.

   (a) Set $v_{et} \leftarrow 0$.

2. Probability Distribution computation
   To understand the behaviour or habit of the users, we have employed a probability mass function (PMF). PMF can be defined as a function that provides the probability that a discrete random variable is exactly equal to some value. Formally this function p: $\boldsymbol{R} \rightarrow [0,1]$ can be defined as:

   $$p_X(x_i) = P(X = x_i)\ for - \infty < x < \infty \quad (1)$$

   And properties of PMF are:

$$\sum p_X(x_i) = 1 \quad (2)$$
$$p_X(x_i) > 0 \quad (3)$$
$$p_X(x_i) = 0\ for\ all\ other\ x \quad (4)$$

   Equation 2 describes the sum of the probabilities associated with each possible value will be always up to 1. According to equation 3, the values of probabilities must be positive. And for other values, the probability will be 0.

   This step will compute probability distribution (PD) for each event across time t with interval $intrvl$for the given type of the day td and assign it to v_{et}
   (a) for each e in E:
      a. compute $p_X(x_i)$

b. $v_{et} \leftarrow p_X(x_i)$

3. Profile creation
At this step, the profile is created by constructing two more PDs by repeating step 1 &2 and assigning weight w_i and merging all the PDs. The reason for creating the PDs is separating most recent, second most recent and previous historical data. In this way, it would be possible to assign a higher weight to most recent patterns as the most recent data will always capture better changes in behaviours. Finally, each weighted vector for each event is concatenated to a matrix M.
(a) ${v_{et}}^2 \leftarrow p_X(x_i)$ and ${v_{et}}^3 \leftarrow p_X(x_i)$
(b) $v_{et}^w \leftarrow (v_{et}^1 \times w_1) + (v_{et}^2 \times w_2) + (v_{et}^3 \times w_3)$ where $w_1 > w_2 \geq w_3$
(c) $M \leftarrow (v_{e_1t}^w \dots v_{e_rt}^w)$

In step one, for each event a vector is initialised with n-dimension. As we have considered five large water consumption categories (e.g., shower, bathtub, washing machine, dishwasher and irrigation), five n-dimensional vectors will be initialised. The dimension of the vectors depends on the interval value. For 15 minutes, there will be 72; for 30 minutes 36; and for 60 minutes interval there will be 18 dimensions. The reason for choosing a vector as the data structure is to measure the Cosine similarity.

Cosine similarity is a measure of similarity between two non-zero vectors to determine how similar the vectors are by calculating the cosine of the angle between them. Mathematically, the cosine of two vectors that are non-zero can be described by following equations.

$$A \cdot B = \|A\|\|B\| cos\theta \tag{5}$$

Given two vectors, A and B, the cosine similarity, $\cos(\theta)$ is measured using a dot product and magnitude as follows:

$$similarity = \cos(\theta) = \frac{A \cdot B}{\|A\|\|B\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} {A_i}^2}\sqrt{\sum_{i=1}^{n} {B_i}^2}} \tag{6}$$

Where A_i and B_i are components of vector A and B, and

$$A \cdot B = \sum_{i=1}^{n} A_i B_i = A_1 B_1 + A_2 B_2 + \cdots + A_n B_n \tag{7}$$

The derived value of similarity ranges from $[-1, 1]$ where -1 completely opposite and to 1 means completely similar. In this scenario, A and B vectors hold the PD for a particular event and the resulting similarity would provide an indication of changes in consumption behaviours based on time.

To capture and understand the timing of water consumption events, a probability mass function (PMF) is used in step two. A PMF is a function that provides the probability that a discrete random variable is equal to some value (Stewart, 2009). The computed probability based on most recent data of each event e across time t at interval $intrvl$ for type of the day td is assigned to designated vector v_{et}.

In the last step, at first, two more vectors for each event are created using PMF based on previous water consumption data. Later, these vectors are combined into one vector for each event by giving more weights to the vector holds the most recent data. Therefore, after assigning a higher weight to the first PD vector and lower weights to the other PD vectors, that is w_1 to v_{et}^1, w_2 to v_{et}^2, and w_3 to v_{et}^3, where $w_1 > w_2 \geq w_3$, a matrix M is constructed by concatenating the vectors for each household.

$$M = \begin{bmatrix} \left((w_1 \times v_{e_1t_1}^1) + (w_2 \times v_{e_1t_1}^2) + (w_3 \times v_{e_1t_1}^3)\right) & \cdots & \left((w_1 \times v_{e_1t_n}^1) + (w_2 \times v_{e_1t_n}^2) + (w_3 \times v_{e_1t_n}^3)\right) \\ \vdots & \ddots & \vdots \\ \left((w_1 \times v_{e_mt_1}^1) + (w_2 \times v_{e_mt_1}^2) + (w_3 \times v_{e_mt_1}^3)\right) & \cdots & \left((w_1 \times v_{e_mt_n}^1) + (w_2 \times v_{e_mt_n}^2) + (w_3 \times v_{e_mt_n}^3)\right) \end{bmatrix} \quad (8)$$

Each row in the matrix represents the PD of one event, whereas each column presents a time at an interval between 6 am and 12 am. Figure 2 describes the steps of the profile creation algorithm.

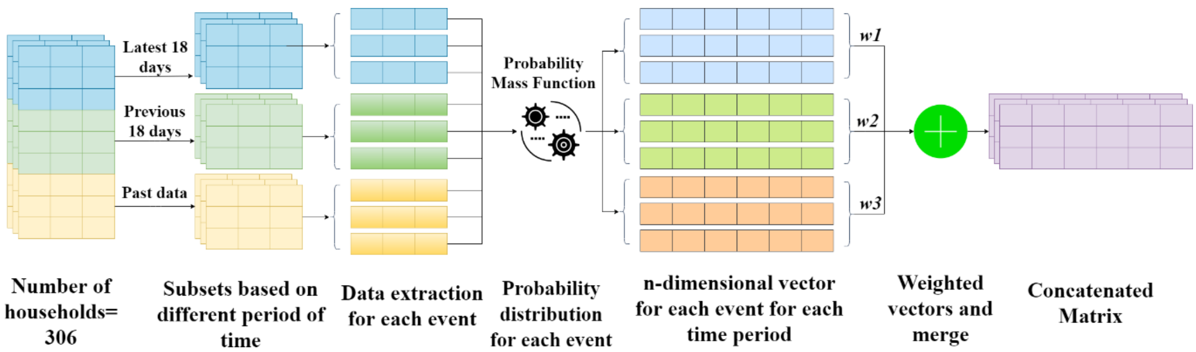


Figure 2. Illustration of the time of use and weighted probability of the use algorithm.

Such a matrix can be used to make recommendations to change a particular water consumption behaviour from a particular time. For instance, if a household seems to always perform irrigation at a particular hour of peak water demand (e.g., 8 am or 5 pm), then recommendations can be made to gradually shift this event outside of the hours of peak demand to avoid low water pressure.

3.3. Feature extraction

A customer profile with a PD that provides a possible time for an event to occur is not sufficient for promoting water-conscious behaviours. This is because more information is required. Therefore, to enhance the profiles and create in-depth user profiles, features must be extracted from the data. Depending on the goal of the water utility and items to recommend, the features to extract may vary. Table 3 lists the extracted features in this study from different events, which can be utilised to develop a personalised recommender system.

To extract these features, we mostly employed statistical methods. For instance, to derive the mean consumed volume of water for showers, bathtubs, washing machines, dishwashers, or toilets, we applied the arithmetic mean using the following equation (9):

$$\overline{x} = \frac{1}{n}\sum x_i = \frac{1}{n}(x_1 + x_2 + \cdots x_n) \quad (9)$$

where $x_1, x_2, \cdots, x_n$ are the volume of each event.

Table 3. Extracted features list from each event

Event	Feature
Shower	Mean consumed volume, mean duration, mode of flow rate, change in volume compared with the previous week, change in duration compared with the previous week, and change in flow rate compared with the previous week
Bathtub	Mean consumed volume, change in volume compared with the previous week
Washing machine	Mean consumed volume, number of loads per week
Dishwasher	Mean consumed volume, number of loads per week
Toilet	Mean consumed volume
Leak	Total volume in the last week
Tap	Mode flow rate

4. RESULTS AND DISCUSSION

4.1 Personalised recommendation system

The extracted features of each end-use category performed in the previous step are the main resource for developing a personalised RS for each household. On the one hand, the mode flow rate of shower events can be used to identify the efficiency of the shower, and recommendations to replace the showerhead can be made accordingly. Similarly, the consumed volume in toilet flushes can be used to infer whether the flush system is a dual flush system because typical single flush systems consume 11 litres per flush in Australia (Business Amenities Fact Sheet, 2009). Replacing the typical system with a 4.5/3 litre dual flush system can save approximately 11,000 litres/year/person, assuming the person uses the toilet four times a day (Business Amenities Fact Sheet, 2009). On the other hand, changes in flow rate or volume or duration can be used to infer the effectiveness of recommendations. For instance, if shower duration reduces compared with the previous week/month after a recommendation is made to reduce shower time, this would indicate that the household is responding to the recommendations. Furthermore, after making a recommendation to replace a showerhead with a water-saving one, any reduction in flow rate or volume compared with the previous week/month would imply the effectiveness of recommendations. Figure S3 is an example of the significance of the water savings that could be achieved from such a recommendation.

4.2 Effective water demand management through behaviour change and flow theory

In positive psychology, flow is defined as a subjective state where people become so completely involved in something that they forget time, fatigue, and everything else except the activity (Csikszentmihalyi et al., 2014). The flow theory of behaviour-changing tasks (Yürüten, 2017) states that if the difficulty of a task matches the capabilities of the person, then his or her engagement will be maximised. This theory can be translated into the RS field as the balance between the user's ability to perform behaviour-changing tasks and the difficulty to perform the recommended tasks. For instance, in our case, let us consider a consumer who performs irrigation or uses a bathtub at a specific time during the hours of peak demand. To shift the event time from a peak to a nonpeak hour of demand, recommendations can be made to shift the event time by 15 or 30 minutes gradually. Suggesting that behaviours be shifted by 1 hour might be difficult for many consumers, and in such cases, the users would lose interest in the system.

In this study, we created profiles at three different intervals (i.e., 15, 30, and 60 minutes) using the proposed algorithm, and performed comparisons among the profiles to identify the most suitable one. Furthermore, we extracted some features from the consumption data to enhance the user profiles, and we noted some very interesting findings.

We compared the profiling at three different intervals and observed that profiling at a lower interval (15 minutes) provided a greater understanding of the behaviour of consumers. Figure S4 a, b, and c represent the total number of each event during weekdays at 60, 30, and 15 minutes, respectively, over a period of 10 months at Site 325. If we consider that the morning peak hour of water demand (Cole and Stewart, 2013) is from 7 AM to 9 AM, then we can see that nearly 50% of shower events occurred from 7:00 AM to 7:29 AM for that particular household. It was only possible to observe this pattern as we performed profiling at 15-minute intervals. Therefore, the profiling interval of 15 minutes provided a greater understanding of water consumption patterns. The implication of this finding is that it can lead to potential effective demand management through changing the behaviour of users. This is achieved by shifting specific water consumption events from a specific time by incorporating flow theory into behaviour change.

4.3 Enhanced demand management thorough understanding customer behaviour changes

We found that profiling based on the PD of different water consumption events could be used to identify, understand, and notice changes in the habit and behaviour of users. To demonstrate this statement, we calculated the PD for three different periods of time. First, we calculated the PD for the latest 18 days, denoted as P1, and then the previous 18 days, denoted as P2, and lastly all other previous days, denoted as P3. An unchanged PD for a particular event in each period would indicate a strong habit, and any change in PD in P1 would indicate recent changes in behaviour compared with previous patterns. Figure 3 depicts the PD of shower events during weekdays across a different period of time for Site 325.

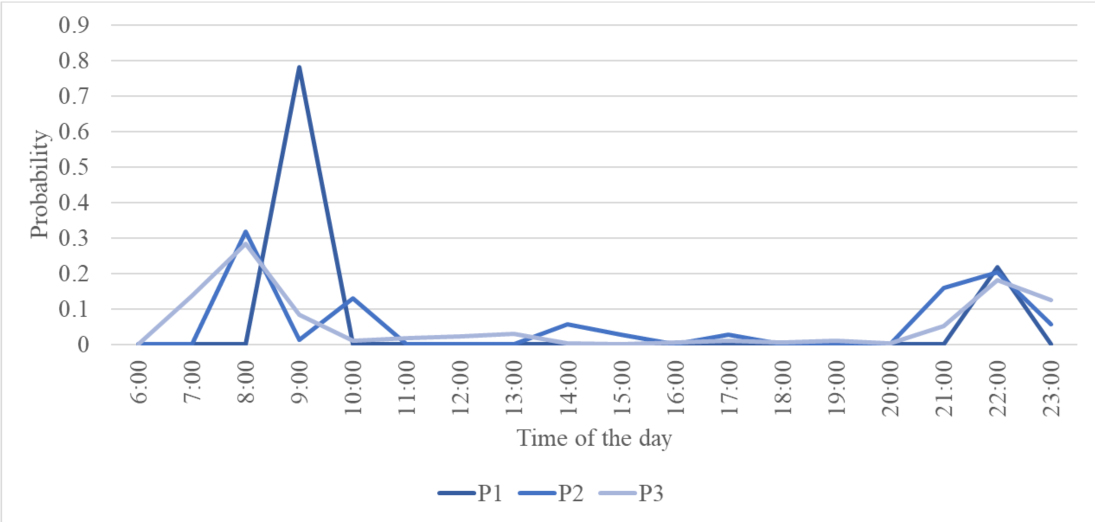


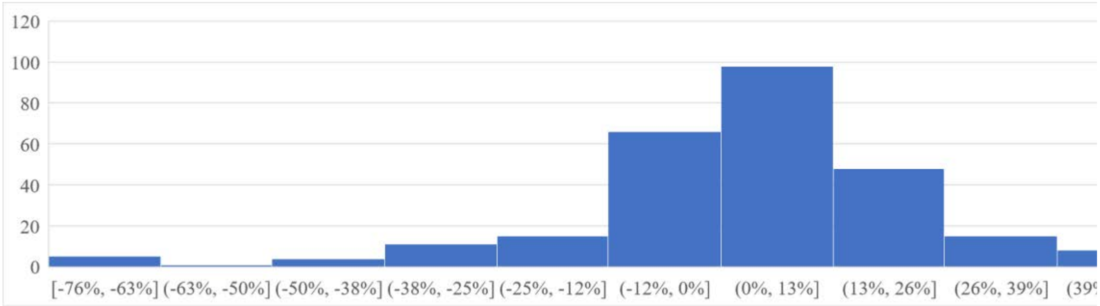
Figure 3. Comparing the probability distribution of shower events during weekdays across different periods of time at Site 325; P1 refers to the recent 18 days, P2 refers to the previous 18 days; and P3 represents all previous days' data.

From Figure 3, we can see that the probability of taking showers at 8 AM was 30% in P2 and P3. However, in the latest 18 days (P1), the shower time changed to 9 AM with the probability of 80%. This clearly indicated the change in shower time compared with previous behaviour patterns. By contrast, the probability of taking a shower at 10 PM almost remained the same, which indicated the strong habit of taking a shower at that time. Similar patterns can be found for other activities using the proposed profiling approach. The 18-day interval was chosen based on a study that reported it took 18–254 days to form a habit (Lally et al., 2010).

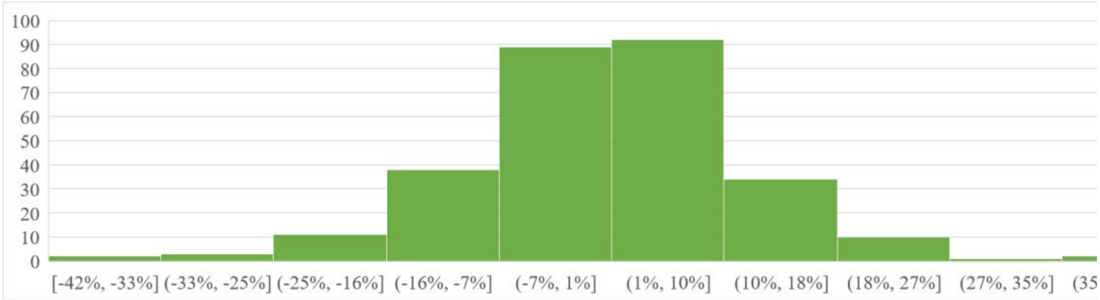
The main implication of this finding is that improved demand management can be ensured through identifying households with strong and flexible behaviour patterns. The weighted PD over three different periods of time will help to identify consumers with flexible demand patterns. For example, if a household exhibits fluctuations in consumption patterns for a particular event that does not have a fixed time for any water consumption activity, then the recommendation system could be developed to target that household because it is the most flexible for behaviour change.

4.4 Opportunities for system enhancement

We extracted a healthy number of interesting features from the data set to enhance profiles further, and concluded that these features provide in-depth analysis and previously unseen insights. For instance, after extracting the feature of weekly change in shower water volume, we could identify those households with a percent increase or decrease in water volume. This insight can be useful for tracking changes in behaviours and recommending relevant and effective activities. Figure 4 a, b represents the histogram of the number of households with a change in shower consumption volume compared with the previous week on weekdays and on weekends respectively.



(A)



(b)

Figure 4. Histogram of the number of households with a change in shower consumption volume compared with the previous week: (a) shower volume change on weekdays; (b) shower volume change on weekends.

For instance, let us assume a household consumes approximately 70 litres of water per shower event over a week before receiving any recommendations. Based on their consumption pattern, a recommendation can be made to reduce their shower water consumption to 60 litres. Then, using the extracted features, we can determine how effective the recommendation is the next week. If the water consumed for showers drops in the following week, this would mean that the recommendation is working properly and has been accepted by the household. However, if no change occurs in the consumption or it increases, this would mean that the recommendation is not effective for that household and further calibrations can be made (i.e., reduce shower water consumption to 65 litres). In this way, it would be possible to obtain implicit preferences of consumers instead of explicit preferences (i.e., through rating or like–dislike).

4.5 Summary of key benefits

The profiling approach introduced in this study overcomes the limitations of existing studies. The anticipated benefits from such profiling are discussed here.

i. ***Identification of habits/behaviour patterns with more detail and accuracy*:** The proposed profiling approach can be helpful for identifying habits/behaviour patterns of consumers in more detail and with enhanced accuracy. On the one hand, the identification of such habits or behaviours depending on the type of day (i.e., weekday or weekend) would encourage consumers to become more aware of or educated about water-conscious behaviours. On the other hand, such insights would provide an enhanced understanding of consumers to the utilities and policy makers, which can be used to determine more effective water conservation programs, campaigns, and education.

ii. ***Tracking of changes in behaviour:*** The proposed profiling approach has the ability to track changes in water consumption behaviour of customers. This would empower them to take control of their water consumption and help determine their progress towards water conservation. In addition, alerting consumers to any deviations from their usual consumption patterns would help them stay on track for sustainable water consumption. For water utilities, tracking any changes in the behaviour of customers would help them understand the various factors that influence water consumption (i.e., type of day, temperature, recommendations for water conservation, education, programs, and campaigns) at a more detailed level.

iii. ***Improvement of demand profiling:*** Demand profiling is critical for utilities and policymakers to understand water consumption patterns, enhance peak water demand management, and reduce water pumping costs as well as greenhouse gas (GHG) emissions. The profiling approach introduced in this study can play a vital role in improving demand profiling by predicting the time and probability of future events. Based on the improved demand profiling and enhanced understanding of water consumption enabled by the proposed approach, utilities and policymakers can introduce flexible tariff plans. Current state-of-the-art water end-use classification systems can classify water end-use at an accuracy rate of 95% (Nguyen et al., 2020; Yang et al., 2018). To improve the accuracy of such systems further, the profiling approach proposed in this study can be adopted. Further accuracy in water end-use would increase customers' acceptance of and trust in such systems, utilities, and policymakers.

iv. ***Grouping of households with similar consumption patterns:*** Households with similar consumption patterns can be grouped at a more detailed level based on the proposed profiling approach. This will help consumers compare their consumption patterns with other customers and increase awareness of water conservation. Grouping households will also help utilities and policymakers identify unique numbers of groups and their characteristics. Based on such insights, targeted water conservation programs and campaigns can be designed to promote water conservation further.

v. ***Foundation for a Recommender System:*** User profiling is an integral part of Recommender Systems. The profiling approach introduced in this study can be considered the foundation for a RS in the water industry for promoting water-conscious behaviours. Such a RS can be highly beneficial for consumers, utilities, and policy makers. Through the RS, consumers would be able to interact directly with water utilities through explicit and implicit preferences. This would result in improvements in customer services/satisfaction. For utilities and policymakers, the recommender system would provide the opportunity to obtain consumer feedback regarding different recommendations and policies in a cost-effective and timely, efficient manner.

Table 4 summarizes the anticipated benefits from the proposed profiling approach for different beneficiaries.

Table 4. Anticipated benefits from the proposed profiling approach for different beneficiaries

Benefit	Beneficiary		
	Customer	Utilities	Policy makers
Identification of habits/behaviour patterns	✓	✓	✓
Tracking of changes in behaviour	✓	✓	
Enhancement of peak water demand management		✓	✓
Reduction in water pumping costs (GHG emissions)		✓	✓
Understanding of water consumption patterns	✓	✓	✓
Prediction of time and probability of future events		✓	
Improvement in customer service/satisfaction	✓	✓	
Flexible tariffs	✓	✓	✓
Increased accuracy of water end-use classification	✓	✓	✓
Grouping of similar households	✓	✓	✓
Basis for a recommender system	✓	✓	✓

4.6 Challenges

In this study, we observed some challenges in user profiling. First, handling vast amounts of high-resolution water consumption data at 5-second intervals was a challenge in terms of processing and analysis even for 306 single-standalone households. This indicates further challenges when DWMs are deployed widely. Second, we observed that the profiling outcome highly depends on the accuracy of water end-use classification because the technique uses water end-use data as the primary input. Therefore, any misclassification in water end-use would result in poor profiling results. Lastly, we observed that households with no water consumption data in recent days (empty households) may not have any meaningful PD for water consumption events. In such scenarios, profiling should be performed carefully and consistency should be maintained.

5. CONCLUSION AND FUTURE WORK

The user profiling technique has been widely adopted in various domains for providing personalised services. However, the water sector is yet to adopt advanced user profiling, which could be used to promote water-conscious behaviours through a recommender system. In this study, a user profiling approach based on water consumption data from residential DWMs was proposed along with a profiling algorithm. Furthermore, profiles were created at different intervals (e.g. 15, 30, and 60 minutes). Our findings suggested that profiling at 15-minute intervals provides better insights regarding the behaviour of consumers. It can also be used to identify habits and track changes in behaviours that are important for promoting and sustaining long-term water-conscious behaviours. We believe our findings will help researchers and practitioners in data analytics and water demand management as well as government policy advisors by providing valuable insights.

Currently, we are working towards a recommender system based on the proposed profiling technique for promoting water-conscious behaviours. As for future work, clustering consumers based on the proposed profiling approach can be performed as it would facilitate grouping consumers at a more detailed level. Furthermore, because the relational database management system used in this study took a long time to execute queries, a data warehouse could be proposed for storing and analysing complex data. A carefully designed data warehouse can handle many complexities in data (Ahmed et al., 2013). In addition, profiling that includes socioeconomic data can be performed to provide further insights. Such a profiling approach can be extended to other industries or resource consumption scenarios, such as profiling of electricity and gas consumption for residential and non-residential customers.

Funding
This research was funded by the Australian Research Council (ARC), grant number LP160100215, which also includes funding from the following industry partners: Yarra Valley Water, City West Water, Southeast Water, and Aquiba.

Declaration of competing interest
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

2009. Business Amenities Fact Sheet. Western Water.
Ahmed, T., Pedersen, T.B., Lu, H., 2013. A data warehouse solution for analyzing rfid-based baggage tracking data, 2013 IEEE 14th International Conference on Mobile Data Management. IEEE, pp. 283-292.
Beal, C., Stewart, R.A., Spinks, A., Fielding, K., 2011. Using smart meters to identify social and technological impacts on residential water consumption. Water Supply 11, 527-533.
Bich-Ngoc, N., Teller, J., 2018. A review of residential water consumption determinants, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Springer, Cham, Switzerland, pp. 685-696.
Burke, R., 2007. Hybrid Web Recommender Systems, The Adaptive Web. Springer, Berlin, Germany, pp. 377-408.
Cardell-Oliver, R., 2013. Water use signature patterns for analyzing household consumption using medium resolution meter data. Water Resources Research 49, 8589-8599.
Cardell-Oliver, R., 2016. A habit detection algorithm (Hda) for discovering recurrent patterns in smart meter time series, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Springer International Publishing, Cham, Switzerland, pp. 109-127.
Cespedes Restrepo, J.D., Morales-Pinzon, T., 2020. Effects of feedback information on the household consumption of water and electricity: A case study in Colombia. J Environ Manage 262, 110315.

Cheifetz, N., Noumir, Z., Samé, A., Sandraz, A.-C., Féliers, C., Heim, V., 2017. Modeling and clustering water demand patterns from real-world smart meter data. Drinking Water Engineering and Science 10, 75-82.
Cole, G., Stewart, R.A., 2013. Smart meter enabled disaggregation of urban peak water demand: precursor to effective urban water planning. Urban Water Journal 10, 174-194.
Cominola, A., Moro, A., Riva, L., Giuliani, M., Castelleti, A., 2016. Profiling residential water users' routines by eigenbehavior modelling.
Cominola, A., Nguyen, K., Giuliani, M., Stewart, R., Maier, H., Castelletti, A., 2019. Data mining to uncover heterogeneous water use behaviors from smart meter data. Water Resources Research.
Csikszentmihalyi, M., Abuhamdeh, S., Nakamura, J., 2014. Flow, Flow and the Foundations of Positive Psychology. Springer Netherlands, Dordrecht, pp. 227-238.
Eke, C.I., Norman, A.A., Shuib, L., Nweke, H.F., 2019. A survey of user profiling: state-of-the-art, challenges, and solutions. IEEE Access 7, 144907-144924.
Gauch, S., Speretta, M., Chandramouli, A., Micarelli, A., 2007. User profiles for personalized information access, The adaptive web. Springer, pp. 54-89.
Gurung, T.R., Stewart, R.A., Beal, C.D., Sharma, A.K., 2015. Smart meter enabled water end-use demand data: platform for the enhanced infrastructure planning of contemporary urban water supply networks. Journal of Cleaner Production 87, 642-654.
Lally, P., Van Jaarsveld, C.H., Potts, H.W., Wardle, J., 2010. How are habits formed: Modelling habit formation in the real world. European journal of social psychology 40, 998-1009.
Leyli-Abadi, M., Same, A., Oukhellou, L., Cheifetz, N., Mandel, P., Feliers, C., Chesneau, O., 2018. Mixture of Non-homogeneous Hidden Markov Models for Clustering and Prediction of Water Consumption Time Series, Proceedings of the International Joint Conference on Neural Networks.
Liu, A., Mukheibir, P., 2018. Digital metering feedback and changes in water consumption – A review. Resources, Conservation and Recycling 134, 136-148.
McKenna, S.A., Fusco, F., Eck, B.J., 2014. Water Demand Pattern Classification from Smart Meter Data. Procedia Engineering 70, 1121-1130.
Montaner, M., López, B., De La Rosa, J.L., 2003. A taxonomy of recommender agents on the internet. Artificial intelligence review 19, 285-330.
Nguyen, K., Stewart, R., Zhang, H., Giurco, D., Blumenstein, M., Rahim, S., 2020. Next Generation Machine Learning for Urban Water Management. Water e-Journal 5, 1-7.
Nguyen, K.A., Sahin, O., Stewart, R.A., Zhang, H., 2016. Water demand forecasting with AUTOFLOW© using State-Space approach, 8th International Congress on Environmental Modelling and Software.
Nguyen, K.A., Stewart, R.A., Zhang, H., 2014. An autonomous and intelligent expert system for residential water end-use classification. Expert Systems with Applications 41, 342-356.
Padulano, R., Del Giudice, G., 2018. A Mixed Strategy Based on Self-Organizing Map for Water Demand Pattern Profiling of Large-Size Smart Water Grid Data. Water Resources Management 32, 3671-3685.
Rahim, M.S., Nguyen, K.A., Stewart, R.A., Giurco, D., Blumenstein, M., 2019. Predicting Household Water Consumption Events: Towards a Personalised Recommender System to Encourage Water-conscious Behaviour, 2019 International Joint Conference on Neural Networks (IJCNN), pp. 1-8.
Rahim, M.S., Nguyen, K.A., Stewart, R.A., Giurco, D., Blumenstein, M., 2020. Machine Learning and Data Analytic Techniques in Digital Water Metering: A Review. Water 12.
Ricci, F., Rokach, L., Shapira, B., 2015. Recommender Systems: Introduction and Challenges, Recommender Systems Handbook. Springer US, Boston, USA, pp. 1-34.

Stewart, R.A., Willis, R., Giurco, D., Panuwatwanich, K., Capati, G., 2010. Web-based knowledge management system: linking smart metering to the future of urban water planning. Australian Planner 47, 66-74.
Stewart, W.J., 2009. Probability, Markov chains, queues, and simulation : the mathematical basis of performance modeling. Princeton University Press, Princeton, N.J.
Willis, R.M., Stewart, R.A., Giurco, D.P., Talebpour, M.R., Mousavinejad, A., 2013. End use water consumption in households: impact of socio-demographic factors and efficient devices. Journal of Cleaner Production 60, 107-115.
Yang, A., Zhang, H., Stewart, R., Nguyen, K., 2018. Enhancing Residential Water End Use Pattern Recognition Accuracy Using Self-Organizing Maps and K-Means Clustering Techniques: Autoflow v3.1. Water 10.
Yürüten, O., 2017. Recommender Systems for Healthy Behavior Change. Swiss Federal Institute of Technology Lausanne.

Appendix A. Supplementary materials

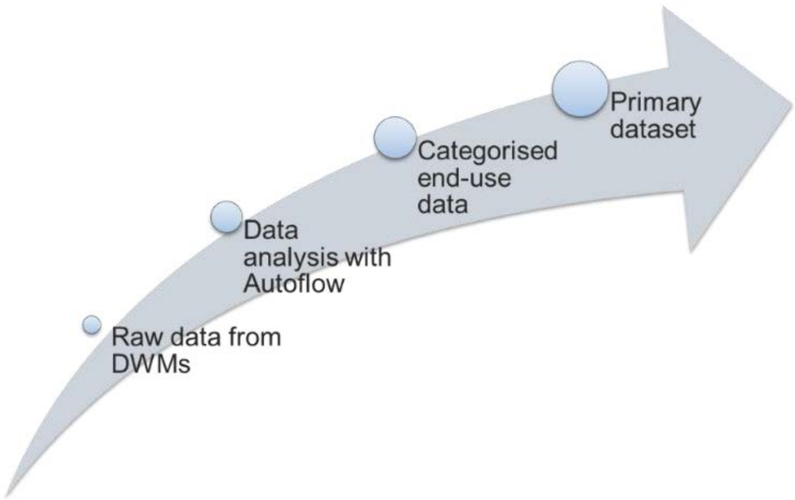
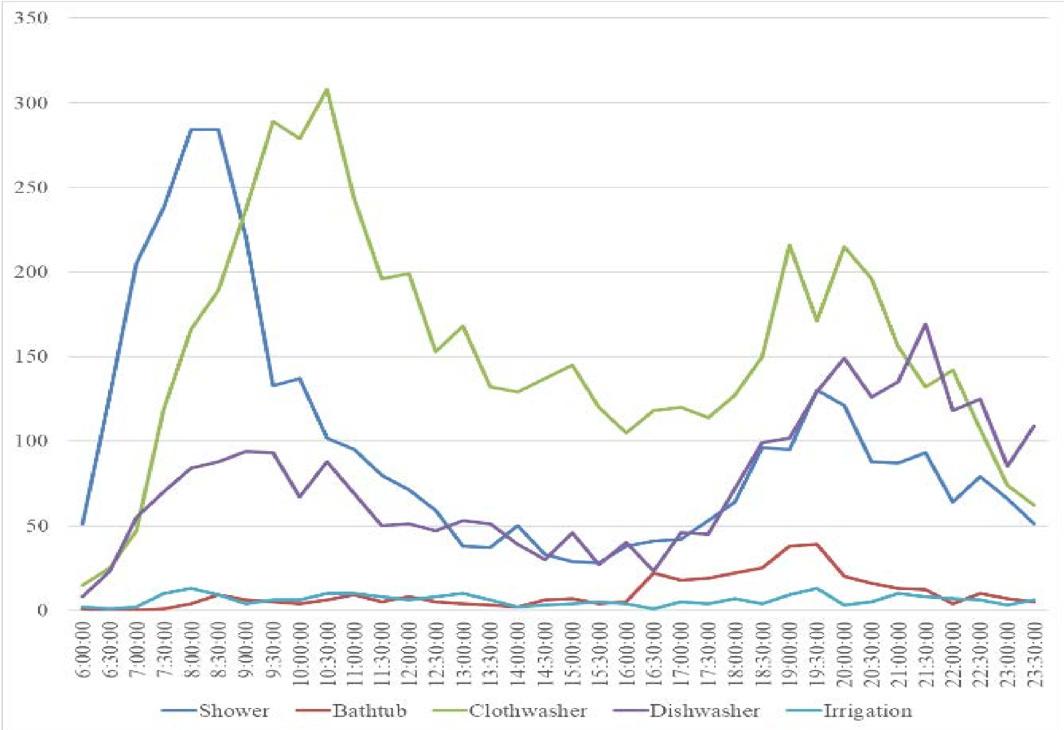
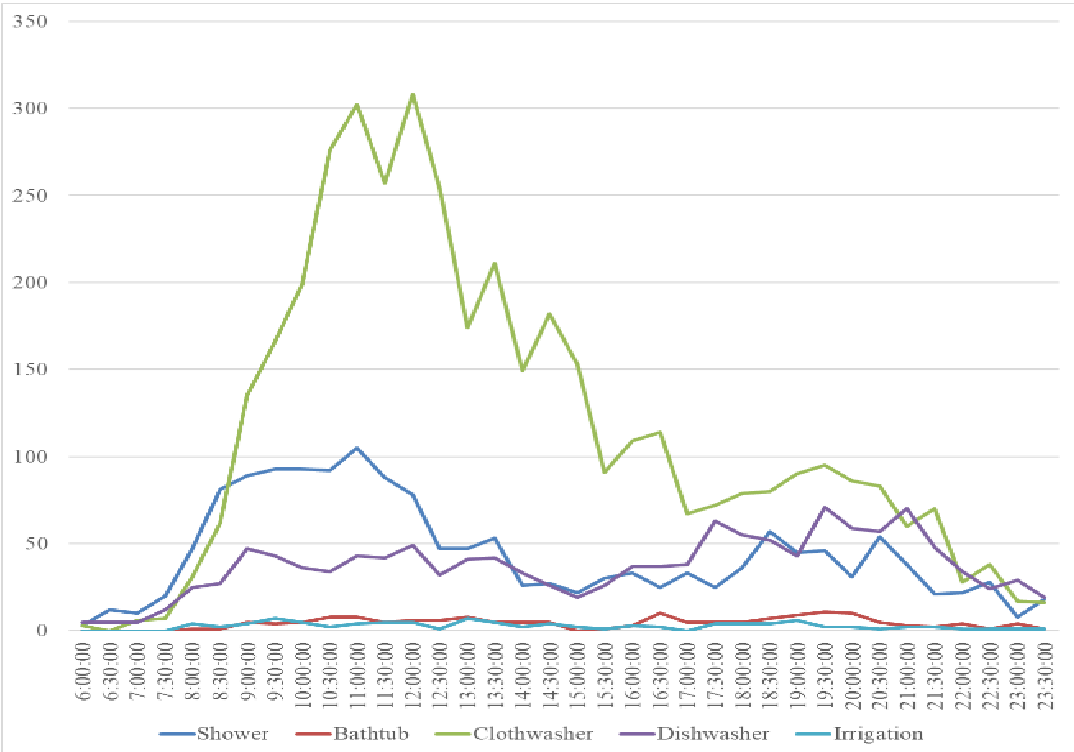


Figure S1. Primary data collection process



(a)



(b)

Figure S2. Count of water end-use events in 30 minutes interval during: a) weekdays; b) weekends.

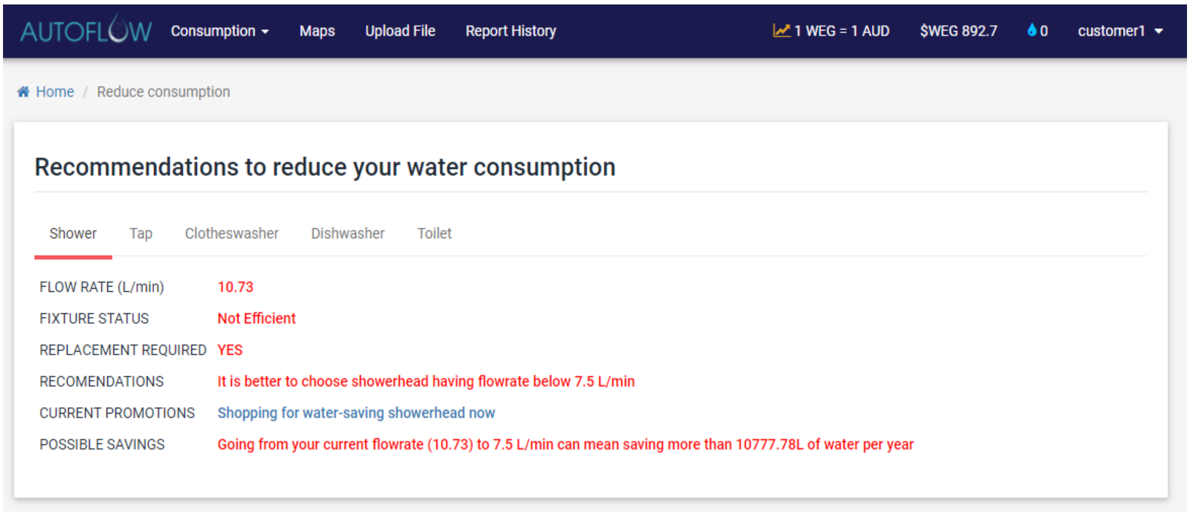
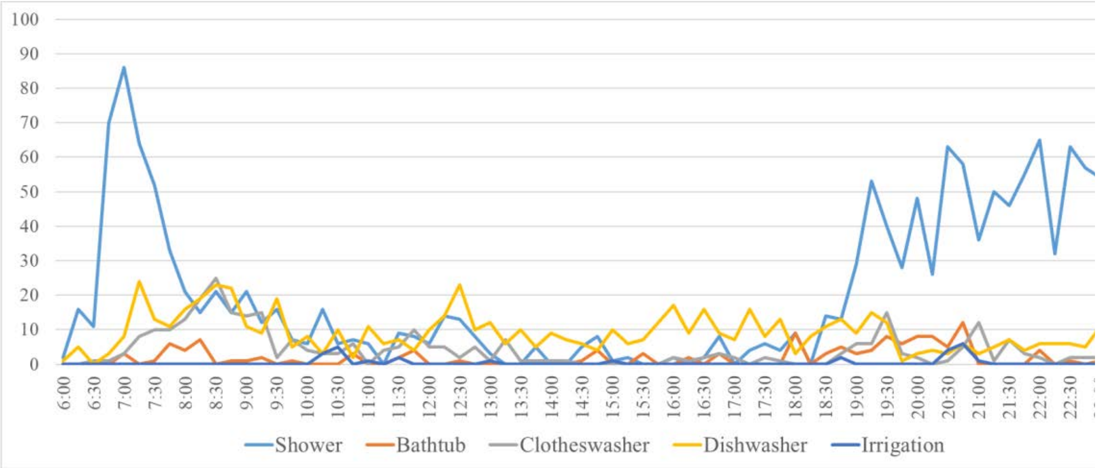
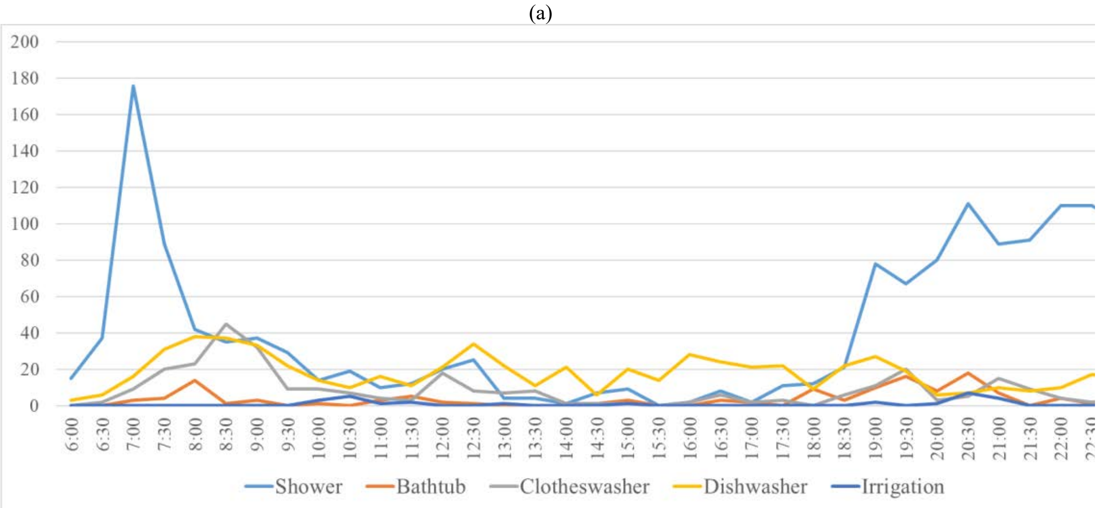


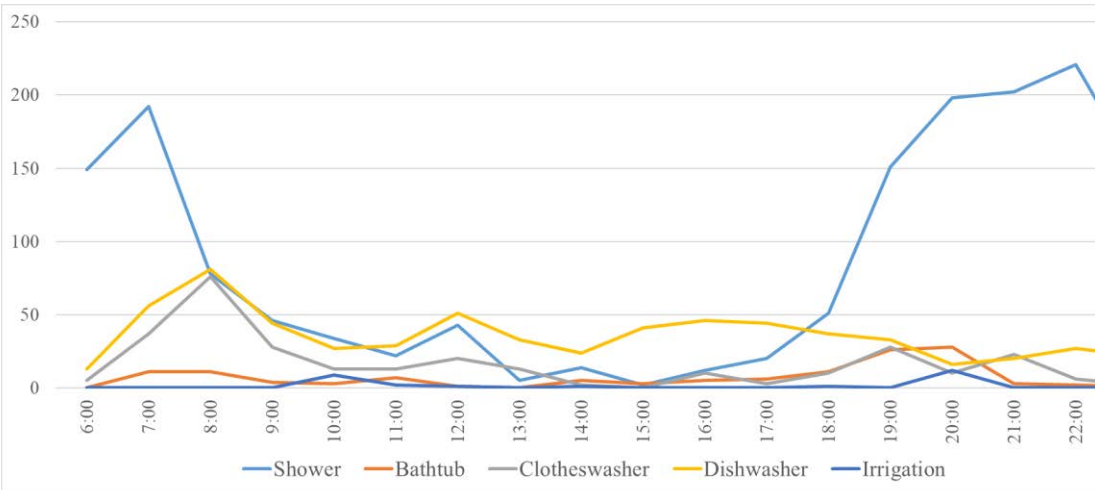
Figure S3. A personalised water recommender system.



(a)



(b)



(c)

Figure S4. Total number of each events at Site 325 at: (a) 60-minute intervals; (b) 30-minute intervals; and (c) 15-minute intervals.

4.3 Paper III (Block C)

A Clustering Solution for Analyzing Residential Water Consumption Patterns

Md Shamsur Rahim
Australian Artificial Intelligence Institute, School of Computer Science
University of Technology Sydney, Sydney, Australia
e–mail: mdshamsur.rahim@student.uts.edu.au

Khoi Anh Nguyen, Rodney Anthony Stewart
School of Engineering and Built Environment
Griffith University, Gold Coast, Australia
e–mail: k.nguyen@griffith.edu.au; r.stewart@griffith.edu.au

Tanvir Ahmed
Department of Computer Science
University of Central Florida, Orlando, Florida, United States of America
e-mail: tanvir.ahmed@ucf.edu

Damien Giurco
Institute for Sustainable Futures
University of Technology Sydney, Sydney, Australia
e–mail: damien.giurco@uts.edu.au

Michael Blumenstein*
Australian Artificial Intelligence Institute, Faculty of Engineering and IT
University of Technology Sydney, Sydney, Australia
e–mail: michael.blumenstein@uts.edu.au

A Clustering Solution for Analyzing Residential Water Consumption Patterns

Abstract: Water utility companies in urban areas face two major challenges: ensuring there is enough water for everyone during prolonged drought and maintaining adequate water pressure during the hours of peak demand. These issues can be overcome by applying data analytics and machine learning to the data gathered from digital water meters. For water conservation and demand management strategies to be effective, utility companies need to gain a better understanding of consumer behaviours, habits and routines. To accomplish this goal, we adapted a clustering approach to reveal residential water consumption patterns within metered data. In the experiment, we used two data sets (engineered features data set as well as the times of use and weighted probabilities of use data set) based on the data collected over 10 months from 306 households in Melbourne, Australia. For the engineered features data set, first, we identified the number of optimal clusters. We then performed extensive experiments to find the best clustering approach in terms of performance evaluation and clustering quality. We chose the hierarchical agglomerative clustering technique based on the nature of the data and the objective of the study. We observed that for the engineered features data set, k-means is the best performing clustering technique after considering performance metrics. For the other data set, we found that the number of clusters varies based on the type of water-consumption event, type of day (i.e., weekday or weekend), profiling interval and probability of use. In addition, we observed that insight into tap-water usage could be used to determine the population's adaptation of hygiene practices in an unprecedented time, such as the COVID-19 pandemic. Finally, we recommend that future clustering studies also employ aligned socio-demographic data and other key features.

Keywords: digital water meters; residential water consumption; clustering; customer segmentation; k-means clustering; hierarchical agglomerative clustering; consumption patterns; data analytics; machine learning.

1. Introduction

Water covers just over 70 per cent of the earth's surface. Still, a recent report identified water scarcity as one of the most significant global risks [1]. This is because *water scarcity* refers to the scarcity of potable water, which is fresh, safe, and easily accessible. Potable water accounts for only 0.014% of all water on earth [2]. In addition, unlike other natural resources, there is no alternative to water, and we cannot live without it. Therefore, any shortage of this resource can interrupt our lives completely. As an example of a community that suffered from a catastrophic drought, in 2017, South African officials coined the term Day Zero to represent the exact time in the future when municipal water would be turned off in Cape Town due to very low reservoir levels [3]. Fortunately, Cape Town escaped this fate by adopting dramatic water conservation, from 235 L per person per day (L/pp/dd) to 50 (L/pp/dd).

As in Cape Town, ensuring there is enough water for everyone during prolonged drought is one of the major challenges for water utility companies in urban areas. Managing demand during peak hours is another challenge for the urban utilities [4]. *Digital water meters* (DWMs) have been identified as a potential solution to water scarcity challenge due to their contribution of data to water conservation and peak demand management [5, 6]. Effective water management strategies require an understanding of the behaviour of consumers and their habits and daily routines. This can be achieved by applying data analytics and machine learning techniques to DWM data [7]. Clustering is among the many data analytics and machine learning techniques that can be used to obtain a better understanding of water consumers' behaviour. Clustering is an unsupervised learning task where the items or objects are grouped based on some inherent similarity among them [8]. After performing clustering, similar objects are assigned in the same group and dissimilar objects are divided into different groups [9]. As clustering helps to identify previously unseen groups from the data, therefore, this technique has gained popularity in different areas such as health, biology, education, business, finance, energy, and water.

Several studies have been performed using clustering based on water-consumption data generated from DWMs. These previous studies can be divided into three categories: (i) customer segmentation, (ii) habit detection and profiling, and (iii) demand profiling. The aim of customer segmentation studies is to understand the water-consumption behaviours of consumers by forming distinct groups of consumers based on their behaviours. The studies under this category are Vieira et al. [10], Garcia et al. [11] and Ji et al. [12], who found 5, 10 and 10 clusters, respectively. Among these studies, only Vieira et al. [10] performed clustering based on disaggregated water-consumption data and features derived from these data. The studies in the category of habit detection and profiling performed clustering based on the identification of habits (i.e., recurring behaviour) and consumption profiles; notable studies are [13-16]. However, none of these studies considered performing clustering based on changes in consumption over time, disaggregated data, type of day (i.e., weekday or weekend) or derived features. Finally, in the demand profiling studies, water consumers are categorised based on their level of water demand [7, 17]. Only Cominola et al. [7] considered disaggregated water consumption for clustering. We have observed various limitations in the existing studies, such as not considering disaggregated consumption data or recent changes in behaviour and that the derived features or patterns are not based on the type of day. However, to fully utilise the data collected from DWMs, these research gaps should be explored. In addition, recent studies [18, 19] have identified the potential of a personalised recommender system for promoting water-conscious behaviour and proposed an advanced customer-profiling approach [20, 21]. Therefore, a suitable clustering study is required that can be used to implement a personalised recommender system for water conservation and demand management.

The primary motivation behind this study is to overcome the limitations of the existing clustering studies by introducing a clustering solution that would provide a better understanding of households' water consumption patterns, and it would be suitable for a recommender system in order to promote water-conscious behaviour [18]. To accomplish this task, we have considered a new and advanced profiling approach that offers in-depth details on consumer water-consumption patterns. This profiling approach is based on disaggregated water-consumption data. It accommodates changes in behaviour with time, separately considers the consumption patterns on weekdays and weekends and includes derived features [20]. Therefore, this approach helps to understand and track changes in behaviour at a household level and to make effective recommendations for promoting water-conscious behaviour. However, not all patterns in water consumption are unique, as they tend to be similar among consumers who belong to the same group. In this scenario, identifying the groups of consumers with similar consumption patterns can improve the performance of finding neighbours in collaborative filtering [22]. In addition, it can be useful for performing social comparisons, as social comparisons are effective in promoting water conservation [23]. Figure 1 shows a broad view of the personalised water-conscious recommender system. In addition, clustering the data gathered from this new profiling approach can reveal previously unseen water-consumption patterns. Therefore, we have chosen to investigate the application of clustering techniques for the purpose of exploring water-consumption patterns based on advanced customer profiling.

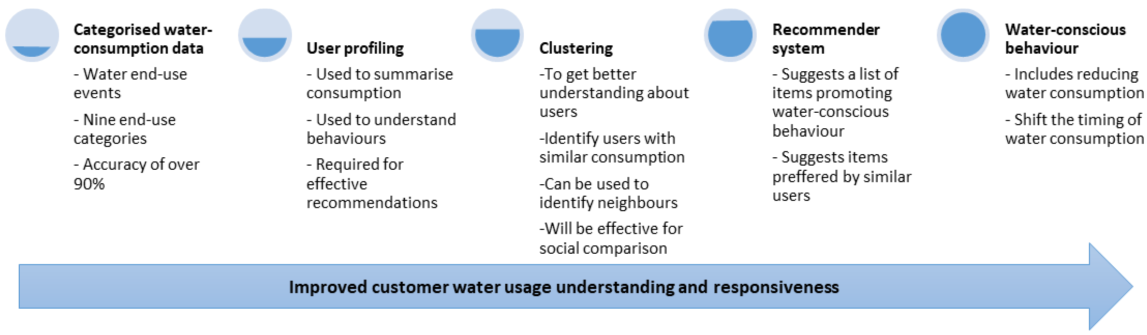


Figure 1. Broad view of the personalised water-conservation system.

In this study, we have performed clustering on two data sets: an *engineered features* (EF) data set and a data set including *times of use and weighted probabilities of use* (TUWPU) of water by consumers. The EF data set was created after performing feature engineering on residential water-consumption data. Based on the existing literature, statistical techniques were used to create 13 new features. Later, with the help of the elbow method [24], the optimal number of clusters was determined, and several clustering techniques were applied to identify the most suitable in terms of running time and clustering quality metrics. For the EF data set, five clusters were found to be optimal, and the characteristics of these clusters were investigated. We concluded that k-means would be the most suitable clustering technique after considering performance and clustering quality metrics. The TUWPU data set provided a weighted probability of an event happening at a particular time at three different intervals (i.e., 15 minutes, 30 minutes and 60 minutes) for five water end-use categories (shower, bathtub, washing machine, dishwasher, gardening) for different day types (i.e., weekday, weekend). In this scenario, we performed hierarchical agglomerative clustering as representing data items in a hierarchy is beneficial to data summarisation and visualisation [24]. Also, different linkage methods were employed to identify the best one, as clustering quality may vary depending on the linkage methods. After performing hierarchical agglomerative clustering, we observed that the optimal number of clusters varied based on the profiling interval, water-consumption event and type of day (i.e., weekday or weekend). However, among the three profiling intervals (i.e., 15 minutes, 30 minutes and 60 minutes), the 15-minute profiling interval outperformed the 30-minute and 60-minute profiling intervals to form clusters based on the similarity of consumer behaviour.

This study addresses many of the research gaps in the existing literature on clustering residential water-consumption data and presents previously unseen and significant findings. The major contributions of this study are as follows:

- It has proposed a detailed clustering solution to effectively apply multiple clustering techniques to analyse water consumption data that overcomes the limitations of existing clustering studies.
- It has identified a suitable clustering technique in terms of run time, scalability and cluster quality metrics for the EF data set.
- It has identified the most suitable profiling interval and linkage method in terms of cluster quality metrics.
- It has observed previously unseen, new, insightful traits about the groups of households that can be used to design future water conservation and demand management programs.
- It has observed that insights from tap-water usage can be used to determine how hand-hygiene practices have adapted to preventing infection from viruses such as COVID-19.

2. Related Work

Several studies have been performed on the clustering of digital data on residential water metering, and they have had different goals. Based on their purpose, these studies can be divided into three categories: customer segmentation, habit detection and profiling and demand profiling. Though it is possible to perform segmentation or clustering based on other data sources such as water-billing data [25], for this study, we have considered any study that utilised data collected from DWMs. We have evaluated related work against four criteria: 1) the ability of the data set used in the study to accommodate recent changes in behaviour, 2) the nature of the consumption data (i.e., aggregated or disaggregated), 3) consideration of the difference in consumption patterns during weekdays and weekends and 4) the inclusion of derived features.

The general aim of the studies we reviewed under the category of customer segmentation was to understand water-consumption behaviours among consumers by forming distinct groups. To understand the overall efficiency of indoor water usage, clustering techniques were applied in a study [10] to compare these techniques and their efficiency patterns with those used by peers. In this study, the authors calculated the averages of the water-use device characteristics and derived an efficiency index. Later, they applied hierarchical agglomerative clustering (HAC) using Ward's method based on the efficiency index, and they found five clusters. However, the data set used in their study was unable to consider recent changes in behaviour or changes in patterns based on the type of day (i.e., weekday or weekend). To understand the behaviour of residential water consumers and to predict water demand, Garcia et al. [11] applied the *k*-means clustering technique to DWM data and found 10 clusters. Similarly, in another study, fuzzy clustering was proposed for analysing urban residential water-consumption data collected at a 15-minute interval [12]. They found 10 groups from 20 users after clustering residents by family structure, job type or lifestyle based on their water-consumption data.

On the other hand, the studies under the category of habit detection and profiling performed clustering by identifying habits (i.e., recurring behaviours) and consumption profiles. Based on the combination of k-means clustering and PCA on hourly consumption data, three profiles were identified by Cominola et al. [14] on the basis of eigenbehaviours (i.e., recurrent water-consumption behaviour). In contrast, clustering using signature patterns was proposed by Cardell-Oliver [15] to explain how households use water. The author found four signature patterns and recommended further research be performed to identify the best way to present the findings to maximise water conservation. Later, Wang et al. [26] proposed a novel algorithm for identifying all routines of variable lengths efficiently. The proposed algorithm expands longer candidate sub-sequences gradually instead of clustering sub-sequences repeatedly. In another study, to identify relevant water-usage profiles, Cheifetz et al. [13] applied a mixture of *k*-means and Fourier regression techniques on hourly consumption data. From residential consumption data, they found three clusters and observed two consumption peaks at 10:00 am and 8:00 pm.

Last, to predict and manage demand, studies on demand profiling categorised water consumers based on their level of demand. It is rare for clustering studies to consider disaggregated water consumption. However, in a recent study, disaggregated water-consumption data were used for customer segmentation. In this study, after extracting eigenbehaviour, dimension-reduction techniques were used: k-means clustering and t-distributed stochastic neighbour embedding. The authors found three main water end-use profile clusters. However, the study did not include derived features or consider recent changes in behaviour or different day types (i.e., weekday or weekend). In another study, various clustering methods (e.g., k-means clustering, dendrograms and self-organising maps) were applied to standardised monthly aggregated time series [17]. They found that a self-organising map with k-means clustering produces the best result with 19 clusters. Table 1 represents the summary of the related studies based on the evaluation criteria mentioned in this section.

Table 1. Summary of the relevant work against evaluating criteria

Category	Study	Considered recent change?	Based on disaggregated data?	Different type of days?	Considered derived features?
Customer Segmentation	[10]	No	Yes	No	Yes
	[11]	No	No	No	No
	[12]	No	No	No	No
Habit detection and profiling	[13]	No	No	No	No
	[14]	No	No	No	No
	[15]	No	No	No	No
	[16]	No	No	No	No
Demand profiling	[7]	No	Yes	Yes	No
	[17]	No	No	No	No

Table 1 shows that most of these studies suffer from the limitations of not having considered recent changes in behaviour, derived features or patterns based on the type of day and of having considered only aggregated data. Though a small number of recent studies had addressed some of these limitations before the current study began, there remained a need to conduct a clustering study on residential DWM data to address these limitations. Using the EF data set and TUWPU data sets, this study has addressed these limitations.

3. Methodology

This study's methodology of performing a comprehensive investigation of residential water-consumption patterns through clustering can be divided into four consecutive steps: 1) initial data collection and processing, 2) data set preparation, 3) experimentation, and 4) result analysis. In Step 1, raw data were collected through DWMs and categorised end-use data were acquired with the help of

Autoflow [6, 27]. In Step 2, data sets for clustering were prepared by employing an advanced customer-profiling technique proposed by Rahim et al. [20, 21]. In Step 3, two clustering methods were applied to identify different groups of customers with similar consumption patterns. Finally, in Step 4, the output generated from the previous step was analysed. Figure 2 illustrates the four steps, and they are described in the following subsections.

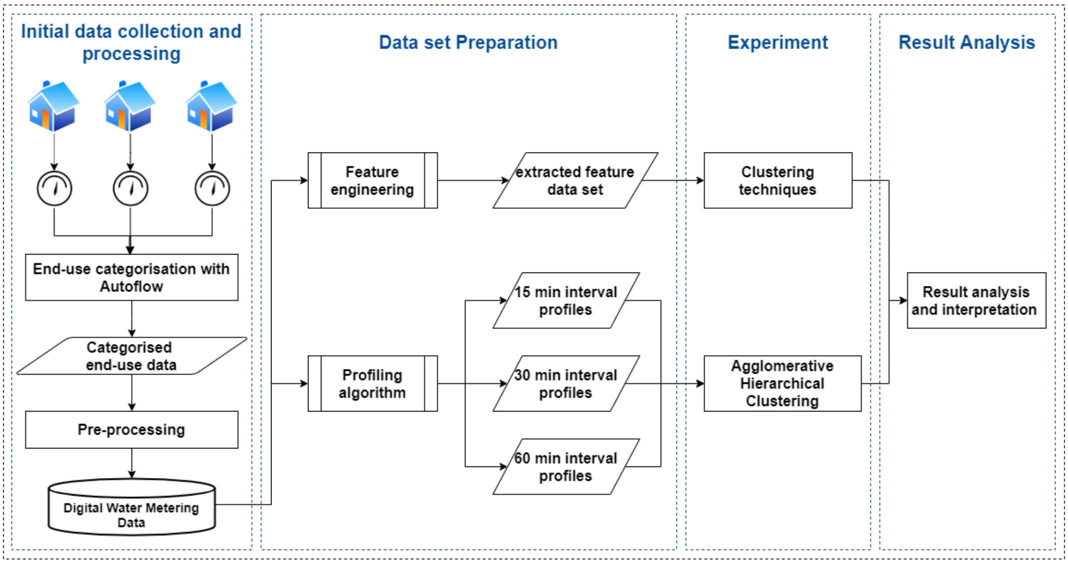


Figure 2. Steps in the methodology.

3.1. Step 1: Initial data collection and preparation

The initial data for this study originated from the high-resolution DWMs installed at 306 single stand-alone household units in various suburbs of Melbourne, Australia. The data set consisted of water-consumption data gathered at 5-s intervals over a period of 10 months (February–December 2010). The collected raw data were then analysed by Autoflow [27, 28], an intelligent metering system that outputs a categorised end-use data set. The categorised end-use data set consisted of nine end-use water categories: taps, dishwashers, leaks, evaporative coolers, clothes washers, showers, toilets, irrigation and bathtubs, with an accuracy of 90%. This was the primary data set for this study.

After the primary data set was prepared, initial pre-processing tasks were performed. First, only the water-consumption events that occurred between 6:00 am and 11:59 pm were considered. The remaining events that occurred between 12:00 am and 5:59 am were discarded due to the insignificant number of events and volume of water consumed during this period. Table 2 summarises the data set after selection for events that occurred between 6:00 am and 11:59 pm.

Next, two end-use categories, evaporative coolers and leaks, were excluded from the data set. Evaporative coolers were excluded because they consume less than one per cent of the total water consumed [29]. Leaks were excluded because a leak is a continuous event, and one long continuous event represents a combination of many small, discrete events recorded at the interval of 5 seconds [20]. Therefore, to reduce computational time and improve scalability, all the leak events were omitted. In the last step of pre-processing, the type of day (weekend or weekday) was extracted from the given date, as water-consumption patterns vary based on this factor [20]. Table 3 describes the data set after this pre-processing step.

Table 2. Summary of data set after selecting events based on time of occurrence

Item	Description
Number of households	306
Data collection duration	2010-02-11 to 2010-10-11
Hours considered	6:00 am – 11:59 pm
Data collection interval	5 seconds
Number of end-use categories	9
Number of events	56,310,576
Total shower events	181,358
Total bathtub events	18,274
Total clothes washer events	30,996
Total dishwasher events	34,524
Total irrigation events	9,975
Total toilet events	1,173,902
Total tap events	4,152,386
Total leak events	50,709,161

Table 3. Description of the attributes in the final dataset

Attribute	Description	Example
Site	Unique identifier for each household in the dataset	Site001, Site002
Start date	Start date of an event	11-Feb-2010
Start time	Start time of an event	08:38:15
End date	End date of an event	11-Feb-2010
End time	End time of an event	08:42:10
Category	Water end-use category	Shower, dishwasher, etc.
Duration	Event duration time	0:03:55
Volume	Quantity (L) of water used in an event	36.81
Max flow	Maximum flow rate (L/min) recorded for an event	10.83
Mode flow	Mode of flow rate (L/min) recorded for an event	9.33
Cyclic event	Multiple intakes of water during one single event	C1 (Cycle 1 for washing machine), D5 (Cycle 5 for dishwasher)
Type of day	A weekend is Saturday or Sunday; a weekday is Monday to Friday (inclusive)	Weekend, weekday

3.2. Step 2: Data set preparation

To accomplish the objectives of this study, two final data sets were derived from the primary data set: (i) the TUWPU data set and (ii) EF data set. The TUWPU data set provides the weighted probability that an event will occur at a particular time at three different intervals (15 minutes, 30 minutes and 60 minutes). In contrast, the EF data set consists of new predictor variables created using feature engineering. The following subsections describe the preparation of these data sets in detail.

3.2.1. **Time of use and weighted probability of use data set**

The TUWPU data set was prepared based on the profile-creation algorithm proposed by Rahim et al. [20]. This algorithm provides profiling at three different intervals: 15 minutes, 30 minutes and 60 minutes. For each profiling interval, a *probability distribution* (PD) was computed across three time periods for each household, for each event across different times and for each type of day. Later, these three PDs were combined after they had each been assigned a different weight. Algorithm 1 represents the steps in profile creation.

Algorithm 1 TUWPU algorithm (S, TD, E, T)

1: **for each** $s \in S$ **do**
2: **for each** $td \in TD$ **do**
3: **for each** $e \in E$ **do**
4: $v_{et}^{1} \leftarrow 0, v_{et}^{2} \leftarrow 0, v_{et}^{3} \leftarrow 0$
5: **for each** $t \in T$ **do**
6: $v_{et}^{1} \leftarrow p_X(x_1)$
7: $v_{et}^{2} \leftarrow p_X(x_2)$
8: $v_{et}^{3} \leftarrow p_X(x_3)$
9: **end for**
10: $v_{et}^{w} \leftarrow (v_{et}^{1} \times w_1) + (v_{et}^{2} \times w_2) + (v_{et}^{3} \times w_3)$
11: **end for**
12: **end for**
13: $M \leftarrow (v_{e_1 t}^{w} \dots v_{e_r t}^{w})$
14: **end for**

Here,
S is the set of households, where $S = \{Site001, Site002, \dots, Site307\}$,
TD is the set of types of day, where $TD = \{weekday, weekend\}$,
E is the set of events, where $E = \{shower, bathtub, clothes\ washer, dishwasher, irrigation\}$,
v_{et} is an n-dimensional vector,
T is the set of times from 6:00 am to 11:59 pm at a 15-minute, 30-minute or 60-minute interval,
$p_X(x)$ is a probability mass function,
v_{etw} is a weighted n-dimensional vector,
w is the weight assigned to PD vectors and
M is the matrix by concatenating all vectors.

At Step 4, three n-dimensional vectors $v_{et}^{1}, v_{et}^{2}, and\ v_{et}^{3}$ are initialised to 0 to calculate and store the probability distribution of event e over a given time period for household s, where each $v_{et} = (v_{et_1}, v_{et_2}, \dots, v_{et_{n-1}}, v_{et_n})$. Here, $v_{et_1}, v_{et_2}, \dots, v_{et_{n-1}}, v_{et_n}$ are the components of vector v_{et}, where each of the component represents the probability of occurring an event e for the given time t. In addition, the number of dimensions depends on the value of the profiling interval. For example, when profiling at the 30-minute interval, there are 35 dimensions, as there will be 35 intervals of 30 minutes in this time period (i.e., 6:00 am, 6:30 am, 7:00 am, …, 11:30 pm). The reason for taking these three vectors is to capture the probability distribution in three different time periods (i.e., latest 18 days, previous 18 days and all previous days). In this way, the first PD vector v_{et}^{1} represents the distribution for the most recent days (i.e., the latest 18 days), the second PD vector v_{et}^{2} represents the previous 18 days, and the last PD vector v_{et}^{3} denotes the PD for all previous historical data. Next, in Step 6-8, a *probability mass function* (PMF) is used to understand the behaviour or habits of the users. PMF is a function that provides the probability that a discrete random variable is exactly equal to some value. Formally this function $p: \boldsymbol{R} \rightarrow [0,1]$ can be defined using equation 1.

$$p_X(x_i) = P(X = x_i)\ for - \infty < x < \infty \quad \text{(Eq. 1)}$$

Equation 1 provides the *probability distribution* (PD) for each event e across time t for the given type of day td and assign it to v_{et}. Equation 1 can be further expanded to equation 2.

$$P(X = x_i) = \frac{|s\ td\ e_i\ t|}{|s\ td\ e_i|} \quad \text{(Eq. 2)}$$

Where $|s\ td\ e_i\ t|$ provides the number of occurrences of an event e_i at a particular time t, for household s, and for the type of day td; $|s\ td\ e_i|$ provides the total of occurrences of an event e_i for household s, and for the type of day td.

Later, in Steps 10, the most recent data were given higher weights as they captured better changes in recent behaviour. Therefore, after assigning a higher weight to the first PD vector and lower weights to the other PD vectors, that is w_1 to v_{et}^{1}, w_2 to v_{et}^{2}, and w_3 to v_{et}^{3}, where $w_1 > w_2 \geq w_3$, a matrix M is constructed by concatenating the vectors for each household.

$$M = \begin{bmatrix} \left((w_1 \times v_{e_1 t_1}^{1}) + (w_2 \times v_{e_1 t_1}^{2}) + (w_3 \times v_{e_1 t_1}^{3})\right) & \cdots & \left((w_1 \times v_{e_1 t_n}^{1}) + (w_2 \times v_{e_1 t_n}^{2}) + (w_3 \times v_{e_1 t_n}^{3})\right) \\ \vdots & \ddots & \vdots \\ \left((w_1 \times v_{e_m t_1}^{1}) + (w_2 \times v_{e_m t_1}^{2}) + (w_3 \times v_{e_m t_1}^{3})\right) & \dots & \left((w_1 \times v_{e_m t_n}^{1}) + (w_2 \times v_{e_m t_n}^{2}) + (w_3 \times v_{e_m t_n}^{3})\right) \end{bmatrix} \quad \text{(Eq. 3)}$$

In equation 3, for the matrix M, each row represents the weighted probability of occurring an event e, where $e \in E = \{shower, bathtub, clothes\ washer, dishwasher, irrigation\}$ and each column represents a time t, where $t \in T = \{6{:}00\ \text{am}, 6{:}30\ \text{am}, 7{:}00\ \text{am}, \dots, 11{:}30\ \text{pm}\}$, or $t \in T = \{6{:}00\ \text{am}, 6{:}15\ \text{am}, 6{:}30\ \text{am}, \dots, 11{:}30\ \text{pm}, 11{:}45\ \text{pm}\}$, or $t \in T = \{6{:}00\ \text{am}, 7{:}00\ \text{am}, \dots, 11{:}00\ \text{pm}\}$, depending on the profiling interval. Figure 3 describes the steps of TUWPU algorithm.

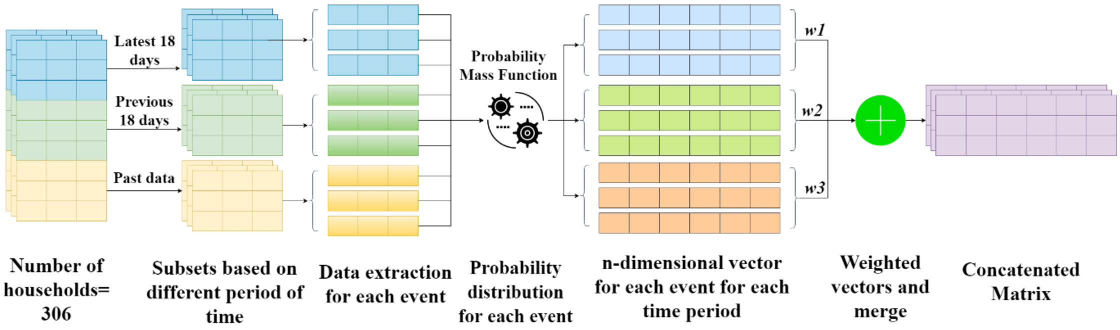


Figure 3. Illustration of the time of use and weighted probability of the use algorithm [20].

3.2.2. Engineered features data set

A feature is a numeric representation of raw data that is used as a predictor variable, which predicts an outcome variable [30]. Features are important in machine learning and data analytics, as they capture salient representations of the data and can be useful for generating new valuable insights. Feature engineering is the process of formulating suitable features using domain knowledge and the given data, model and task [30]. The success or failure of an analytics project depends highly on identifying appropriate features [31]. Therefore, feature engineering can be considered as a primary task in such projects, and a large amount of time is spent on it. Figure 4 illustrates the position of feature engineering in a typical analytics project, based on Zheng and Casari [30].

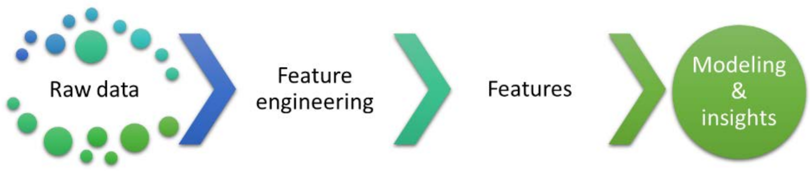


Figure 4. The position of feature engineering in the analytics workflow.

The data collected in this study were raw in nature. Thus, to generate actionable insights, it was necessary to conduct feature engineering based on domain knowledge. First, we analysed existing literature on residential water conservation and then identified which features would be required to promote water conservation and could be used to group users based on their water-consumption data. Based on the literature, we identified several events and their attributes that were used to create new features [32-35]. Figure 5 summarises these events and their attributes.

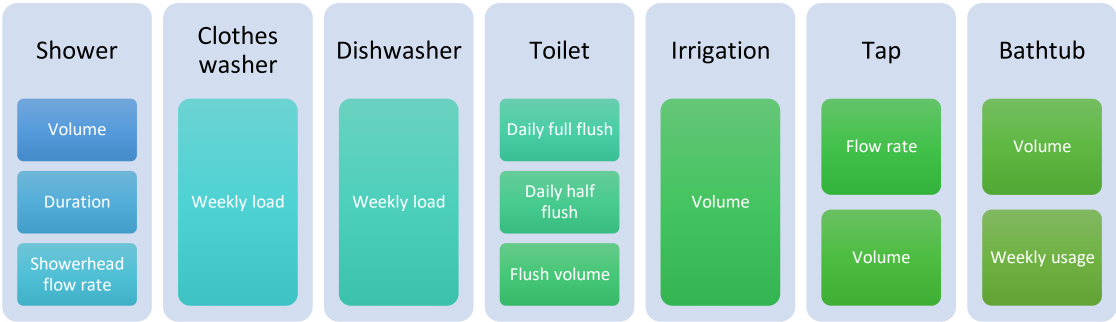


Figure 5. The events and their attributes used for feature engineering in this study.

The findings from the literature show that statistical techniques are useful in creating these features and achieving the target of water conservation by identifying users with similar consumptions. Table 4 describes the created features, creation techniques and their importance in water conservation. We discarded total water consumption/day as several studies had used this feature. Otherwise, clusters would be created based on average water consumption per day, and it would not create new patterns.

Table 4. Description of the features derived from feature engineering

Event	Derived feature name	Description	Goal
Shower	Mean of shower volume	Mean of the volume of water used in the shower for each household (litre/shower): $$\overline{sv} = \frac{1}{n}\left(\sum_{i=1}^{n} sv_i\right)$$ here, $\overline{sv}$ is the mean of water used in shower, sv_i is the volume of water used in shower *i*, n is the total number of shower events.	Promoting water conservation by reducing the volume of water used in the shower
	Mean of shower duration	Mean of the duration of the shower for each household (minute/shower): $$\overline{sd} = \frac{1}{n}\left(\sum_{i=1}^{n} sd_i\right)$$ here, $\overline{sd}$ is the mean of shower duration, sd_i is the duration of shower *i*, n is the total number of shower events.	Promoting shorter shower to conserve more water
	Mean of max flow for shower	Mean of the max flow of shower for each household (litre/minute/shower): $$\overline{mfs} = \frac{1}{n}\left(\sum_{i=1}^{n} mfs_i\right)$$ here, $\overline{mfs}$ is the mean of maximum water flow in shower, mfs_i is the maximum water flow during shower *i*, n is the total number of shower events.	Identifying the efficiency of showerhead to reduce water consumption
Clothes washer	Mean of weekly clothes washer load	Mean of number of weekly clothes washer loads per household (number of load/week): $$\overline{wcl} = \frac{1}{n}\left(\sum_{i=1}^{n} wcl_i\right)$$ here, $\overline{wcl}$ is the mean of weekly clothes washer loads, wcl_i is the of number of clothes washer loads in week *i*, n is the total number of weeks.	Identifying the scope of water conservation by reducing weekly clothes washer load.
Dishwasher	Mean of weekly dishwashing load	Mean of number of weekly dishwashing loads per household (number of load/week): $$\overline{wdl} = \frac{1}{n}\left(\sum_{i=1}^{n} (wdl_j)\right)$$ here, $\overline{wdl}$ is the mean of weekly dishwasher loads, wdl_i is the of number of dishwasher loads in week *i*, n is the total number of weeks.	Identifying the scope of water conservation by reducing weekly dishwasher load.
Toilet	Mean of daily full flush	Mean of number of daily full flushes per household (number of full flushes/day):	For discouraging the usage of full flushes after

		$$\overline{dff} = \frac{1}{n}\left(\sum_{i=1}^{n} dff_i\right)$$ here, $\overline{dff}$ is the mean of daily full flush, dff_i is the of number of full flushes in day *i*, n is the total number of days.	comparison with half flushes. (4:1=full flushes: half flushes)
	Mean of daily half flush	Mean of number of daily half flushes per household (number of half flushes/day/household): $$\overline{dhf} = \frac{1}{n}\left(\sum_{i=1}^{n} dhf_i\right)$$ here, $\overline{dhf}$ is the mean of daily half flush, dhf_i is the of number of half flushes in day *i*, n is the total number of days.	For encouraging the usage of half flushes after comparison with full flushes. (4:1=full flushes: half flushes)
	Mean of flush volume	Mean of the volume of water used in flush per household (litre/flush): $$\overline{fv} = \frac{1}{n}\left(\sum_{i=1}^{n}(fv_j)\right)$$ here, $\overline{fv}$ is the mean of water volume used in flush, fv_i is the of amount of water volume used in flush *i*, n is the total number flushes.	To determine the water efficiency of the toilet (i.e., star) To determine users with a single flush system
Irrigation	Mean of irrigation volume	Mean of the volume of water used in irrigation for each household (litre/irrigation): $$\overline{irrv} = \frac{1}{n}\left(\sum_{i=1}^{n} irrv_i\right)$$ here, $\overline{irrv}$ is the mean of water volume used in irrigation, $irrv_i$ is the of amount of water volume used in irrigation *i*, n is the total number irrigation events.	To understand water consumption patterns in irrigation.
Tap	Mean of flow rate	Mean of the flow rate of tap for each household (litre/minute/tap): $$\overline{mft} = \frac{1}{n}\left(\sum_{i=1}^{n} ft_i\right)$$ here, $\overline{mft}$ is the mean of water flow in taps, ft_i is the water flow during tap event *i*, n is the total number of tap events.	To determine the efficiency of the tap.
	Mean of volume	Mean of the volume of water used in the tap for each household (litre/day): $$\overline{tv} = \frac{1}{n}\left(\sum_{i=1}^{n} tv_i\right)$$ here, $\overline{tv}$ is the mean of water used in taps, tv_i is the water used during tap event *i*, n is the total number of tap events.	Can be used to understand tap water consumption patterns
Bathtub	Mean of Volume	Mean of the volume of water used in bathtub for each household (litre/bathtub/household): $$\overline{btv} = \frac{1}{n}\left(\sum_{i=1}^{n} btv_i\right)$$ here, $\overline{btv}$ is the mean of water volume used in bathtub, btv is the of amount of water volume used in bathtub event *i*, n is the total number bathtub events.	Can be used to understand bathtub water consumption patterns
	Mean of Weekly count	Mean of the number of weekly bathtub usage per household (bathtub events/week/household): $$\overline{wbt} = \frac{1}{n}\left(\sum_{i=1}^{n} wbt_j\right)$$ here, $\overline{wbt}$ is the mean of weekly bathtub events, wbt_i is the of number of bathtub events in week *i*, n is the total number of weeks.	Can be used to understand bathtub usage frequency

3.3. Experiments

Because the types and goals of the two data sets are different, we applied two different approaches. As the EF data set is less complex and easier to explain, summarise and visualise, therefore, we first determined the optimal number of clusters for this data set using the elbow method [24] and k-means clustering [36, 37]. Before determining the optimal number of clusters, a dimension reduction technique was applied on the data set. Later, we applied k-means [36, 37], affinity propagation [38], mean shift [39], spectral [40], birch [41] and hierarchical agglomerative clustering [42] techniques to compare their performance in terms of running time and clustering quality. On the other hand, the TUWPU data set is more complex than the EF data set as it provides outputs at three different intervals (i.e., 15 minutes, 30 minutes and 60 minutes) for five different water end-use categories (shower, bathtub, washing machine, dishwasher, gardening) for different day types (i.e., weekday, weekend). As representing data items in a hierarchy is beneficial to data summarisation and visualisation, therefore, we chose hierarchical agglomerative clustering [42] for the TUWPU data set, as this method forms clusters iteratively by dividing the patterns using a bottom-up approach [8]. In addition, we applied different linkage methods and compared their performance to identify the suitable one for hierarchical agglomerative clustering. The following subsection discusses these approaches and their applications in our scenario briefly.

3.3.1. Clustering the EF data set

Determining the appropriate number of clusters is an important task as it controls the granularity of cluster analysis and some clustering algorithms require that parameter [24]. One of the popular and simple methods is to determine the number of clusters as $\sqrt{\frac{n}{2}}$ for a data set with n points. However, this method would provide a greater number of clusters for a big data set. Therefore, we chose another popular method called the elbow method to determine the appropriate number of clusters. First, we applied a dimension reduction technique and then applied the elbow method with the k-means clustering technique and identified the optimal number of clusters. Subsequently, we applied other clustering techniques to compare the clustering quality and run time. As the ground truth was not available, instead of using extrinsic methods, we applied two intrinsic methods, namely silhouette coefficient and the Calinski-Harabasz index. After determining the best clustering technique, we identified the traits of each cluster.

Dimension reduction is the process of reducing the number of attributes while preserving as much possible variance in the original data set as possible. In this experiment, we applied principal component analysis (PCA) as the dimension reduction technique. Principal Component Analysis (PCA) is an algorithm that minimises the dimension of the data set by preserving most of the variance [43]. This task is accomplished by identifying the dimensions that explain the maximum variance in the data [44]. In PCA, new variables known as principal components are constructed and are linear combinations of the initial variables. The first principal component explains the largest proportion of the variance; the second principal component explains the second largest proportion, and so on. Therefore, a smaller number of principal components leads to lower variance. This is why, to choose the optimal number of principal components, 95% explained variance should be used instead of total variance [43]. After applying the dimension reduction technique, we chose the k-means clustering algorithm with the elbow method to determine the optimal number of clusters.

k-means clustering is one of the most well-known algorithms and is easy to understand and easy to implement [36, 37]. The algorithm divides the data set into k groups, where the value of k is defined by a user. Each k presents the centroid of a cluster and is typically assigned randomly [9]. The objective function J is defined in Equation 4.

$$minimise\ J = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| x_i^{(j)} - c_j \right\|^2 \quad \text{(Eq. 4)}$$

where $\left\| x_i^{(j)} - c_j \right\|^2$ is the distance between a data point $x_i^{(j)}$ and centroid c_j [8]. The instances within the same group are similar to each other, and instances from different groups are different from each other [24]. Therefore, we chose this algorithm to identify households with similar water-consumption patterns. However, determining the optimal numbers for k is a difficult task for this algorithm. To overcome this challenge, we used the elbow method [24].

The elbow method is based on the finding that the sum of within-cluster variance for each cluster decreases with an increasing number of clusters [24]. This occurs because having more clusters helps to identify more groups of items with similar characteristics. However, with too many clusters, there might not be enough of a reduction in the sum of within-cluster variances because partitioning a well-integrated cluster into two clusters provides a reduction by a very small margin. Therefore, by plotting the curve of the sum of within-cluster variances for different values of k, where $k > 0$, the correct number of clusters can be identified from the first or most significant turning point of the curve. This process is defined mathematically in Equation 5.

$$minimise(\sum_{k=1}^{K} W(C_k)) \quad \text{(Eq. 5)}$$

where C_k is the k^{th}cluster, and $W(C_k)$ is the within-cluster variation. The sum of within-cluster variation $W(C_k)$ is defined in Equation 6.

$$W(C_k) = \sum_{x_i \in C_k} (x_i - \mu_k)^2 \quad \text{(Eq. 6)}$$

where x_i is a data point that belongs to the cluster C_k, and μ_k is the mean of points assigned to the cluster C_k. Equation 7 represents the final equation.

$$minimise(\sum_{k=1}^{K} \sum_{x_i \in C_k} (x_i - \mu_k)^2) \quad \text{(Eq. 7)}$$

We observed that for the EF data set, four principal components explained 95% of the variance in the data set. Figure 6 shows each number of principal components and its corresponding ratio of explained variance.

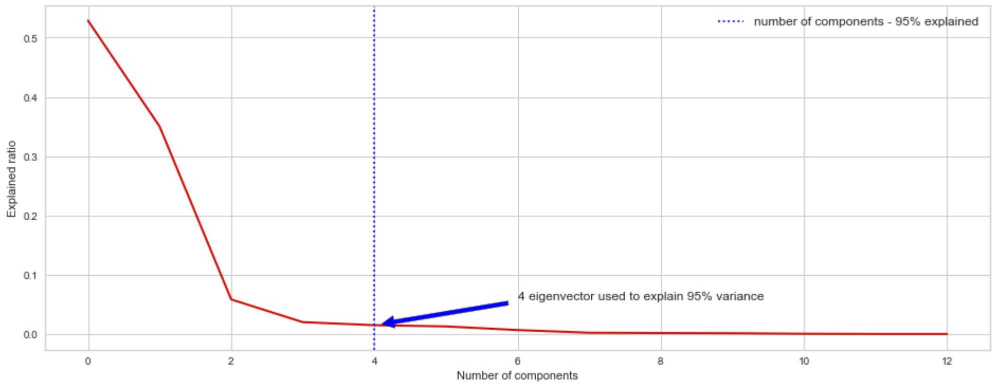


Figure 6. Ratio of explained variance for the different number of principal components for the EF data set.

After identifying the optimal number of principal components, we applied the elbow method to determine the optimal number of clusters. The elbow method used in this study was implemented by selecting the optimal number of clusters by using k-means clustering with a varying number of clusters. After calculating the sum of within-cluster variances for different values of k with four principal components and clustering model k-means, where $0 < k < 20$, we found that $k = 5$ is the optimal number of clusters, as the most significant turning point occurred there. Figure 7 plots the sum of the within-cluster variances for different values of k after PCA was performed. From the figure, we concluded that the number of optimal clusters is five as the elbow appears there.

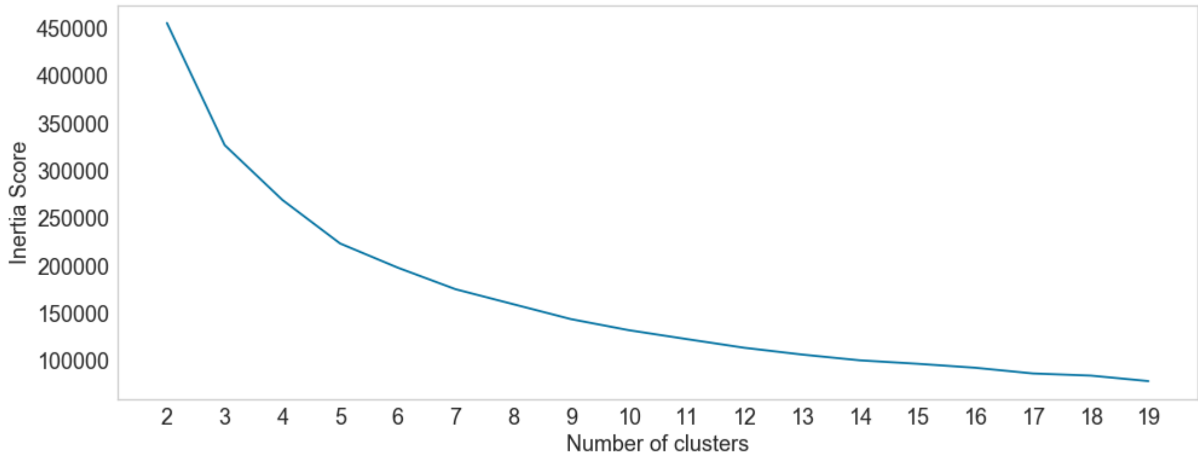


Figure 7. Plotted curve of the within-cluster variances for different values of k after PCA was performed.

For further confirmation regarding the optimal value of k, we applied a silhouette plot for different values of k. Silhouette plots display the silhouette coefficient for each data point from the clusters. Figure S2 displays the silhouette plots of k-means for 306 households in 3,4,5 and 6 clusters. From these plots, we concluded between cluster 4 and 5 in figure S2 (b) and S2 (c); cluster 5 would be the optimal number of clusters as households were more uniform among the clusters compared to cluster 4.

After identifying the optimal number of clusters, for the next step, we applied different clustering techniques on the transformed EF data set. Besides k-means, other clustering techniques included affinity propagation [38], mean shift [39], spectral [40], birch [41] and hierarchical agglomerative clustering [42]. In affinity propagation, clusters are created by sending messages between data points until convergence [42]. Mean shift is a centroid-based algorithm that aims to find blobs in a smooth density of data points [39, 45]. Spectral clustering performs dimension reduction before clustering in fewer dimensions by using the eigenvalues of the similarity matrix of data points [45]. On the other hand, the birch clustering technique works by building a clustering feature tree (CFT) for the data points. Lastly, in hierarchical agglomerative clustering, initially, all the data items are considered separate clusters, and they iteratively merge with one of the closest clusters until there is only one larger cluster, which becomes the hierarchy's root. After running these algorithms on the transformed EF data set, we recorded the run time to measure the performance in terms of running time. In addition, we also checked the scalability of these algorithms by executing them in different sizes to get an average-case performance. Furthermore, as the ground truth labels for the transformed EF data set were not available, therefore, to measure the clustering quality, the silhouette coefficient and Calinski-Harabasz indexes were recorded for each technique. The silhouette coefficient is defined for each sample using two scores a, b and silhouette coefficient s can be defined using equation 8.

$$s = \frac{b-a}{max(a,b)} \quad \text{(Eq. 8)}$$

Here, a is the mean distance between one sample and other points from the same class, and b is the mean distance between one sample and other points from the next nearest cluster.

Calinski-Harabasz index, also known as the Variance Ratio Criterion, is the ratio of the sum of between-cluster dispersion and of within-cluster dispersion for all clusters [45]. It can be defined using Equation 9.

$$s = \frac{tr(B_k)}{tr(W_k)} \times \frac{n_e-k}{k-1} \quad \text{(Eq. 9)}$$

where E is the data set, n_e is the size of the data set, k is the number of clusters, $tr(B_k)$ and $tr(W_k)$ are the trace between the group dispersion matrix and within-cluster dispersion matrix, respectively.

After completing the experiment, we observed that performance and clustering quality metrics varied depending on the clustering technique. For instance, in terms of performance in time taken to cluster, agglomerative clustering took the least amount of time, and the mean shift took the greatest amount of time. For measuring the quality of clustering, in the case of silhouette score, the mean shift outperformed other techniques followed by affinity propagation and k-means, and when using the Calinski-Harabaz index, k-means provided the best result. Table 5 presents the data collected from this experiment. After looking carefully at the data, we noticed that even though the mean shift technique performed better with the silhouette score, however, it performed poorly in terms of time taken and when using the Calinski-Harabaz index. K-means provided the best result with the Calinski-Harabaz index score, third-best with the silhouette score and average in terms of time performance. It seemed that k-means would be a comparatively better choice in terms of these matrices. However, before taking the final decision, we also observed the scalability of these techniques.

To understand the scaling performance of the clustering techniques for the EF data set, we ran experiments several times for different data sizes. Initially, we started with 10 households, and with a step size of 10, we completed all the 306 households. Figure 8 represents the result of the experiment by drawing a point at the mean of the time taken to create clusters and the error bar. From there, we concluded that mean shift was the worst-performing clustering technique in terms of scaling performance, and k-means showed good performance. As detailed earlier, we identified that k-means performed better in clustering quality and it was also scalable; therefore, we decided to further refine the clusters using k-means clustering.

Table 5. Performance and clustering quality metrics of different clustering techniques for the EF data set

Clustering Algorithm	Time taken (in seconds)	Silhouette score	Calinski-Harabaz index score
K-means	0.048	0.2999	**156.6242**
Affinity Propagation	0.140	0.3061	153.5803
Mean Shift	0.970	**0.4478**	81.0378
Spectral Clustering	0.046	0.2514	130.5797
Birch	0.020	0.2776	132.8396
Agglomerative Clustering	**0.004s**	0.2549	129.9911

To understand the scaling performance of the clustering techniques for the EF data set, we ran experiments several times for different data sizes. Initially, we started with 10 households, and with a step size of 10, we completed all the 306 households. Figure 8 represents the result of the experiment by drawing a point at the mean of the time taken to create clusters and the error bar. From there, we concluded that mean shift was the worst-performing clustering technique in terms of scaling performance, and k-means showed good performance. As detailed earlier, we identified that k-means performed better in clustering quality and it was also scalable; therefore, we decided to further refine the clusters using k-means clustering.

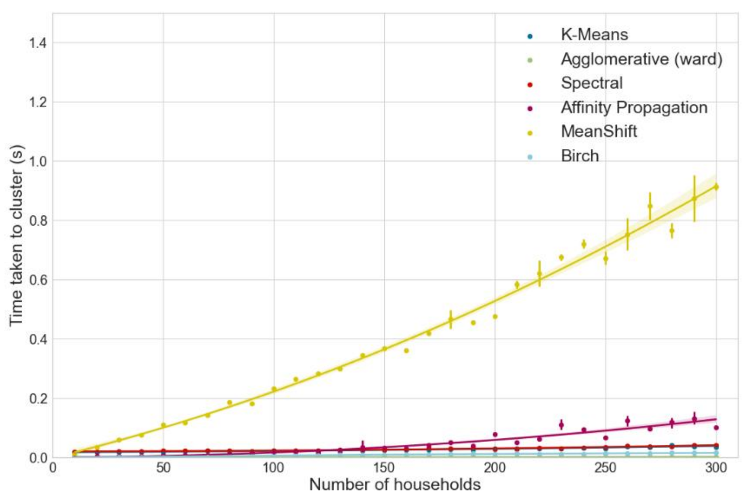


Figure 8. Scaling performance of different clustering techniques for the EF data set.

After performing clustering on the EF data set using the k-means algorithm, we moved to the TUWPU data set to group users based on their consumption time for each event. To accurately interpret the result, we deployed HAC because representing data items as a hierarchy is beneficial to data summarisation and visualisation [24]. The following subsection discusses the method in detail.

3.3.2. **Clustering the TUWPU data set**

Hierarchical agglomerative clustering is a variation on hierarchical clustering where the hierarchical decomposition is performed in either a bottom-up or a top-down approach [24]. In this method, initially, all the data items are considered separate clusters, and they iteratively merge with one of the closest clusters until there is only one larger cluster, which becomes the hierarchy's root. After measuring the distance between two clusters, a linkage criterion is used to merge the two clusters into one. In this study, to measure the distance between two clusters, we used Euclidean distance as Equation 8.

$$d(p,q) = \sqrt{\sum_{i=1}^{n}(p_i - q_i)^2} \quad \text{(Eq. 8)}$$

where p and q are two points in Euclidean n-space.

Linkage methods determine the distance between clusters as a function of pairwise distances between data points [46]. Several linkage methods are currently available, each with its own characteristics. We explored most of the common linkage methods and determined the most suitable method by comparing the clustering quality metrics and dendrograms generated by HAC. The details of these methods, along with their notations, are discussed in this section.

Table 5. Notations and their descriptions for different linkage methods.

Notation	Description
d_{mj}	Distance between clusters *m* and *j*
m	Merged cluster
d_{kj}	Distance between clusters *k* and *j*
d_{lj}	Distance between clusters *l* and *j*
d_{kl}	Distance between clusters *k* and *l*
N_j	Number of data items in cluster *j*
N_k	Number of data items in cluster *k*
N_l	Number of data items in cluster *l*
N_m	Number of data items in cluster *m*

Single-linkage method: In this method, the distance between two clusters is the minimum distance between one item from one cluster and another item from another cluster [47]. After the distance between all clusters is calculated, the two closest clusters are merged into one larger cluster. However, one of the limitations of this method is that it suffers from the chaining effect and has the tendency to produce irregular or widened clusters [47, 48]. Equation 9 can be used to calculate the minimum distance.

$$d_{mj} = min(d_{kj}, d_{lj}) \quad \text{(Eq. 9)}$$

Complete-linkage method: In the complete-linkage method, the distance between two clusters is the maximum distance among all pairwise distances between observations in the two clusters [47]. This method can produce tightly bound or compact clusters [49]. Equation 10 can be used to determine the maximum distance.

$$d_{mj} = max(d_{kj}, d_{lj}) \quad \text{(Eq. 10)}$$

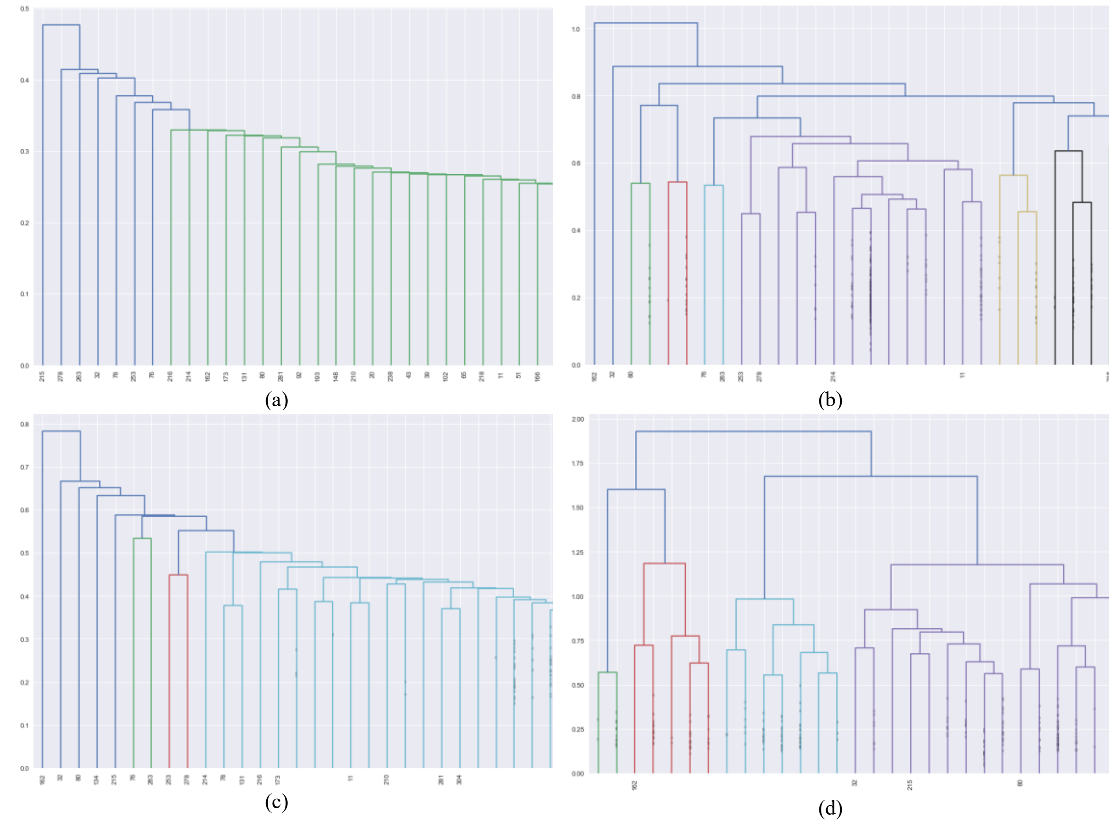
Average-linkage method: In this method, the distance between two clusters is the average distance between the pairs of observations in the two clusters [50]. Equation 11 can be used to derive this average distance.

$$d_{mj} = \frac{N_k d_{kj} + N_l d_{lj}}{N_m} \quad \text{(Eq. 11)}$$

Ward's linkage method: The aim of Ward's linkage method is to minimise the total within-cluster variance. To this end, a pair of clusters to be merged on each iteration is chosen such that it produces the minimum within-cluster variance. Here, the distance between two clusters is the sum of squared deviations from the observations to the centroids [50] and can be calculated using equation 12.

$$d_{mj} = \frac{(N_j+N_k)d_{kj}+(N_j+N_l)d_{lj}-N_j d_{kl}}{N_j+N_m} \quad \text{(Eq. 12)}$$

Clustering may vary depending on the profiling interval and linkage method used. Figure 9 shows the dendrograms of the shower event during weekdays with the 30-minute interval for different linkage methods. From that figure, we can see that the number of clusters varies depending on the linkage methods as the distance cut-off value differs. Therefore, after applying the above linkage methods on the TUWPU data set, we measured the clustering quality using the silhouette coefficient and the Calinski-Harabasz index as well as the distance cut-off value to determine the best profiling interval and linkage method.



(a) (b) (c) (d)

Figure 9. Dendrograms of the shower event during weekdays with the 30-minute interval for a) single linkage, b) complete linkage, c) average linkage, d) Ward's linkage. Here, the x-axis represents households, and the y-axis represents the distance.

After performing the experiments, all results were collected and analysed carefully. The following section discusses the results and findings.

4. Results and Discussion

The previous sections described the comprehensive experiments on the two data sets derived from residential DWMs. This section presents our findings and discusses their implications for promoting water-conscious behaviour. To visualise the outcomes of this study, we have applied various visualisation tools such as Power BI [51] and Seaborn [52].

4.1. Engineered features data set: Findings and discussion

Identifying the optimal number of clusters and traits for each cluster is beneficial for water utility companies to better understand their customers and devise effective water-conservation awareness programs. To this end, after applying the *k*-means clustering algorithm to the EF data set, we identified the optimal number of clusters to be five and determined the characteristics of each cluster. Among the 306 households, the number of households in each cluster varies. For instance, Clusters 1, 2, 3, 4 and 5 contain 61, 124, 59, 15 and 47 households, respectively. Figure 10 illustrates the number of households in each cluster.

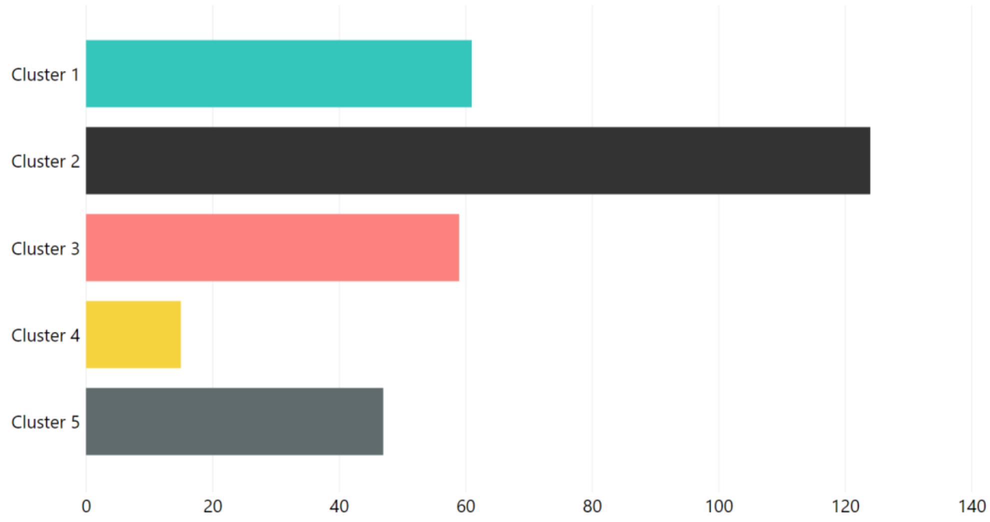


Figure 10. The number of households by cluster for the EF data set.

Table 6. Summary of the traits of clusters based on the arithmetic mean of different events

Event	Clusters					Feature
	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	
Shower	**40.19**	47.37	47.79	44.67	**55.63**	Mean of shower volume (L)
	5.68	5.66	5.89	5.51	5.83	Mean of shower duration (min)
	8.83	10.82	10.54	10.70	**11.93**	Mean of max flow for shower (L/min)
Toilet	8.48	8.14	14.29	**20.73**	8.15	Mean of daily full flush (L)
	6.44	6.06	12.20	**21.93**	6.57	Mean of daily half flush (L)
	5.33	5.36	5.24	5.40	5.21	Mean of flush volume (L)
Clothes washer	2.95	2.95	3.46	**6.40**	2.68	Mean of weekly clothes washer load
Dishwasher	3	3	6	**8**	3	Mean of weekly dishwashing load
Tap	4.13	4.33	4.53	4.49	4.55	Mean of flow rate (L/min)
	47.02	41.40	**87.02**	**161.68**	41.76	Mean of volume (L)
Bathtub	0.82	1.98	4.14	4.53	3.72	Weekly count
Irrigation	190	190	190	160	170	Mean of irrigation volume (L)

After determining the optimal number of clusters from the EF data set, we investigated the characteristics of each cluster and found some significant, valuable and previously unseen insights. To summarise the results, we applied summary statistics on each feature. Summary statistics are commonly used to present the largest amount of information in the simplest possible manner. In this study, we employed a five-number summary that provides values for the median, first quartile, third quartile, minimum and maximum based on the observations of a sample. These values can be used to quickly compare several sets of observations. In addition, we also calculated the arithmetic mean of these features to understand and report central tendencies. Table 6 presents the arithmetic mean of different events for each cluster to denote their characteristics.

Table 6 shows that several aspects are distinctive between these clusters. For instance, Cluster 1 can be differentiated from other clusters by its lower mean of shower volume and lower mean of maximum shower flow. In contrast, Cluster 5 has the highest value of mean shower volume and maximum shower flow. Another relevant finding is that the mean of daily flushes, the volume of tap water used, and the number of weekly clothes-washer loads for Cluster 4 is the highest compared to other clusters. Even though socio-demographic data are not available for this study, we assume based on previous studies that households with elderly people and infants might fall in this cluster.

Another finding from the results is the relationship between the number of flushes and tap-water volume. We observed that households with a higher number of daily flushes tend to use more water from taps. Figure 11 shows that Cluster 4 has the highest number of average daily full and half flushes as well as the highest amount of tap water compared to other clusters. On the other hand, Cluster 2 has the lowest number of full flushes and used the least amount of tap water.

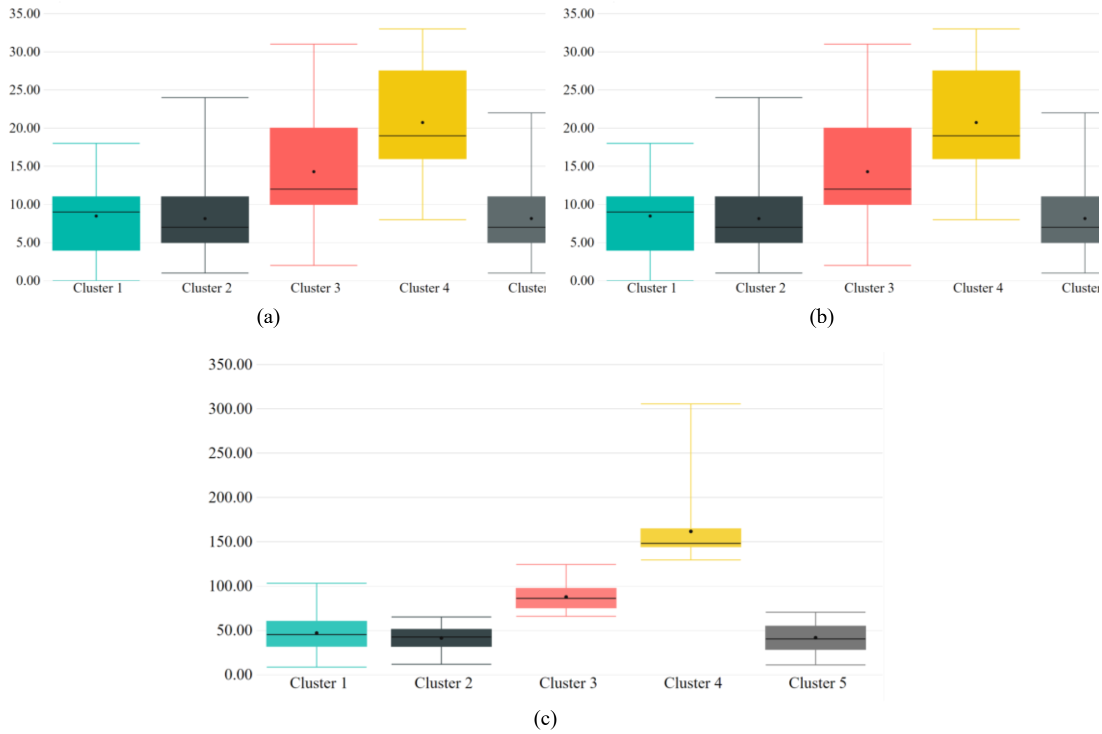


Figure 11. Comparison among different clusters in terms of (a) mean of daily full flushes, (b) mean of daily half flushes and (c) mean of tap water used daily.

This might not be surprising at first because people typically use tap water to wash their hands after using the toilet. However, during the current and unprecedented global outbreak of the novel coronavirus disease 2019 (COVID-19), insights from tap-water usage can be used to understand the adaptation of hand-hygiene practices among the population to prevent infection from the COVID-19 virus [53]. For instance, the water consumption patterns of tap-water usage from before and during the pandemic could be compared to determine the acceptance of hand hygiene among different groups of consumers with the inclusion of socio-demographic data and could be useful to policymakers and health experts in deciding their next actions. For instance, after performing clustering on the tap-water usage data collected before and during the pandemic, the difference in the mean of the tap water used daily among the clusters could be used to help understand changes in hand hygiene practices. As the data set used in this study was collected before (i.e., 2010) the COVID-19 pandemic, it is not possible to undertake changes in this present study. Through applying our approach with a 2020 high resolution data set, this type of evaluation could be conducted in the future.

5. Time of use and weighted probability of use data set: Findings and discussion

The TUWPU data set addresses the aspects of demand management by characterising the patterns of behaviour and habits of water consumption by providing the probability that a specific water-use event will occur at a particular time, for various intervals throughout the day. With the application of HAC, the results from this data set show that the participating households fall under different clusters based on the type of day, event and interval. For instance, for the shower event, one household can be in different clusters for its behaviours and habits depending on the type of the day (i.e., weekday or weekend) and profiling interval (15-minute, 30-minute or 60-minute). Figure 12 shows the result of applying HAC on 306 households for the shower event and the three profiling intervals.

The heat map associated with the dendrograms represents the intensity of the probability of a shower occurring in a household at a particular time. A higher probability represents a strong habit that is a higher chance of the same event reoccurring at that particular time. From these dendrograms, it is possible to identify households with similar consumption patterns that fall under the same clusters, and the intensity or magnitude of a habit occurring at a particular time. Such clustering based on the probability of an event occurring at a particular time can be valuable to identify households with similar patterns, and this information can be used for demand management. For instance, a recommendation to shift an event outside of the hours of peak demand can be made through a personalised recommender system because shifting the timing of water consumption on the customer end presents a viable cost-reduction opportunity for water utility companies [54]. However, making recommendations to shift an event time by 60 minutes might not be effective from the perspective of behaviour theory. In this case, profiling even at a 15-minute interval can identify a higher probability of an event occurring at a particular time; therefore, recommendations can be made to shift event time by 15 minutes. However, depending on the objectives of the utilities and types of events they are interested in examining, a 30-minute or 60-minute profiling interval can be chosen for clustering. For such scenarios, the dendrograms generated from HAC provide the flexibility to determine the number of clusters after performing the analysis. The number of optimal clusters can be determined based on the distance between the two newly merged clusters, where a gap in the distance means that the two clusters do not belong in one cluster. In other words, a larger distance means a greater difference between the two clusters for the chosen features. However, there is no golden rule that will determine the optimal number of clusters from dendrograms. The results could vary based on the nature of the problem and the data set. Therefore, for the TUWPU data set, we set the distance cut-off value at 70% from the final merge to understand the differences based on the event, type of day and profiling interval. Table 7 presents a summary of the number of clusters for different events at the three different profiling intervals.

Table 7. Estimated number of clusters for different events, type of day at different profiling interval

Event	**Type of day**	**Number of clusters for interval**		
		15 min	**30 min**	**60 min**
Shower	Weekday	5	4	3
	Weekend	6	3	3
Clothes washer	Weekday	7	5	4
	Weekend	5	5	5
Dishwasher	Weekday	6	2	3
	Weekend	3	4	3
Bathtub	Weekday	4	3	3
	Weekend	5	4	4
Irrigation	Weekday	4	4	6
	Weekend	6	5	5

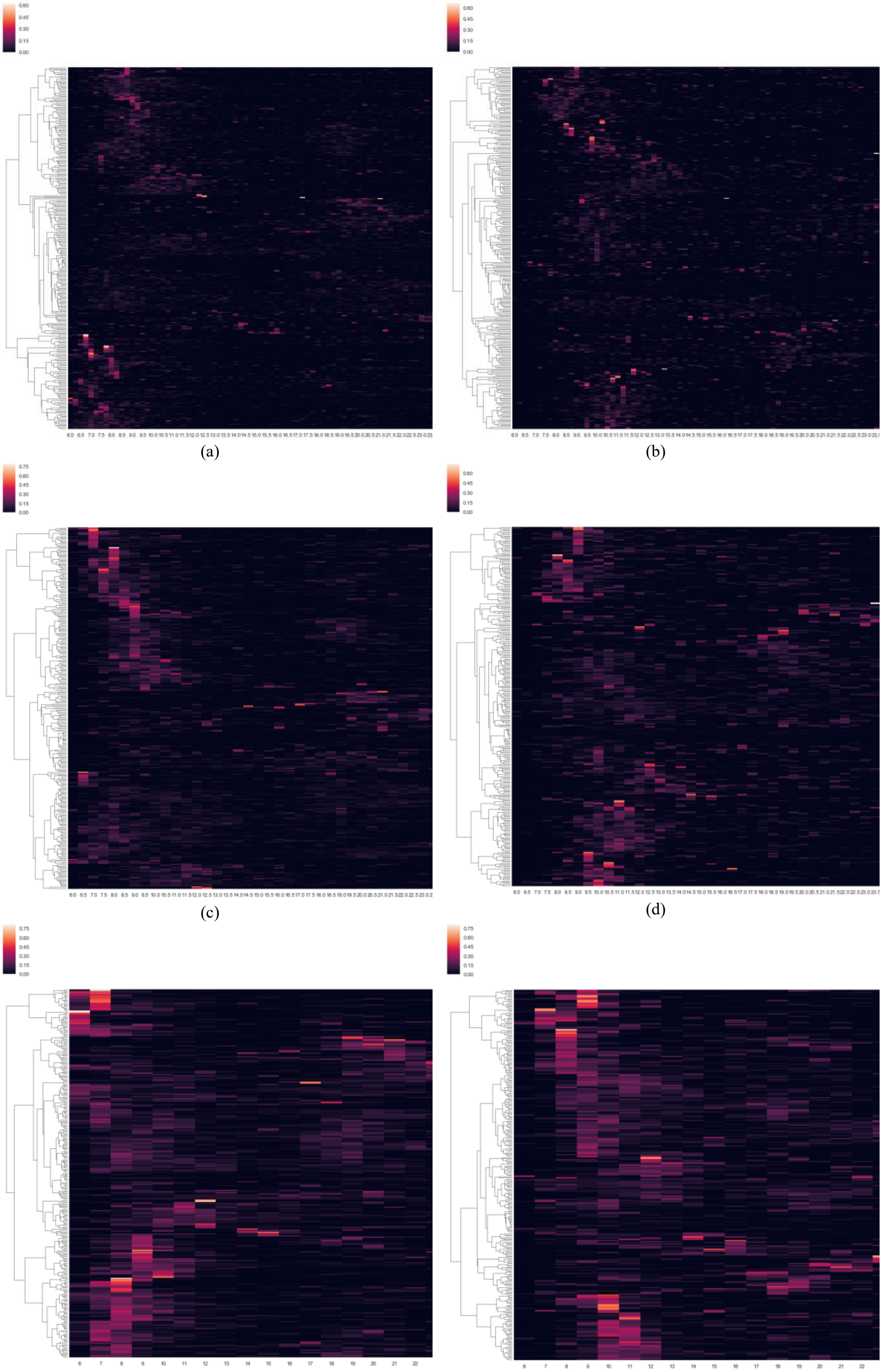
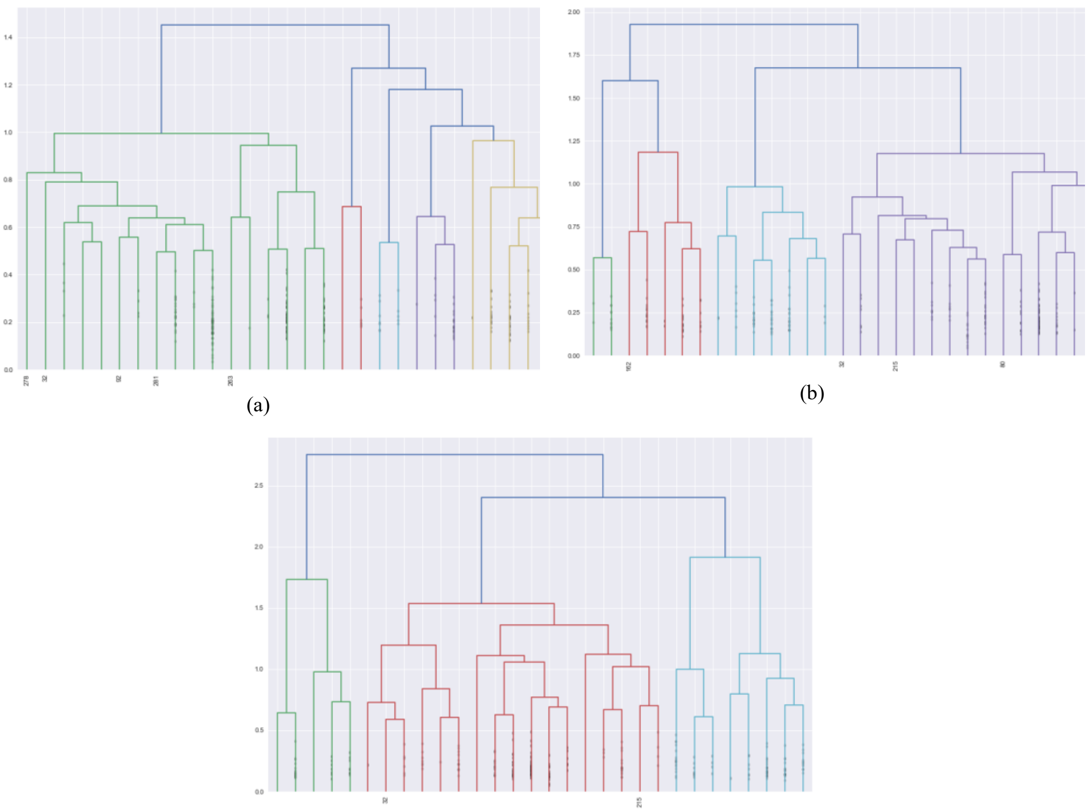


Figure 12. Result of applying HAC to 306 households for the shower event at a) a 15-minute interval on weekdays, b) a 15-minute interval on weekends, c) a 30-minute interval on weekdays, d) a 30-minute interval on weekends e) a 60-minute interval on weekdays and f) a 60-minute interval on weekends. Here, the x-axis represents the times for different intervals and the y-axis represents the household numbers.

In this experiment, we also sought to determine the most suitable profiling interval among the three intervals. The distance cut-off value can be a good measure for this task, as previously mentioned since a larger distance means a greater difference between the two clusters based on the chosen features. Based on this property, we have examined the distance cut-off values for these three profiling intervals and observed that the 15-minute profiling interval performs merging at the lowest-distance cut-off values compared to the two other intervals. This indicates that the households in each cluster for the 15-minute interval are more similar than those in the two other intervals. We also observed that when using the silhouette coefficient, the average linkage method outperformed other linkage methods at a 15-minute profiling interval. Conversely, when using the Calinski-Harabasz index, we observed that Ward's method outperforms the other linkage methods. This finding aligns with a previous study, where Ward's method also outperformed other linkage methods [55]. Table S1 represents the clustering quality analysis for different linkage methods at different intervals. As two linkage methods performed better at two intervals in two metrics, therefore, to determine the best one, we investigated the characteristics of these two metrics. We observed that the silhouette coefficient provides a score between –1 and +1, where a score towards -1 represents incorrect clustering, and a score towards +1 represent highly dense clustering. For the Calinski-Harabasz index, a higher value represents better-defined clusters. For the TUWPU data set at a 60-minute profiling interval, the Calinski-Harabasz index scores were not as high to represent better-defined clusters. As the silhouette coefficient provides a better understanding of the quality of clustering for the TUWPU, therefore, we concluded that a 15-minute profiling interval is the best profiling interval and the average linkage method is the best linkage method for the TUWPU data set. Therefore, we conclude that the 15-minute profiling interval outperforms the 30-minute and 60-minute profiling intervals. Figure 13 shows a sample comparison between different profiling intervals for the shower event on weekdays. From this figure, we can conclude that the number of clusters varies depending on the profiling interval, and the 15-minute profiling interval provides the best result since it merges at the lowest-distance cut-off values. From this type of dendrogram, it is possible to identify groups of households that have a similar probability of an event occurring at a particular time, which has implications for improved water demand management and profiling.



(a) (b)

(c)

Figure 13. Comparison among different profiling intervals for the shower event on weekdays using dendrograms: a) 15-minute profiling interval, b) 30-minute profiling interval and c) 60-minute profiling interval.

The insights found from the TUWPU data set after performing HAC might be valuable to water utility companies, researchers and policymakers in better understanding the behaviours and habits for the given sample. Such insights can be used to determine more effective water-conservation programs, as well as manage demand to reduce operational costs and improve the overall efficiency of the system. Furthermore, the outcomes from this study can be used in a recommender system for promoting water-conscious behaviours, by identifying households with similar patterns and suggesting activities that are popular among their neighbours (i.e., similar households).

6. Conclusions and Future Work

Water scarcity is an ongoing concern for many cities, placing ever more importance on water conservation. Water utilities are facing a significant challenge to ensure that there is enough water for everyone during prolonged drought. In addition, efficiently providing sufficient water pressure during hours of peak demand is another challenge that troubles some utility companies. Data collected from DWMs allow them to overcome these challenges. However, understanding consumers and their water usage behaviour are essential in devising effective strategies for water conservation and demand management.

In this study, we have performed a comprehensive analysis of residential DWM data by clustering consumers by consumption patterns. Two data sets were prepared, and clustering techniques were applied based on the nature of the data and the objective of the clustering. For the EF data set, we found that the optimal number of clusters is five and k-means is the most suitable clustering technique, thus we applied the k-means clustering technique. Later, we discussed the traits of these clusters. For the TUWPU data set, we applied HAC using different linkage methods and found that the optimum number of clusters varies based on the profiling interval, type of water-consumption event and type of day (i.e., weekday or weekend). We also observed that 15-minute profiling is the most suitable profiling interval in terms of forming clusters based on the similarity of consumer behaviour. The study findings would be beneficial to policymakers and researchers in the water sector. As the data set used in this study was collected across only 10 months, we propose that future work further validates the clustering methods using longer-period datasets from multiple locations. This study was designed based solely on DWM data. However, the inclusion of socio-demographic data could offer more relevant and valuable insights. To this end, further investigation of clustering should be conducted on DWM data with socio-demographic data. In addition, future studies should also focus on feature engineering, as the inclusion of new data should provide more opportunities to create new features. Lastly, future studies can be conducted to understand the characteristics of different groups of consumers in terms of adopting hand hygiene practices.

Funding

This research was funded by the Australian Research Council (ARC), grant number LP160100215, which also includes funding from the following industry partners: Yarra Valley Water, City West Water, Southeast Water, and Aquiba.

References

1. *The Global Risks Report*. 2019, World Economic Forum: Geneva, Switzerland.
2. Behera, B.K., P.K. Rout, and S. Behera, *Move Towards Zero Hunger*, in *Move Towards Zero Hunger*. 2019, Springer. p. 1-35.
3. LaVanchy, G.T., M.W. Kerwin, and J.K. Adamson, *Beyond 'Day Zero': insights and lessons from Cape Town (South Africa).* Hydrogeology Journal, 2019. **27**(5): p. 1537-1540.
4. Nguyen, K.A., et al., *Water demand forecasting with AUTOFLOW© using State-Space approach*, in *8th International Congress on Environmental Modelling and Software*. 2016.

5. Liu, A. and P. Mukheibir, *Digital metering feedback and changes in water consumption – A review.* Resources, Conservation and Recycling, 2018. **134**: p. 136-148.
6. Nguyen, K.A., et al., *Re-engineering traditional urban water management practices with smart metering and informatics.* Environmental Modelling & Software, 2018. **101**: p. 256-267.
7. Cominola, A., et al., *Data mining to uncover heterogeneous water use behaviors from smart meter data.* Water Resources Research, 2019.
8. Saxena, A., et al., *A review of clustering techniques and developments.* Neurocomputing, 2017. **267**: p. 664-681.
9. Rahim, M.S. and T. Ahmed, *An initial centroid selection method based on radial and angular coordinates for K-means algorithm*, in *2017 20th International Conference of Computer and Information Technology (ICCIT)*. 2017, IEEE. p. 1-6.
10. Vieira, P., C. Jorge, and D. Covas, *Efficiency assessment of household water use.* Urban Water Journal, 2018. **15**(5): p. 407-417.
11. Garcia, D., et al., *Big Data Analytics and Knowledge Discovery Applied to Automatic Meter Readers*, in *Real-time Monitoring and Operational Control of Drinking-Water Systems*. 2017, Springer International Publishing: Cham, Switzerland. p. 401-423.
12. Ji, Y., et al. *Analysis of urban residential water consumption based on smart meters and fuzzy clustering*. in *Proceedings - 15th IEEE International Conference on Computer and Information Technology, CIT 2015, 14th IEEE International Conference on Ubiquitous Computing and Communications, IUCC 2015, 13th IEEE International Conference on Dependable, Autonomic and Secure Computing, DASC 2015 and 13th IEEE International Conference on Pervasive Intelligence and Computing, PICom 2015*. 2015.
13. Cheifetz, N., et al., *Modeling and clustering water demand patterns from real-world smart meter data.* Drinking Water Engineering and Science, 2017. **10**(2): p. 75-82.
14. Cominola, A., et al., *Profiling residential water users' routines by eigenbehavior modelling.* 2016.
15. Cardell-Oliver, R., *Water use signature patterns for analyzing household consumption using medium resolution meter data.* Water Resources Research, 2013. **49**(12): p. 8589-8599.
16. Solanas, J.L. and M.R. Cussó, *Multivariate consumption profiling (MCP) for intelligent meter systems: a methodology to define categories and levels.* Water Supply, 2010. **10**(5): p. 710-720.
17. Padulano, R. and G. Del Giudice, *A Mixed Strategy Based on Self-Organizing Map for Water Demand Pattern Profiling of Large-Size Smart Water Grid Data.* Water Resources Management, 2018. **32**(11): p. 3671-3685.
18. Rahim, M.S., et al. *Predicting Household Water Consumption Events: Towards a Personalised Recommender System to Encourage Water-conscious Behaviour*. in *2019 International Joint Conference on Neural Networks (IJCNN)*. 2019.
19. Rahim, M.S., et al., *Machine Learning and Data Analytic Techniques in Digital Water Metering: A Review.* Water, 2020. **12**(1).
20. Rahim, M.S., et al., *Advanced household profiling using digital water meters.* Journal of Environmental Management, 2021. **288**: p. 112377.
21. Rahim, M.S., et al., *Digital Water Meters for Advanced Residential Customer Profiling to Promote Water Conservation*.
22. Guo, G., J. Zhang, and N. Yorke-Smith, *Leveraging multiviews of trust and similarity to enhance clustering-based recommender systems.* Knowledge-Based Systems, 2015. **74**: p. 14-27.

23. Schultz, W., *Social comparison as a tool to promote residential water conservation.* Frontiers in Water, 2019. **1**: p. 2.
24. Han, J., M. Kamber, and J. Pei, *10 - Cluster Analysis: Basic Concepts and Methods*, in *Data Mining (Third Edition)*, J. Han, M. Kamber, and J. Pei, Editors. 2012, Morgan Kaufmann: Boston. p. 443-495.
25. Bolorinos, J., N.K. Ajami, and R. Rajagopal, *Consumption change detection for urban planning: Monitoring and segmenting water customers during drought.* Water Resources Research, 2020. **56**(3): p. e2019WR025812.
26. Wang, J., R. Cardell-Oliver, and W. Liu, *An incremental algorithm for discovering routine behaviours from smart meter data.* Knowledge-Based Systems, 2016. **113**: p. 61-74.
27. Nguyen, K.A., R.A. Stewart, and H. Zhang, *An autonomous and intelligent expert system for residential water end-use classification.* Expert Systems with Applications, 2014. **41**(2): p. 342-356.
28. Nguyen, K., et al., *Next Generation Machine Learning for Urban Water Management.* Water e-Journal, 2020. **5**(1): p. 1-7.
29. Stewart, R.A., et al., *Web-based knowledge management system: linking smart metering to the future of urban water planning.* Australian Planner, 2010. **47**(2): p. 66-74.
30. Zheng, A. and A. Casari, *Feature engineering for machine learning: principles and techniques for data scientists*. 2018: " O'Reilly Media, Inc.".
31. Domingos, P., *A few useful things to know about machine learning.* Communications of the ACM, 2012. **55**(10): p. 78-87.
32. Willis, R.M., et al., *Alarming visual display monitors affecting shower end use water and energy conservation in Australian residential households.* Resources, Conservation and Recycling, 2010. **54**(12): p. 1117-1127.
33. Jorgensen, B.S., et al., *Aligning theory and measurement in behavioral models of water conservation.* Water Policy, 2015. **17**(4): p. 762-776.
34. Makki, A.A., et al., *Revealing the determinants of shower water end use consumption: enabling better targeted urban water conservation strategies.* Journal of Cleaner Production, 2013. **60**: p. 129-146.
35. Willis, R., et al., *End use water consumption in households: impact of socio-demographic factors and efficient devices.* Journal of Cleaner Production, 2013. **60**: p. 107-115.
36. Lam, D. and D.C. Wunsch, *Chapter 20 - Clustering*, in *Academic Press Library in Signal Processing*, P.S.R. Diniz, et al., Editors. 2014, Elsevier. p. 1115-1149.
37. MacQueen, J. *Some methods for classification and analysis of multivariate observations*. in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*. 1967. Oakland, CA, USA.
38. Dueck, D., *Affinity propagation: clustering data by passing messages*. 2009: Citeseer.
39. Cheng, Y., *Mean shift, mode seeking, and clustering.* IEEE transactions on pattern analysis and machine intelligence, 1995. **17**(8): p. 790-799.
40. Ng, A.Y., M.I. Jordan, and Y. Weiss, *On spectral clustering: Analysis and an algorithm.* Advances in neural information processing systems, 2002. **2**: p. 849-856.
41. Zhang, T., R. Ramakrishnan, and M. Livny, *BIRCH: A new data clustering algorithm and its applications.* Data Mining and Knowledge Discovery, 1997. **1**(2): p. 141-182.
42. Sasirekha, K. and P. Baby, *Agglomerative hierarchical clustering algorithm-a.* International Journal of Scientific and Research Publications, 2013. **83**: p. 83.
43. Jolliffe, I.T. and J. Cadima, *Principal component analysis: a review and recent developments.* Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 2016. **374**(2065): p. 20150202.

44. Ringnér, M., *What is principal component analysis?* Nature Biotechnology, 2008. **26**(3): p. 303-304.
45. *Clustering*. [cited 2021 28/03]; Available from: https://scikit-learn.org/stable/modules/clustering.html.
46. Anderson, B., *Pattern Recognition: An introduction*. 2019: Scientific e-Resources.
47. Jain, A.K., M.N. Murty, and P.J. Flynn, *Data clustering: a review.* ACM computing surveys (CSUR), 1999. **31**(3): p. 264-323.
48. Nagy, G., *State of the art in pattern recognition.* Proceedings of the IEEE, 1968. **56**(5): p. 836-863.
49. Baeza-Yates, R.A., *Introduction to Data Structures and Algorithms Related to Information Retrieval*. 1992.
50. Minitab. *Linkage methods for Cluster Observations*. 2019 [cited 2020 15/06/2020]; Linkage methods for Cluster Observations]. Available from: https://support.minitab.com/en-us/minitab/18/help-and-how-to/modeling-statistics/multivariate/how-to/cluster-observations/methods-and-formulas/linkage-methods/.
51. Lachev, T. and E. Price, *Applied Microsoft Power BI: Bring your data to life!* 2018: Prologika Press.
52. Waskom, M., et al., *Seaborn: statistical data visualization.* URL: https://seaborn. pydata. org/(visited on 2017-05-15), 2014.
53. WHO, W.H.O., *Water, sanitation, hygiene and waste management for COVID-19: technical brief, 03 March 2020*. 2020, World Health Organization.
54. House, L.W. and J.D. House, *Shifting the timing of customer water consumption.* Journal-American Water Works Association, 2012. **104**(2): p. E82-E92.
55. Milligan, G.W. and M.C. Cooper, *Methodology review: Clustering methods.* Applied psychological measurement, 1987. **11**(4): p. 329-354.

Appendix A. Supplementary materials

Table S1. Clustering quality analysis for different linkage methods at different intervals

Event	Type of day	Linkage method	15 minutes interval		30 minutes interval		60 minutes interval	
			Silhouette score	Calinski-Harabaz index score	Silhouette score	Calinski-Harabaz index score	Silhouette score	Calinski-Harabaz index score
Shower	Weekday	average	**0.448506853**	8.527780474	0.389057434	5.959777119	0.308926746	9.180718971
		complete	0.372250317	7.911656809	0.122297591	15.51640278	0.239199749	17.06467639
		single	0.427971072	5.783713894	0.363332933	4.466860595	0.316601415	6.925762188
		ward	0.064585126	16.48461465	0.085704233	23.35527664	0.131193544	**38.18387255**
	Weekend	average	**0.387995798**	4.969430565	0.350154662	4.66748129	0.257707829	6.091642695
		complete	0.279966194	6.599699951	0.169437378	7.613812463	0.146911062	14.73137115
		single	0.368262744	4.756962776	0.314006442	4.490504089	0.194980232	2.840232588
		ward	0.054635753	10.45767908	0.05713935	17.53167742	0.108168696	**29.74925878**
Bathtub	Weekday	average	**0.400183455**	6.846952511	0.373110812	10.51186476	0.348201581	14.25684781
		complete	0.312699229	8.869961496	0.304695483	16.15385684	0.370836873	22.35718599
		single	0.367654212	3.758340884	0.327866855	5.689559169	0.258029231	4.217600053
		ward	0.295876159	12.46834084	0.20133816	18.6062978	0.200069113	**34.67335942**
	Weekend	average	**0.461686018**	9.093022361	0.438843043	10.83575144	0.414427679	21.86735394
		complete	0.395908944	9.271544696	0.386051994	17.86323387	0.367811647	35.01426593
		single	0.401628429	4.850188956	0.395706999	3.740394846	0.276570993	3.006134119
		ward	0.355189864	16.13754607	0.297756946	23.48946723	0.383651228	**42.70918856**
Washing machine	Weekday	average	0.283630271	5.064187824	0.281818241	5.817987705	0.288801034	10.27674789
		complete	**0.29916421**	7.429594973	0.120457566	10.90812392	0.134126337	22.390375
		single	0.238873969	2.720056885	0.239247221	3.989578345	0.175324957	3.37178633
		ward	0.111866528	9.842611968	0.097792561	17.79484028	0.154892491	**37.58881767**
	Weekend	average	**0.260427015**	3.210386585	0.23152945	6.461625555	0.214952245	6.408004666
		complete	0.123262971	4.77034938	0.174960716	14.50455964	0.17552722	33.89088256
		single	0.247076743	2.702268244	0.199163182	3.059571985	0.119874696	2.78488543
		ward	0.124873229	11.97837429	0.165163019	22.23208569	0.191602706	**43.52112063**
Dishwasher	Weekday	average	**0.483852879**	6.373322196	0.431564782	6.462600815	0.40715529	12.21027227
		complete	0.40210014	9.269771751	0.371042259	16.13988977	0.357417046	19.70524038
		single	0.469878616	4.885980105	0.380708779	3.998327193	0.375620594	4.019743989
		ward	0.233929965	13.20546727	0.27033453	21.9835683	0.326593531	**36.57076642**
	Weekend	average	**0.45981917**	5.360905564	0.45143523	9.36519768	0.41383852	9.280031376
		complete	0.390831996	8.243764828	0.40428604	10.2314595	0.375240228	13.80016539
		single	0.447383489	4.144558588	0.410585569	5.613689998	0.318238961	4.497633016
		ward	0.395046889	11.10108041	0.292073135	15.70810369	0.326733657	**25.72856855**
Irrigation	Weekday	average	0.423492042	5.902786329	**0.455812809**	10.23541785	0.393684221	17.54319606
		complete	0.281541077	9.600119056	0.432439422	13.19444842	0.374403278	17.73854126
		single	0.386334541	5.577191876	0.377082886	8.177564031	0.206190944	4.65424353
		ward	0.237771667	10.14647307	0.291315195	15.8600189	0.208685034	**24.22976972**
	Weekend	average	0.400884363	4.061446995	**0.404678698**	7.747732313	0.383231008	16.97523022
		complete	0.288535336	6.555188651	0.173686917	13.3046751	0.32607151	18.24875331
		single	0.394944274	4.373974615	0.349175947	4.308400813	0.206098349	2.74666702
		ward	0.241033094	9.71990254	0.230537101	16.01134383	0.251962115	**28.24239568**

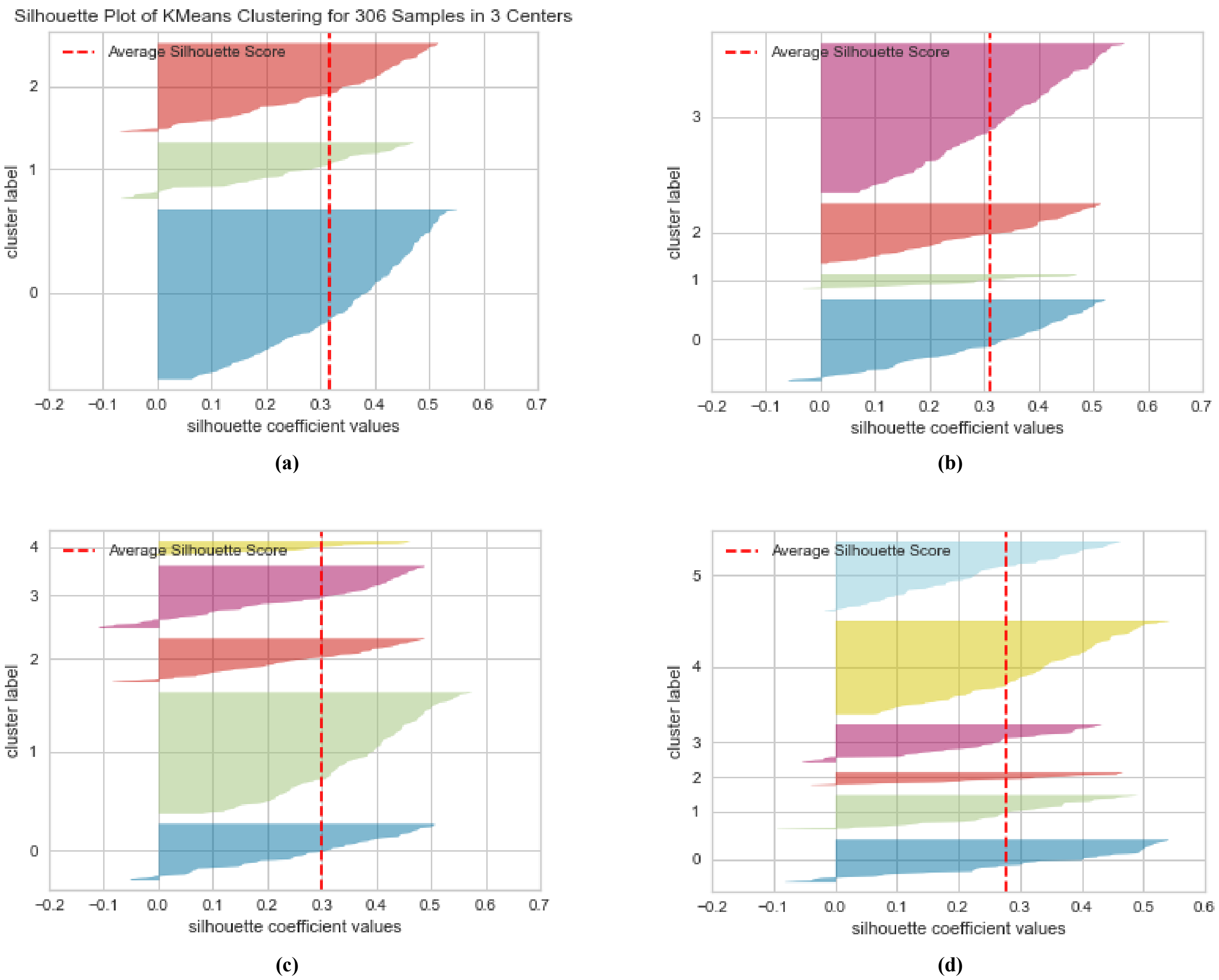


Figure S2. Silhouette plot of k-means clustering for 306 households in (a) 3, (b) 4, (c) 5, and (d) 6 clusters.

4.4 Paper IV (Block D)

Recommender systems in behaviour change interventions: A generalised model with applications in water conservation

Md Shamsur Rahim
Australian Artificial Intelligence Institute, School of Computer Science
University of Technology Sydney, Sydney, Australia
e-mail: mdshamsur.rahim@student.uts.edu.au

Khoi Anh Nguyen, Rodney Anthony Stewart
School of Engineering and Built Environment
Griffith University, Gold Coast, Australia
e-mail: k.nguyen@griffith.edu.au; r.stewart@griffith.edu.au

Damien Giurco
Institute for Sustainable Futures
University of Technology Sydney, Sydney, Australia
e-mail: damien.giurco@uts.edu.au

Michael Blumenstein
Australian Artificial Intelligence Institute, Faculty of Engineering and IT
University of Technology Sydney, Sydney, Australia
e-mail: michael.blumenstein@uts.edu.au

Recommender systems in behaviour change interventions: A generalised model with applications in water conservation

Abstract: Behaviour change interventions are aimed at influencing the psychological variables that result in behaviour. Over the past decade, online intervention methods have become increasingly common due to the wide adoption of the Internet. As part of the range of online approaches, computer-tailored interventions have been investigated for changing behaviour yet remain underused. To overcome the limitations of computer-tailored interventions, recommender systems (RSs) are considered a potential approach. Moreover, many studies have reported the success of RSs in behaviour change interventions. However, no generic model has been proposed that would provide guidelines for designing a successful intervention using RSs. In the water industry, the application of RSs in behaviour change interventions for promoting water-conscious behaviour is a new research topic. This study presented a generic model based on behaviour, an RS, and user engagement theories and created a prototype of a RS for behaviour change interventions. A survey questionnaire was conducted to understand people's attitude towards the system, and the results indicated a positive attitude among the participants. The findings of this study indicated the acceptance and potential impact of the system in saving water.Future research directions for the proposed model and the RS for promoting water-conscious behaviour are also outlined.

Keywords: behaviour change interventions, behaviour change methods, recommender systems, water conservation, user engagement.

1. Introduction

Behaviour change can be challenging for anyone, whether in the long or short term. To overcome this challenge, behaviour change scientists have proposed many effective methods (Peters, De Bruin & Crutzen 2015). Such methods employ theory-based change processes to influence the psychological variables that result in a behaviour (Peters 2018). Due to the wide adoption of technology and the Internet, many of these methods have investigated the effectiveness of online and computer-tailored interventions over the past decade (Short et al. 2015). Several systematic review studies have reported the efficacy of online behaviour change interventions and provided insights into the factors associated with intervention success (Brouwer et al. 2011; Davies et al. 2012; Kuijpers et al. 2013; Webb et al. 2010). These online/web/Internet-based behaviour change interventions mainly deliver information about changing behaviour via the Internet. On the other hand, computer-tailored interventions create a personally tailored intervention based on an individual's data gathered through questionnaires. In such cases, a computer program analyses the behaviour and motivational characteristics of the participant from the collected data and then sends tailored feedback for the intervention (Cheung et al. 2019). Although computer-tailored interventions are cost-effective and their effectiveness at changing behaviour has been demonstrated, they are underused due to the following limitations: a lack of engagement, which results in a high dropout rate; minimal or incorrect use of recommendations; and few logins. To foster the user experience (UX) in tailored behaviour change programs, recommender systems (RSs) are considered a promising approach (Cheung et al. 2019).

An RS is defined as an intelligent system that combines software tools and technologies to recommend a list of items that are most likely of interest to the user (Burke 2007; Resnick & Varian 1997). Today, the applications of RSs are not only limited to increasing sales; they are widely adopted in various domains such as the entertainment, content, e-commerce, services, and social domains (Montaner, López & De La Rosa 2003; Ricci, Rokach & Shapira 2015). In an RS, the user refers to the entity to whom the recommendations are made, and the item is denoted as the product that is recommended to the user. Generally, the goal of an RS is to assist individuals who lack the experience or competence to select a potential item from an overwhelming number of alternatives offered by a service provider (Ricci, Rokach & Shapira 2015). As RSs can assist individuals in selecting potential items from many options, many studies in recent years have investigated their potential in behaviour change interventions.

Although, it can be argued that any applications of RSs can be considered as an attempt to change behaviour, however, the focus of this research is to only consider those applications that explicitly investigated the applications of RSs in behaviour change interventions. Therefore, in this study, we excluded the applications of RSs as these applications are more focused on suggesting items by predicting a list of items of interest. Based on our criteria, the recent applications of RSs in behaviour change interventions mainly fall into two categories: (i) public health and (ii) resource conservation. Although many studies have investigated the application of RSs in behaviour change interventions, no generic model exists that can be used to investigate the integration of RSs with behaviour change interventions. Since RSs and behaviour change interventions are two different disciplines, a generic model can reduce the time and effort required to design such a study for students, researchers, and practitioners. In addition, a generic model can be adopted in new behaviour change intervention studies, such as those aimed at changing people's water consumption behaviour as well as demand patterns for promoting water-conscious behaviour.

For many urban water utilities, the two ongoing challenges influenced by householder behaviour are water scarcity during prolonged drought and demand management during the hours of peak demand to avoid low water pressure. Among the numerous initiatives, the rollout of digital water meters (DWMs) is a potential solution for overcoming these challenges. Several studies have utilised data from residential DWMs, which can be categorised into five categories: (1) water demand forecasting; (2) socioeconomic analysis; (3) behaviour analysis; (4) water event categorisation; and (5) water-use feedback (Rahim et al. 2020). From these existing studies, it has been identified that the impact on water conservation depends on the level of personalisation: water conservation programs with higher personalisation achieve a greater reduction in water consumption (Rahim et al. 2020). This finding indicates the application of an RS in the water industry for promoting water-conscious behaviour would be novel.

According to Kok (2018), the major challenges with any behaviour change studies include the following: "*1) the correct identification of the change objectives (and thereby the evaluation outcomes), 2) the selection and application of appropriate behavior change methods in an intervention, and 3) adequate implementation of the intervention.*" The inclusion of RSs in this context for behaviour change makes it even more challenging. Though there are studies that investigated the applications of RSs in behaviour changes, to the best of the authors' knowledge, there is a lack of a generalisable model to integrate RSs successfully in behaviour change intervention. In this study, we proposed a generic model to address the research gap in the integration of RSs in behaviour change interventions. The framework was developed based on behaviour, an RS, and user engagement theories. Later, we designed the prototype of an RS based on the proposed framework for promoting water-conscious behaviour among residential customers. To verify the acceptance and potential impact of the RS, we surveyed 300 participants. After examining the survey responses, we concluded that their attitude towards the RS for promoting water-conscious behaviour was positive, which indicated the early acceptance and potential success of the RSs.

The main contributions of this paper are as follows:

a) A generic model is proposed that can be used to design any behaviour change interventions using an RS;
b) A prototype RS is designed for the water industry for promoting water-conscious behaviour; and
c) The acceptance and potential impact of the RS are verified through a survey.

The remainder of the paper is structured as follows: Section 2 discusses the related works, Section 3 presents the proposed model and its application in water conservation, Section 4 provides the result of the experiment, Section 5 discusses the findings, and finally, Section 6 draws the conclusions.

2. Related works

In recent years, the application of RSs in behaviour change interventions has become popular. However, to the best of the authors' knowledge, no research has addressed the generic approach of how RSs can be integrated with behaviour change interventions. Therefore, in this section, we critically analyse the studies that have applied RSs in behaviour change intervention as well as other closely related studies.

As previously mentioned, the applications of RSs in behaviour change interventions are mainly categorised into two categories: (i) public health and (ii) resource conservation. The public health studies have investigated the application of RSs in health programs (e.g., smoking cessation, nutrition, and mental health) by changing behaviours. The resource conservation studies have aimed to understand the impact of recommendations on resource (i.e., energy) consumption through modifying participants' behaviour. For the sake of relevancy, we only critically analysed studies that have explicitly used RSs for behaviour change interventions.

The applications of RSs in public health for behaviour change interventions can be categorised into three types: (1) generic health promotion, (2) disease relief, and (3) disease prevention (Cheung et al. 2019). The first category of RS application in public health for behaviour change interventions is generic health promotion. Generic health promotion interventions provide recommendations to the targeted population for changing their behaviour to improve their health. A case-based reasoning RS was proposed by Hidalgo et al. (2014) to improve the habits and knowledge about patients' diseases based on their recorded data and physician preferences. In this approach, the rules were first created by the physicians based on previous patients with the same disease. Then, once the patients provided the data, recommendations were created by comparing the data with historical data within the rules. In another study, a knowledge-based based RS was introduced for integrating physical activity in the daily lives of elderly people (Palumbo et al. 2020). In a similar study, an RS was proposed based on similar people and ontology trees to recommend leisure-time physical activities among Japanese adults (Sami et al. 2008).

The second category of RS application in public health for behaviour change interventions is disease relief. Such studies focus on the application of RSs for relieving disease or pain by changing patients' behaviour. For instance, to relieve diabetes, the application of RSs has been investigated by Norouzi et al. (2018) and Hammer, Kim, and André (2010). Specifically, Norouzi et al. (2018) proposed a comprehensive knowledge-based dietary RS. Moreover, highly personalised recommendations for patients' long-term health based on their short-term preferences were generated using a knowledge-based hybrid RS by Hammer, Kim, and André (2010). Other notable applications of RSs in disease relief include for hypertension by Radha et al. (2016), for depression by Rohani et al. (2020) and Zhang et al. (2019), for smoking cessation by Adams et al. (2014) and Sadasivam et al. (2016), and finally for treating lower back pain by Esteban et al. (2014).

The third category of RS application in public health for behaviour change interventions is disease prevention. Such studies aim to prevent diseases by changing peoples' behaviour using RSs. As a part of disease prevention, to reduce people's weight, a collaborative RS was proposed by Hales et al. (2016) that included diet, physical activity, and weight tracking features. Furthermore, for recommending healthy meals and recipes Espín, Hurtado, and Noguera (2016) proposed a hybrid RS and Ge et al. (2015) introduced a collaborative RS.

The second and final category of RS applications in behaviour change interventions is resource conservation. Such studies aim to investigate the application of RSs in behaviour change to conserve resources such as electricity. All of these studies have focused on saving electricity using RSs. For instance, Onile, Belikov, and Levron (2020) presented the concept of an RS that would generate consumer-oriented recommendations. Moreover, deep reinforcement learning was used in an RS for reducing energy consumption in buildings by Wei et al. (2020). Lastly, a hybrid RS was proposed by (Sardianos et al. 2020) for changing habits related to energy consumption in an office environment. Table 1 summarises the findings of the relevant studies.

Table 1. Summary of the findings of relevant studies

Category	Intervention type	Subcategory	Authors	Recommender system approach	Theoretical basis of behaviour change	Proposed generalisable approach?
Health	Generic health promotion	-	(Hidalgo et al. 2014)	Case-based reasoning	×	No
		-	(Palumbo et al. 2020)	knowledge-based	Social Cognitive Theory	No
		-	(Sami et al. 2008)	Other	×	No
	Disease relief	diabetes	(Norouzi et al. 2018)	knowledge-based	×	No
			(Hammer, Kim & André 2010)	Knowledge-based hybrid	×	No
		hypertension	(Radha et al. 2016)	Other	×	No
		depression depression	(Rohani et al. 2020)	content-based (CB)	Behavioral Activation	No
			(Zhang et al. 2019)	Other		No
		smoking cessation	(Adams et al. 2014)	collaborative filtering	×	No
			(Sadasivam et al. 2016)	Hybrid	×	No
		Lower back pain	(Esteban et al. 2014)	hybrid	×	No
		weight loss	(Giabbanelli & Crutzen 2015)	Collaborative	Health games	No
	Prevention	weight loss	(Hales et al. 2016)	Collaborative	Social Cognitive Theory, Ecological Perspective – Interpersonal Level	No
		Nutrition	(Espín, Hurtado & Noguera 2016)	hybrid	×	No
			(Ge et al. 2015)	Collaborative	×	No
Resource conservation	Energy	Commercial building	(Onile, Belikov & Levron 2020)	content-based (CB)	Social comparison	No
		Office	(Sardianos et al. 2020)	hybrid	×	No
		Commercial building	(Wei et al. 2020)	Other	×	No

After analysing the applications of RSs in generic health promotion interventions, we concluded that the applications of RSs are gaining momentum in behaviour change interventions. However, these studies share a common limitation, namely that they do not provide any generalised approach or model that can be used to design a new study. As the existing studies are aimed to solve specific problems, they have not provided any guideline or generic framework that can be adapted to behaviour change interventions using RSs. Moreover, many of the studies have not considered any behaviour change theories while integrating RSs in behaviour change interventions. Therefore, there is a clear research gap of a generic model that addresses the integration of RSs with behaviour change theories.

3. Proposed model and its application in water conservation

Although the applications of RSs are common in many behaviour change intervention studies, no common framework exists for designing any type of behaviour change intervention using RSs. This section presents a framework based on behaviour theories, RS theories, and user engagement theories that can be adopted in any behaviour change intervention studies using RSs. Furthermore, to measure the effectiveness of the proposed framework, we applied it in water conservation and conducted a survey to understand participants' feedback. Depending on the theories covered or the nature of the tasks in the framework, the proposed framework can be categorised into four sections: (1) behaviour theories, (2) intervention design using RSs, (3) user engagement theories, and (4) deployment and evaluation. Figure 1 illustrates the different components of the proposed model and the relevant literature to inform its development is discussed in the following sub-sections.

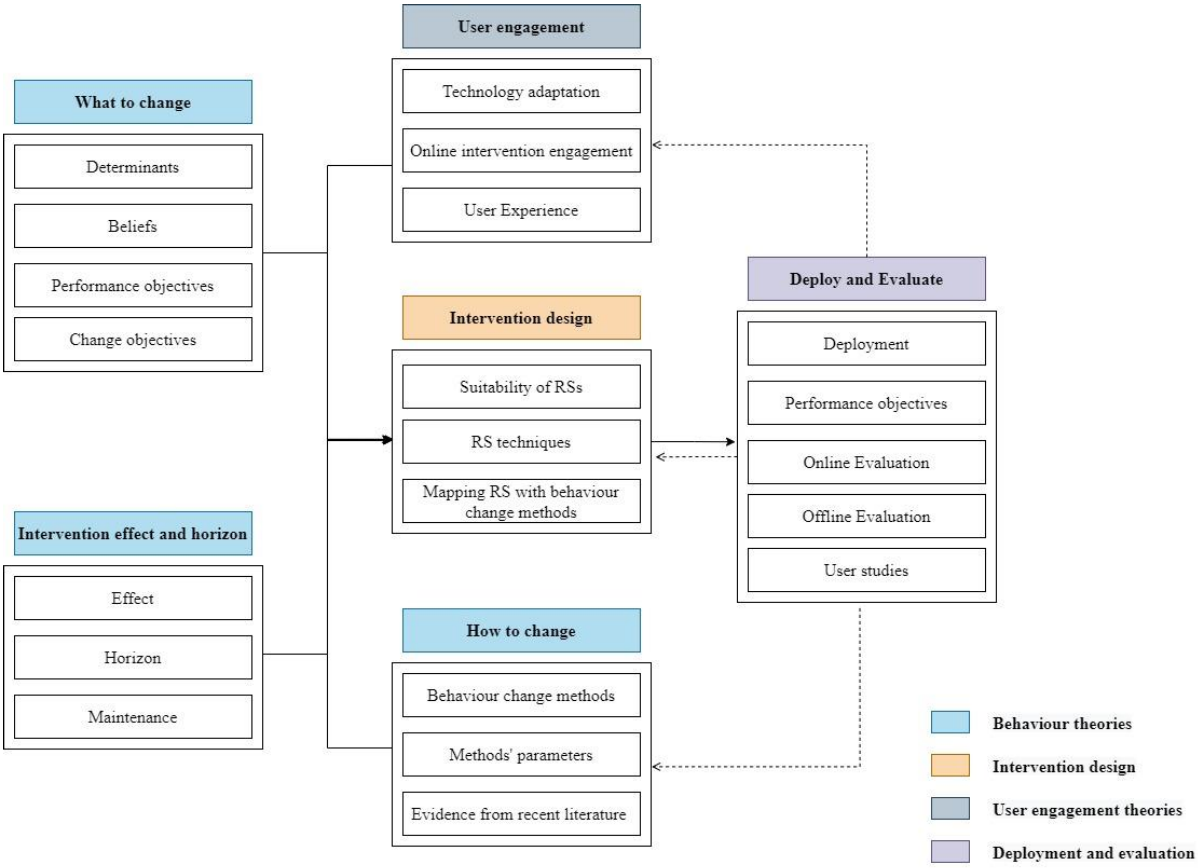


Figure 1. The proposed generic model of behaviour change interventions using RSs.

3.1 Behaviour theories

The first section of the proposed framework is based on behaviour theories. This section covers the knowledge required to conduct a behaviour change intervention experiment, which is divided into three subsections: (1) what to change, (2) intervention effect and horizon, and (3) how to change.

3.1.1 What to change

Before starting any behaviour change intervention study, one must identify what needs to change. As no method directly operates on behaviour, it is important to identify the relevant determinants and beliefs in any steps of a behaviour change intervention (Peters 2018). Along with the determinants and beliefs, one must also determine the performance objectives; however, identifying them incorrectly could lead to intervention failure. Here, we provide an overview of the determinants, beliefs, and performance objectives, and then we discuss them in the context of promoting water-conscious behaviour.

Determinants and beliefs were defined by Peters (2018) as follows: “determinants are convenient categories of functionally similar or functionally related sub-entities which we call beliefs.” There are several ways to identify determinants and beliefs (i.e., systematic reviews, interviews, and surveys). Based on the theories of behaviour, behavioural determinants include intention, self-efficacy, perceived behavioural control, outcome expectancy, response-reinforcement contingencies, and discriminative stimuli (Michie et al. 2008). In general, behaviours consist of preparatory and/or sub-behaviours, which can be limited (van Empelen & Kok 2006); however, it not might be feasible to compile an exhaustive list (Peters 2018). Therefore, understanding which preparatory and/or sub-behaviours are to be influenced is important as the determinants can vary depending on these behaviours. According to Peters (2018), any behaviour change intervention “has several behavioural objectives, each of which concerns performance of a preparatory or sub-behaviour,” which are defined as the performance objectives (Bartholomew et al. 2006). An intervention is said to be effective when all performance objectives have been achieved. The next paragraph discusses the use of these determinants, beliefs, and performance objectives for promoting water-conscious behaviour.

The promotion of water-conscious behaviour has two aspects: the first involves changing the consumer's water consumption behaviour to conserve more water, and the second involves shifting the timing of an event for demand management. From the perspective of habitual behaviour, the determinants of water-conscious behaviour are attitude, awareness, social norms, self-efficacy, and intention. Based on the theory of reasoned action proposed by Fishbein (Fishbein 1979), Kantola et al. (Kantola, Syme & Campbell 1982) found that subjective norms were the most influential determinants of the intention to conserve water. Another essential theory to consider for determining determinants is practice theory (Rouse 2007). The main idea is that people do not consume water directly but rather do so to achieve a social practice (Shove, Pantzar & Watson 2012) where the participations participate for longer or shorter time, occasionally or regularly and there could be several reasons depending on the context such as cleaning or cooking. According to Gram-Hanssen (2014), four elements influence practices: (1) embodied habits, (2) institutional knowledge, (3) engagements, and (4) technologies. In Figure 2, we present a sample matrix of our findings regarding determinants, beliefs, and performance objectives for promoting water-conscious behaviour.

Target population: DWM residences Target behaviour: Conserving water and shifting time	Determinant 1: Attitude	Determinant 2: Subjective norms	Determinant 3: Self-efficacy
PO1: Decide to conserve water	C01.1.1 Expresses that water conservation is beneficial	CO1.2.1 Describes others' approval of taking a shorter shower	
PO2: Replace showerhead with a water-efficient one	CO2.1.1 Expresses that an efficient showerhead can save water		CO2.3.1 Replaces showerhead

Figure 2. Sample matrix of change objectives for promoting water-conscious behaviour.

The matrix presented in Figure 2 is called the matrix of change objectives (Bartholomew et al. 2006), where the rows represent performance objectives, the columns represent determinants, and the cells represent change objectives. For each water consumption event, to conserve and shift the timing, there are separate performance objectives. After identifying the determinants, beliefs, and performance objectives, the next step is to identify the intervention effect and horizon.

3.1.2 Intervention effect and horizon

As part of a behaviour change intervention, it is crucial to determine the intervention effect and horizon. By intervention effect, we mean whether the behaviour change is temporary, and by horizon, we mean how long the behaviour change would last. For instance, during a severe drought and the highest level of water restrictions, people might need to change their behaviour to not using sprinklers at any time. Such behaviour change can be temporary (i.e., only during a severe drought), but it should be effective immediately. By contrast, taking a shorter shower could be a permanent behaviour change in the long term. Most behaviour change interventions effectively achieve temporary behaviour change and behaviour change maintenance is rarely accomplished (Kwasnicka et al. 2016); therefore, it is essential to consider the intervention effect, horizon, and behaviour change maintenance. Although the concept of behaviour maintenance theory requires further development, findings from a systematic literature review suggested that at least one sustained motivator is required by individuals for maintaining behaviour, which may include behaviour enjoyment, satisfaction, self-determination, or an experience of behavioural congruence (Kwasnicka et al. 2016).

For water conservation and demand management, some behaviours require a one-time or temporary change (e.g., replacing one's showerhead with a water-efficient one, fixing leaks, and not using sprinklers). However, some behaviours need to be maintained in the long term (e.g., taking a shorter shower and using a washing machine at full load). Therefore, the intervention effect and horizon should be determined for each performance objective. For behaviour change maintenance, we designed our intervention in such a way that it could track changes in a behaviour. The details of the intervention design are discussed in Section 3.2. The following subsection discusses how to change behaviours based on intervention methods and techniques.

3.1.3 How to change

Once the intervention effect and horizon have been identified with determinants, beliefs, and performance objectives, the next step is to determine how to change the behaviour, which includes identifying the intervention methods. To be successful, it is essential for a behaviour change method to target a determinant that predicts behaviour and for it to be able to change that determinant after translating it into a practical application while maintaining effectiveness (Kok et al. 2016). To determine a suitable behavioural change method, an intervention mapping (IM) taxonomy can be used. The IM taxonomy proposed by Kok et al. (2016) can be consulted to identify the initial list of change methods for each determinant. For instance, to increase knowledge, the initial list of methods may include chunking, which uses stimulus patterns, and discussion, which presents a summary of the material. Moreover, to change the habitual behaviour, the initial list of methods may contain deconditioning, counterconditioning, implementation intentions, cue altering, and stimulus control among others. After identifying the potential methods, the next step is to determine the feasibility of each method from their definition and parameter from the IM taxonomy (Kok et al. 2016). While identifying the suitable methods based on their definitions and parameters, it is also crucial to collect evidence from the recent related literature because the effectiveness of the parameters and strength of the change methods may vary in new literature. In addition, new studies might have identified the emerging method and which methods to avoid (Kok et al. 2016).

We identified the suitable change methods and their parameters for each determinant for promoting water-conscious behaviour. For instance, to change awareness and risk perception towards water-conscious behaviour, we identified that a consciousness-raising method based on the precaution adoption process model (Weinstein, Sandman & Blalock 2020) and the trans-theoretical model (Prochaska & Marcus 1994) would be a suitable method. Table 2 presents a summary of selected change methods for changing water consumption behaviour. A recent study on water conservation concluded that the effectiveness of water conservation varies depending on the four communication strategies adopted from psychological literature (Seyranian, Sinatra & Polikoff 2015), which were (1) social norms, (2) social identity, (3) personal identity, and (4) the knowledge deficit approach. In that study, for social norms, water consumption was compared with a neighbour using social approval (i.e., a sad face for consumption above the mean or a happy face for consumption below the standard). For social identity, short communications on water consumption behaviour were performed using highly inclusive language. In the case of personal identity, it was the same as social identity except that personal identity language was used instead of highly inclusive language. Finally, the knowledge deficit model provided a series of water-saving tips. Based on the findings, it was concluded that social norms, social identity, and personal identity methods provide greater water conservation. In another study, alarming visual display monitors were used to give feedback on shower water consumption (Willis et al. 2013; Willis et al. 2010). Although water consumption was reduced by 27% in terms of volume in the shower in the beginning, it reverted to previous levels after four months. This finding implies the necessity of behaviour change maintenance in water conservation. Lastly, in another notable study, Liu et al. (2017; 2017) investigated the impact of water conservation feedback through paper and an online portal. The findings demonstrated that online portal feedback was more effective than paper-based feedback for reducing water consumption.

3.2 Intervention design using RSs

In this section, we discuss the characteristics of RSs, different techniques, and algorithms and how they can be integrated into behaviour change interventions. Next, we present our argument for selecting RSs in water conservation and a suitable technique with an appropriate algorithm.

3.2.1 RSs and their characteristics

An RS is a combination of software tools and techniques that usually suggests a list of top-n items that are most likely of interest to a particular user or a group of users with similar tastes or preference (Burke 2007; Resnick & Varian 1997; Ricci, Rokach & Shapira 2015). In an RS, the user is referred to as the entity to whom the recommendations are made, and the item is denoted as the product that is recommended. The concept of an RS is to infer user interests from user-item preference data (Aggarwal 2016). Depending on the platform and types of items, preferences can be divided into two categories, namely explicit and implicit preferences. Explicit preference data include users' unambiguous interactions with things such as user ratings, reviews, or votes. By contrast, implicit preferences include users' interactions with items such as clicks, views, purchases, or follows. For instance, on an e-commerce website, when a user purchases a particular item, it can then be marked as an implicit preference. When the user expresses his/her preference regarding a product through rating/voting and/or review, then it is considered an explicit preference.

Regardless of the different forms of expression of user preferences, each RS is built with specific aims that benefit both the user and service provider. From a user perspective, the primary objective of an RS is to assist individuals with less experience or less competence to select a potential item from an overwhelming number of alternatives that service providers offer (Ricci, Rokach & Shapira 2015). For instance, a restaurant RS can recommend a personalised shortlist of potential restaurants to its users from thousands of restaurants. This type of personalised recommendation is beneficial to different users or user groups for its ability to provide tailored suggestions according to their needs. To achieve personalisation in RSs, the user profiling technique can be adopted. User profiling can be defined as the "system of collecting, organizing and inferring the user profile information," and a user profile is a "summary of the user's interest, characteristics, behaviours, and preferences" (Eke et al. 2019).

Based on the aforementioned characteristics of RSs, we concluded that RSs may be effective in behaviour change interventions where the items are a list of activities or messages that need to be suggested to entities or users based on their preferences for performance objectives. To understand user preferences, explicit preferences can be gathered in ratings, reviews, or votes, whereas implicit preferences can be derived from user activities or actions (e.g., following a healthy diet according to recommendations and performing suggested physical activities). To achieve personalisation and store preferences, the user profile information may vary depending on the behaviour change interventions.

For promoting water-conscious behaviour, we observed that an RS could be integrated into a behaviour change intervention. With the help of advanced household profiling (Rahim et al. 2021), it would be possible to obtain the characteristics, behaviours, and habits of the members of a household based on which recommended items are suggested in terms of activities (e.g., taking a shorter shower, replacing the showerhead with a water-efficient one, and shifting the time of a particular event). We proposed that such a system would collect both explicit and implicit preferences. Explicit preference data would be extracted from ratings of the suggested activities, whereas implicit preference data would be extracted by monitoring and analysing water consumption behaviour.

3.2.2 Identifying suitable RS techniques

For successful behaviour change interventions using RSs, it is essential to identify a suitable RS technique. Depending on the nature of the application, the developer of an RS can locate the appropriate technique. In a traditional approach proposed by Burke (Burke 2007), the methods applied in RSs can be distinguished into five classes. This section provides an overview of the techniques used in RSs, and then presents the most suitable technique for water conservation along with the justification.

Content-based: This type of RS learns to suggest items to the user that are similar to the items preferred by them in the past. The system matches the attributes of the user with the characteristics of things, and the similarity among items is derived from the associated features of the compared items (Aggarwal 2016b). For instance, if a user bought a science fiction book in the past, a content-based RS would recommend another science fiction book by inferring that the user is interested in sci-fi books.

Collaborative filtering: Collaborative filtering is considered the most widely deployed technique in RSs (Ricci, Rokach & Shapira 2015). The original idea behind this technique is to recommend items to a user that have been preferred by other users with similar preferences or tastes in the past (Goldberg et al. 1992). The similarity in preference or taste is calculated based on the similarity in the feedback of the users in the past.

Demographic: The idea behind this type of RS is to provide recommendations based on the demographic profile of the user (Bobadilla et al. 2013). Using such information, it is possible to make more relevant and accurate recommendations. For instance, news content can be recommended based on the reader's age and educational background.

Knowledge-based: This technique achieves a high level of customisation for recommendations in a specific domain by incorporating a knowledge base that encodes constraints or similarity metrics of the relevant domain knowledge (Aggarwal 2016a). These systems calculate how much the recommendations (i.e., solutions to a problem) match the user's needs (problem description) through a similarity function (Ricci, Rokach & Shapira 2015).

Hybrid: Hybrid RSs are a combination of any two of the aforementioned techniques. Such a system attempts to use the advantages of technique A to fix the disadvantages of technique B while also benefiting from the latter's advantages (Ricci, Rokach & Shapira 2015). There are several ways to develop a new hybrid RS by combining two or more basic recommendation techniques.

Behaviour change intervention designers must identify the most suitable RS technique after carefully analysing the domain of the application and nature of the intervention. For promoting water-conscious behaviour, we concluded that a hybrid of content- and knowledge-based RSs would be the most suitable technique. The justification behind this conclusion was the nature of the domain of the application and intervention. We aimed to suggest a top list of activities to users that are preferred by other users. We did not wish to recommend any activities that are not preferred by the user but preferred by neighbours or similar users. For instance, shifting the shower time might not be preferred by a user; however, it might be preferred by the neighbours. Alternatively, a household might have flexible water consumption timing; therefore, they might prefer recommendations related to the time-shifting of water consumption events (e.g., shifting the time for using the washing machine or taking a bath). Because of these considerations, we selected a content-based RS, but providing only content-based recommendations might not be helpful. To make the recommendations more effective, the knowledge-based technique might be required for constraints. For instance, to make recommendations regarding the replacement of efficient appliances, encoding constraints from a knowledge base is required. Therefore, for the application of an RS in promoting water-conscious behaviour, we concluded that a hybrid of content-based and knowledge-based RSs would be the most suitable.

3.2.3 Mapping RSs with behaviour change methods

Once the characteristics of the RSs are understood and a suitable recommender technique has been identified, the next step is to map the different components of the RSs with behaviour change methods. Through this mapping, one can determine how the behaviour change methods can be translated into practical applications using the different components of RSs. Here, it should be noted that there could be other complementary elements for translating the change methods into practice. The definition and parameters of each change method can be used to determine the mapping with different features. For instance, for the tailoring method, user profiling can be applied to provide customised feedback.

In this step, we mapped the elements of the RS with the suitable change methods for promoting water-conscious behaviour identified in subsection 3.3.1. This mapping revealed that all the identified methods could be translated into practical applications using different elements of the RS. For instance, individualisation and tailoring methods can be solved using household profiling, belief selection, persuasive communication, and guided practice methods, which can be converted into practical application using the recommended items. Table 2 presents the mapping of RS elements and behaviour change methods for water conservation and demand management.

3.3 User Engagement

User engagement is one of the critical factors for a successful intervention using RSs. We observed that there are three aspects of RSs in behaviour change interventions from a user engagement perspective: (1) technology adaptation, (2) online intervention, and (3) UX. First, as interventions using RSs require technologies (e.g., mobile, web, Internet, and computer technologies), it is vital to understand how and why individuals adopt innovations. Such an understanding has implications for superior outcomes from an intervention using an RS. Second, as most of RSs offer online recommendations, it is also essential to understand the theories behind effective online interventions and apply them accordingly. Lastly, the UX aspect ensures that the interaction with the system provides a pleasant and engaging experience for users. In the following subsections, we discuss these aspects and then present their application in the design of an RS for water conservation.

3.3.1 Technology adaptation

The understanding of how and why individuals accept and use information technology is a mature stream of information system research (Venkatesh, Davis & Morris 2007). This understanding is crucial for the success of an intervention program as it would drive the technology selection process depending on the target population. Several models have been proposed, mainly from psychological and sociological theories. After reviewing eight adaptation models, Venkatesh et al. (2003) proposed a model named the unified theory of acceptance and use of technology (UTAUT). In the UTAUT model, four constructs were theorised as direct determinants of acceptance and usage behaviour, namely (1) performance expectancy, (2) effort expectancy, (3) social influence, and (4) facilitating conditions. In addition, four moderators (i.e., gender, age, experience, and voluntariness of use) of key relationships were formulated (Venkatesh et al. 2003). Later, the UTAUT was extended to the UTAUT2 (Venkatesh, Thong & Xu 2012), which added three more constructs (i.e., hedonic motivation, price value, and habit) into the UTAUT. The UTAUT and UTAUT2 models have implications for assessing the likelihood of adopting new technology and understanding the drivers behind acceptance (Venkatesh et al. 2003), which would be essential for measuring the early success of any intervention study using an RS.

To promote water-conscious behaviour using an RS in an intervention, we carefully designed the program by taking technology adaptation theories into consideration. For instance, in the case of performance expectancy, according to Venkatesh et al. (2003), gender and age moderate behaviour intention, where the effect will be more substantial for young men. Based on these findings, we designed the RS in a way that it could be adopted for elderly populations as well as irrespective of gender.

3.3.2 Online intervention

Online behaviour change interventions have become common over the past decade because of the growing popularity of the Internet (Short et al. 2015). User engagement was one of many factors behind the success of these intervention programs. Here, engagement refers to the quality of UX quantified by increased attention, positive affect, satisfaction, and mastery (O'Brien & Toms 2008). In most cases, the development of online interventions is guided by the theories of behaviour change that focus on psychological determinants. However, it is important to consider the determinants behind user engagement in online intervention as it helps to overcome shortcomings such as poor retention rates, few logins, and lower use of intervention features (Davies et al. 2012; Kelders et al. 2012; Short et al. 2015). To achieve this goal, a model of user engagement in online interventions was proposed by Short et al. (2015). The proposed model was based on three types of factors related to (1) the individual's environment, (2) the individual, and (3) the intervention. In the model, the environment consists of external factors that enable intervention use, such as time, Internet access, and online environment. For the individual, the determinants are the characteristics of the user related to the relevance of the interventions, including online behaviour, Internet self-efficacy, expectation, affect, demographics, psychosocial factors, and current and past behaviour. Finally, under interventions, the determinants include novelty, tailoring, self-monitoring, praise, rewards, reminders, aesthetics, sorting cues, the strength of the argument, and credibility (Short et al. 2015). Any online intervention using an RS should focus on these determinants during the design stage to ensure intervention success.

For our intervention in water consumption behaviour using an RS, we decided to deliver the contents online. Therefore, we focused on the determinants of effective online engagement. For instance, we designed the intervention in such a way that all the recommended items were custom-tailored, users can monitor by themselves, and they are rewarded when they follow the recommendations.

3.3.3 User experience

UX has become a popular term in human–computer interaction (HCI) over the past decade (Hassenzahl & Tractinsky 2006). Norman and Nielson, two well-known usability researchers, defined UX as "all aspects of the end-users interaction with the company, its services, and its products" (Norman & Nielsen). Although the meaning of UX varies widely, the main aim of UX is to meet the user's need without any issues or difficulties. With the wide adaptation of RSs in several application areas, UX has become critical in driving user engagement through encouraging continuous interaction. A recent user study outlined several UX guidelines for recommended content to increase user engagement and satisfaction (Harley 2018). Some of these guidelines were prioritising individualised recommendations, stating the data source, allowing users to fine-tune recommendations, and updating recommendations quickly and often. In that study, in separate recommendations based on categories, the participants suggested that personalised recommendations were effective when they were prioritised over generic items. Stating the source of the data provides insight to users and also adds credibility to the recommended items, which is essential for user engagement. Separating the recommendations based on categories was appreciated by users because it helped to separate the items into smaller chunks instead of lumping all contents into a single category. Lastly, the study concluded that some users appreciated having a method for fine-tuning the recommendations and seeing the reflection immediately (Harley 2018). All of these findings provide clear guidelines for how to use RSs in behaviour change interventions.

We reflected on the UX guidelines for RSs when designing the prototype of our RS for promoting water-conscious behaviour. In our proposed design, all recommendations are personalised and given the highest priority. For each recommended item, the data source is provided with an explanation. Finally, to fine-tune the recommendations, a rating for each content is available, which enables the recommendations to be updated immediately.

3.4 Deployment and evaluation

The last steps in interventions using RSs are deployment and evaluation. In this phase, the RS is deployed by adhering to all the insights from the previous phases. For the deployment, both mobile and web platforms can be used depending on the target population and the aim of the intervention. Before the deployment, historical data can be collected for comparison with the data after the intervention for evaluation purposes.

There are two main ways to evaluate the performance of an interventions using an RS. One is from the behaviour change domain, namely performance objectives and the second is based on the information system domain. Regarding performance objectives, the intervention is considered successful when all the performance objectives have been achieved. In the information system domain, the performances of RS techniques is determined using any of the following metrics: (1) offline experiments, (2) user studies, and (3) online evaluation (Shani & Gunawardana 2011). In offline experiments, a pre-collected user choosing or rating data set is used to simulate the behaviour of users. Using offline experiments, it is possible to compare a wide range of algorithms without any interactions from real users. User studies are another evaluation metric for RSs, where a set of the test subject is recruited to perform several tasks that require interaction with RSs. Users and their interactions are observed and recorded using quantitative measurements, and qualitative questions can be asked to understand their feedback on the user interface. Lastly, in online evaluations, A/B testing is used, where a small amount of traffic is sent to a different, separate recommendation engine, and user interaction is recorded to determine the optimal one. Intervention studies can select any of these evaluation methods from the two domains to measure the effectiveness of an RS in a behaviour change interventions.

In our case, we believed that offline evaluation was not possible because of the absence of pre-collected rating data, the reason for which is that the concept of RSs in water conservation is very new. Along with the pre-recorded rating data, water consumption data must also be collected by simulating users' behaviour, which is a time-consuming process since the data must be recorded over at least one year to avoid variations in consumption due to seasonal effects. Again, online evaluation is also time-consuming as recommender engines must be evaluated for at least one year to avoid variations in consumption due to seasonal effects. For the initial result, only user studies seemed suitable. Therefore, to understand the impact of our RS in promoting water-conscious behaviour, we decided to perform a user study through a survey questionnaire. We hypothesised that if the participants exhibit a positive attitude towards the prototype personalised RS for promoting water-conscious behaviour, then based on behaviour change theories, the behaviour change intervention will be effective. In the following section, we discuss the experiment involving a survey questionnaire.

The generic model proposed in this study for behaviour change interventions using RSs is categorised into four sections: (1) behaviour theories, (2) intervention design using RSs, (3) user engagement theories, and (4) deployment and evaluation based on the theories or nature of tasks. Though these theories or tasks are well-known in their respective domain, however, to the best of authors knowledge, they were never arranged together to address the integration of behaviour change interventions with RSs. This is the main novelty of the proposed model.

4. Prototype of the RS and experiment

In this section, we first present the prototype of the RS for promoting water-conscious behaviour based on the proposed model introduced earlier. Then, we describe the experiment that we performed using a survey questionnaire, which was to validate the prototype by measuring people's attitude towards it.

The overall design of the RS was based on behaviour change methods and UX. First, we identified that the RS had six elements: (1) recommended items, (2) profiling, (3) explanation behind the recommendation, (4) incentive, (5) items preferred by neighbours, and (6) user engagement; moreover, they could be mapped with different behaviour change methods. For instance, recommended items or activities can be mapped with belief selection, persuasive communication, nudging, consciousness raising, and guided practice methods. Table 2 presents the complete mapping of RS elements with different behaviour change methods.

Once the mapping of the RS elements was completed with the behaviour change methods, we then developed a prototype of the RS as part of a web application. The main idea behind the design of the RS was to deliver custom-tailored recommendations or suggestions that would promote water conservation and shift the time of consumption using the advanced household profiling approach (Rahim et al. 2021). The advanced household profiling approach provides better understanding regarding water consumption behaviour and habit and it can identify changes in behaviour. Therefore, such approach acts as the foundation of delivering personalised recommendations. For each recommended item, the source of the data and an explanation were proposed for consciousness raising. To connect behaviour with a consequence and to prompt, repeat, and rehearse as part of reinforcement and guided practice methods consecutively, an incentive in the form of reward points was introduced. To encourage taking others' perspective as part of the shifting perspective method, recommendations preferred by neighbours (i.e., similar users) and popular recommendations were also included in the prototype. Figure 3 illustrates the prototype of the RS.

Lastly, to ensure user engagement, the design of the RS was based on technology adaptation, online intervention, and UX theories. For instance, as part of the UX guidelines, we proposed including a star rating for each suggestion to fine-tune and provide immediate feedback. Using the star rating, the customers can rate the relevancy of the recommended items that can be used to fine-tune the recommendations. Once the prototype of the RS was completed, the next step was the experiment for understanding people's attitude towards the RS for promoting water-conscious behaviour. This is because attitude is a crucial determinant in behaviour change interventions.

The experiment was performed using a survey questionnaire. The main objective of the survey was to understand people's attitude towards the prototype of RS for water conservation and demand management. The target population was anyone living in Australia and over the age of 18 years old. As the sampling technique, we employed random stratified sampling, and the sample design and size were estimated using Equation 1.

$$Sample\ size = \frac{\frac{z^2 \times p(1-p)}{e^2}}{1+\left(\frac{z^2 \times p(1-p)}{e^2 N}\right)} \quad (1)$$

where N is the population size, e is the margin of error, and z is the z-score for desired confidence level.

Table 2: Mapping of RS elements and behaviour change methods for water conservation and demand management

RS element	Method	Behaviour theory	Definition	parameter
user engagement	Participation	Diffusion of Innovations Theory; Theories of Power; Organizational Development Theories; Models of Community Organization;	Assuring high-level engagement	The willingness of the promoter and convenor; Motivations and skills of the participants
Recommended items	Belief selection	Theory of Planned Behaviour; Reasoned Action Approach	Using messages for strengthening positive beliefs and weakening negative belief	current attitudinal, normative and efficacy beliefs
Recommended items	Persuasive communication	Communication-Persuasion Matrix; Elaboration Likelihood Model; Diffusion of Innovations Theory;	Guiding towards adopting an idea, attitude or action	Relevancy of the messages.
Profiling	Tailoring	Trans-Theoretical Model; Precaution Adoption Process Model; Protection Motivation Theory; Communication-Persuasion Matrix;	Matching the intervention to previously measure characteristics	Tailoring variable
Profiling	Individualization		Communication according to individual progress	Personal communication
Incentive	Reinforcement	Theories of Learning; Social Cognitive Theory	Connect a behaviour with a consequence that increases participation	Tailer to individual
Recommended items	Nudging	Theories of Automatic, Impulsive and Habitual Behaviour;	Presenting a choice of alternatives to make the desired choice easy, automatic or default one.	Autonomy, freedom of choice, awareness
Recommended items and explanations	Consciousness-raising	Precaution-Adoption Process Model; Trans-Theoretical Model;	Information, feedback or confrontation about the causes, consequence, and alternatives.	Feedback, confrontation
Items preferred by neighbours	Shifting perspective	Theories of Stigma and Discrimination;	We are encouraging taking the perspective of the other.	Imitation; imaginary competence
Recommended items, Incentive	Guided practice	Social Cognitive Theory; Theories of Self-Regulation	Prompting to repeat and rehearse the behaviour	

As the recommendations would be made at the household level, we first needed to measure the population size, in this case the number of households. According to the data from the Australian Bureau of Statistics, the projected number of households in Australia during 2021 is 10,100,500 (ABS 2015). With an error tolerance rate of 6%, the equation provided the required number of respondents, namely 189 and 267 for 90% and 95% confidence intervals, respectively. To collect the responses, an online questionnaire was prepared that contained 10 questions. These questions can be broadly divided into two types: (1) feedback on general RS usage and (2) attitudes to the prototype RS for promoting water-conscious behaviour. A five-point Likert scale was used to capture the responses for most of the questions (i.e., 1 = Strongly disagree, 2 = Disagree, 3 = Neither agree nor disagree, 4 = Agree, and 5 = Strongly Agree). Table 3 presents the questions that were used for the online questionnaire survey.

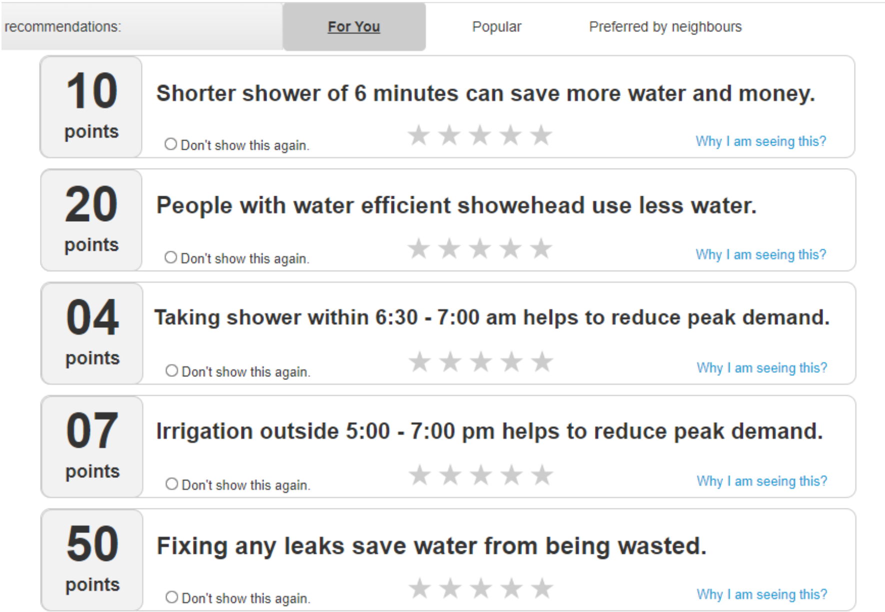


Figure 3.: Prototype of the RS for promoting water-conscious behaviour

Table 3: Questions in the survey questionnaire

Question #	Question
1	Thinking about when you have interacted with a recommender system, what sort of things do you like about these systems?
2	Thinking about when you have interacted with a recommender system, what sort of things do you dislike about these systems?
3	How often do you follow the recommended items either online or offline because of the recommender system?
4	Do you think recommending activities for water conservation based on your preferences would be beneficial for you?
5	Please choose the statement that best reflects your opinion on the following question: I think personalized water conservation tips would help me to save more water.
6	Please choose the statement that best reflects your opinion on the following question: The more personalized suggestions I get, the less effort I would have to spend in finding ways to conserve water.
7	Please choose the statement that best reflects your opinion on the following question: I think such a system would help me to identify the scopes (shorter shower, shifting shower time) to conserve water and managing demand.
8	Please choose the statement that best reflects your opinion on the following question: Overall, such a system would help me to build a positive attitude towards water conservation.
9	Please choose the statement that best reflects your opinion on the following question: I find the incentive points for each recommended activity to conserve water useful/motivational.
10	How important are the following features to you for a recommender system to promote water conservation? a. Personalised suggestions b. Usability of the system c. Interaction d. Popular recommended activites e. Points/incentives for following suggestions

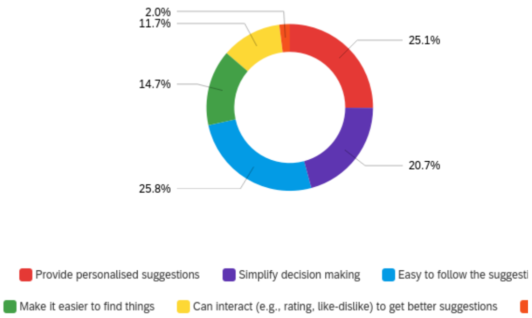
We did not collect any sociodemographic data as we assumed that they would have a minimal impact on the outcome of the study at this level. The Australia-wide survey was conducted in May 2021. Once the required number of responses had been received, the survey was closed, and we applied descriptive statistics to analyse the responses. The following section presents the findings from the analysis.

5. Results and discussion

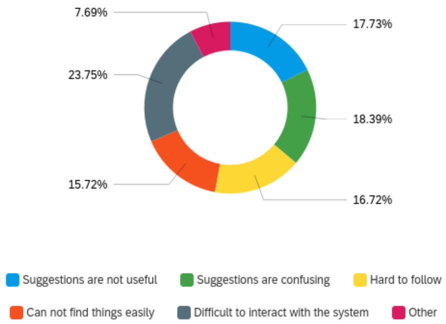
This section presents the results from performing descriptive statistics on the survey responses and discusses the findings. Depending on the nature of the questions, the results are divided into three sections: (1) feedback on general RS usage; (2) attitude towards the prototype RS for promoting water-conscious behaviour; and (3) important features of the RS. The first section aimed to provide feedback on the overall usage of RSs; the second section included results about respondents' attitudes towards the prototype RS; and the last section provides results from investigating the features that respondents considered important for promoting water-conscious behaviour using an RS. The following subsections present the results and discussion on these three types of results.

5.1 Feedback on general RS usage

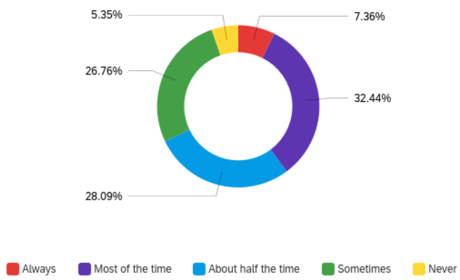
At the beginning of the survey, the concept of RSs was introduced to the participants. Then, the first three questions (Questions 1–3) presented in Table 3 were asked to understand the respondents' feedback regarding the general usage of RSs. The first question aimed to understand what the respondents like about interacting with RSs with the following options: Provides personalised suggestions, Simplifies decision making, Easy to follow the suggestions, Makes it easier to find things, Can interact (e.g., rating and like/dislike) to obtain better suggestions, and Other with an option to enter the desired text. From Figure 4 (a), we can see that the majority of the respondents like easy-to-follow suggestions (25.8%), followed by personalised recommendations (25.1%) and simplified decision making (20.7%) as part of using RSs. This indicates that items recommended in an RS should be easy to follow, personalised, and should simplify decision making. On the other hand, interaction with the system (11.7%) and making it easier to find things (14.7%) did not attract many responses.



(a)



(b)



(c)

Figure 4. Responses on general RS usage: (a) things people like about RSs, (b) things people do not like about RSs, and (c) how often people follow recommendations.

The second question in this section attempted to understand what people do not like about RSs with the following options: Suggestions are not useful, Suggestions are confusing, Hard to follow, Cannot finds things easily, Difficult to interact with the system, and Other. Based on the responses from the survey presented in Figure 4(b), many of the respondents find it difficult to interact with RSs (23.75%). This indicates that future RS studies should focus on the UX aspects of RSs. Under the option of "Other" (7.69%), most of the respondents expressed their concerns regarding the security, relevance, and information overload (e.g., recommended ads by various competitors) of RSs.

The last question in this section asked participants how often they follow recommended items in online or offline. The findings revealed that over 94% of participants follow the recommendations at least some of time either online or offline. Only 5.35% never follow any recommendations. This finding had greater implications for the RS for promoting water-conscious behaviour. Figure 4(c) illustrates the responses for this question.

5.2 Attitude to the prototype RS for promoting water-conscious behaviour

At the beginning of the second section, the prototype RS for promoting water-conscious behaviour was presented to the respondents. Then, they were asked Questions 4–9 presented in Table 3. As mentioned earlier, this section aimed to understand people's attitude towards the prototype RS for promoting water-conscious behaviour. To achieve this goal, these questions asked to what degree the respondents agreed or disagreed with the following topics of personalised recommendations: Beneficial for water conservation, Helpful for saving more water, Require less effort to find the scope of water conservation, Helpful for identifying further scope for water savings, Build a positive attitude towards water conservation, and Useful or motivational with incentive points. For ease of analysis, all the five-level Likert items were converted into a numeric scale (e.g., 1 = Strongly disagree, 5 = Strongly agree) any level of disagreement was considered a negative attitude and agreement was considered a positive attitude. In addition, neither agree nor disagree was marked as a neutral sentiment. To summarise the results, Figure 5 presents a Gantt chart that displays the sentiments of respondents for each question.

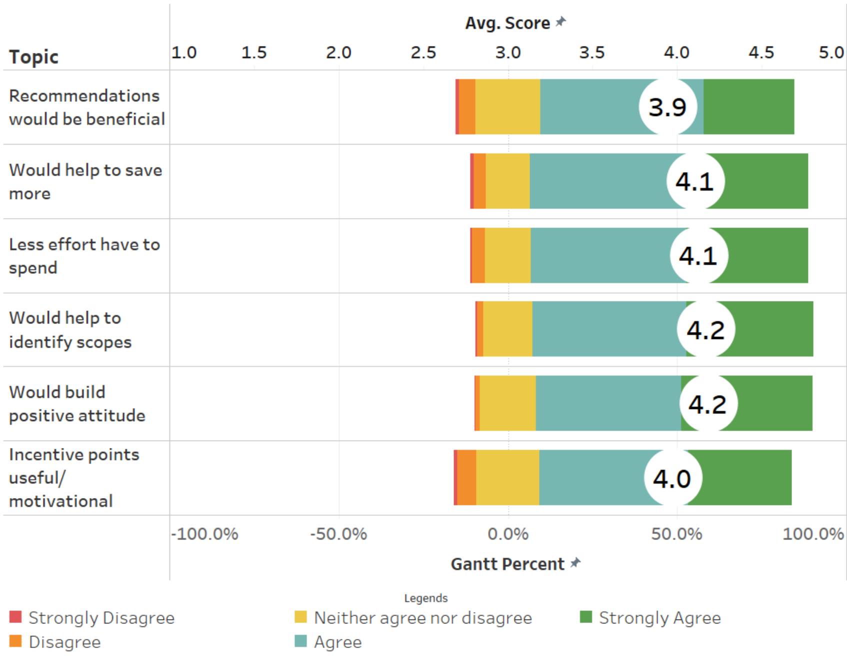


Figure 5. Sentiment of respondents' attitude towards the prototype RS for water conservation.

From Figure 5, we concluded that the overall sentiment of the participants towards the prototype RS was highly positive. The amount of negative sentiment was highly nonsignificant. Most importantly, a majority of the participants believed that such a system would help them to build a positive attitude towards water conservation. Attitude is one of the determinants behind behaviour change interventions according to many behaviour theories; therefore, this is a significant finding because it validated the potential impact of the RS before a pilot study was performed. Figure 5 also displays the average score for each question on a scale of 1 to 5.

5.3 Important features of the RS

The goal of the last section of the survey was to identify the level of importance of each feature of the RS to the participants. Therefore, this section contained only one question, with five features in the rows of a matrix table and a Likert scale in the columns to determine the level of importance. Figure 6 illustrates the level of sentiment towards each feature of the RS with the average score.

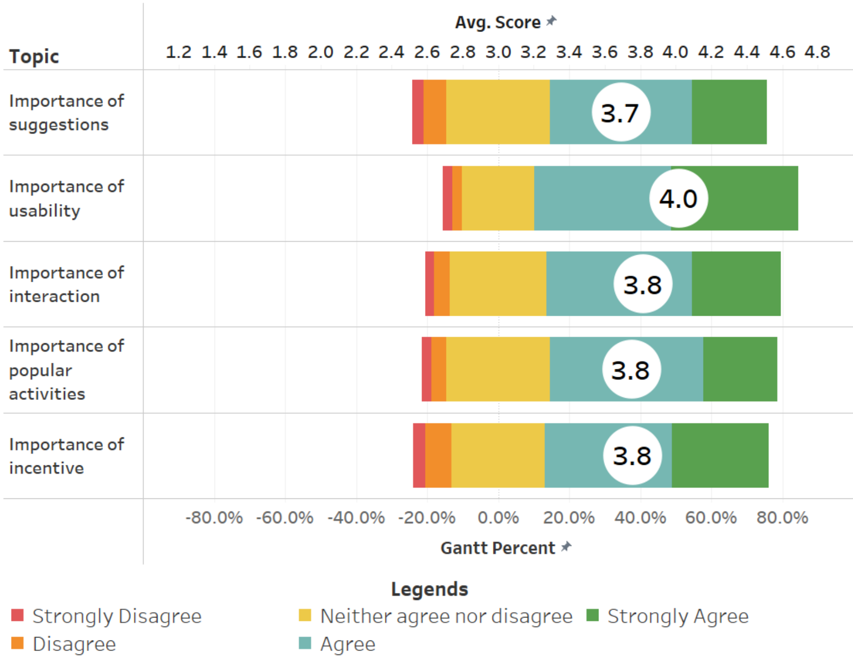


Figure 6. Sentiment towards the important features of the RS.

From Figure 6, we concluded that the usability of the RS was one of the most important features to the users. This finding is consistent with the findings reported in Section 5.1. Surprisingly, the personalised suggestions received less importance compared with other features. One reason behind this result could be because of the novelty of such feature; most of the people are still not familiar with the advantages and benefits of personalised suggestions in changing behaviour. Also, it can be noted that a good percentage (<66%) of the sample participated in this survey do not follow recommendations made online most of the time which could be another reason behind receiving lower importance compared to other features. Another reason behind this result may be that many people are not aware of how recommendations are made. Therefore, when designing RSs, future studies should focus on how the recommendations are made.

6. Conclusion and future work

The application of RSs in behaviour change interventions has become popular in recent years. However, until now, no model existed that could be used for performing any intervention studies with RSs. In this study, we introduced a model based on behaviour, RSs, and user engagement theories that addresses this research gap. Later, we applied the proposed model to designing an intervention program for promoting water-conscious behaviour using an RS. A prototype RS was developed, and respondents' attitude towards the system was measured through a survey questionnaire. After analysing the responses, we concluded that the survey participants exhibited a positive attitude towards the personalised RS for promoting water-conscious behaviour. In addition, the survey responses revealed that the usability of and interaction with the system seemed more important compared to the importance of the suggestions. This finding suggests that the usability and interactivity of such a such is crucial for successful interventions. Most importantly, we observed that the participants think this type of system would help them to build positive attitude towards water conservation, which implies the potential success of the system should it be deployed.

Findings from this study have implications for addressing water demand management. The proposed model and prototype of the RS can be developed further and deployed into the real world to fully utilise the data generated from digital water meters. The model and prototype of RS can be used to change consumers' behaviour to save more water and reduce peak-hours water demand in the long term. Further, the findings have implications for water demand professionals, customers, and researchers. The outcomes from this research should provide better understanding of consumer's behaviour and their response to the personalised recommendations which is crucial to the water demand professionals for designing new policies and programs. Through personalised recommendations, it is possible to improve customer satisfaction and it offers customers a one stop place to know how to save water and get incentives for following the recommendations. For researchers, this study offers a new approach for behaviour change interventions that can be applied in any filed using RSs.

As for future work, this study recommends investigating the applications of the proposed model in different intervention studies using RSs and reporting the findings. In the case of application in water conservation, we plan to perform a UX study to improve and finalise the design of the prototype and perform a pilot study for measuring the effectiveness of the system as well as the model.

References

ABS 2015, *Household and family projects, Australia, 2011 to 2036*, Australian Bureau of Statistics, viewed 15 May 2021, <https://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/3236.02011%20to%202036?OpenDocument>.

Adams, R.J., Sadasivam, R.S., Balakrishnan, K., Kinney, R.L., Houston, T.K. & Marlin, B.M. 2014, 'PERSPeCT: collaborative filtering for tailored health communications', pp. 329-32.

Aggarwal, C.C. 2016a, 'Knowledge-based recommender systems', *Recommender systems*, Springer, pp. 167-97.

Aggarwal, C.C. 2016b, *Recommender systems*, Springer.

Bartholomew, L.K., Parcel, G.S., Kok, G., Gottlieb, N.H., Schaalma, H.C., Markham, C.C., Tyrrell, S.C., Shegog, R.C., Fernández, M.C. & Mullen, P.D.C. 2006, *Planning health promotion programs: an intervention mapping approach*, Jossey-Bass.

Bobadilla, J., Ortega, F., Hernando, A. & Gutiérrez, A. 2013, 'Recommender systems survey', *Knowledge-based systems*, vol. 46, pp. 109-32.

Brouwer, W., Kroeze, W., Crutzen, R., de Nooijer, J., de Vries, N.K., Brug, J. & Oenema, A. 2011, 'Which intervention characteristics are related to more exposure to internet-delivered healthy lifestyle promotion interventions? A systematic review', *Journal of medical Internet research*, vol. 13, no. 1, p. e2.

Burke, R. 2007, 'Hybrid Web Recommender Systems', *The Adaptive Web*, Springer, Berlin, Germany, pp. 377-408.

Cheung, K.L., Durusu, D., Sui, X. & de Vries, H. 2019, 'How recommender systems could support and enhance computer-tailored digital health programs: a scoping review', *Digital health*, vol. 5, p. 2055207618824727.

Davies, C.A., Spence, J.C., Vandelanotte, C., Caperchione, C.M. & Mummery, W.K. 2012, 'Meta-analysis of internet-delivered interventions to increase physical activity levels', *International Journal of Behavioral Nutrition and Physical Activity*, vol. 9, no. 1, p. 52.

Eke, C.I., Norman, A.A., Shuib, L. & Nweke, H.F. 2019, 'A survey of user profiling: state-of-the-art, challenges, and solutions', *IEEE Access*, vol. 7, pp. 144907-24.

Espín, V., Hurtado, M.V. & Noguera, M. 2016, 'Nutrition for Elder Care: a nutritional semantic recommender system for the elderly', *Expert Systems*, vol. 33, no. 2, pp. 201-10.

Esteban, B., Tejeda-Lorente, Á., Porcel, C., Arroyo, M. & Herrera-Viedma, E. 2014, 'TPLUFIB-WEB: A fuzzy linguistic Web system to help in the treatment of low back pain problems', *Knowledge-Based Systems*, vol. 67, pp. 429-38.

Fishbein, M. 1979, 'A theory of reasoned action: some applications and implications'.

Ge, M., Elahi, M., Fernaández-Tobías, I., Ricci, F. & Massimo, D. 2015, 'Using tags and latent factors in a food recommender system', pp. 105-12.

Giabbanelli, P.J. & Crutzen, R. 2015, 'Supporting self-management of obesity using a novel game architecture', *Health informatics journal*, vol. 21, no. 3, pp. 223-36.

Goldberg, D., Nichols, D., Oki, B.M. & Terry, D. 1992, 'Using collaborative filtering to weave an information tapestry', *Communications of the ACM*, vol. 35, no. 12, pp. 61-70.

Gram-Hanssen, K. 2014, 'New needs for better understanding of household's energy consumption–behaviour, lifestyle or practices?', *Architectural Engineering and Design Management*, vol. 10, no. 1-2, pp. 91-107.

Hales, S., Turner-McGrievy, G.M., Wilcox, S., Fahim, A., Davis, R.E., Huhns, M. & Valafar, H. 2016, 'Social networks for improving healthy weight loss behaviors for overweight and obese adults: A randomized clinical trial of the social pounds off digitally (Social POD) mobile app', *International journal of medical informatics*, vol. 94, pp. 81-90.

Hammer, S., Kim, J. & André, E. 2010, 'MED-StyleR: METABO diabetes-lifestyle recommender', pp. 285-8.

Harley, A. 2018, *UX Guidelines for Recommended Content*, Nielsen Norman Group, viewed 15 May 2021, <https://www.nngroup.com/articles/recommendation-guidelines/>.

Hassenzahl, M. & Tractinsky, N. 2006, 'User experience-a research agenda', *Behaviour & information technology*, vol. 25, no. 2, pp. 91-7.

Hidalgo, J.I., Maqueda, E., Risco-Martín, J.L., Cuesta-Infante, A., Colmenar, J.M. & Nobel, J. 2014, 'glUCModel: A monitoring and modeling system for chronic diseases applied to diabetes', *Journal of biomedical informatics*, vol. 48, pp. 183-92.

Kantola, S., Syme, G. & Campbell, N. 1982, 'The role of individual differences and external variables in a test of the sufficiency of Fishbein's model to explain behavioral intentions to conserve water', *Journal of Applied Social Psychology*, vol. 12, no. 1, pp. 70-83.

Kelders, S.M., Kok, R.N., Ossebaard, H.C. & Van Gemert-Pijnen, J.E. 2012, 'Persuasive system design does matter: a systematic review of adherence to web-based interventions', *Journal of medical Internet research*, vol. 14, no. 6, p. e152.

Kok, G. 2018, 'A practical guide to effective behavior change: How to apply theory-and evidence-based behavior change methods in an intervention'.

Kok, G., Gottlieb, N.H., Peters, G.-J.Y., Mullen, P.D., Parcel, G.S., Ruiter, R.A., Fernández, M.E., Markham, C. & Bartholomew, L.K. 2016, 'A taxonomy of behaviour change methods: an intervention mapping approach', *Health psychology review*, vol. 10, no. 3, pp. 297-312.

Kuijpers, W., Groen, W.G., Aaronson, N.K. & van Harten, W.H. 2013, 'A systematic review of web-based interventions for patient empowerment and physical activity in chronic diseases: relevance for cancer survivors', *Journal of medical Internet research*, vol. 15, no. 2, p. e37.

Kwasnicka, D., Dombrowski, S.U., White, M. & Sniehotta, F. 2016, 'Theoretical explanations for maintenance of behaviour change: a systematic review of behaviour theories', *Health psychology review*, vol. 10, no. 3, pp. 277-96.

Liu, A., Giurco, D. & Mukheibir, P. 2017, 'Advancing household water-use feedback to inform customer behaviour for sustainable urban water', *Water Science and Technology: Water Supply*, vol. 17, no. 1, pp. 198-205.

Liu, A., Giurco, D., Mukheibir, P., Mohr, S., Watkins, G. & White, S. 2017, 'Online water-use feedback: household user interest, savings and implications', *Urban Water Journal*, vol. 14, no. 9, pp. 900-7.
Michie, S., Johnston, M., Francis, J., Hardeman, W. & Eccles, M. 2008, 'From theory to intervention: mapping theoretically derived behavioural determinants to behaviour change techniques', *Applied psychology*, vol. 57, no. 4, pp. 660-80.
Montaner, M., López, B. & De La Rosa, J.L. 2003, 'A taxonomy of recommender agents on the internet', *Artificial intelligence review*, vol. 19, no. 4, pp. 285-330.
Norman, D. & Nielsen, J., *The Definition of User Experience (UX)*, Nielsen Norman Group, viewed 15 May 2021, <https://www.nngroup.com/articles/definition-user-experience/>.
Norouzi, S., Ghalibaf, A.K., Sistani, S., Banazadeh, V., Keykhaei, F., Zareishargh, P., Amiri, F., Nematy, M. & Etminani, K. 2018, 'A mobile application for managing diabetic patients' nutrition: A food recommender system', *Archives of Iranian medicine*, vol. 21, no. 10, pp. 466-72.
O'Brien, H.L. & Toms, E.G. 2008, 'What is user engagement? A conceptual framework for defining user engagement with technology', *Journal of the American society for Information Science and Technology*, vol. 59, no. 6, pp. 938-55.
Onile, A.E., Belikov, J. & Levron, Y. 2020, 'Innovative Energy Services for Behavioral-Reflective Attributes and Intelligent Recommender System', IEEE, pp. 242-6.
Palumbo, P., Cattelani, L., Fusco, F., Pijnappels, M., Chiari, L., Chesani, F. & Mellone, S. 2020, 'A Recommender System for Behavioral Change in 60-70-year-old Adults', <http://ceur-ws.org/Vol-2804/paper5.pdf>.
Peters, G.-J. 2018, 'A practical guide to effective behavior change: How to identify what to change in the first place'.
Peters, G.-J.Y., De Bruin, M. & Crutzen, R. 2015, 'Everything should be as simple as possible, but no simpler: towards a protocol for accumulating evidence regarding the active content of health behaviour change interventions', *Health Psychology Review*, vol. 9, no. 1, pp. 1-14.
Prochaska, J.O. & Marcus, B.H. 1994, 'The transtheoretical model: Applications to exercise'.
Radha, M., Willemsen, M.C., Boerhof, M. & IJsselsteijn, W.A. 2016, 'Lifestyle recommendations for hypertension through Rasch-based feasibility modeling', pp. 239-47.
Rahim, M.S., Nguyen, K.A., Stewart, R.A., Giurco, D. & Blumenstein, M. 2020, 'Machine Learning and Data Analytic Techniques in Digital Water Metering: A Review', *Water*, vol. 12, no. 1.
Rahim, M.S., Nguyen, K.A., Stewart, R.A., Giurco, D. & Blumenstein, M. 2021, 'Advanced household profiling using digital water meters', *Journal of Environmental Management*, vol. 288, p. 112377.
Resnick, P. & Varian, H.R. 1997, 'Recommender systems', *Communications of the ACM*, vol. 40, no. 3, pp. 56-8.
Ricci, F., Rokach, L. & Shapira, B. 2015, 'Recommender Systems: Introduction and Challenges', *Recommender Systems Handbook*, Springer US, Boston, USA, pp. 1-34.
Rohani, D.A., Quemada Lopategui, A., Tuxen, N., Faurholt-Jepsen, M., Kessing, L.V. & Bardram, J.E. 2020, 'MUBS: A personalized recommender system for behavioral activation in mental health', pp. 1-13.
Rouse, J. 2007, 'Practice theory', *Philosophy of anthropology and sociology*, Elsevier, pp. 639-81.
Sadasivam, R.S., Borglund, E.M., Adams, R., Marlin, B.M. & Houston, T.K. 2016, 'Impact of a collective intelligence tailored messaging system on smoking cessation: the Perspect

randomized experiment', *Journal of medical Internet research*, vol. 18, no. 11, p. e285.

Sami, A., Nagatomi, R., Terabe, M. & Hashimoto, K. 2008, 'Design of Physical Activity Recommendation System', pp. 148-52.

Sardianos, C., Varlamis, I., Dimitrakopoulos, G., Anagnostopoulos, D., Alsalemi, A., Bensaali, F., Himeur, Y. & Amira, A. 2020, 'Rehab-c: Recommendations for energy habits change', *Future Generation Computer Systems*, vol. 112, pp. 394-407.

Seyranian, V., Sinatra, G.M. & Polikoff, M.S. 2015, 'Comparing communication strategies for reducing residential water consumption', *Journal of Environmental Psychology*, vol. 41, pp. 81-90.

Shani, G. & Gunawardana, A. 2011, 'Evaluating recommendation systems', *Recommender systems handbook*, Springer, pp. 257-97.

Short, C., Rebar, A., Plotnikoff, R. & Vandelanotte, C. 2015, 'Designing engaging online behaviour change interventions: a proposed model of user engagement'.

Shove, E., Pantzar, M. & Watson, M. 2012, *The dynamics of social practice: Everyday life and how it changes*, Sage.

van Empelen, P. & Kok, G. 2006, 'Condom use in steady and casual sexual relationships: Planning, preparation and willingness to take risks among adolescents', *Psychology & Health*, vol. 21, no. 2, pp. 165-81.

Venkatesh, V., Davis, F. & Morris, M.G. 2007, 'Dead or alive? The development, trajectory and future of technology adoption research', *Journal of the association for information systems*, vol. 8, no. 4, p. 1.

Venkatesh, V., Morris, M.G., Davis, G.B. & Davis, F.D. 2003, 'User acceptance of information technology: Toward a unified view', *MIS quarterly*, pp. 425-78.

Venkatesh, V., Thong, J.Y. & Xu, X. 2012, 'Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology', *MIS quarterly*, pp. 157-78.

Webb, T., Joseph, J., Yardley, L. & Michie, S. 2010, 'Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy', *Journal of medical Internet research*, vol. 12, no. 1, p. e1376.

Wei, P., Xia, S., Chen, R., Qian, J., Li, C. & Jiang, X. 2020, 'A Deep-Reinforcement-Learning-Based Recommender System for Occupant-Driven Energy Optimization in Commercial Buildings', *IEEE Internet of Things Journal*, vol. 7, no. 7, pp. 6402-13.

Weinstein, N.D., Sandman, P.M. & Blalock, S.J. 2020, 'The precaution adoption process model', *The Wiley Encyclopedia of Health Psychology*, pp. 495-506.

Willis, R.M., Stewart, R.A., Giurco, D.P., Talebpour, M.R. & Mousavinejad, A. 2013, 'End use water consumption in households: impact of socio-demographic factors and efficient devices', *Journal of Cleaner Production*, vol. 60, pp. 107-15.

Willis, R.M., Stewarta, R.A., Panuwatwanich, K., Jones, S. & Kyriakides, A. 2010, 'Alarming visual display monitors affecting shower end use water and energy conservation in Australian residential households', *Resources, Conservation and Recycling*, vol. 54, no. 12, pp. 1117-27.

Zhang, T., Jarrad, G., Murphy, S.A. & Bidargaddi, N. 2019, 'A smartphone-based behavioural activation application using recommender system', pp. 250-3.

Exegesis References

Abu-Bakar, H., Williams, L. & Hallett, S.H. 2021, 'An empirical water consumer segmentation and the characterisation of consumption patterns underpinning demand peaks', *Resources, Conservation and Recycling*, vol. 174.

Bich-Ngoc, N. & Teller, J. 2018, 'A review of residential water consumption determinants', *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Series A review of residential water consumption determinants Conference Paper, vol. 10964 LNCS, Springer, Cham, Switzerland, pp. 685-96, <https://link.springer.com/content/pdf/10.1007%2F978-3-319-95174-4_52.pdf>.

Bolorinos, J., Ajami, N.K. & Rajagopal, R. 2020, 'Consumption change detection for urban planning: Monitoring and segmenting water customers during drought', *Water Resources Research*, vol. 56, no. 3, p. e2019WR025812.

Burke, R. 2007, 'Hybrid Web Recommender Systems', *The Adaptive Web*, Springer, Berlin, Germany, pp. 377-408.

Cahn, A., Katz, D. & Ghermandi, A. 2020, 'Analyzing Water Customer Preferences for Online Feedback Technologies in Israel: A Prototype Study', *Journal of Water Resources Planning and Management*, vol. 146, no. 4, p. 06020002.

Cardell-Oliver, R. 2013, 'Water use signature patterns for analyzing household consumption using medium resolution meter data', *Water Resources Research*, vol. 49, no. 12, pp. 8589-99.

Carvalho, T.M.N. & de Assis de Souza Filho, F. 2021, 'Variational Mode Decomposition Hybridized With Gradient Boost Regression for Seasonal Forecast of Residential Water Demand', *Water Resources Management*, vol. 35, no. 10, pp. 3431-45.

Cespedes Restrepo, J.D. & Morales-Pinzon, T. 2020, 'Effects of feedback information on the household consumption of water and electricity: A case study in Colombia', *J Environ Manage*, vol. 262, p. 110315.

Cheung, K.L., Durusu, D., Sui, X. & de Vries, H. 2019, 'How recommender systems could support and enhance computer-tailored digital health programs: a scoping review', *Digital health*, vol. 5, p. 2055207618824727.

Cominola, A., Giuliani, M., Castelletti, A., Fraternali, P., Gonzalez, S.L.H., Herrero, J.C.G., Novak, J. & Rizzoli, A.E. 2021, 'Long-term water conservation is fostered by smart meter-based feedback and digital user engagement', *npj Clean Water*, vol. 4, no. 1.

Cominola, A., Moro, A., Riva, L., Giuliani, M. & Castelleti, A. 2016, 'Profiling residential water users' routines by eigenbehavior modelling'.

Daminato, C., Diaz-Farina, E., Filippini, M. & Padrón-Fumero, N. 2021, 'The impact of smart meters on residential water consumption: Evidence from a natural experiment in the Canary Islands', *Resource and Energy Economics*, vol. 64.

Eke, C.I., Norman, A.A., Shuib, L. & Nweke, H.F. 2019, 'A survey of user profiling: state-of-the-art, challenges, and solutions', *IEEE Access*, vol. 7, pp. 144907-24.

Emmanuel A. Donkor, S.M.A.T.A.M.R.S.a.J.A.R., P.E.4 2014, 'Urban Water Demand Forecasting: Review of Methods and Models', *JOURNAL OF WATER RESOURCES PLANNING AND MANAGEMENT*.

The Global Risks Report 2019, World Economic Forum, Geneva, Switzerland.

Gourmelon, N., Bayer, S., Mayle, M., Bach, G., Bebber, C., Munck, C., Sosna, C. & Maier, A. 2021, 'Implications of experiment set-ups for residential water end-use classification', *Water (Switzerland)*, vol. 13, no. 2.

Guo, G., Zhang, J. & Yorke-Smith, N. 2015, 'Leveraging multiviews of trust and similarity to enhance clustering-based recommender systems', *Knowledge-Based Systems*, vol. 74, pp. 14-27.

Gurung, T.R., Stewart, R.A., Beal, C.D. & Sharma, A.K. 2016, 'Smart meter enabled informatics for economically efficient diversified water supply infrastructure planning', *Journal of Cleaner Production*, vol. 135, pp. 1023-33.
Han, J., Kamber, M. & Pei, J. 2012, '10 - Cluster Analysis: Basic Concepts and Methods', in J. Han, M. Kamber & J. Pei (eds), *Data Mining (Third Edition)*, Morgan Kaufmann, Boston, pp. 443-95.
Hauber-Davidson, G. & Idris, E. 2006, 'Smart water metering', *Water*, vol. 33, no. 3, pp. 56-9.
Jessoe, K., Lade, G.E., Loge, F. & Spang, E. 2021, 'Residential water conservation during drought: Experimental evidence from three behavioral interventions', *Journal of Environmental Economics and Management*, vol. 110, p. 102519.
Kok, G. 2018, 'A practical guide to effective behavior change: How to apply theory-and evidence-based behavior change methods in an intervention'.
LaVanchy, G.T., Kerwin, M.W. & Adamson, J.K. 2019, 'Beyond ‘Day Zero’: insights and lessons from Cape Town (South Africa)', *Hydrogeology Journal*, vol. 27, no. 5, pp. 1537-40.
Lee, D. & Derrible, S. 2020, 'Predicting Residential Water Demand with Machine-Based Statistical Learning', *Journal of Water Resources Planning and Management*, vol. 146, no. 1.
Liu, A., Giurco, D. & Mukheibir, P. 2015, 'Motivating metrics for household water-use feedback', *Resources, Conservation and Recycling*, vol. 103, pp. 29-46.
Liu, A., Giurco, D. & Mukheibir, P. 2016, 'Urban water conservation through customised water and end-use information', *Journal of Cleaner Production*, vol. 112, pp. 3164-75.
Liu, A., Giurco, D. & Mukheibir, P. 2017, 'Advancing household water-use feedback to inform customer behaviour for sustainable urban water', *Water Science and Technology: Water Supply*, vol. 17, no. 1, pp. 198-205.
Liu, A., Giurco, D., Mukheibir, P. & White, S. 2016, 'Detailed water-use feedback: A review and proposed framework for program implementation', *Utilities Policy*, vol. 43, pp. 140-50.
Liu, A. & Mukheibir, P. 2018, 'Digital metering feedback and changes in water consumption – A review', *Resources, Conservation and Recycling*, vol. 134, pp. 136-48.
Luby, I.H., Polasky, S. & Swackhamer, D.L. 2018, 'U.S. Urban Water Prices: Cheaper When Drier', *Water Resources Research*, vol. 54, no. 9, pp. 6126-32.
McKenna, S.A., Fusco, F. & Eck, B.J. 2014, 'Water Demand Pattern Classification from Smart Meter Data', *Procedia Engineering*, vol. 70, pp. 1121-30.
Meyer, B.E., Nguyen, K., Beal, C.D., Jacobs, H.E. & Buchberger, S.G. 2021, 'Classifying Household Water Use Events into Indoor and Outdoor Use: Improving the Benefits of Basic Smart Meter Data Sets', *Journal of Water Resources Planning and Management*, vol. 147, no. 12.
Moglia, M., Cook, S. & Tapsuwan, S. 2018, 'Promoting Water Conservation: Where to from here?', *Water*, vol. 10, no. 11.
Montaner, M., López, B. & De La Rosa, J.L. 2003, 'A taxonomy of recommender agents on the internet', *Artificial intelligence review*, vol. 19, no. 4, pp. 285-330.
Nguyen, K.A., Sahin, O., Stewart, R.A. & Zhang, H. 2016, 'Water demand forecasting with AUTOFLOW© using State-Space approach', paper presented to the *8th International Congress on Environmental Modelling and Software*.
Nguyen, K.A., Stewart, R.A. & Zhang, H. 2014, 'An autonomous and intelligent expert system for residential water end-use classification', *Expert Systems with Applications*, vol. 41, no. 2, pp. 342-56.
Nguyen, K.A., Stewart, R.A., Zhang, H. & Jones, C. 2015, 'Intelligent autonomous system for residential water end use classification: Autoflow', *Applied Soft Computing*, vol. 31, pp. 118-31.

Nguyen, K.A., Stewart, R.A., Zhang, H., Sahin, O. & Siriwardene, N. 2018, 'Re-engineering traditional urban water management practices with smart metering and informatics', *Environmental Modelling & Software*, vol. 101, pp. 256-67.

Peters, G.-J. 2018, 'A practical guide to effective behavior change: How to identify what to change in the first place'.

Peters, G.-J.Y., De Bruin, M. & Crutzen, R. 2015, 'Everything should be as simple as possible, but no simpler: towards a protocol for accumulating evidence regarding the active content of health behaviour change interventions', *Health Psychology Review*, vol. 9, no. 1, pp. 1-14.

Pickering, C. & Byrne, J. 2014, 'The benefits of publishing systematic quantitative literature reviews for PhD candidates and other early-career researchers', *Higher Education Research & Development*, vol. 33, no. 3, pp. 534-48.

Rahim, M.S., Nguyen, K., Stewart, R., Giurco, D. & Blumenstein, M., 'Digital Water Meters for Advanced Residential Customer Profiling to Promote Water Conservation'.

Rahim, M.S., Nguyen, K.A., Stewart, R.A., Giurco, D. & Blumenstein, M. 2019, 'Predicting Household Water Consumption Events: Towards a Personalised Recommender System to Encourage Water-conscious Behaviour', pp. 1-8.

Rahim, M.S., Nguyen, K.A., Stewart, R.A., Giurco, D. & Blumenstein, M. 2020, 'Machine Learning and Data Analytic Techniques in Digital Water Metering: A Review', *Water*, vol. 12, no. 1.

Rahim, M.S., Nguyen, K.A., Stewart, R.A., Giurco, D. & Blumenstein, M. 2021, 'Advanced household profiling using digital water meters', *Journal of Environmental Management*, vol. 288, p. 112377.

Randall, T. & Koech, R. 2019, 'Smart water metering technology for water management in urban areas', *Water e-Journal*, vol. 4, no. 1.

Resnick, P. & Varian, H.R. 1997, 'Recommender systems', *Communications of the ACM*, vol. 40, no. 3, pp. 56-8.

Reynaud, A., Pons, M. & Pesado, C. 2018, 'Household water demand in Andorra: Impact of individual metering and seasonality', *Water (Switzerland)*, vol. 10, no. 3.

Ricci, F., Rokach, L. & Shapira, B. 2015, 'Recommender Systems: Introduction and Challenges', *Recommender Systems Handbook*, Springer US, Boston, USA, pp. 1-34.

Sahin, O., Stewart, R.A. & Porter, M.G. 2015, 'Water security through scarcity pricing and reverse osmosis: a system dynamics approach', *Journal of Cleaner Production*, vol. 88, pp. 160-71.

Schultz, P.W., Messina, A., Tronu, G., Limas, E.F., Gupta, R. & Estrada, M. 2014, 'Personalized Normative Feedback and the Moderating Role of Personal Norms: A Field Experiment to Reduce Residential Water Consumption', *Environment and Behavior*, vol. 48, no. 5, pp. 686-710.

Schultz, W. 2019, 'Social comparison as a tool to promote residential water conservation', *Frontiers in Water*, vol. 1, p. 2.

Stewart, R.A., Nguyen, K., Beal, C., Zhang, H., Sahin, O., Bertone, E., Vieira, A.S., Castelletti, A., Cominola, A., Giuliani, M., Giurco, D., Blumenstein, M., Turner, A., Liu, A., Kenway, S., Savić, D.A., Makropoulos, C. & Kossieris, P. 2018, 'Integrated intelligent water-energy metering systems and informatics: Visioning a digital multi-utility service provider', *Environmental Modelling & Software*, vol. 105, pp. 94-117.

Tanverakul, S.A. & Lee, J. 2015, 'Impacts of metering on residential water use in California', *American Water Works Association*, vol. 107, no. 2, pp. E69-E75.

Water Conservation Report 2018, Sydney Water, Sydney.

Willis, R.M., Stewarta, R.A., Panuwatwanich, K., Jones, S. & Kyriakides, A. 2010, 'Alarming visual display monitors affecting shower end use water and energy conservation in

Australian residential households', *Resources, Conservation and Recycling*, vol. 54, no. 12, pp. 1117-27.

Appendix A Information sheet and consent form for online survey

ETH21-5861 - Survey on Personalised recommendations for water conservation

What is the research study about?
The purpose of this online survey is to investigate the feedback and attitude towards a prototype of a recommender system for promoting water-conscious behaviour. It may take 5-10 minutes to complete the survey.

Who is conducting this research?
My name is Md Shamsur Rahim and I am a student at UTS. My supervisor is Professor Michael Blumenstein, Associate Dean (Research Strategy & Management), Faculty of Engineering & Information Technology, Michael.Blumenstein@uts.edu.au.

Inclusion/Exclusion Criteria
Before you decide to participate in this research study, we need to ensure that it is ok for you to take part. Inclusion criteria:

- You must be over 18 years old.
- Currently living in Australia

Do I have to take part in this research study?

Participation in this study is voluntary. It is completely up to you whether or not you decide to take part. If you decide to participate, I will invite you to complete an online questionnaire. You can change your mind at any time and stop completing the surveys without consequences.

Are there any risks/inconvenience?
We don't expect this questionnaire to cause any harm or discomfort, however, if you experience feelings of distress as a result of participation in this study you can let the researcher know and they will provide you with assistance.

What will happen to information about me?
Access to the online questionnaire is via https://utsau.au1.qualtrics.com/jfe/form/SV_e59gwlUfyKOBExM. Submission of the online questionnaire/s is an indication of your consent. By clicking the https://utsau.au1.qualtrics.com/jfe/form/SV_e59gwlUfyKOBExM you consent to the research team collecting and using personal information about you for the research project. All this information will be treated confidentially. All contact details information will be removed from the final data set. Your information will only be used for the purpose of this research project.
We plan to publish the results in a journal to indicate the response towards the concept of personalised recommender system for water-conscious behaviour. Information only related to the concept will be disclosed in a summarised form to gain better understanding.

This survey is provided through Qualtrics. The survey responses will be stored on behalf of UTS by Qualtrics and extracted for use by UTS. Any personal information will not otherwise be disclosed unless you have provided express consent, or where required or permitted by law. Further information on privacy can also be found in the Qualtrics privacy statement.

If you wish to access or correct information held about you in relation to this survey, contact UTS at the details below. Note that information that is anonymous may not be identified for access or correction.

What if I have concerns or a complaint?

If you have concerns about the research that you think I or my supervisor can help you with, please feel free to contact us on School of Computer Science, 81-113, Broadway, Ultimo NSW 2007Australia, mdshamsur.rahim@student.uts.edu.au.

If you would like to talk to someone who is not connected with the research, you may contact the Research Ethics Officer on 02 9514 9772 or Research.ethics@uts.edu.au and quote this number ETH21-5861

Appendix B Survey Questionnaire

Video 1 Please watch the below video. Click next when you are done to continue with the survey.

Concept These questions ask for your feedback on the concept of a system. Your answers will help us understand the strengths and weaknesses of our proposed system.

A **Recommender System** is a system that uses stored user preferences to locate, choose and suggest items (recommend items) that will be of interest to a person. For example, if you are shopping for a digital camera, on Amazon, Amazon may recommend you some camera lenses that go with your camera, or some other cameras that may interest you.

We propose to develop a similar system for promoting water-conscious behaviours which will suggest to consumers a list of activities. These activities will be custom tailored based on your behaviour patterns and will be changing based on your feedback.

Q1 Thinking about when you've interacted with a recommender system, what sort of things do you like about these systems?

- ○ Provide personalised suggestions
- ○ Simplify decision making
- ○ Easy to follow the suggestions
- ○ Make it easier to find things
- ○ Can interact (e.g., rating, like-dislike) to get better suggestions
- ○ Other ____

Q2 Thinking about when you've interacted with a recommender system, what sort of things do you dislike about these systems?

- ◯ Suggestions are not useful
- ◯ Suggestions are confusing
- ◯ Hard to follow
- ◯ Can not find things easily
- ◯ Difficult to interact with the system
- ◯ Other ____

Q3 How often do you follow the recommended items either online or offline because of the recommender system?

- ◯ Always
- ◯ Most of the time
- ◯ About half the time
- ◯ Sometimes
- ◯ Never

Q4 Do you think recommending activities for water conservation based on your preferences would be beneficial for you?

- ◯ Definitely yes
- ◯ Probably yes
- ◯ Might or might not
- ◯ Probably not
- ◯ Definitely not

Q5 Please choose the statement that best reflects your opinion on the following question: I think personalised water conservation tips would help me to save more water. (pick one)

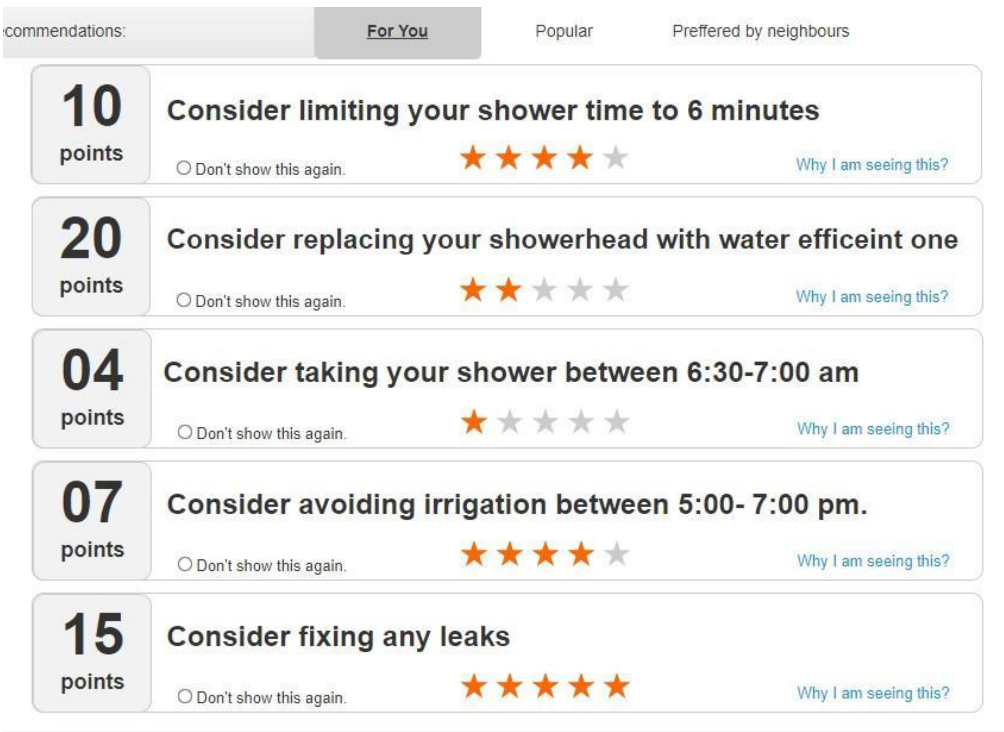
- ◯ Strongly agree
- ◯ Somewhat agree
- ◯ Neither agree nor disagree
- ◯ Somewhat disagree
- ◯ Strongly disagree

Instruction Below is a reminder of the concept we are asking you to evaluate:

A **Recommender System** is a system that uses stored user preferences to locate, choose and suggest items (recommend items) that will be of interest to a person. For example, if you are shopping for a digital camera, on Amazon, Amazon may recommend you some camera lenses that go with your camera, or some other cameras that may interest you.

We propose to develop a similar system for promoting water-conscious behaviours which will suggest to consumers a list of activities. These activities will be custom tailored based on your behaviour patterns and will be changing based on your feedback.

In this example, you will see a picture from the recommender system for promoting water-conscious behaviour. Please note that this is not a complete screenshot but only a part of the screen.



Q6 Please choose the statement that best reflects your opinion on the following question: The more personalised suggestions I get, the less effort I would have to spend in finding ways to conserve water.

◯ Strongly agree

◯ Somewhat agree

◯ Neither agree nor disagree

◯ Somewhat disagree

◯ Strongly disagree

Q7 Please choose the statement that best reflects your opinion on the following question: I think such a system would help me to identify the scopes (shorter shower, shifting shower time) to conserve water and managing demand.

◯ Strongly agree

◯ Somewhat agree

◯ Neither agree nor disagree

◯ Somewhat disagree

◯ Strongly disagree

Q8 Please choose the statement that best reflects your opinion on the following question: Overall, such a system would help me to build a positive attitude towards water conservation. (pick one)

◯ Strongly agree

◯ Somewhat agree

◯ Neither agree nor disagree

◯ Somewhat disagree

◯ Strongly disagree

Q9 Please choose the statement that best reflects your opinion on the following question: I find the incentive points for each recommended activity to conserve water useful/motivational.

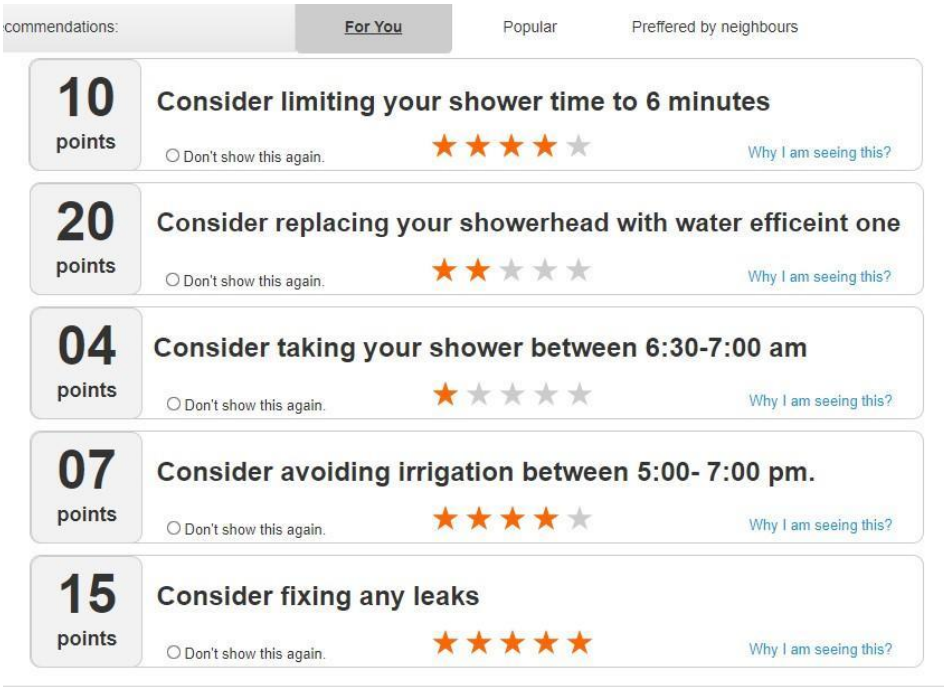
- Strongly agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Strongly disagree

Instruction Below is a reminder of the concept we are asking you to evaluate:

A Recommender System is a system that uses stored user preferences to locate, choose and suggest items (recommend items) that will be of interest to a person. For example, if you are shopping for a digital camera, on Amazon, Amazon may recommend you some camera lenses that go with your camera, or some other cameras that may interest you.

We propose to develop a similar system for promoting water-conscious behaviours which will suggest to consumers a list of activities. These activities will be custom tailored based on your behaviour patterns and will be changing based on your feedback.

In this example, you will see a picture from the recommender system for promoting water-conscious behaviour. Please note that this is not a complete screenshot but only a part of the screen.



Q10 How important are the following features to you for a recommender system to promote water conservation?

	Not important at all	Less important	Somewhat important	Important	Very important
Personalised suggestions	○	○	○	○	○
Usability of the system	○	○	○	○	○
Interaction (e.g., rating) with the system	○	○	○	○	○
Popular recommended activities	○	○	○	○	○
Points/incentive for following suggestions	○	○	○	○	○

Appendix C Copyright release for published material

Publication I

About Water

mdpi.com/journal/water/about

water

Aims

Water (ISSN 2073-4441) is an international and interdisciplinary open-access journal covering all aspects of water, including water science, technology, management and governance. It publishes original research papers, critical reviews and short communications. There is no restriction on the length of the papers.

Full methodical and/or experimental details must be provided for research articles. We encourage scientists to publish their research in as much detail as possible. Computed data or files regarding the full details of the experimental procedure or model set-up, if unable to be published as part of the main manuscript, can be deposited as supplementary material.

Scope

The scope of Water encompasses the following research areas:

- **Water resources management:** water resources systems; monitoring, remediation and protection of water resources; water resources planning; adaptive management; water demand management; national and international water policy; water economics.
- **Water governance:** institutional arrangements; water law; water rights; property regimes; trans-boundary water issues; water conflict; water politics; water security; history of water resources.
- **Hydrology & hydraulics:** catchment hydrology; modelling and remote sensing; climate change & land use change impacts; surface-groundwater interactions; soil water; aquifers; river & groundwater flow; peak and low flows; drought; variability & uncertainty.
- **Water scarcity:** water allocation & use; water recycling and reuse; reservoirs; rainwater harvesting; river basin closure; groundwater depletion; drought management.
- **Flood risk:** flood modelling; flood impacts; flood risk management; flood protection; flood resilience; flood retention basins; hazard zoning; flood insurance.
- **Water quality:** physics, chemistry and biology of water; point-source and diffuse water pollution; transport processes; nutrients; pesticides; metals; pharmaceuticals; erosion and sediment transport; hydro-geochemistry and contaminant hydrogeology.

- **Water & wastewater treatment:** water purification; treatment of domestic and industrial wastewater; water reuse; constructed wetlands; treatment wetlands.
- **Urban water management:** urban drainage; storm water management; sewerage; local water storage; permeable pavement systems.
- **Water footprint assessment:** water consumption & pollution along supply chains; virtual water trade; international water dependencies; sustainability, equitability and efficiency of water use; corporate water disclosure & stewardship; water risk.
- **Water-food:** rain-fed and irrigated agriculture; water productivity, irrigation efficiency; aquaculture; hydroponics; aquaponics.
- **Water-energy:** water use in energy production; energy use in water supply and treatment; hydropower; integrated water-carbon footprint studies.
- **Water-human development:** drinking water supply and sanitation; effects of contaminated water on health; waterborne diseases; water and poverty; water constraints to growth.
- **Water-ecosystems:** aquatic ecology; wetlands; environmental water requirements; degradation of aquatic ecosystems; preservation; restoration; environmental impacts of dams and other infrastructure; invasive species.

MDPI Publication Ethics Statement

Water is a member of the Committee on Publication Ethics (COPE). MDPI takes the responsibility to enforce a rigorous peer-review together with strict ethical policies and standards to ensure to add high quality scientific works to the field of scholarly publication. Unfortunately, cases of plagiarism, data falsification, inappropriate authorship credit, and the like, do arise. MDPI takes such publishing ethics issues very seriously and our editors are trained to proceed in such cases with a zero tolerance policy. To verify the originality of content submitted to our journals, we use iThenticate to check submissions against previous publications. MDPI works with Publons to provide reviewers with credit for their work.

Book Reviews

Authors and publishers are encouraged to send review copies of their recent related books to the following address. Received books will be listed as *Books Received* within the journal's *News & Announcements* section.

MDPI
St. Alban-Anlage 66
CH-4052 Basel
Switzerland

E-mail:

Copyright / Open Access

Articles published in *Water* will be Open-Access articles distributed under the terms and conditions of the Creative Commons Attribution License (CC BY). The copyright is retained by the author(s). MDPI will insert the following note at the end of the published text:

Reprints

Reprints may be ordered. Please contact for more information on how to order reprints.

Announcement and Advertisement

Announcements regarding academic activities such as conferences are published for free in the *News & Announcements* section of the journal. Advertisement can be either published or placed on the pertinent website. Contact e-mail address is .

Editorial Office

Ms. Lulu Ma

Managing Editor

For further MDPI contacts, see here.

Publication II

CCC | Marketplace™

Elsevier Science & Technology Journals - License Terms and Conditions

This is a License Agreement between Md Shamsur Rahim ("You") and Elsevier Science & Technology Journals ("Publisher") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Elsevier Science & Technology Journals, and the CCC terms and conditions.

All payments must be made in full to CCC.

Order Date	12-Jun-2021	Type of Use	Republish in a thesis/dissertation
Order License ID	1125414-1		
ISSN	0301-4797	Publisher	ACADEMIC PRESS
		Portion	Chapter/article

LICENSED CONTENT

Publication Title	Journal of environmental management	Rightsholder	Elsevier Science & Technology Journals
Article Title	Advanced household profiling using digital water meters	Publication Type	Journal
		Start Page	112377
		Volume	288
Date	01/01/1973	URL	http://www.journals.elsevier.com/journal-of-environmental-management/
Language	English		
Country	United Kingdom of Great Britain and Northern Ireland		

REQUEST DETAILS

Portion Type	Chapter/article	Rights Requested	Main product
Page range(s)	1-10	Distribution	Worldwide
Total number of pages	10	Translation	Original language of publication
Format (select all that apply)	Electronic	Copies for the disabled?	No
Who will republish the content?	Author of requested content	Minor editing privileges?	No
Duration of Use	Life of current edition	Incidental promotional use?	No
Lifetime Unit Quantity	Up to 499	Currency	AUD

NEW WORK DETAILS

Title	Using Data Analytics and Artificial Intelligence-based techniques to promote water-conscious behaviour through personalized recommendations	Institution name	University of Technology Sydney
		Expected presentation date	2021-06-18
Instructor name	Md Shamsur Rahim		

ADDITIONAL DETAILS

Order reference number	N/A	The requesting person / organization to appear on the license	Md Shamsur Rahim

REUSE CONTENT DETAILS

Title, description or numeric reference of the portion(s)	Publication II	**Title of the article/chapter the portion is from**	Advanced household profiling using digital water meters
Editor of portion(s)	Rahim, Md Shamsur; Nguyen, Khoi Anh; Stewart, Rodney Anthony; Giurco, Damien; Blumenstein, Michael	**Author of portion(s)**	Rahim, Md Shamsur; Nguyen, Khoi Anh; Stewart, Rodney Anthony; Giurco, Damien; Blumenstein, Michael
Volume of serial or monograph	288	**Issue, if republishing an article from a serial**	N/A
Page or page range of portion	112377	**Publication date of portion**	2021-06-15

PUBLISHER TERMS AND CONDITIONS

Elsevier publishes Open Access articles in both its Open Access journals and via its Open Access articles option in subscription journals, for which an author selects a user license permitting certain types of reuse without permission. Before proceeding please check if the article is Open Access on http://www.sciencedirect.com and refer to the user license for the individual article. Any reuse not included in the user license terms will require permission. You must always fully and appropriately credit the author and source. If any part of the material to be used (for example, figures) has appeared in the Elsevier publication for which you are seeking permission, with credit or acknowledgement to another source it is the responsibility of the user to ensure their reuse complies with the terms and conditions determined by the rights holder. Please contact permissions@elsevier.com with any queries.

CCC Republication Terms and Conditions

1. Description of Service; Defined Terms. This Republication License enables the User to obtain licenses for republication of one or more copyrighted works as described in detail on the relevant Order Confirmation (the "Work(s)"). Copyright Clearance Center, Inc. ("CCC") grants licenses through the Service on behalf of the rightsholder identified on the Order Confirmation (the "Rightsholder"). "Republication", as used herein, generally means the inclusion of a Work, in whole or in part, in a new work or works, also as described on the Order Confirmation. "User", as used herein, means the person or entity making such republication.

2. The terms set forth in the relevant Order Confirmation, and any terms set by the Rightsholder with respect to a particular Work, govern the terms of use of Works in connection with the Service. By using the Service, the person transacting for a republication license on behalf of the User represents and warrants that he/she/it (a) has been duly authorized by the User to accept, and hereby does accept, all such terms and conditions on behalf of User, and (b) shall inform User of all such terms and conditions. In the event such person is a "freelancer" or other third party independent of User and CCC, such party shall be deemed jointly a "User" for purposes of these terms and conditions. In any event, User shall be deemed to have accepted and agreed to all such terms and conditions if User republishes the Work in any fashion.

3. Scope of License; Limitations and Obligations.

   3.1. All Works and all rights therein, including copyright rights, remain the sole and exclusive property of the Rightsholder. The license created by the exchange of an Order Confirmation (and/or any invoice) and payment by User of the full amount set forth on that document includes only those rights expressly set forth in the Order Confirmation and in these terms and conditions, and conveys no other rights in the Work(s) to User. All rights not expressly granted are hereby reserved.

   3.2. General Payment Terms: You may pay by credit card or through an account with us payable at the end of the month. If you and we agree that you may establish a standing account with CCC, then the following terms apply: Remit Payment to: Copyright Clearance Center, 29118 Network Place, Chicago, IL 60673-1291. Payments Due: Invoices are payable upon their delivery to you (or upon our notice to you that they are available to you for downloading). After 30 days, outstanding amounts will be subject to a service charge of 1-1/2% per month or, if less, the maximum rate allowed by applicable law. Unless otherwise specifically set forth in the Order Confirmation or in a separate written agreement signed by CCC, invoices are due and payable on "net 30" terms. While User may exercise the rights licensed immediately upon issuance of the Order Confirmation, the license is automatically revoked and is null and void, as if it had never been issued, if complete payment for the license is not received on a timely basis either from User directly or through a payment agent, such as a credit card company.

3.3. Unless otherwise provided in the Order Confirmation, any grant of rights to User (i) is "one-time" (including the editions and product family specified in the license), (ii) is non-exclusive and non-transferable and (iii) is subject to any and all limitations and restrictions (such as, but not limited to, limitations on duration of use or circulation) included in the Order Confirmation or invoice and/or in these terms and conditions. Upon completion of the licensed use, User shall either secure a new permission for further use of the Work(s) or immediately cease any new use of the Work(s) and shall render inaccessible (such as by deleting or by removing or severing links or other locators) any further copies of the Work (except for copies printed on paper in accordance with this license and still in User's stock at the end of such period).

3.4. In the event that the material for which a republication license is sought includes third party materials (such as photographs, illustrations, graphs, inserts and similar materials) which are identified in such material as having been used by permission, User is responsible for identifying, and seeking separate licenses (under this Service or otherwise) for, any of such third party materials; without a separate license, such third party materials may not be used.

3.5. Use of proper copyright notice for a Work is required as a condition of any license granted under the Service. Unless otherwise provided in the Order Confirmation, a proper copyright notice will read substantially as follows: "Republished with permission of [Rightsholder's name], from [Work's title, author, volume, edition number and year of copyright]; permission conveyed through Copyright Clearance Center, Inc. " Such notice must be provided in a reasonably legible font size and must be placed either immediately adjacent to the Work as used (for example, as part of a by-line or footnote but not as a separate electronic link) or in the place where substantially all other credits or notices for the new work containing the republished Work are located. Failure to include the required notice results in loss to the Rightsholder and CCC, and the User shall be liable to pay liquidated damages for each such failure equal to twice the use fee specified in the Order Confirmation, in addition to the use fee itself and any other fees and charges specified.

3.6. User may only make alterations to the Work if and as expressly set forth in the Order Confirmation. No Work may be used in any way that is defamatory, violates the rights of third parties (including such third parties' rights of copyright, privacy, publicity, or other tangible or intangible property), or is otherwise illegal, sexually explicit or obscene. In addition, User may not conjoin a Work with any other material that may result in damage to the reputation of the Rightsholder. User agrees to inform CCC if it becomes aware of any infringement of any rights in a Work and to cooperate with any reasonable request of CCC or the Rightsholder in connection therewith.

4. Indemnity. User hereby indemnifies and agrees to defend the Rightsholder and CCC, and their respective employees and directors, against all claims, liability, damages, costs and expenses, including legal fees and expenses, arising out of any use of a Work beyond the scope of the rights granted herein, or any use of a Work which has been altered in any unauthorized way by User, including claims of defamation or infringement of rights of copyright, publicity, privacy or other tangible or intangible property.

5. Limitation of Liability. UNDER NO CIRCUMSTANCES WILL CCC OR THE RIGHTSHOLDER BE LIABLE FOR ANY DIRECT, INDIRECT, CONSEQUENTIAL OR INCIDENTAL DAMAGES (INCLUDING WITHOUT LIMITATION DAMAGES FOR LOSS OF BUSINESS PROFITS OR INFORMATION, OR FOR BUSINESS INTERRUPTION) ARISING OUT OF THE USE OR INABILITY TO USE A WORK, EVEN IF ONE OF THEM HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. In any event, the total liability of the Rightsholder and CCC (including their respective employees and directors) shall not exceed the total amount actually paid by User for this license. User assumes full liability for the actions and omissions of its principals, employees, agents, affiliates, successors and assigns.

6. Limited Warranties. THE WORK(S) AND RIGHT(S) ARE PROVIDED "AS IS". CCC HAS THE RIGHT TO GRANT TO USER THE RIGHTS GRANTED IN THE ORDER CONFIRMATION DOCUMENT. CCC AND THE RIGHTSHOLDER DISCLAIM ALL OTHER WARRANTIES RELATING TO THE WORK(S) AND RIGHT(S), EITHER EXPRESS OR IMPLIED, INCLUDING WITHOUT LIMITATION IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. ADDITIONAL RIGHTS MAY BE REQUIRED TO USE ILLUSTRATIONS, GRAPHS, PHOTOGRAPHS, ABSTRACTS, INSERTS OR OTHER PORTIONS OF THE WORK (AS OPPOSED TO THE ENTIRE WORK) IN A MANNER CONTEMPLATED BY USER; USER UNDERSTANDS AND AGREES THAT NEITHER CCC NOR THE RIGHTSHOLDER MAY HAVE SUCH ADDITIONAL RIGHTS TO GRANT.

7. Effect of Breach. Any failure by User to pay any amount when due, or any use by User of a Work beyond the scope of the license set forth in the Order Confirmation and/or these terms and conditions, shall be a material breach of the license created by the Order Confirmation and these terms and conditions. Any breach not cured within 30 days of written notice thereof shall result in immediate termination of such license without further notice. Any unauthorized (but licensable) use of a Work that is terminated immediately upon notice thereof may be liquidated by payment of the Rightsholder's ordinary license price therefor; any unauthorized (and unlicensable) use that is not terminated immediately for any reason (including, for example, because materials containing the Work cannot reasonably be recalled) will be subject to all remedies available at law or in equity, but in no event to a payment of less than three times the Rightsholder's ordinary license price for the most closely analogous licensable use plus Rightsholder's and/or CCC's costs and expenses incurred in collecting such payment.

8. Miscellaneous.

    8.1. User acknowledges that CCC may, from time to time, make changes or additions to the Service or to these terms and conditions, and CCC reserves the right to send notice to the User by electronic mail or otherwise for the purposes of notifying User of such changes or additions; provided that any such changes or additions shall not apply to permissions already secured and paid for.

    8.2. Use of User-related information collected through the Service is governed by CCC's privacy policy, available online here:https://marketplace.copyright.com/rs-ui-web/mp/privacy-policy

    8.3. The licensing transaction described in the Order Confirmation is personal to User. Therefore, User may not assign or transfer to any other person (whether a natural person or an organization of any kind) the license created by the Order Confirmation and these terms and conditions or any rights granted hereunder; provided, however, that User may assign such license in its entirety on written notice to CCC in the event of a transfer of all or substantially all of User's rights in the new material which includes the Work(s) licensed under this Service.

    8.4. No amendment or waiver of any terms is binding unless set forth in writing and signed by the parties. The Rightsholder and CCC hereby object to any terms contained in any writing prepared by the User or its principals, employees, agents or affiliates and purporting to govern or otherwise relate to the licensing transaction described in the Order Confirmation, which terms are in any way inconsistent with any terms set forth in the Order Confirmation and/or in these terms and conditions or CCC's standard operating procedures, whether such writing is prepared prior to, simultaneously with or subsequent to the Order Confirmation, and whether such writing appears on a copy of the Order Confirmation or in a separate instrument.

    8.5. The licensing transaction described in the Order Confirmation document shall be governed by and construed under the law of the State of New York, USA, without regard to the principles thereof of conflicts of law. Any case, controversy, suit, action, or proceeding arising out of, in connection with, or related to such licensing transaction shall be brought, at CCC's sole discretion, in any federal or state court located in the County of New York, State of New York, USA, or in any federal or state court whose geographical jurisdiction covers the location of the Rightsholder set forth in the Order Confirmation. The parties expressly submit to the personal jurisdiction and venue of each such federal or state court.If you have any comments or questions about the Service or Copyright Clearance Center, please contact us at 978-750-8400 or send an e-mail to support@copyright.com.

v 1.1