Leveraging Potentials of Big Data for Better Decision-Making and Value Creation in Nonprofit Organisations

by **Idrees Alsolbi**

Thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

in Information Systems

under the supervision of

Dr. Mukesh Prasad
Australian Artificial Intelligence Institution, School of Computer Science, Faculty of Engineering and Information Technology, University of Technology Sydney
and
co-supervised by
Prof. Renu Agarwal
Business School, University of Technology Sydney
and
Prof. Bhuvan Unhelkar
College of Business, University of South Florida, Florida, USA

University of Technology Sydney
Faculty of Engineering and Information Technology
School of Computer Science
April 2023

CERTIFICATE OF ORIGINAL AUTHORSHIP

I, *Idrees Alsolbi* declare that this thesis, is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the *School of Computer Science*, *Faculty of Engineering and Information Technology.* 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

DATE: 4TH APRIL, 2023

ABSTRACT

In Nonprofit Organisations, analysing and understanding donor behaviour remain critical and challenging. While big data and machine learning techniques promise technical solutions to address this problem, how to design and build an intelligent decision support system based on these technologies remains unclear. The literature reveals that nonprofit organisations are deficient in using various data analytics due to a lack of expertise, low financial budgets and insufficient awareness of data analytics capabilities that enable those organisations to be data-driven and decision-making beneficiaries. Therefore, analysing and understanding donor behaviour remain critical challenges for nonprofit organisations. To address these research gaps, the researcher adopted a design science framework which helped to create an artefact (an intelligent decision support system) to analyse donor behaviour in nonprofit organisations. In addition, the framework led to the creation a design theory of the artefact which guides the design process and generalises the design requirements of such analytical and decision-making solutions for NPOs. The results show that (by analysing public big data sets of donors from different sources) certain variables are essential to analyse donor behaviour in nonprofit organisations. These variables are the total amount of donations, the number of donations, gender, age, social level of income, educational level, and the frequency of donations. Furthermore, these variables assist the researcher in choosing the appropriate analysis model, from classification to predictions, and deciding the most beneficial machine learning techniques that generate a useful analysis for nonprofit organisations. The researcher aims to provide a theoretical foundation design for developing an intelligent decision support system for analysing donor behaviour. The research contributes to decision support and data analytics research

by presenting the capabilities of data analytics and machine learning techniques in the context that face the difficulty of understanding donor behaviour. Finally, it contributes to the literature by producing descriptive and predictive analytics models to support nonprofits for leveraging applications of data analytics and big data awareness.

Dedication

To my parents and my family who supported me and suffered through living without me during this journey. To my wife, who stood next to me during every stage of this journey, with her support and encouragement. To everyone who cared and took action to help me achieve this dream of obtaining a Doctor of Philosophy degree.

I acknowledge the support I have received for my research through the provision of the Saudi Ministry of Education Scholarship (Saudi Arabia Cultural Mission in Australia), an Umm Al-Qoura University Scholarship Program.

Foremost, I would like to express my sincere gratitude to my supervisors Dr Mukesh Prasad, Professor Renu Agarwal, and Professor Bhuvan Unhelkar, for supporting my PhD study and research journey, and for their patience, motivation, enthusiasm, and immense knowledge. Their guidance helped me throughout the research and writing of this thesis. Most importantly, I would like to thank my family: my parents Nasser and Athbah, my wife Aliah, and my brothers and sisters, for their unconditional support, both financial and emotional, throughout my whole PhD journey. Last but not least, I would like to thank collaborators, scholars, and data science experts from external universities and nonprofit organisations who provided me with critical advice and provided suggestions at every stage of this journey. I thank all panel experts from three candidature assessments at the University of Technology Sydney for their comments and guidance.

Idrees Alsolbi

March 2023 @UTS

Glossary of Acronyms

NPO	Non-profit organisation
DSS	Decision Support System
AI	Artificial Intelligence
ML	Machine Learning
IS	Information Systems
CS	Computer Science
RG	Research Gap
RO	Research Objective
RQ	Research Questions
SEP	Scientific Evolutionary Paths
DSR	Design Science Research
DP	Design Principle
DF	Design Feature
DR	Design Requirement

TABLE OF CONTENTS

List of Figures

List of Tables

CHAPTER 1

INTRODUCTION

Nonprofit organisations (NPOs) are entities created with the intention of providing a service or benefit to the public for the purpose of advancing a specific mission, rather than generating a profit for private ownership or shareholders. This type of organisation is typically sustained by donations, grants, and other forms of philanthropy (Lewis, 2021). NPOs strive to serve the public good as opposed to generating revenue. NPOs can also be defined as "third sector", "voluntary or giving sector", and "not-for-profits" (Anheier, 2005).
NPOs differ from other businesses and companies in that they are private, organised as institutions independent of the government, self-governing (they manage and control their practices and missions), and the profits they acquire are used to benefit other organisations (Anheier, 2005). There are several examples of NPOs, such as museums, schools, universities, health centres, human rights organisations, religious organisations, charities, youth centres and humanitarian aid. Anheier (2005) notes that NPOs' missions cover individual activities and the values and motivations that drive people to engage in activities to benefit society, the environment, and cultural heritage through charities, philanthropy, volunteering, and giving. Three types of financial resources can be used by NPOs: monetary, in-kind, and labour, both paid and unpaid.
In addition, Monetary resources include grants, contributions, sales revenue, and service fees (Anheier, 2005). However, in different nations, most NPOs benefit financially from grants and service fees for managing fundraising campaigns (Anheier, 2005). NPOs often have a variety of diverse revenue sources that can be categorised,

and their revenue structure is more complex than that of for-profit businesses and government entities (Anheier, 2005). According to Anheier (2005), examples of categories of the revenue sources are:

- Origin: (marketing, public sector, organisation, individual).
- Intent: (gifts or goods, exchanges or transfers).
- Kind: (time, services, monetary or in-kind).
- Formality: (informal donations or contract-based exchange).
- Restrictions: (any fund has restrictions).
- Source: (donations, fees of users, sales of services or goods).

As claimed by Anheier (2005), source or origin is most utilised and includes:

- Public payments, which include grants and contracts.
- Private giving such as operating foundations.
- Private charges or fees of the NPO's services and investments.
- Formality: (informal donations or contract-based exchange).
- Restrictions: (any fund has restrictions).
- Source: (donations, fees of users, sales of services or goods).

Nonprofit organisations can generate income through a variety of sources, including donations and grants (Weisman, 2020), corporate sponsorships (Tejada, 2020), fundraising events (Smith, 2021), individual membership dues (Barker, 2019), government funding (Gutierrez, 2021), corporate matching gifts, merchandise sales (Barker, 2019), and royalties from intellectual property (Wang, 2020). The revenue sources of NPOs vary in different nations. For example, only 9.5% of NPOs' revenue in Australia comes from public donations, with the government providing 33.5% of an NPO's income (Commission, 2010). Australia's public donation percentage is higher than that of Germany, France, and New Zealand, but unsurprisingly, lower than that of the United States. During 2014-15 in Australia, the direct value added by the charity sector was $71.8 billion (Charities and for-profits Commission, 2021). It is noted that donations rose by 8% to $12.7 billion annually in recent years (Charities and for-profits Commission, 2021). NPOs, according to (Anheier, 2005), are now a substantial social and economic force. According to Rathi and Given (2017), NPOs have a sizable impact on national economies. For instance, in Australia, the economy benefited by almost $43 million in 2007 (Commission, 2010).

NPOs can significantly influence society by attracting donors and volunteers and establishing strong relationships with clients to pursue NPO interests. Donors support the missions of NPOs in different ways, such as by giving money, gifts, and time. Donors also volunteer their time and experience at various events. According to a study done by Dietz and Keller (2016), donors gave around $260 billion to American fundraising campaigns. Private donations represent a significant portion of NPO funding in the United States, which annually accounts for more than 10% of the GDP, or Gross Domestic Product (Farrokhvar et al., 2018). Dietz and Keller (2016) reports that individuals donate because of their deep passion or belief in the specific NPO's needs. It is believed that certain factors impact peoples' intentions towards donating, such as income, educational level, and previous giving history (Farrokhvar et al., 2018).

1.1 Research Background

Donor behaviour is an area of research within the field of philanthropy which seeks to understand the attitudes and decision-making processes of individuals and organisations when making charitable donations. Donor behaviour is the act of willingly offering money, products, or time to someone in need in order to improve the recipient's well-being (Zhou et al., 2012). Donor behaviour plays a pivotal role in determining the success of an NPO, as it offers an understanding of how and why donors are motivated to contribute to a particular cause.

Donor behaviour research can offer valuable insights to organisations on how to interact with donors and enhance fundraising initiatives. This includes exploring the factors that influence why people choose to donate, which causes they support, how their donations impact the charity or cause they support, and how their motivations for giving change over time. There are several theories which have been proposed to explain donor behaviour, such as the theory of planned behaviour (Ajzen, 1991), social exchange theory (Etzioni, 1968), and the principle of reciprocal altruism (Trivers, 1971). Several empirical studies have applied these theories to donor behaviour and can provide valuable insights into donors' motivations and decision-making processes.

Donor behaviour research can be employed to better comprehend donor motivations and preferences, providing insight into how NPOs can engage better and retain donors. For example, research has revealed that donors are motivated by recognition, appreciation, and a sense of making a meaningful contribution to society (Lott and Brinkerhoff, 2008). This information can be utilised to create more efficient donor retention strategies and craft personalized messages that resonate with donors. Additionally, comprehending donor behaviour can provide insight into optimising fundraising efforts, such as by targeting donors who are more likely to give, or by adjusting the frequency of fundraising campaigns (McKinney and Chatterjee, 2008).

Interestingly, one may argue about the similarity of donor behaviour to consumer behaviour in other fields. Donor behaviour is the action of providing financial aid to causes or organisations, while consumer behaviour is the action of purchasing products or services from a business. The relationship between donor behaviour and consumer behaviour is strong because both involve decision-making processes related to the allocation of resources.

When making decisions about how to use their resources, individuals are engaging in either consumer behaviour or donor behaviour. Both involve a mix of emotional and cognitive aspects, such as understanding the product or cause and gauging its subjective value. The key difference between the two is that consumer behaviour relates to the purchase of goods and services for personal use, while donor behaviour involves giving money or other resources to a specific cause. Both involve the consideration of potential rewards, be it satisfaction from a purchase or the feeling of having made a meaningful contribution. Ultimately, however, both forms of behaviour are driven by a desire to improve one's life's.

A recent study introduced "prosocial consumer behaviour", which can include doing an action that aids or benefits an individual or a group, but it can also apply to more broad actions that benefit society as a whole (White et al., 2020). Prosocial consumer behaviour, ethical purchasing, cause-related involvement, charity giving, various donation behaviours (such as organ and blood donation), volunteering, and consumer advocacy or activism are only a few examples of prosocial consumer behaviours (White et al., 2020). Nevertheless, in this research, the term donor behaviour is used.

Nevertheless, in this research, the term "donor behaviour" is used. Donor behaviour is determined by a range of external and internal forces that may affect an individual's choices when it comes to giving. Despite the variety of influential factors on donor behaviour, understanding the fundamentals of donor behaviour is crucial, according to Li and Wu (2019). Understanding and analysing donor behaviour can assist NPOs in increasing marketing and fundraising efficiency. Today's NPOs focus not only on gaining donations but also on knowing donors' habits, which leads NPOs to authentically interact with their donors and learn how they resonate with them (Te, 2019). One important behaviour is the return, or intention to donate for a second time. Te (2019) reported that 19% donors donate for a second time. However, (Sargeant and Jay, 2014) mentions that targeting appropriate donors to charities and improving communications with them remain critical for NPOs.

There are certain behaviours, including donor intentions to donate time or money, donor profiling (their demographic, financial, and social characteristics), donor retention (who is likely to donate or volunteer again) (Te, 2019), donor engagement, donor communications, and volunteer engagement, which require a deeper understanding. Consequently, the advancement of technologies, capabilities of data analytics, and machine learning (ML) techniques can help understand donor behaviour. Some influential determinants on donor behaviour towards contributing included donor education level, gender, age, population, household income, and ethnicity (Farrokhvar et al., 2018). By analysing these behaviours, NPOs improve their chances of increasing their current financial support and interaction with outgoing donors for potential opportunities for repeat donation activity (Dunford, 2016).

However, Bopp et al. (2017) claims that managing information in NPOs is challenging. Maxwell et al. (2016) added that if the data is not well-collected and organised, NPOs will not benefit from the available data to draw insights and conclusions. To enable the usage of the data, NPOs should adopt strategies that clarify employees' understanding and utilise the data to achieve the organisations' tasks goals and mission (Maxwell et al., 2016). According to Mayer

(2019), the relationship between data analytics and technology in NPOs is uncertain. Data analytics allows monitoring and evaluating specific situations by identifying significant issues, influencing policy through evidence, improving fundraising capabilities, and understanding donor behaviour, as cited by Bopp et al. (2017); Maxwell et al. (2016); Mayer (2019); Rathi and Given (2017). For this reason, involving the capable technologies of data analytics and big data analytics can solve various problems and provide meaningful information for organisational management.

1.1.1 Data and Big Data Analytics in NPOs

Data technology is evolving rapidly to address global needs. Professionals and practitioners from for-profit organisations are being urged to consider data analytics applications to maximise their potential and increase productivity. This rapid development has made professionals from various for-profit organisations aware of the potential of applying big data analytics to their business operations. In contrast, NPOs have paid scant attention to the applications of big data technology in their activities. According to a survey by Everyaction (2018) about the usage of big data, 90% of NPOs collect data from different sources, but only half of this data is utilised. As noted by Bopp et al. (2017), many NPOs store their data on spreadsheets and process it manually rather than digitising it, which is more efficient. However, Mayer (2019) notes major issues with manual processing such as privacy violations and unintended discrimination, which NPOs can no longer ignore. However, there are recent calls to apply such analytics to NPOs' activities to analyse donors, for example with consideration of privacy and security for the dataMayer (2019); Johnson (2015).

The role of data analytics in NPOs' activities is significant; it can assist these organisations in monitoring, evaluating, and determining barriers to their success, and can provide meaningful visualisations to support decision-makers. Big data can provide comprehensive information about donors, members, and NPO managers, and make organisations more efficient (Mayer, 2019). Many organisations prioritise big data analytics because it can support productivity, reduce costs, and improve customer experiences (Kalema and Mokgadi, 2017). Big data analytics can assist managers, leadership teams, and decision-makers. Most importantly, big data provides a quantitative basis for resource deployment (Klievink et al., 2017). Big data also assists non-governmental agencies and the general public by contributing to innovation and cost-effectiveness (Kassen, 2018).

NPOs can significantly benefit from the integration of data analytics and big data, due to technologies' abilities to optimize efficiency, pinpoint areas of improvement, and gain insights into NPOs' operations. Data analytics and big data allow large and complex datasets to be organised and interpreted to generate insights that would otherwise be lost (Kaur et al., 2017). Furthermore, big data technologies can identify new strategies to optimise fundraising and improve decision-making and resource allocation (Kaur et al., 2017). By leveraging data analytics and big data technologies, NPOs can gain deeper insights into their operations and more accurately target resources, which will ultimately lead to more efficient and effective operations.

1.1.2 Definition and Characteristics of Big Data

By definition, big data is a massive amount of data in different forms and comes from various resources (Klievink et al., 2017; Mahmoud and Yusif, 2012) and cannot be handled by standard data processing technologies (Klievink et al., 2017). According to (Mayer, 2019), big data in NPOs has three characteristics: Volume (size of the data), Variety (data comes from various), and Velocity (the speed at which data is being created or stored) (Kalema and Mokgadi, 2017; Costa and Santos, 2017). Other scholars include two additional characteristics: Value (can the data be trusted to generate insights?) and Veracity (the quality or credibility of the data) (Mahmoud and Yusif, 2012; Shah et al., 2017).

1.2 Research Context

The nature of fundraising is evolving due to donors using social media more frequently and being exposed to a wide variety of campaigns. Therefore, NPOs need to deeply understand donor behaviour to involve more individuals for continuous support. A thorough understanding of long-term intentions as part of the overall donation experience can help nonprofits implement consumer-driven marketing strategies to attract and retain donors (Kashif and Zarkada, 2015). Nevertheless, NPOs need time, expertise, and managerial resources to assess and offer useful input on opportunities for improvement. Most importantly, NPOs need to engage in research activities, knowl-

edge exchange, and evaluation and assessment of present business procedures (Commission, 2010).
On the contrary, data analytics generates useful insights and analytics and visualises through artificial intelligence (AI) DSS using ML techniques. In this research context, and according to a study conducted by Dietz and Keller (2016), donors are divided into three categories: Giving (money, goods and services, making purchases, and so forth); Doing (volunteering, attending events, serving in a leadership role, and so on); and Communicating (Spreading the word, advocating, following on social media, and staying informed).
In this research context, the focus is on giving money and partaking in volunteer activities. This focus is to obtain such useful insights which may lead in future to consider other behaviour of donors, such donors' communication. Also, considering the time frame and data availability, donors' communication is excluded from this research. It has been found that donors who give money and time are older, married, and have children; they live in well-established, more affluent neighbourhoods and have solid financial backgrounds from earning higher incomes, receiving valuable gifts, and inheriting money (Shehu et al., 2015). Therefore, to narrow the scope of this research, we classified donors (who give money) and volunteers (who do activities) under donor behaviour to build predictive and descriptive analyses that help NPOs make better decisions.
Various big data sources are considered for building descriptive and predictive analyses:

1- Literature data represents our core database for establishing a systematic review of a research survey. Large-scale databases are reliable for finding relevant documents, including books, book chapters, peer-reviewed journal articles, conference papers, conference proceedings, academic reports, and online reports. Scopus, Web of Science and ProQuest are common databases containing big literature on data analytics applications.

2- Public datasets of donors and volunteers (these are demographic data which contain useful descriptions of donors and volunteers to help create useful analyses and predictions). These datasets contain millions of rows and hundreds of columns which contain relevant data. Census organisations, both public and private, gather demographic data for a variety of purposes, including research, marketing, and environmental and human development, which can be incorporated into datasets. Government bodies can use this data, such as population and employment data, along with its related data fields like density, ethnicity, and gender, to plan for infrastructure initiatives, such as roads, hospitals and law enforcement (Techpedia, 2020).

Most importantly, the nature of donations and volunteering analysis in this research are considered under "charitable donations", which is slightly different from other types of donations such as "political donations". Political donations are financial contributions made to a political organisation or individual running for office to assist their campaign activities. These donations may be provided by individuals, businesses, labour unions, and other organisations. Political donations present unique issues of corporate governance that set them apart from charitable donations. Charitable donations are typically given with the aim of benefiting a cause or organisation, but political donations confer a more direct benefit, as the party receiving the donation can potentially return the favour when they hold power (Vernon et al., 2001).

This research motivates small and medium-sized NPOs to be data-driven and to avoid any pitfalls of donor and volunteer relationships. Moreover, this research aspires to provide insights into the capabilities of data analytics applied to NPOs' big data as the research carries out various descriptive and predictive analyses. Most importantly, the current study is driven by the need to emphasise the significance of data analytics in analysing donor behaviour in NPOs. Moreover, the current study is driven to develop an artificial intelligence, or AI-enabled DSS to help NPOs understand donor behaviour using ML techniques.

Design theory is another angle of the research context, which helps scholars understand the main requirements of designing AI-enabled DSS for analysing behaviours. Design theory is meant to provide such instantiation of the design process of the DSS (Peffers et al., 2007). The AI-enabled DSS is a type of DSS relying on AI and applied techniques.

This research lies at the intersection of three disciplines (Figure 1.1), which are:

1- The core problem for NPOs is how to analyse donor behaviour by relying on advanced capabilities of big data and ML techniques.

2- With **Information Systems (IS)**, the research methodology involves a design science approach to guide the design of a solution to the problem in (1).

3- Computer Science (CS) uses big datasets from certain databases to review the literature to determine the research gaps by applying bibliometric analysis. Bibliometrics, often called "scientometrics", is a quantitative tool for measuring and mapping existing research in a scientific area (Trivedi, 2019).
A bibliometric analysis is a study that aims to uncover the fundamental structure of a field's research (Shukla et al., 2020). It analyses research trends in depth, evaluates science as a productive knowledge system, and is trustworthy and objective (Esfahani et al., 2019). There are different applications for bibliometric analysis, however; Biblometrix, an associated tool with an R-package (Aria and Cuccurullo, 2017), was created in response to the perception that scientometric analysis is difficult and unclear.
The Bibliometrix software allows researchers to import data from Scopus, Web of Science, PubMed, and Cochrane databases, and to perform bibliometric analysis (Aria and Cuccurullo, 2017). It includes data metrics for author and country collaboration analysis, word analysis, author productivity analysis, and author collaborations. Limited research has been conducted to explain the evolution of various scientific subjects around data analytics, to identify emerging trends, and to assess the performance and influence of countries, regions, scholars, and research organisations on specific fields of science (Trivedi, 2019; Shukla et al.,

2020; Esfahani et al., 2019).

Meanwhile, the solution to the research problem implies data analytics and ML techniques using big datasets of donors and volunteers. As disciplines, IS and CS lead to the creation of an instantiation for the research problem. The big datasets used in this research contain documents and records from the literature, demographic information about donors and volunteers, such as their age, gender, education level, and history of donations and volunteering.

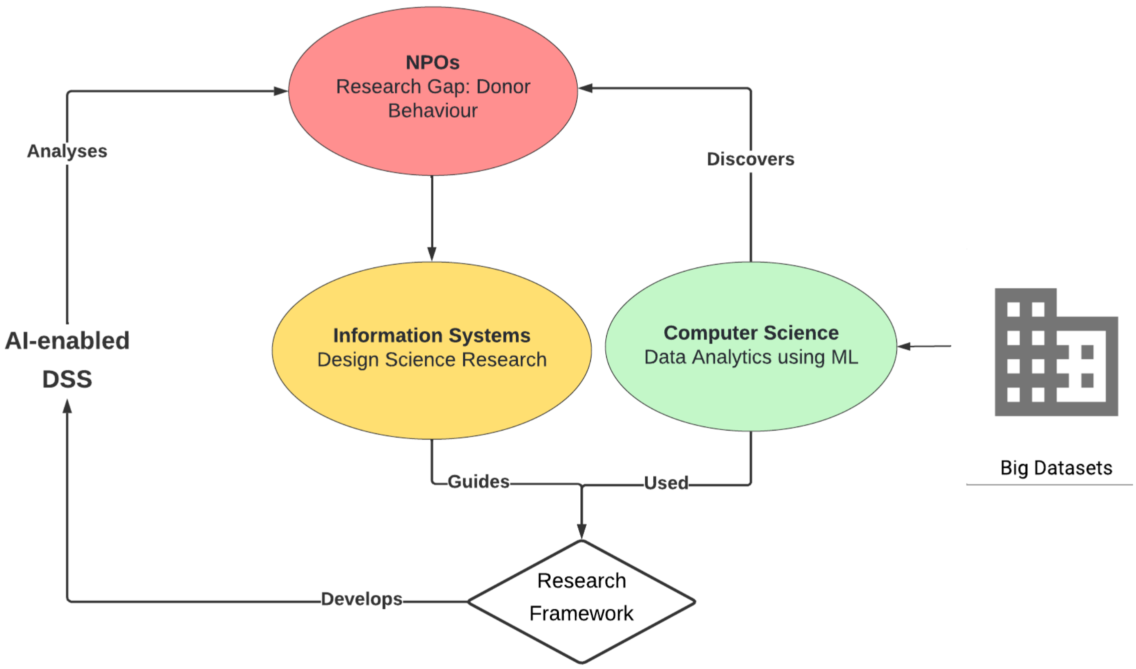


Figure 1.1: The Intersections of the Research

1.3 Problem Statement

To ensure that the research problem is validated, the search in the literature review was updated regularly on common databases such as Scopus, Web of Science, and ProQuest. For instance, the researcher kept the records of documents updated on related topics of analysing donor behaviour until February 2023. However, the current literature review highlights the observation that data analytics, ML techniques, and DSSs have not been applied, despite their comprehensive involvement in NPOs. We found certain research gaps (RGs) that need to be void, which are:

- **RG1**: There is scant research on a design theory for an analytical solution to donor behaviour. Design theory is a set of principles and features that guide designers to create desired systems. It delivers functional explanations within a simple structure to explain a generalised solution (Baskerville and Pries-Heje, 2010). With a respectful consideration to the studies of analysing and understanding donors behaviours done by (Korolov et al., 2016; Farrokhvar et al., 2018), there is limited evidence of constructing a design theory on the created predictive, descriptive models in these studies. Also, the designed DSS developed by (Barzanti et al., 2017) lacks statements about the design principles and features, which may require extensions in the future. Hence, designing a technical solution without a guideline or guidance may result in difficulty in generalising the outcomes of that solution. In a similar designed solution for a decision-making problem,Rhyn and Blohm (2017a) developed a design theory to assist with semi-automated information processing and decision-making on crowdsourcing platforms through the use of text mining and machine learning algorithms. Although their study covers a research gap in another area, it would provide a core base for constructing the design theory of a DSS that follows the reported steps.

- **RG2**: There are limited attempts to design a DSS for analysing donor behaviours around their intentions of donating to NPOs. A call was made by Maxwell et al. (2016)'s study that a DSS can benefit NPOs by identifying challenges to conducting activities, and supporting cultures to promote sustainable decision-making. Despite the availability of tools and technologies, some NPOs still fail to implement a DSS, analyse and use the data to decide according to Maxwell et al. (2016). Thus, there is a lack in the literature to highlight and discuss the implications of DSS in NPOs, especially in analysing donor behaviour. There are only a few attempts have been made of case studies to predict donor intentions towards donations (Farrokhvar et al., 2018; Arnott and Pervan, 2012). However, each study applied a single ML technique which may have resulted in a bias. It is argued that analysing donors and volunteers requires a trial of different ML techniques to help identify the correlations between the influential factors on their donations and volunteering activities. Hence, using technologies for analysing donors' data can significantly increase return on investment by raising donations from donors, fostering solid relationships, and cutting expenses

(Nabar, 2021).

1.4 Research Scope

This research aims to assist NPOs in developing a DSS to help them understand the influential factors surrounding donors and their contributions of time and money. The study is scoped to create a design framework and identify the constructs, methods, and analytics models to guide and solve the proposed problems. A deeper understanding of donor behaviour and factors that promote donations is crucial in a world increasingly dependent on NPOs to support community development and growth, particularly in light of budget scarcity. Applying data analytics using different ML techniques enables NPOs to interpret, understand, and predict donor behaviours.

This research project attempts to fill this void by exploring the intentions to donate and volunteer by applying data analytics of the expected actions of donors and volunteers and evaluating the generated models. Descriptive and predictive analysis is carried out through the use of statistical analysis. While charities cannot monitor who donates, they can monitor their own actions to create the best chances of success. This shows that statistical modelling can play a powerful role in organisations that rely on charitable donations from their constituents. Further, involving design theories is effective management for a design problem (such as applying data analytics to NPOs donors' data to understand their behaviours) to support researchers with foundational theories to deploy and evaluate the usage patterns of related instantiation.

1.5 Research Questions

The researcher initiates two major research questions (RQs) corresponding to the two research gaps discussed earlier to create a design theory that helps build a DSS for analysing donor behaviour. The research questions help design AI-enabled DSS and create the design theory for building this system. For the creation of a design theory, and a DSS, a systematic literature review and a scientific framework are conducted to ensure the validity of the research outcomes.

- **RQ1. Corresponding to RG1, what is the design theory that guides the development of an AI-enabled DSS to analyse donor behaviour in NPOs?**
  Design theory is a set of design principles (DPs) and design features (DFs) (Meth et al., 2015). Design theory includes a set of testable propositions and knowledge justificatory grounded in decision support theory (Gregor and Hevner, 2013). Hence, RQ1 will be divided into sub-questions:
  RQ1.1 What design principles should an AI-enabled DSS follow to analyse donor behaviour in NPOs?
  RQ1.2 What design features should an AI-enabled DSS follow to analyse donor behaviours in NPOs?
  The answer to the above two sub-questions can be found by crafting a design theory of the AI-enabled DSS based on theoretical and non-theoretical sources to create and design an artefact AI-enabled DSS for analysing donor behaviours).

- **RQ.2 Corresponding to RG2, what are the main func-**

tionalities of an AI-enabled DSS for analysing donor behaviours in NPOs?

The main features of an AI-enabled DSS may vary depending on the objectives and the requirements. Hence, RQ2 is divided into two sub-questions for better answering:

- **RQ2.1: Upon our understanding of different descriptive data models, what is the best model that reveals the influential factors on donor intentions to donate or volunteer?**

  Descriptive models were generated, such as classification and clusters to understand the donor's data and create insights for choosing the best features to answer the following research questions. The output of the question is used as input in RQ.2.2 and will be a component of the AI-enabled DSS.

  RQ2.2: Who among the previous donors or volunteers is likely to donate or volunteer in the future?

  This question predicts donor intentions to recruit or become frequent donors by understanding the factors that affect donor intentions which include geographical residence, income, age, and gender. These factors were mentioned in the previous literature. Defining these factors helps to predict return donors. Using data analytics techniques, the answers to this question are used in the AI-enabled DSS component (e.g., a predictive model of likely people who will donate/volunteer in future).

1.6 Research Aims and Objectives

This research has two main objectives and associated aims to find the answers to the above research questions (RQs). Aim 1 is to answer RQ.1 successfully, and Aim 2 is to answer RQ.2. Figure 1.2 presents the associations between the proposed research questions and research objectives. Each aim consists of certain research objectives (ROs) as follows:

Aim 1: To create a new design science theory of designing AI-enabled DSS for analysing donor behaviour in NPOs.

- RO 1.1: Determining the DPs of AI-enabled DSS to analyse donor behaviour.
- RO 1.2: Determining the DFs of AI-enabled DSS to analyse donor behaviour.

Aim 2: To develop an artefact (AI-enabled DSS) to analyse donor behaviour in NPOs..

This aim can be achieved by developing descriptive and predictive models. A combination of the developed models would form the contents of the AI-enabled DSS for analysing donor behaviour in NPOs.

- RO 2.1: Building descriptive models using ML techniques to define the influential factors on donors and volunteers towards donating money and time.
- RO 2.2: Building predictive models using ML techniques to predict donor and volunteers' intentions to donate money and time.

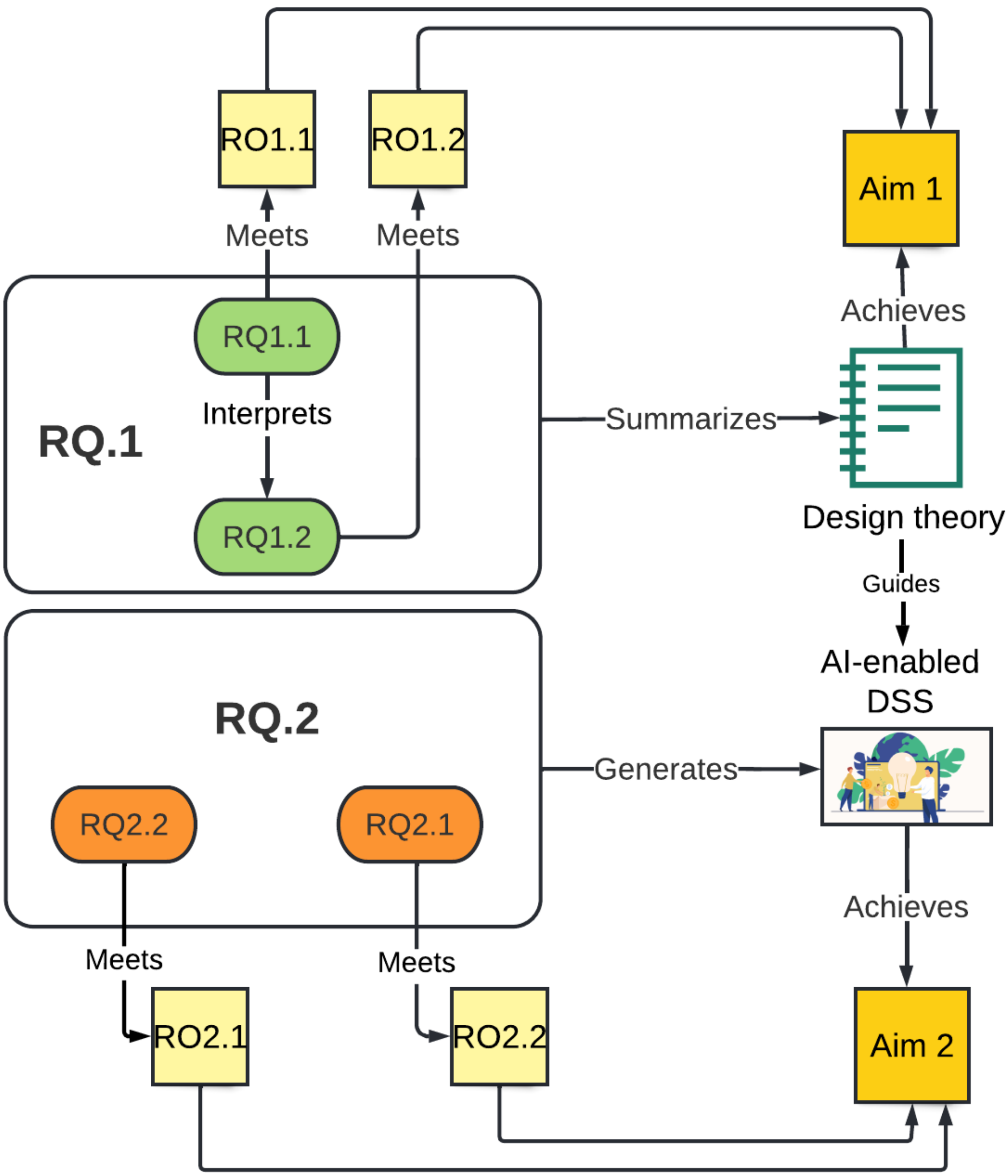


Figure 1.2: Mapping of Research Questions, Research Aims and Objectives

1.7 Research Methodology

To achieve the research aims and objectives and solve the research questions, the researcher applied a design science research (DSR) framework which employed qualitative and quantitative research methods. However, comprehensive reviews, including a systematic review, taxonomy, and a bibliometric analysis survey, have been conducted to validate the research gaps and help achieve the research objectives and aims. Another part of the research methodology applies a design science process model adopted from Peffers et al. (2007). The research methodology is divided into three main parts:

1.7.1 Systematic Literature Review

A systematic literature review is conducted to acquire a deep understanding of the applications of data analytics in NPOs, which is presented in Chapter 2 of this thesis. Also, to identify the associations between prevalent data analysing techniques in NPOs, the researcher incorporated quantitative analysis (using Bibliometrix R-tool), co-term analysis, scientific evolutionary pathways (SEP), and identified the research topic changes over time. The outcomes of bibliometric analysis assisted in knowing the research trends around data analytics applications in NPOs and provided a clear overview of what scholars mainly focus on to void any research gaps. The systematic review helped find the research gaps, form a taxonomy of data analytics applications in NPOs, and will drive future directions of future research. The outcomes of the systematic review also helped us choose a research framework to solve the research problem. Most importantly, the systematic literature review has been accepted by the Asia Pacific Journal of Information Systems (APJIS).

1.7.2 Design Science Research

Design Science Research (DSR) seeks to improve the understanding of the phenomena of IS by developing information technology (IT) artefacts (Lawrence et al., 2010). Design Science Research (DSR) seeks to improve the understanding of the phenomena of IS by developing IT artefacts ((Lawrence et al., 2010). The artefact is meant to solve a research problem and achieve the research goals. There is a demand for better artefacts to enable individuals to work effectively in solving a research problem (Lawrence et al., 2010). This research context concerns defining design principles and features for building an AI-enabled DSS in NPOs.

The researcher found that the design science process model developed by Peffers et al. (2007) suits the research aims and helps to answer the research questions. The framework is illustrated in the next section. We adopted the DSR process model introduced by (Peffers et al., 2007) to solve the proposed questions and achieve the research aims. The DSR process model (Peffers et al., 2007) helps construct a design theory for a DSS and data analytics on NPOs' available data using ML techniques. Figure 1.3 illustrates an AI-enabled DSS design science framework (adapted from Peffers et al. (2007). One of the advantages of the DSR approach is an iterative and incremental process (Hevner et al., 2004), which requires conducting at least three iterations (Rhyn and Blohm, 2017b).

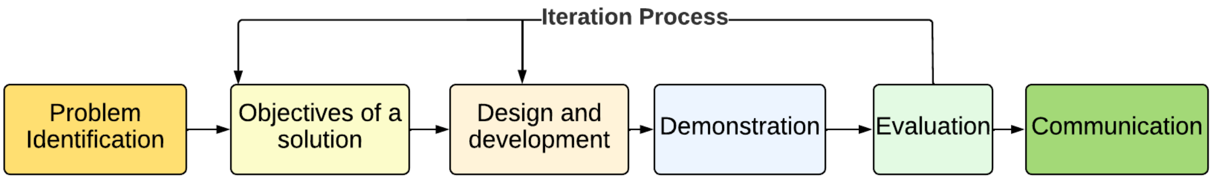


Figure 1.3: Design Science Research Process Model of Peffers et al. (2007)

The main benefit of the DSR methodology is that it allows researchers to create artefacts that can be used to solve real-world problems while providing practitioners with a better understanding of the problem and its solutions. Additionally, this approach allows researchers to develop frameworks and models that can be used as a starting point for further research and development in the field (Fisher, 2016).

1.7.3 Research Methodology Justifications

DRS helped in this research by providing transparency of meta requirements for developing an AI-enabled DSS. Since no existing study exists in the literature to highlight the required guidelines for designing an artefact to analyse donor behaviour, DSR guides the development of an AI-enabled DSS. Furthermore, DSR has three main iterations that ensure the achievement of objectives, the effectiveness of the demonstration, and the validation of the artefact. DSR has a validation phase which involves experts testing and validating the designed prototype.

Finally, DSR has a communication stage which involves publishing the results, design theory, guidelines, and recommendations. By communicating, the theoretical design knowledge for establishing an AI-enabled DSS for donor behaviour analysis would be captured, which is an addition to the research field of NPOs. We strongly believe that the DSR approach is assisting us in building a reliable DSS for analysing donor behaviour in NPOs. Not only this, but it also provides researchers and scholars who share our interests to continue developing our proposed artefact for more effectiveness and sustainability in NPOs.

1.7.4 Ethics Approval

The UTS Human Research Ethics Committees (HRECs) classified this research as low-risk. The researcher obtained the ethics approval by 22 March 2021, UETH21-5802. (See Appendix B). This approval is for data collection and evaluation as a prime part of the main parts of the research methodology.

1.8 Research Contributions

This study is expected to have significant implications for NPOs, as it will enable them to better understand the dominant donors and volunteers in their communities and identify the key predictors of charitable giving and volunteering. This knowledge can be leveraged to tailor their outreach strategies and cultivate relationships with potential donors in ways that are more closely aligned with donor motivations for giving. Furthermore, this study may have broad implications for future research, as it suggests that disparities in the literature may be resolved by grouping donors based on income level or other criteria.

The study's contribution falls into two categories: theoretical and practical. Firstly, we provide a theoretical basis (answers to RQ.1) for developing an AI-enabled DSS for future directions and implications. Secondly, the research makes a practical contribution to the academic literature by implementing ML techniques (answers to RQ.2) and using big datasets of donors and volunteers to apply useful descriptive and predictive analysis. These analyses support mechanisms for decision-making and may serve as added-value proposals for NPO missions or means of improving the efficiency and effectiveness of internal data processing. In particular, the main contributions listed in detail are:

1- **Introducing a new design theory to design a DSS for analysing donor behaviours that contributes to the IS literature.**
A design theory is a perspective set of explanations to show how to do an activity to achieve a goal (Vaishnavi et al., 2019). Hence, the derived design theory can guide developers in defining the necessary system features and reducing development activities. This would increase the probability of success. By doing this contribution, the RG1 will be addressed.

2- **Developing an AI-enabled DSS that analyses donor behaviour towards donating to and volunteering for NPOs**.
A dashboard is created to link and visualise the descriptive and predictive analyses of donors and volunteers. This dashboard is interactive, provides information, and guides the user to explore them effectively. This contribution aims to void RG2. The level of

contributions in DSR is introduced in the framework of (Chadha et al., 2015). Accordingly, our developed AI-enabled DSS is considered a new solution for a known problem (analysing donor behaviour). Most importantly, Vaishnavi et al. (2019) notes that if the research contributes some type of improvement, then the research output is at a high level of generalization, according to the framework of (Gregor and Hevner, 2013). In addition, the created artefact, system, or model lead to the creation of knowledge and novel solution (Vaishnavi et al., 2019). Hence, the researcher believes that the proposed solution, AI-enabled DSS, presents a novel artefact and contributes significantly to the research field according to the framework proposed in Figure 1.4.

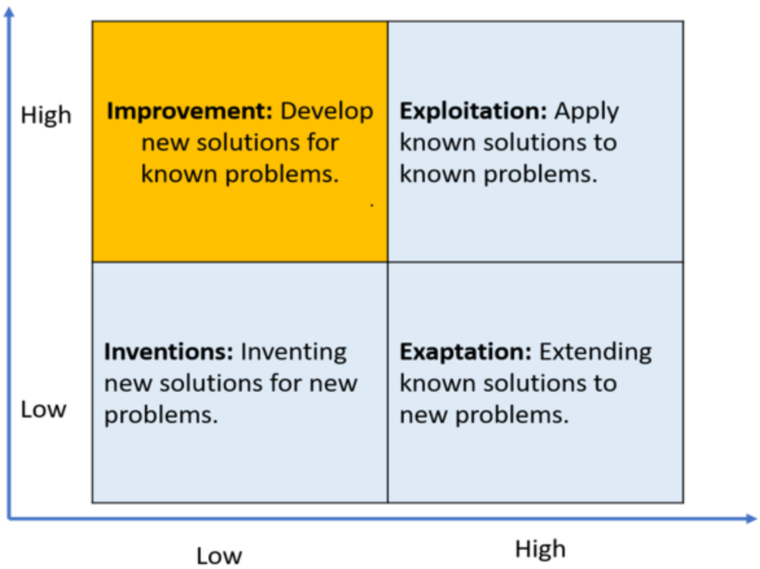


Figure 1.4: Knowledge Maturity by Gregor and Hevner (2013)

3- **Creating descriptive and predictive analyses to under-**

stand and analyse donor behaviour using a variety of ML techniques.

Since no study in the literature tried and applied different ML techniques to analyse donor behaviour (RG2), the researcher proposes other analytical models to predict, describe, and provide insights into donors and volunteers. This contribution is distinct from previous studies because of the ML techniques and involvement in presenting the results in a dashboard of AI-enabled DSS. To the best of our knowledge, this contribution has not been shared in the past. Therefore, researchers and scholars may find this contribution as an initial step for extending future work on analysing donor behaviour in NPOs, using ML techniques.

4- **The research supports grounding and viability via involving NPOs managers, data scientists, and decision-making experts.**

To ensure the grounding, some experts provided useful evaluations and concentrated on the proposed design's effectiveness, functionality, usability, and viability during the research activities. The significance of this research lies in its interdisciplinary contribution that integrates the nexus between ML, IS, and NPOs. In other words, this research is different from the previous studies in two central lines of work: (a) developing an AI-enabled DSS capable of analysing donor behaviour to produce descriptive analysis and accurate predictions and (b) applying DSR process model framework to create a design theory of developing an AI-enabled DSS in NPOs.

1.9 Thesis Structure

Table 1.1 summarises the main work and data sources of the thesis chapters. Figure 1.5 also presents the thesis's structure and the published papers' inclusion. The chapters in this thesis proceed as follows:

* **Chapter One: (Introduction)**
  The first chapter introduces the research context, the research problem, the research questions, aims and objectives, methodology, contribution, and significance, and finally, the research structure of the thesis.

* **Chapter Two: (Systematic Literature Review)**
  The second chapter presents a taxonomy and a systematic review of data analytics applications in NPOs. The contents of the chapter have been published in two journals highlighting the importance of applying data analysis, big data applications, and decision-making solutions in NPOs, with coverage of the literature, the applied research methods, research gaps and future directions. The two published papers in this chapter represent a theoretical and practical investigation of the literature and the research problem.

* **Chapter Three: (Research Methodology)**
  The third chapter covers the research methodology, including the research design, framework, and methods. This chapter

further explains the data collection procedures and the data analysis method used to conduct this study. Finally, this chapter describes the stages involved in designing and building an artefact for analysing donor behaviour in NPOs. Mainly, the chapter presents the contents of one published paper that comprehensively covered the research methodology: a design science process model framework.

* **Chapter Four: (Data Analysis and Artefact Deployment)** This chapter thoroughly examines the three iterations reported in the research methodology chapter. This chapter primarily presents the methods used in each iteration, including preliminary analysis, data processing, and modelling, and demonstrates the process of developing an AI-enabled DSS for analysing donor behaviour. The three iterations are presented following the design science process model for consistency and better development of the artefact. The chapter presents the contents of one paper (under submission stage) and also presents a conceptual design of an AI-enabled DSS for analysing donor behaviour in NPOs and its evaluation results.

* **Chapter Five: (Research Results and Discussion)** This chapter presents the results of our research, explains the framework's output, and highlights the key findings of the problem's solution. This chapter aims to present the results of the evaluations, the outputs, and the effectiveness of the solution. A deep discussion follows to address the findings of the research questions, with consideration of the research questions, the re-

search aims and objectives, and how we achieved each.

* **Chapter Six: (Conclusion)**
  The final chapter provides an overview of the research contributions and discusses the implications of big data in NPOs, future directions, and recommendations. This chapter also highlights the research limitations.

Chapter	**Main Work**	**Data Sources**
Systematic Review	Synthesising the literature on data analytics in NPOs using bibliometric analysis. Also, reviewing the literature and developing a taxonomy of data analytics applications in NPOs	Scopus, Web of Science, and ProQuest Databases
Research Methodology	Presenting the phases of the DSR process model framework to develop an AI-enabled DSS in NPOs	Literature
Data Analysis and Artefact Development	Introducing the processes of developing the AI-enabled DSS through three different iterations to enhance the development process	Public datasets of donors and volunteers
Research Results and Discussion	Presenting the results obtained after conducting each iteration	Public datasets of donors and volunteers
Conclusion	Presenting an overview of the research outcomes, lessons learnt and highlighting limitations and future directions of this research	Literature and thesis outcomes

Table 1.1: Summary of main work and data sources of the thesis chapters

Figure 1.5: Research Process and Thesis Components Structure

1.10 Chapter One Summary

This introductory chapter has presented the overarching theme of the research, which is to build a DSS for analysing donor behaviour in NPOs. The introduction chapter covered the research context and background and discussed the significance of this research. The research aims, the research questions, the research objectives, the research methodology and the research contribution were also discussed.

CHAPTER 2

Systematic Literature Review

Copyright/credit/reuse notice: This chapter has materials which are published as
Different Approaches of Bibliometric Analysis for Data Analytics Applications in Nonprofit Organisations (Alsolbi et al., 2022b)
A Systematic Review and Taxonomy of Data Analytics in Nonprofit Organisations (Alsolbi et al., 2023)

2.1 Introduction

NPOs use data to resolve problems and create new service business opportunities (Hou and Wang, 2017; Johnson, 2015; Mayer, 2019). The use of data is only partially applied in NPOs to obtain such beneficial insights of NPOs' stakeholders. Many NPOs store their data on spreadsheets and process it manually instead of digitising it (Bopp et al., 2017). The relationship between data, analytics, and technology in NPOs remains uncertain (Mayer, 2019). Data analytics enables monitoring and evaluation of specific situations by determining significant issues, influencing policy through evidence, enhancing fundraising capabilities, and studying donor behaviour (Bopp et al., 2017; Maxwell et al., 2016; Mayer, 2019; Rathi and Given, 2017). The gap between the available resources and actual usage of data analytics by NPOs is considerable (Johnson, 2015).

Although many studies have been conducted on data analytics applications in various domains, this systematic review specifically focuses on applications of data technology by NPOs. Our study of the literature in the NPO sector discovered only reviews had been found on investigating the usage of data analytics applications in NPOs. One systematic literature review by Gupta et al. (2019) focuses on big data and the humanitarian supply chain, particularly on information systems and operations management. Their review provides an understanding of the role played by humanitarian supply chain organisations to facilitate better decision-making.

Another review classifies big data utilisation in health, social networking, and governmental and public agencies (Mohamed et al., 2020). This second review relates data-related research to frameworks of big data applications, shows the trends in the research on big data tools, and gives researchers an insight into the most recent research activities. These reviews do not consider NPO-specific issues in sufficient depth. First, the reviewed papers only dealt with big data applications in different humanitarian and public sectors and did not precisely focus on NPOs. Second, the methodology of the previous reviews did not include any textual analysis of the papers that enables researchers to identify big data research trends in NPOs. Third, the two studies rarely presented a separate section that discussed the implications of data analytics in NPOs, or considered the challenges, the advantages, and any proposed and alternative solutions for NPOs. Despite their limitations, these two literature review studies (Gupta et al., 2019; Mohamed et al., 2020) provided the starting point of our study.

The systematic literature review differs from earlier studies as it aims to (1) use a developed taxonomy to investigate and synthesise the literature on data analytics as adopted and applied by NPOs, (2) provide insights into publication trends in the field of NPOs through bibliometric analysis of the selected documents, and (3) encourage researchers to study and NPOs' management to adopt and apply data technologies. As a result, this study provides four specific contributions to the field by (1) developing a new taxonomy of data analytics research in NPOs and identifying issues that need addressing, (2) focusing on the challenges, requirements, and technical tools associated with applying data analytics in NPOs based on the review of

25 publications found in three databases: Scopus, Web of Science, and ProQuest, (3) demonstrating the usefulness of data analytics applications by NPOs by means of case, and (4) presenting a future research agenda by identifying opportunities and directions based on the implications of the findings and their applicability to other domains, such as decision-making systems. Based on our literature study and understanding of it, this investigation into the applications of data analytics in NPOs is unique as it combines a systematic review with a taxonomy and bibliometric analysis of the literature.

This chapter is organised as follows: Section 2.2 describes the research methodology, the adopted research strategy, and the process used to select the documents. Following that, a bibliometric analysis tool and keywords analysis are used to evaluate the chosen articles in terms of most published and cited. Then, a developed taxonomy of data analytics publications about NPOs is presented to guide the answers to the research questions. Section 2.3 presents the results and discussions of each selected article, including discussions of the research methods applied, frameworks, main ideas, significance, and limitations. Sections 2.4, and 2.5 present research gaps and future directions to address these gaps, and the summary of the chapter.

2.2 Systematic Review Methodology

2.2.1 Definition of Data and Big Data Analysis

Data analytics is defined as the usage of computer systems to analyse different scales of data sets to support decision-making (Thomas, 2020). As a field of study, data analytics has adopted aspects from different disciplines including machine learning, artificial intelligence (AI), systems theory, operations research, and pattern recognition (Thomas, 2020). Hidden information, patterns, and structures for gaining new insights can be discovered through data analytics (Thomas, 2020).

Data analytics projects have several stages: assessing and selecting data, cleaning and filtering data, visualising and analysing data, and finally, interpreting and evaluating the results (Thomas, 2020). Value from data analytics applications is enhanced by using methods and models from the fields of statistics, machine learning, and AI (Thomas, 2020). On the other hand, big data consists of massive amounts of data in different forms and from various sources (Kassen, 2018; Klievink et al., 2017) and is difficult to handle by standard data processing technologies (Klievink et al., 2017). The features of big data prompted scholars to consider new methods of analysing it, and to conduct further research on NPOs (Litofcenko et al., 2020). Therefore, big data analytics involves the use of specialized tools and techniques to analyse large datasets in order to extract insights, while data analytics is a broader field which encompasses the use of algorithms, statistical analysis, and machine learning to analyse data (Ahmed et al., 2019). Consequently, big data analytics and data analytics differ in the type of data sets used and the analytical approaches employed.

2.2.2 Types of Data Analytics

There are three different types of analytics and associated techniques (Anitha and Patil, 2018), as summarised in Table 2.1.

1. Descriptive analytics mines huge amounts of data to discover hidden patterns stored in repositories. Descriptive analytics can be used, for example, to find similarities in customer behaviour. Techniques to generate descriptive analysis include regression of correlated variables and visualising the data by presenting it as charts.

2. Predictive analytics improves decision-making by predicting future trends. Predictive analysis is conducted by combining massive amounts of data from various resources. Techniques such as regressions, time series, decision trees, and random forest, among others, can be applied to generate a predictive model of historical data.

3. Prescriptive analytics helps professionals assess the impact of different decisions in the decision-making process. Prescriptive analytics uses optimisation, numerical modelling, and simulation techniques. Anitha and Patil (2018) emphasise that in prescriptive analysis, data is collected continuously to re-predict actions intended to increase the prediction accuracy. All three types of analytics help organisations establish informed decision-making systems (Pyne et al., 2016). According to the level of analytics, there are certain techniques for conducting descriptive, predictive, and prescriptive analysis.

Type of Analytics	Techniques
Descriptive	Regression analysis, data modelling, visualization, online analytical processing (OLAP)
Predictive	Time series, naïve Bayes, Bayes Networks, discriminant analysis, decision trees, random forest, CART, clusters
Prescriptive	Optimization, simulation, decision trees, fuzzy rule-based, neural networks

Table 2.1: Three Types of Data Analytics

2.2.3 Methodology: The Systematic Review Process

This section discusses the process and the methodology followed for the systematic literature review. To conduct a systematic review of the literature, researchers must establish a system that allows for efficient detection and searching within articles relevant to the review (Watson, 2015). Hence, we adopted the process for systematic quantitative literature reviews developed by Pickering and Byrne (2014). Therefore, the process was modified to include bibliometric analysis of the collected articles. The bibliometric analysis analyses science as a productive knowledge system by providing a thorough analysis of research trends (Trivedi, 2019). Bibliometric analysis helped discover the most common themes and keywords associated with applications of data analytics in NPOs.

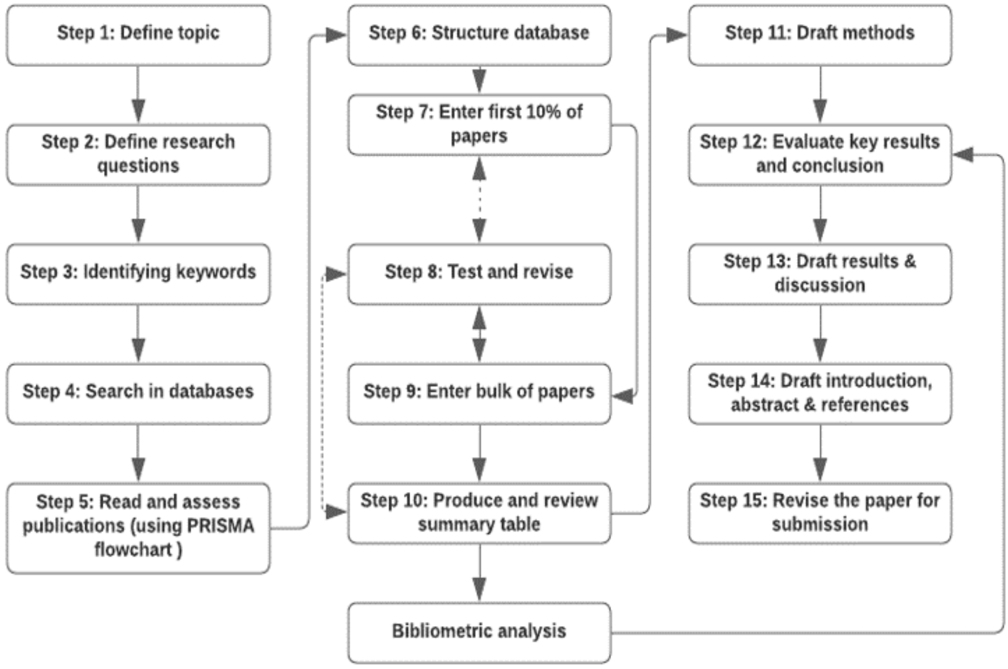


Figure 2.1: The Modified Systematic Quantitative Approach (adapted from Pickering and Byrne (2014))

2.2.3.1 Step 1: Define the Topic

The review process began by stating the topic, aims, and contributions, as described in the previous section. Next, research questions were formulated, keywords for the search were determined, and the databases to be searched for relevant articles were selected (steps 2, 3, and 4, respectively).

2.2.3.2 Step 2: Define Research Questions

Three research questions were formulated to guide the research and meet the objectives of this study:

- RQ.1: What are the proposed frameworks for adopting and applying data analytics in NPOs?
- RQ.2: What type of data analytics is utilised for NPO activities and missions?
- RQ.3: What challenges and barriers do NPOs face when adopting and implementing data analytics for their business operations and missions?

2.2.3.3 Step 3: Identifying the Keywords

This study deals with two independent domains: CS (data analytics) and business (NPOs). Table 2.2 lists the keywords used to search for documents. Each group in this table represents a relevant keyword. Group A consists of keywords relevant to data concepts, and Group B keywords pertain to the nonprofit domain. In each group, the researchers include the wildcard (*) to provide more results, especially from scientific databases. Keywords from Groups A and B are joined using the operator “AND” in the search field, and the operator “OR” is used within the same group for a similar keyword; for example, (Big Data) AND (Nonprofit OR Nonprofitable). The keywords were set to search in the fields of Titles, Abstracts, and Keywords. The completed searching string is shown in Appendix A.

2.2.3.4 Step 4: Database Search

The search was conducted based on all possible matches between the two categories of keywords during a ten-year period from 2010 to 2021 in the Scopus, ProQuest, and Web of Science databases. Research

Group A	Group B
“Big Data”	“Nonprofi*” Nonprofit/Nonprofitable
“Data-driven”	“Non-Profi*” (Non-Profit/Non-Profit/Non-profitable Non-Profitable)
“Data analy*” (Analytics or Analysis or analysing)	“Not*Profi*” (Not-profitable/Not Profitable)
“Predictive analy*” (Analytics or Analysis)	“Not-for-profit”
	“Charit*” (Charity/Charities/Charitable)
	“Fundraising”
	“Donations”
	“Donors”

Table 2.2: Keywords Table for Articles Search

and studies conducted on data analytics applications in NPOs during this ten-year period were sought. According to Khoo-Lattimore et al. (2019), these libraries are comprehensive, used mostly by researchers, and contain more peer-reviewed papers. In addition, these databases are considered the most comprehensive and powerful available (Yang et al., 2017). Other databases, such as Google Scholar and arXiv, were excluded from our search because they provided pointers to publications already contained in our searched databases or they included non-peer-reviewed papers. We concentrated on peer-reviewed articles to obtain evidence of the research's quality. Additionally, the literature review aimed to discover the most recent scholarly works on data analytics applications in NPOs. Recently, NPOs adopted and have become more data-driven in fulfilling their missions (Hou and Wang, 2017).

2.2.3.5 Step 5: Reading and Assessment

The literature search was conducted on all three databases to ensure a comprehensive review. Also, to ensure clarity, choices were based on several criteria: the articles had to be written in English, peer-reviewed, and published in journals. These selection criteria were reported in two studies (Mishra et al., 2018; Gupta et al., 2019). The PRISMA flowchart (Page et al., 2021) was used in the systematic review to ensure better reporting and gather the relevant documents during the search process. The PRISMA flowchart shown in Figure 2.2 consists of four main stages described below:

Identification Stage

The identification stage yielded 7,215 documents after a search combined the keywords in Group A and Group B. The selection criteria (only English, CS subject area, peer-reviewed, and academic journal articles published from 2010 to 2021, and the domain of CS) were not applied at this stage. The 7,215 documents were comprised of 3,327 articles from Scopus, 3,137 articles from ProQuest, and 751 articles from Web of Science.

Screening Stage

In this stage, 66 duplicate documents were removed, leaving 7,149 articles screened against the selection criteria. After choosing only English-language, peer-reviewed journal articles published from 2010 to June 8, 2021, in the domain of CS, 49 publications were eligible for further consideration. Books, book chapters, reviews, conference papers, and surveys were excluded.

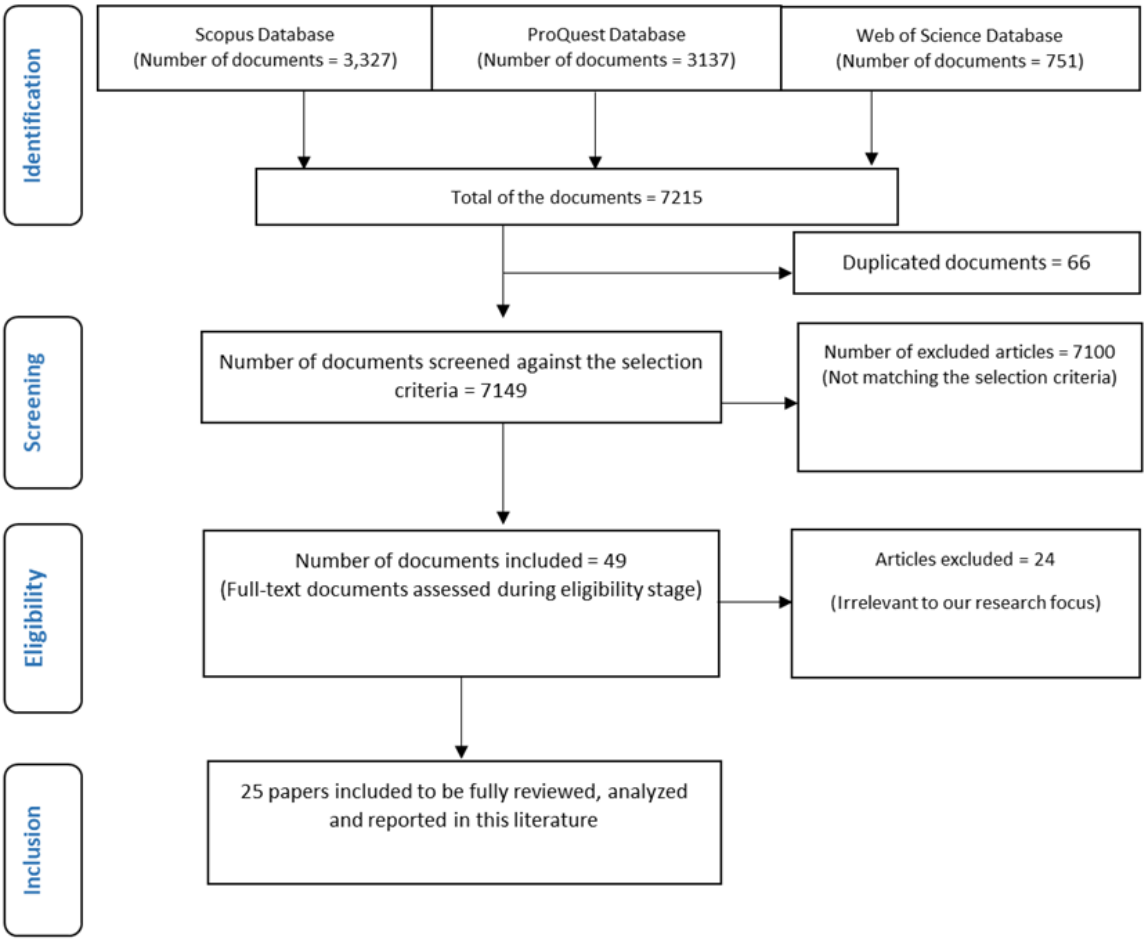


Figure 2.2: PRISMA Flowchart

Eligibility Stage

The full texts of 49 peer-reviewed articles obtained from the three databases were downloaded, and their abstracts, keywords, and conclusions were analysed to determine the relevance of the contents. The most challenging part of this process was determining whether the research focus of each paper was relevant to the authors' research questions. Irrelevant articles that did not help answer the research questions were excluded. For example, in *New Scientist*, Rutkin (2015) provides only an overview of the blood donation prediction project, but the paper does not describe the research methods or means, give the data sources, or specify the project's target audience. Hence, it was excluded; it offered no relevant material that could contribute to our research. In total, 25 publications were chosen for closer examination.

Inclusion Stage

In this inclusion stage, the 25 chosen papers were assessed for their relevance to the topic and to this study's research questions. The full text of each article was read, the references were checked, and all references were extracted into Research Information Systems (RIS) files to be stored in Endnote for data management.

2.2.3.6 Steps 6 to 10

Steps 6 to 10 involved structuring the database, entering the data, and producing a summary table. A database is established (Step 6) to enter essential information for each of the 25 articles. Steps 7, 8, and 9 are entering the papers sequentially, testing and revising them, and finalising all required fields in the database. In Step 10, a summary of the 25 articles with the essential information is produced. The summary included essential information about each selected article, such as scope, main work, methods applied, and any limitations found.

2.2.3.7 Steps 11 to 15

Steps 11 to 15 concerned drafting the systematic review's main parts: methods, results, introduction, and abstraction. Most importantly, as the systematic review approach was modified, the results from the bibliometric analysis were evaluated in Step 12. Steps 13 to 15 included more writing activities and drafting the systematic review, to prepare it for submission to a journal.

2.2.4 Bibliometric Analysis

We conducted a data pre-processing procedure to obtain the inputs for co-occurrence and Scientific Evolutionary Pathways (SEP) analyses. Specifically, we used the natural language function integrated into VantagePoint software to extract terms using a process of inclusion and exclusion (Zhang et al., 2014) was applied to the extracted terms to accomplish term cleaning and consolidation. This subsection incorporates quantitative analysis (using Bibliometrix R-tool), co-term analysis and scientific evolutionary pathways (SEP) to identify the associations between prevalent data analysing techniques and different types of NPOs, and identified the research topic changes over time. By applying those analyses: this paper can profile the research landscape of this field and generates the following insights:

2.2.4.1 Bibliometrix R-tool

The bibliometric analysis ensured that all publications were comprehensively examined and aimed to discover more insights (for example, keyword frequency) about the research of data analytics applications in NPOs. Bibliometric analysis is a complex process (Esfahani et al., 2019), thereby motivating leading data scientists (Aria and Cuccurullo, 2017) to develop a bibliometric package in R language. Their Bibliometrix R-tool supports only the Scopus and Web of Science databases used for our search.

Trivedi (2019) conducted a study on data analytics using Bibliometrix R tool. The results are to provide policymakers, researchers, and people in agriculture with a comprehensive understanding of agricultural big data research worthy of future study. Esfahani et al. (2019) also conducted a bibliometric analysis to look into the current state and progress of scientific studies on the impact of social networks on big data and the use of big data to model the behaviours of social network users. In another bibliometric study (Shukla et al., 2020), total publications, total citations, and citation per paper are among the performance measures calculated.

Although the previous studies (Trivedi, 2019; Shukla et al., 2020; Esfahani et al., 2019) have conducted a number of metrics and indicators that measure the trends of research in their specific area, which are not related to NPOs, their work provided a broad analysis over a long duration, but it is difficult to identify development trends, evolutionary paths of keywords and emerging topics over different time periods in data analysis applications in NPOs.

The Bibliometrix R-package was chosen because it provided several types of metrics for keywords analysis, authors' scientific publications, collaborations of countries and authors, and a conceptual structure (Aria and Cuccurullo, 2017). The Bibliometrix R-package was installed for a variety of analyses.

2.2.4.2 Conceptual Structure

The Bibliometrix R-tool generates a conceptual structure of a framework involving a co-occurrence of words (Zarei and Jabbarzadeh, 2019). The words can be extracted from titles, abstracts, and authors' keywords. In this analysis, two types of conceptual structures were chosen: the thematic map (Figure 2.3), and co-occurrence of keywords (Figure 2.4) used by most authors in their publications.

Thematic Map shows the co-word analysis by clustering keywords which are relevant to the main themes of this study. According to the analysis method reported in a bibliometric study (Aria and Cuccurullo, 2017), a thematic map has a vertical axis representing certainty (e.g., the strength between different themes) and a horizontal axis representing density (the strength of the links within the clusters). Figure 2.3 depicts: Quadrant 1: The upper right shows well-discussed and developed themes. For example: “business intelligence” and “challenges.” Quadrant 2: The lower right shows basic themes discussed in the papers. It indicates that the “big data” theme has not been discussed adequately. Quadrant 3: The upper left shows well-developed themes that do not interfere with other themes, such as “privacy” and “crowdfunding.” Quadrant 4: The lower left shows poorly developed themes; studies of these are emerging, such as “data analysis,” “nonprofit organisations,” and “business analytics.”

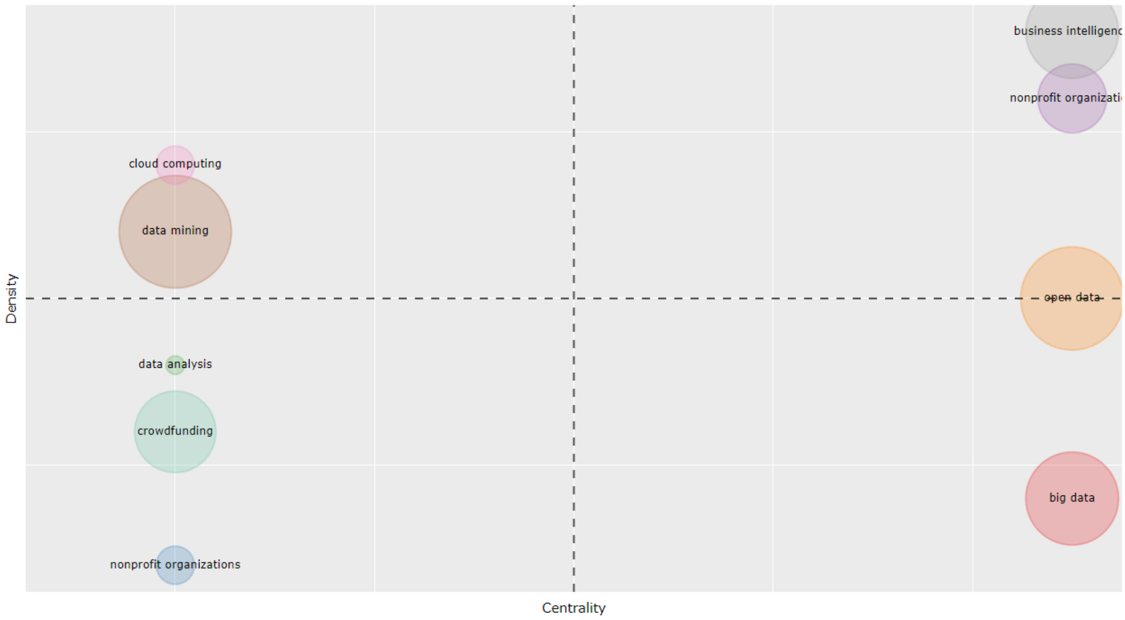


Figure 2.3: Thematic Map of Keyword Clusters

Co-occurrence Network depicts the relationship between keywords of scientific and technical topics (Aria and Cuccurullo, 2017). In Figure 2.4, three clusters are shown in three different colours. For clarity of visualisation, 50 words were used for clustering. Keywords such as “data,” “analytics,” “big,” “non-profit,” and “organisations” were used in the same articles. Another cluster contains words frequently appearing in the same paper, such as “data-driven,” “charitable,” “sector,” and “behaviour.” The third cluster consisted of keywords such as “study,” “case,” “text,” and “donations.”

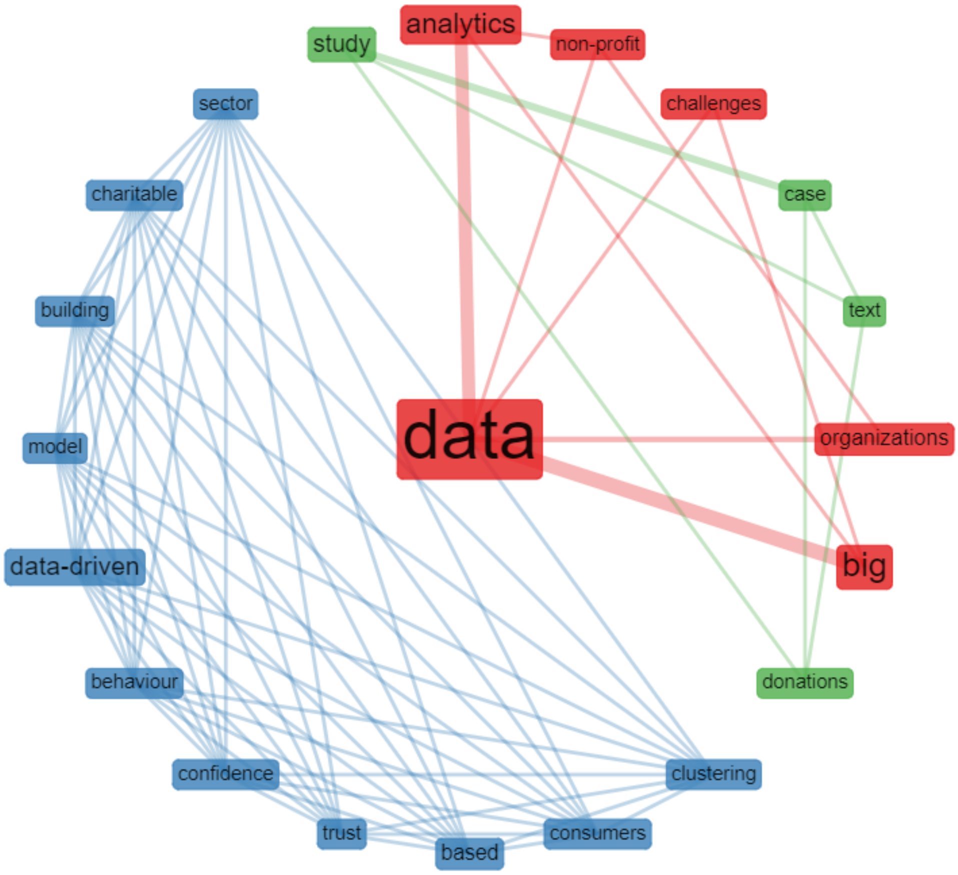


Figure 2.4: Co-occurrence Network for Keywords

2.2.4.3 Co-term Analysis and SEP

Co-term analysis is a classical topic analysis method in bibliometrics (Wang et al., 2010; Callon et al., 1983). It adopts the assumption that two terms that appear in the same context and may share similar semantic meanings. The collection of such term co-occurrences can constitute a co-term network represented as $G = (V, E)$, where V denotes the set of terms and E denotes the co-occurrences between them.

Interestingly, Figure 2.5 yields a bird's eye view of (1) the frequently mentioned data analysis methods in the field and the key types of NPOs, represented by the arc length of terms, and (2) how strong data analysis methods are related to different types of NPOs, represented by the width of ribbons connecting two terms. By observing the length of those terms, we could identify that organisation management-related and statistics-based data analysing methods, such as document analysis, knowledge management, data analysis, decision-making, and logistic regression, are frequently seen in relevant NPOs publications.

Intriguingly, data analysis methods in the CS domain also appear on this map, including DSS, predictive model and machine learning, even though they seem less prevalent than the top ones. The frequency ranking of those methods indicates that current NPOs still prefer to apply those robust, classical and organisation management-related methods to real-world implementations. Nevertheless, the utilisation of novel CS-based methods is still an emerging trend. Next, we focus on the linking strength of different NPOs data analysis method pairs. From the NPOs perspective of healthcare-related NPOs such as nonprofit hospitals and public health, the dominating techniques are mostly derived from statistical analysis like logistic regression and multivariate logistic regression, representing the typical data analysis methods in the medical domain.

However, in other public affair-related NPOs such as the public sector or public administration, knowledge management seems to take a more crucial role in the associated studies. From the technique perspective, we could see document analysis is notably involved with civil society; in such cases, document analysis is widely used to identify the impacts and roles of civil society using multiple case studies(Anaf et al., 2017; Powell and Wittman, 2017). Some AI-related techniques like a predictive model and machine learning tend to have stronger associations with the healthcare domain including nonprofit hospitals and public health, indicating the emergence of those techniques in the healthcare sector.

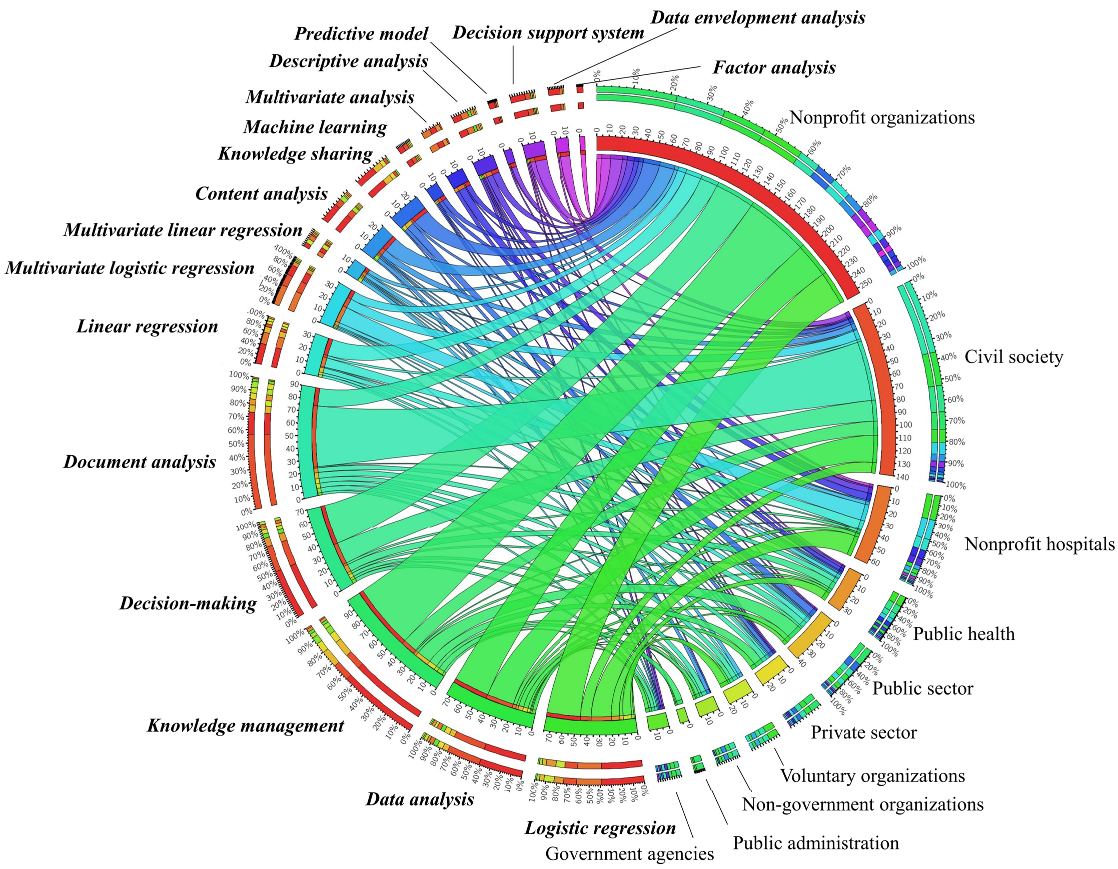


Figure 2.5: Co-term Analysis

SEP is a research topic tracking method developed by (Zhang et al., 2017). It is used to identify the changes in research attention in time-labelled streaming documents. The basic assumption of this method is that scientific novelty derives from the accumulative changes and recombination of existing knowledge (Fleming, 2001; Fleming and Sorenson, 2004). We employed this method to identify the topic changes at a macro level. The definitions and stepwise explanations of the SEP are stated below.

Algorithm design: Initially, the SEP represents every document with scientific terms and aligns all documents with a term-document matrix; in such case, every document can be represented as a term vector with the entire vocabulary as the feature space. By separating the documents into consecutive time slices, SEP measures the drifts of documents and assigns documents to topics generated in different years to indicate the topics and their changes. We generated a topic evolutionary with 87 nodes and 86 edges by applying SEP to the extracted terms. Every node represents a research term and each directed edge represents the predecessor-descendant relationship between the connected nodes, time labels in the brackets indicate when the topic was proposed. With the aid of Gephi (Mathieu et al., 2009), a community detection algorithm is used in the SEP network and partitioned into five topic communities with different colours, as shown in Figure 2.6.

Concept definition: A topic is defined as a collection of documents that share semantically similar research content, denoted as T. The centroid of a topic is represented by the mean vector of all corresponding document vectors in the topic, denoted as c. The radius of a topic is defined as the largest Euclidean distance of all the documents in the topic to the centroid, defined as r, using standard 5 steps processes suggested in the study of Zhang et al. (2017).

The outcome of SEP is a visualising map with nodes representing the topics and directed edges linking those topics representing the descendent-predecessor relationships between connected topics. The topic communities could respectively be concluded as *#1 data processing (pink), #2 public health and health management (green), #3 hospital management (orange) and #4 public affair and knowledge management (blue)*. The four communities represent the evolving patterns and divergence of research attention in this field over time. Tracing back to 1973, the initial topic data processing *[1973-1990]* indicates the emergence of data utilisation needs and data analysis methods in NPOs at the early stage.

This topic derives a pink topic community *#1 data processing*, in which we could observe topics focusing on classical and traditional statistical methods *(linear regression [2011], logistic regression [1991-2000])* and classical organisational roles *(program managers [2014], policy makers [2017])*. Even the topics at the end of those branches, which represent the recent changes, are still related to basically organisational or data analysis concepts such as *decision-making [2004-2006], decision-makers [2016],* and *quantitative data analysis [2020]*. Hence, we summarise that this community profiles a fundamental and traditional pathway of how the data analysis methods are applied in NPOs.

Derived from community #*1*, community #*2* *public health* and *health management* have a clear emphasis on the healthcare sector. Some of those topics reveal the healthcare data sources like *empirical data [2015]* and *medical records [2018]*, while some others reflect realistic problems that use data analytics to solve in the medical domain such as stem cell *stem cell [2011]*, *diabetes patients [2015]*, *life satisfaction [2014]*, *labour costs [2016]*, *guideline recommendations [2019]*, and so on. The remaining topics mostly refer to specific data analysis methods. Interestingly, except for the data analysis methods identified in community #1, a few AI-related topics like predictive model *[2020]* and artificial neural networks *[2020]* emerge from which we could have a glimpse of artificial intelligence implementation for data analysis in the healthcare domain. However, the limited number of such topics also indicates this trend is still in its infancy.
Community #*3* typically inherits the healthcare attribute of community #*2* and focuses on a more specific domain of hospital management. Topics in this community cover hospital attributes *[public hospital [2013]*, *nonprofit hospitals [2016]*, *nonprofit status [2019]* and *general hospital [2020])*, evaluating metrics *(hospital efficiency [2018]*, *clinical outcomes [2018]* and *pressure ulcers [2020]* (Pressure ulcer incidence rate is an important evaluation metric for clinical nursing quality)], policy-related issues (*private insurance [2017]* and *Affordable Care Act [2017]*), and data analysis methods used on hospital management issues. From the technical perspective, data analysing techniques applied in this topic community are still highly coupled with the fundamental ones in community #1.

Lastly, community #2 additionally derives another community #4 public affair and knowledge management. Different from community #2, community #4 expands the scope of topics to a wider public affair sector. Politics and charity-related topics (*charitable organisations [2015], constituency building [2016]* and *donors [2020]*) are frequently seen in this community, accoupling with a branch emphasising knowledge management concerns in organisations (*knowledge sharing [2016], knowledge donation [2019]* and *knowledge management [2020]*). The appearance of another AI-related topic, machine learning *[2018],* indicates AI's application has also started to emerge in those public sectors.

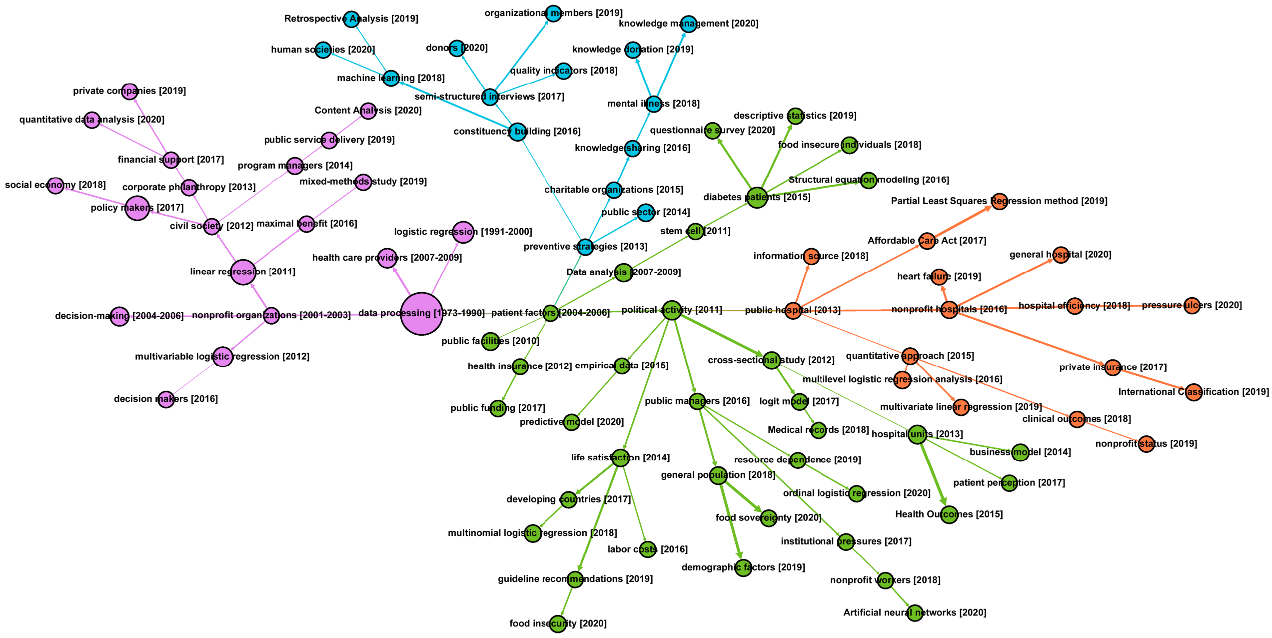


Figure 2.6: Scientific Evolutionary Pathways (SEP)

Overall, "big data" seems a popular term in various research fields. However, the term "data analytics" is not well-involved in NPO-related research. This indication inspires scholars interested in researching big data applications in the field of NPOs. For example, further research is needed to understand how big data can be used to align the various stakeholders of humanitarian organisations (Gupta et al., 2019).

Although the research was conducted on three databases, the obtained results are subject to be scalable by involving more databases and visualization tools. For example, there are other effective visualisations tools such as Gephi, Pajek, VOSviewer, and Bibexcel (Donthu et al., 2021). Therefore, useful results may be generated to provide answers to specific answers to research questions. Moreover, other bibliometric analysis techniques could be applied, such as citation analysis. When two publications appear in the reference list of another article, they are linked in a co-citation network. Scholars can benefit from co-citation analysis by learning about theme clusters in addition to locating the most important publications (Donthu et al., 2021). Therefore, co-citation analysis is appropriate for business academics looking to identify key works and theoretical basis (Donthu et al., 2021).

Thus, the bibliometric analysis applied in this research provided a comprehensive roadmap toward practical implications of data analytics, its usage, and its practical application. Broadly, the study's results may contribute to comparing the trends of applying data analytics in different categories of NPOs such as charities, healthcare, pharmaceutical, health centers, educational institutions, and social agencies. The comparison may urge scholars to seek the possible directions of obtaining the advantages of applying data analytics and enhancing decision-making in NPOs.

The bibliometric analysis ultimately contributed to an overview of the chosen topic of study by laying out a research area's structure, utilising multiple indicators (documents, journals, authors, and keywords) and comprehending the underlying relationships between them. These statistics provide a comprehensive overview of the field's research and might be helpful, particularly to early researchers in the field for conducting further research in NPOs.

2.2.5 Taxonomy of Found Studies

We identified critical topics in this study based on our examination of the primary data analytics applications for NPOs. Our three research questions' goals were to provide overviews of (1) proposed frameworks for applying data analytics in NPOs, (2) applied data analytics in NPOs, and (3) challenges that prevent NPOs from applying data analytics. To answer each research question comprehensively, we extracted information by identifying the focus of each study, understanding the research methodology, and highlighting the key findings. We then categorised the topics into the study's three research questions. Table 2.3 presents the first part of the taxonomy of the 25 studies. For each research question, we identified relevant answers from the literature as follows:

1. RQ1: Frameworks, approaches, and guidelines for applying and adopting analytics in NPOs.
2. RQ2: Methods, techniques, technology examples, and cases studies of data analytics in NPOs.
3. RQ3: Solutions and alternatives to tackle issues of applying data analytics in NPOs.

Table 2.3 provides guidance in answering the three research questions with a comprehensive discussion in Section 3 of this review.

RQ #	**Authors**	**Theme of study**
RQ1	Henriksen-Bulmer et al. (2019)	Propsoing frameworks to tackle issues in NPOs such as collecting, accessing information, and privacy.
	Montalvo-Garcia et al. (2020)	
	Chianese et al. (2017)	
RQ2	Fredriksson (2018)	Clarifying the importance of data analytics applications and decision-making in NPOs.
	Litofcenko et al. (2020)	Conducting experiments of data analytics applications in NPOs
	Wang et al. (2019)	
	Bambha et al. (2020)	Predicting donors probabilities using a special datasets (A case study of a special datasets).
	De Vries et al. (2015)	Segmenting donors using quantitative datasets, through applying different techniques (different case studies applied).
	Dag et al. (2017)	
	Eigner et al. (2017)	
	Kassen (2018)	Presenting examples of data analytics applications in NPOs, with highlights on the importance of decision-making in NPOs, considering policies and governance.
	Kline and Dolamore (2020)	
	Maxwell et al. (2016)	
	Wang et al. (2010)	
	Patel et al. (2017)	
	Ryzhov et al. (2016)	An application to understand how mailing design analysis can help in cultivating disaster donors.
RQ3	Rathi and Given (2017)	Suggesting solutions for tackling challenges such as privacy and access of information of data analytics applications in NPOs.
	Shapiro and Oystrick (2018)	
	Lněnička and Komárková (2015)	
	Prakash and Singaravel (2015)	
	Hou and Wang (2017)	
	Johnson (2015)	
	Witjas-Paalberends et al. (2018)	

Table 2.3: Taxonomy of Topics Researched in Data Analytics Applications in NPOs

Table 2.4 is the second part of the taxonomy, which summarises all 25 relevant articles, clarifying the scope, main work, and research methods used in each study. According to Table 2.4, there are few studies that demonstrate the capabilities of data analytics on the decision-making process in NPOs through real examples and case studies (Bopp et al., 2017; Hou and Wang, 2017; Kassen, 2018; Kline and Dolamore, 2020; Maxwell et al., 2016; Mayer, 2019; Shah et al., 2017; Bambha et al., 2020; Fredriksson, 2018; Patel et al., 2017). Notably, other studies tackled the challenges of applying data analytics in NPOs (Witjas-Paalberends et al., 2018; Prakash and Singaravel, 2015). Table 2.4 aims to summarise all 25 relevant articles [1], including the scope, authors, main work, the research and technology methods used, and the analytics techniques applied. This summary helped us discuss the findings in the next section, and identify the research gaps and future directions of research in Section 2.4.

[1]The relevancy of the collected articles from the literature is shown in Appendix A

Scope	Authors	Main Work
Big Data	Witjas-Paalberends et al. (2018)	Explaining the challenges and best practises in managing big data-driven healthcare innovations by public-private partnerships in the Netherlands.
	Fredriksson (2018)	Showing examples of how big data is used in practice in the public sector
	Patel et al. (2017)	Concentrating on how big data and analysis may support good e-governance
	Mayer (2019)	Exploration of using big data in NPOs (examples and cases studies)
	Chianese et al. (2017)	Proposing an integrated approach integrating BI, big data, and the Internet of Things (IoT) and developing an Associative in-memory technology information system
Business analytics	Ryzhov et al. (2016)	Identifying designs that have a huge effect on the results of a fundraising campaign through the study of large-scale datasets

Data Analytics	Wang et al. (2019)	Collecting and analysing call centre data of an NPO for optimization
	De Vries et al. (2015)	Giving the nonprofit sector an easy-to-understand segmentation method based on a novel unsupervised learning algorithm.
	Montalvo-Garcia et al. (2020)	Proposing a data analytics methodology for small and medium-sized NPOs
	Johnson (2015)	Review of data analytics and information technology research and practise, with a focus on community-based organisations.
	Hou and Wang (2017)	Understanding how civic data hackathons can produce useful data analytics for nonprofits' data-driven efforts.
	Litofcenko et al. (2020)	Classifying NPOs
	Dag et al. (2017)	Predicting the 1-, 5-, and 9-year patient's graft survival following a heart transplant surgery.
	Shapiro and Oystrick (2018)	Proposing a framework model containing three elements: accessibility, reliability, and adaptability.

	Wang et al. (2010)	A summary of the current and existing data mining technologies with examples in NPOs.
Data Privacy	Prakash and Singaravel (2015)	Proposing a personalised anonymization approach for protecting sensitive data
	Henriksen-Bulmer et al. (2019)	Presenting Data Protection Impact Assessment (DPIA) framework devised as part of the case study.
Decision-Making	Maxwell et al. (2016)	Investigates how data-driven decision-making (DDDM) activities and culture are perceived in organisations.
	Eigner et al. (2017)	Determining the concerns, behaviours, and requirements for patient release in relation to the risk of readmission and the information provided.
	Bopp et al. (2017)	Presenting the importance of data in mission-driven businesses' monitoring and assessment procedures.
	Bambha et al. (2020)	Generating probabilities using US population projections.

	Kline and Dolamore (2020)	Analysing a case study using the cultural dimension of the organisation.
Information and knowledge management	Rathi and Given (2017)	Investigating the use of tools and technologies for knowledge management (KM).
Open Data	Kassen (2018)	Studying open data to meet various interests of stakeholders in different sectors.
	Lněnička and Komárková (2015)	Processing data on Apache Hadoop distributed system.

Table 2.4: Summary of All 25 Articles

2.3 The Systematic Review Findings

The literature on data analytics in NPOs evaluated in our systematic literature review enabled us to identify and analyse the practice's investigated aspects, classify existing frameworks to support it, and comprehending how to address the problems of using data analytics. As a result, we answered each research question together with the relevant answer from the 25 connected papers. To provide a deep understanding of the current usage of data analytics in NPOs, each research question was followed by a discussion and content analysis of the literature. Content analysis is a systematic method that allows users to make correct inferences from a vocal, graphical, or structured text to define and measure certain occurrences systematically and objectively (Downe-Wamboldt, 1992). The following subsections, formed from the research questions, discuss the findings of these articles which can provide a deep understanding of the current usage of data analytics in NPOs.

2.3.1 Q.1: What are the proposed frameworks for adopting and applying data analytics in NPOs?

The literature presents three studies that propose frameworks for adopting, adapting, and implementing data analytics in NPOs. A three-step framework describes the usage of spreadsheets in developing and implementing an accessible, adaptable, reliable, and sustainable system for evaluation practices (Shapiro and Oystrick, 2018). Some NPOs improved their evaluation practices after becoming aware of data analytics (Shapiro and Oystrick, 2018). Spreadsheets can be an easy way to collect demographic information about service users, handle caseload specifications, monitor attendance, and coordinate customer satisfaction data. Moreover, spreadsheets are an easily available, adaptable, and accurate method of data entry and analysis. However, NPOs can use other tools, such as Google Sheets and Open Office (Shapiro and Oystrick, 2018).

The three-step framework of Shapiro and Oystrick (2018) is sustainable for evaluation purposes when all three elements are properly integrated into a data analysis system. The three-step process advocated by this study comprises:

- Accessibility: Ensuring the system captures and produces adequate information to meet NPOs' requirements.
- Adaptability: Ensuring the NPOs can integrate data and analysis of their data.
- Reliability: Data can be obtained and processed consistently over time.

A second framework, which uses a data collection technique, is presented for nonprofit small- and medium-sized enterprises (SMSEs) (Montalvo-Garcia et al., 2020). The architecture of this methodology was built on the cross-industry standard process for data mining (CRISP-DM) as a reference system, as defined by the Software Process Engineering Metamodeling (SPEM). CRISP-DM is distinguished by its simplicity, flexibility, and low implementation costs. The proposed methodology aims to reduce the effort required for implementing data analytics, reducing the costs, and minimising the complexity

(Montalvo-Garcia et al., 2020).
The privacy of data is addressed by the European Union's General Data Protection Regulation (GDPR) which came into effect in May 2018 (Henriksen-Bulmer et al., 2019). These authors explain that the GDPR covers the scope of the data, gives more rights to individuals, establishes data collection protocols, and requires organisations to justify their data collection. The work of Henriksen-Bulmer et al. (2019) is significant, as it explains how charities and small and medium-sized enterprises (SMEs) should implement GDPR in a structured way.

They established several steps in their application of GDPR using a case study (two managers, 29 staff and volunteers):

1. Data holdings: To understand how the charity stores and processes the data.

2. Analysis of data holdings: To better understand the data.

3. GDPR Process Guidance: To ensure GDPR compliance by assessing the practices and the processes of the data.

4. The DPIA Data Wheel: To assess any privacy risks.

GDPR is suitable for NPOs whose staff might lack technical and analytical skills. The only critical point is that these regulations and policies presented in GDPR might not apply to non-European NPOs. This algorithm is easy and can be applied by anyone who specialises in data mining and data privacy techniques. This framework may assist NPOs in tackling the challenges of data privacy and data accessibility in NPOs (as discussed in RQ3 below). However, NPOs have different missions and different characteristics; this framework might not apply to all NPOs globally. Table 2.5 summarises the three frameworks reported to adopt and apply data analytics in NPOs associated with the relevant data scope.

Scope	Framework's purpose
Data analytics	A data analytics methodology for small and medium-sized NPOs (Montalvo-Garcia et al., 2020)
Data analytics	A model containing three elements: accessibility, reliability, and adaptability (Shapiro and Oystrick, 2018)
Data privacy	Data Protection Impact Assessment (DPIA) framework (Henriksen-Bulmer et al., 2019)

Table 2.5: A Summary of Applying Data Analytics in NPOs Frameworks

2.3.2 RQ.2: What type of data analytics is being applied for NPO activities and missions?

The answer requires the exploration of different data analytics techniques applied to NPO activities and missions. NPO missions vary in their objectives, such as attracting donors, predicting donations, and improving the decision-making process. There are different applications of large and medium-sized data analytics for different research problems related to NPOs and other charitable sectors. For example, NPOs have business problems, such as attracting donors, predicting donations, and managing financial resources. Different data analytics techniques were applied, for different research problems related to NPOs missions, and are summarised in Figure 2.7.

Not surprisingly, the process of analysing data in NPOs is known as "knowledge discovery in databases" (Mayer, 2019). The process involves data capture and cleansing, aggregation, data mining, and

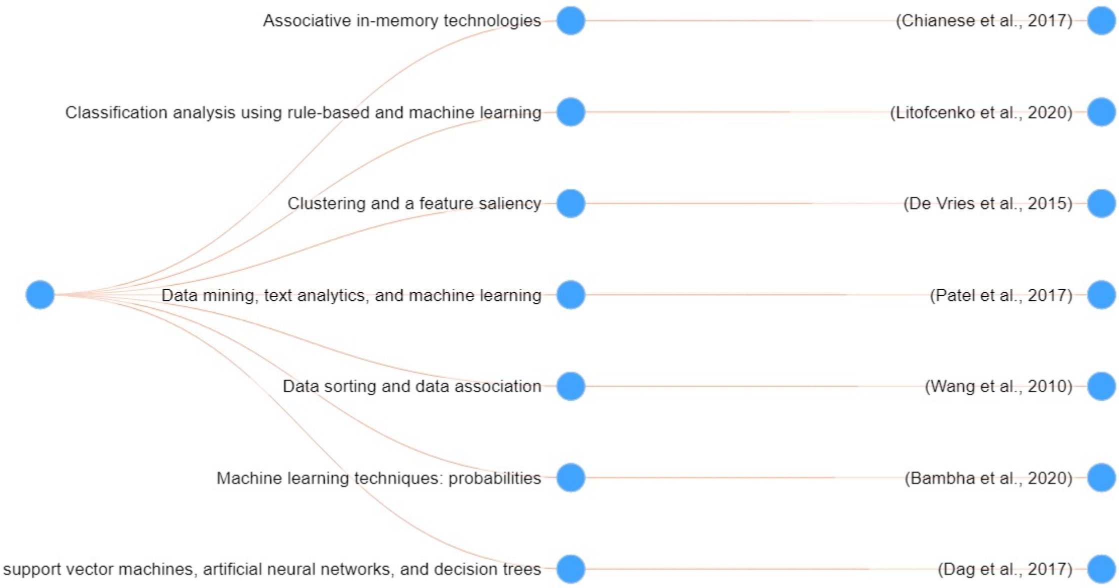


Figure 2.7: Summary of Different Applied Analytics for Different Problems and Missions in NPOs

interoperation. Two data analytic techniques applicable to NPO's activities are data mining and text analytics (Patel et al., 2017). Data mining is a common type of analytics used by NPOs because it can predict customer behaviours (Wang et al., 2010). It involves statistical analytics and machine learning technologies to extract hidden patterns from the data, which are difficult to extract manually (Wang et al., 2010). Data mining based on historical data can optimise the decision-making process (Wang et al., 2010).

Rule-based classification uses textual data to organise Austrian NPOs according to their missions, such as education and research, health, and religion (Litofcenko et al., 2020). International Classification of Nonprofit Organisations (ICNPO) categorises NPOs according to economic growth, political, cultural, and legal structures, and their scale, scope, and function (Salamon and Anheier, 1996). The rule-based method is a type of classification (semi-structured requires manual creation of IF-THEN rules) that can generate data-driven rules. This classification method solves the ongoing research issue that needs quantitative information on the activities and missions of NPOs, but administrative data is not readily accessible. Long, high-quality texts detailing the activities of NPOs are often unavailable, and there are limited human resources (Litofcenko et al., 2020). This ICNP classification approach is significant as it classifies NPO missions; the framework is appropriate for any number of NPOs.

A study predicting the survival of heart transplant patients was conducted by Dag et al. (2017). The researchers applied data mining methods to discover and analyse raw data from the transplant databases of the United Network for Organ Sharing, which is classified as an NPO in the United States. The study aimed to predict the probability of patient survival after heart transplant surgery. Four analytical models were used: artificial neural networks, decision trees, support vector machines, and logistic regression. The researchers adopted a hybrid data analytics methodology comprised of five stages: data preparation, classification, model assessment, a fusion of variables, and differentiation of variables. The authors claimed the decision tree model produced the best classification outcomes. This paper demonstrated the capability of big data analytics to generate valuable information in the field of health research.

In a similar approach, and demonstrating another application of data analytics by NPOs, an application of data analytics is conducted on organ donors who are individuals waiting for organ transplant donors (dd-SoT), and dd-SoT recipients (Bambha et al., 2020). The study provides a statistical analysis of organ donations, the dd-SoT waiting list, and deceased organ donors in the United States to ascertain (1) a deceased organ donor, (2) dd-SoT organs and (3) the organ recipient required for dd-SoT. They used datasets and UNOS. Their results show that individuals are more likely to require dd-SOT in the United States than to become organ donors following death. This study provides predictive analysis using an NPO dataset in the field of health.

An information system designed exploiting associative in-memory technologies to analyse the behaviours of visitors to social networks (e.g., Twitter's feedback from visitors) for an NPO (Chianese et al., 2017). The integrated approach combines business intelligence, big data, and IoT. Data, such as georeferenced tweets, was collected from social media. The outcomes were: new types of visitor profiles that consider the importance of the digital content provided to tourists and the effect of physical constraints on potential routes, relations between local and territorial visits, and strong support for social business intelligence.

Another research proposes a simply understandable segmentation strategy for the nonprofit sector, based on a unique, unsupervised clustering technique (MST-kNN), followed by a feature saliency method, the CM1 score (De Vries et al., 2015). The MST-kNN clustering method identifies groups of related responders with similar traits. A specific distance or similarity (correlation) metric is used to measure these related qualities. A sample of more than 15,000 responses was collected and analysed to reveal donor segmentations. The strategy used in this study provides a basis for nonprofit organisations to better cluster, segment, comprehend, and target their donor base (De Vries et al., 2015). The clusters discovered in this study provide a first look at the distinct donor segments of the Australian market. This research helps charities identify how donors behave and may be applicable to some NPOs.

Some studies underscore the importance of data analytics in decision-making processes. The decision-making process of NPOs needs growing data, which necessitates efficient management(Fredriksson, 2018). However, managing information in NPOs is a challenging task (Bopp et al., 2017). NPOs will not be able to draw insights and conclusions if the data is not well-collected and organised (Maxwell et al., 2016). To enable data usage, NPOs should adopt strategies that clarify employees' understanding of how to utilise data to achieve their organisation's tasks and mission (Kline and Dolamore, 2020). The literature presents several examples and case studies to demonstrate the importance of using data to drive decision-making by NPOs.

A study investigated how Family League, a large NPO in Baltimore, Maryland, collects, organises, and utilises data from the perspective of organisational culture (Kline and Dolamore, 2020). The researchers explained the importance of an organisation's values and found that organisational values may vary from one NPO to another. They learned that an internal culture could define and integrate data into the decision-making process. In another interesting study, (Bopp et al., 2017) conducted interviews with 13 mission-driven organisations to examine the role of data in evaluating and monitoring practices in NPOs. Their results showed that data did not empower these 13 NPOs. They identified three negative consequences of failing to evaluate, monitor, and empower data-driven decision-making: data drift, erosion of autonomy, and data fragmentation. These three impacts gave NPOs less control of data, and data is the key to successful decision-making.

Another study was conducted using a qualitative approach to determine the factors affecting decision-making regarding patient discharge from the hospital (Eigner et al., 2017). Focus groups were conducted at an Australian NPO hospital to gather data. This study aimed to analyse data about factors influencing patient discharge decisions. The results indicate a growing interest in data analytics and applications to enable faster and more accurate decision-making in healthcare. Although this study concludes that data analytics is a valuable tool for decision-making, this research's value is limited because the authors interpreted the results; they have not been evaluated using a data analytics tool.

2.3.3 RQ.3: What are the common challenges that NPOs face when adopting and applying data analytics?

Studies of data applications in the NPO literature are limited. This limitation means that NPOs face major challenges when adopting and applying data analytics. Investigating these challenges and considering them under different sub-headings provides researchers with a comprehensive overview of the existing literature. Ongoing challenges in applying data analytics in NPOs are evident. For example, the literature shows NPO leaders and managers face some significant obstacles, such as a lack of financial resources, ensuring the privacy of data, accessibility of data to authorised personnel, and technical skills required to manage and analyse their data (Gamage, 2016). The challenges are summarised below under different subheadings.

2.3.3.1 Financial and Technical Resources

NPOs spend less than 2% of their budget on information technology (IT) infrastructures, such as hardware and software, and only 36% on technical training (Hackler and Saxton, 2007). Selecting the best data analytics tools to respond to their enormous demands creates a significant challenge for NPOs (Johnson, 2015). Moreover, the study by (Rathi and Given, 2017) revealed that NPOs invest in low-cost applications to capture and store data, and may utilise cloud-based computing, such as Google Docs and Microsoft OneDrive (Rathi and Given, 2017). Technical constraints prevent NPOs from making the best use of software, hardware, and professional expertise (Johnson, 2015). However, affordable technologies exist in the form of cloud solutions and large-scale clusters to facilitate big data processing which can improve the performance of these organisations (Lněnička and Komárková, 2015).

MongoDB is one solution that does not require a complex distributed system. Patel et al. (2017) used MongoDB to store the data in an application of data analysis in the public sector for better e-governance. The study is useful as it shows the effective application of a tool to store large-scale data for NPOs. MongoDB is a common-purpose distributed system with features to ensure availability, scalability, and compliance with the most demanding data protection and privacy requirements (MongoDB, n.d.).

Several software applications, analytics methods, and visualisation tools inspire, show promise and can assist NPO researchers and practitioners in achieving their social missions (Johnson, 2015). The list of technologies and analytics tools are affordable and easy to use by NPOs professionals (Johnson, 2015). One limitation of web-based technologies reported in the list of Johnson (2015) is that they have been developed for NPOs in the United States and might not be applicable or beneficial to worldwide NPOs. No details are given regarding the ability of these solutions to handle massive amounts of data, which is another limitation.

2.3.3.2 Data Privacy

Since NPOs collect data from different sources, including the public, data privacy and accessibility issues present significant challenges to applying data technology (Costa and Santos, 2017; Gamage, 2016). NPOs collect sensitive and private data such as funds, medical research, and donor data, which require an anonymisation mechanism that restricts data access (Prakash and Singaravel, 2015). To overcome the challenge of securing data privacy in non-governmental organisations,Prakash and Singaravel (2015) proposed a privacy framework that used a top-down greedy algorithm with five stages:

1. Store all data on one single database.

2. Modify the data through processing such as aggregation and sampling.

3. Apply data mining algorithms.

4. Hide sensitive data.

5. Use data privacy techniques, such as blocking.

This algorithm is easy and can be applied by anyone who specialises in data mining and data privacy techniques. However, NPOs have different missions and different characteristics; this framework might not be applicable to all NPOs globally.

2.3.3.3 Data Accessibility

The literature shows an ongoing debate around the policies for accessing public data by organisations. NPOs have to ensure full accessibility to the required data (Hackler and Saxton, 2007). Gaining access to any type of data by any organisation involves formal procedures, and all staff must adhere to company policy. It is essential to provide guidelines for data property management and opting-out principles for public data (Witjas-Paalberends et al., 2018). Accessing NPOs datasets helps potential supporters, participants, clients, volunteers, and staff learn about specific business problems, and assist researchers and journalists in evaluating information about NPOs (Mayer, 2019). However, previous studies have only given examples of data access processes in NPOs; the literature did not discuss strategies, principles, and rules for accessing data.

2.3.3.4 Human Resources

Several researchers noted that a lack of NPO employees with adequate technical skills was a major challenge. Maxwell et al. (2016) claimed that NPOs struggle to invest in developing employee skills and maintaining a technically skilled workforce. Hou and Wang (2017) highlighted the critical challenge posed by the lack of knowledge among staff in applying analytics to NPOs' data. Developing personnel skills, particularly in data science specialities like analytics, is crucial to improving NPOs' performance and viability (Shah et al., 2017). Moreover, NPOs may rely on volunteers to provide technical support, which reveals another human resource limitation (Hackler and Saxton, 2007).

Providing adequate training to extrinsically and intrinsically motivated staff facilitates knowledge transfer and technical skill development (Witjas-Paalberends et al., 2018). Experts who were interviewed suggested best practices to implement when addressing the challenges of applying big data analytics in public-private partnerships (PPPs) (Witjas-Paalberends et al., 2018). NPOs can use these suggestions as guidelines for seeking solutions. However, the study by Witjas-Paalberends et al. (2018) is limited to the Netherlands. The study also examined the applications of big data from a holistic perspective that focused on PPS in the health sector. Figure 2.8 summarises the reported solutions for each category of challenges and considers all the drawbacks of applying data analytics.

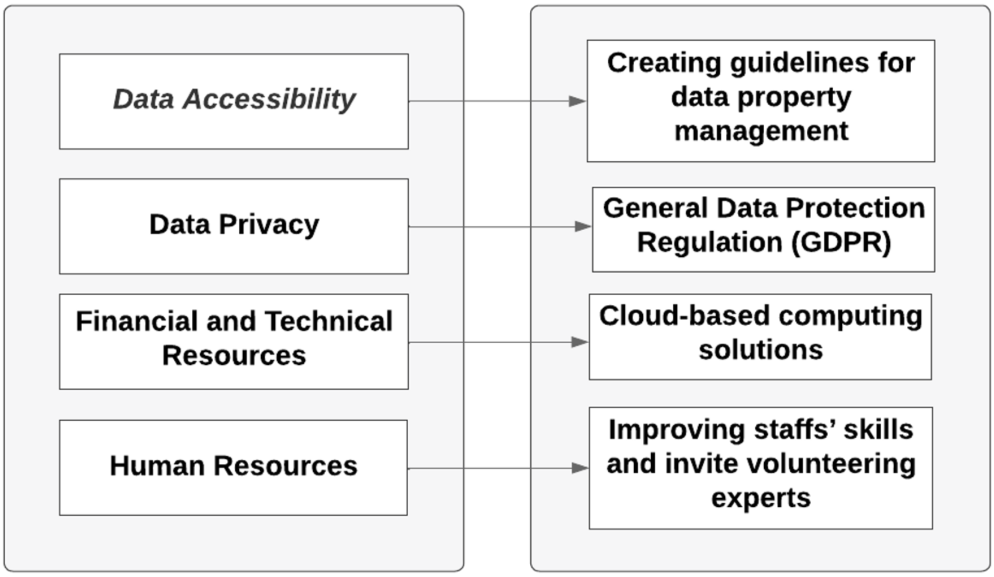


Figure 2.8: Summary of Suggested Solutions to Data Analytics Challenges in NPOs

2.4 Research Gaps and Future Directions

Research conducted on the adoption and application of data analytics in NPOs uncovered two specific gaps. These gaps are based on the findings in Section 2.3, and the analysis of the documents' records, citations, and keywords in subsection 2.4.

2.4.1 Research Gaps

First, analysing donor behaviour remains both critical and challenging for many NPOs, which is evident from the lack of literature on the topic. Future research can explore the application of data analytics to analyse donor behaviours. For example, NPOs can investigate and understand donor behaviours, and enhance their decision-making process,using machine learning and data mining techniques such as regression and classification. Therefore, a DSS to analyse donor behaviour in nonprofit organisations needs to be designed. However, little is known about designing a DSS, due to certain limitations of NPOs, such as technical experience and technical resources.

Second, NPOs may face critical obstacles, technical concerns, and a lack of sufficient resources when adopting data (as discussed in Sub-section 3.3). Challenges have been discovered while applying data analytics in NPOs; developing theoretical frameworks on how NPOs can adopt data analytics may resolve their critical issues. These frameworks could address privacy concerns and the NPO requirements that data analytics tools need to meet. Understanding and predicting users' perceptions (e.g., NPO managers and scholars) can overcome some of these challenges. For example, the technology acceptance model (TAM) (Davis, 1985) can provide insights and indications concerning the perception of data in NPOs. Case studies conducted in several countries could validate the framework to ensure its effectiveness, accuracy, and generalisability.

2.4.2 Future Directions

Recently, a design science research process framework adopted from Peffers et al. (2007) has started to address the research gap; filling this gap may help NPOs analyse donor behaviour (Alsolbi et al., 2022a). The framework helps design and build a DSS for analysing donor behaviour in nonprofit organisations (Alsolbi et al., 2022c). The framework contains six phases and three iterations to conduct specific actions towards building the DSS. The output framework's output is a design theory and an artefact (an AI-enabled DSS for analysing donor behaviours in NPOs). The design theory is the instantiation of the DSS we are building. It creates guidelines for further developments of an AI-enabled DSS, which will build a descriptive and predictive analysis of donors using machine learning techniques. Since this framework is under development, there could be some limitations of data features or attributes that help analysts perform the analysis intensively.

Investigating factors that influence the adoption of data analytics by NPOs may address the second research gap and offer a future direction to researchers. technology acceptance For example, applying TAM will enable researchers to study challenges and barriers facing NPOs which prevent NPOs from applying data analytics. It will also reveal how aware scholars of NPOs are of applying data analytics and data concepts. Currently, there does not appear to be a study attempting to determine user adoption of data analytics by NPOs. Investigating the influence of using quantitative or qualitative research methods to apply data analytics in NPOs offers another future research direction. These research methods will answer a range of questions about the characteristics of NPOs, including size, service types, scope, and intensity, which will help develop theories that provide adequacy of information for investing data analytics in NPOs (Johnson, 2015). Also, these methods could help practitioners of data analytics become aware of NPOs data analytics needs. However, these studies should be conducted in the literature under a collaboration between NPOs and industrial sectors who may volunteer their experience to assist in applying data analytics.

2.5 Chapter Summary

The researchers adopted the explorative methodology to conduct the systematic literature review. This systematic literature review aims to (1) investigate and synthesise the literature on data analytics as adopted and applied in nonprofits, based on a developed taxonomy of the literature, (2) provide perspectives into research trends in the field of NPOs via bibliometric analysis and a taxonomy of the discovered documents, and (3) encourage researchers to study and NPOs' management to adopt and apply data technologies in practice. Twenty-five peer-reviewed journal articles were examined in-depth to determine how and where data has been applied in NPOs and offer

a comprehensive view of this technology and its current applications. It was found that data is adopted to some extent in NPOs to improve organisational effectiveness. Nonetheless, the review and discussion of findings revealed two research gaps that require future studies to fill the void and contribute to the knowledge of data applications. Advanced analytics can use data analytics techniques to investigate donor behaviour and predict intentions, motivations, and engagement in NPOs. Frameworks and appropriate analytical models are essential to address technical challenges, privacy, and data accessibility. This study's contribution (1) focuses on challenges, specifications, and technical tools associated with applying data analytics in NPOs, (2) demonstrates the effectiveness of data analytics applications in NPOs using case studies, and (3) presents future research opportunities and directions based on the implications of the findings and their applicability to other domains, such as decision-making systems.
The systematic review identifies implications for data analytics researchers and practitioners. The results of the review show that applications of data analytics are limited in NPOs due to known challenges, such as financial constraints and technical experience. As a result, empirical studies may identify and propose solutions that practitioners can use to help NPOs to develop technical solutions that tackle these challenges. The potential to share lessons about applying data analytics, which may help NPOs worldwide overcome the challenges reported in the literature, is a final and beneficial implication.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

Many factors influence donor behaviour that requires the use of understanding donors (Li and Wu, 2019). Understanding donors and analysing their behaviour can assist NPOs in increasing efficiency of marketing and fundraising (Weinger, 2019). Donor behaviour includes donor intentions to offer either time or money, frequency (returning to give time or money after an initial donation), donor engagement, donor communications, and volunteer engagement (Shehu et al., 2015; Dietz and Keller, 2016). These donor behaviours can be understood better by the NPOs using technology, data science, and ML (Dag et al., 2017). ML has the potential to aid NPOs (nonprofit organisations) in several ways. Processes can be automated through, applying ML, permitting NPOs to concentrate more on accomplishing their core mission. Large datasets can also be analysed to identify trends and patterns that can inform decision-making (Deng et al., 2014). In addition, ML can improve customer service at NPOs by recognizing customer preferences and providing faster and more accurate responses to their inquiries (Li et al., 2018). Finally, ML can offer recommendations for donors and volunteers to inform decision-making. ML enables NPOs to allocate resources more effectively for multiple tasks (Chu et al., 2019).

Given our focus on NPO donor giving and doing, this research aims to (1) create a design science theory and (2) develop an artefact (AI-enabled DSS) to analyse donor behaviours. This chapter illustrates the components of the research methodology and the research aims to be achieved. The optimal research methodology should always be

adopted to accomplish research goals and objectives. The research methodology and data-gathering procedures chosen depended on various criteria, including the areas of research, aims, problems, and questions.
This chapter thoroughly explains and rationalises the chosen research approach, covering the methodology, data collection procedures, tools and applications, and data analysis methods. Moreover, by following the DSR process model framework presented by (Alsolbi et al., 2022a), this chapter presents a conceptual design that provides abstract and occasionally insufficient answers. A conceptual design is intended to meet all user and consumer needs from a functional, economic, technological, and other perspective (Horváth, 2000). This chapter is organised as follows: Section 3.2, which covers an introduction to the DSR, and the DSR process model framework. Each phase in the DSR process model framework is explained and clarified.

3.2 Design Science Research Methodology

Design Science is the process of designing artefacts and scientific investigations to answer a specific problem (Johannesson and Perjons, 2014). Design science develops and assesses IT artefacts to address specific organisational issues (Hevner et al., 2004). Constructs made from software, hardware, systems, or models are called artefacts (Hevner et al., 2004). The artefact must be innovative, productive, or valuable in resolving a previously unsolved or known issue (Hevner et al., 2004). The artefact might range from simple instantiations to greater efforts in the context of final design theories, implemented software or algorithms (Hevner et al., 2004). As Hevner et al. (2004) conveyed the characteristics of DSR methodology, this research follows DSR methodology.

3.2.1 Research Process Model

Creating the artefact should draw on existing theories and bodies of knowledge during the search process for a solution to a specified problem, (Peffers et al., 2007). Meanwhile, the study's findings must be effectively communicated to the appropriate audience (Hevner et al., 2004). The DSR represents an incremental and iterative process (Hevner et al., 2004). The iterative cycles also imply constant reflection and abstraction (Meth et al., 2015), which are necessary foundations for developing a design theory and artefact. Design theory describes how an artefact should be constructed to achieve the desired initiatives and results (Meth et al., 2015). Thus, the DSR process presented by Peffers et al. (2007) suited our research aims. The research process model developed by (Peffers et al., 2007) provides a useful synthesised general model that builds on other approaches (Gregor and Hevner, 2013).

Furthermore, we find this process model to be consistent with our research aims, which include: (1) identifying the problem, (2) defining the objectives of a solution, (3) design and development, (4) demonstration, (5) evaluation, and (6) communication. Considering the process model of Peffers et al. (2007), our research process model (Figure 3.1) (1) identified the problem through analysing the literature (2) found and formed the objectives of a solution, (3) developed

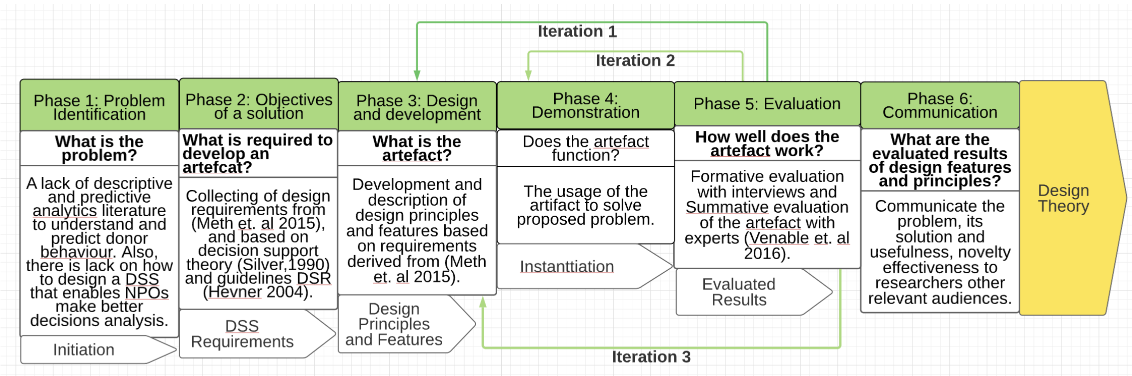


Figure 3.1: Designing AI-enabled DSS Process Model Based on Peffers et al. (2007) to Analyse Donor Behaviour in NPOs

a conceptual design, (4) built an artefact as an instantiation of the problem in a(further study, (5) evaluated the design of the artefact conceptually and practically, and (6) communicated the problem, the solution, and usefulness of the solution to researchers and other audience. Moreover, three iterations are conducted to ensure a variety of evaluation methods and to seek more validity for the artefact's design.

3.2.2 Phase 1: Problem Identification

This phase identifies a research problem and the importance of solving the proposed problem. Such instantiations demonstrated ML's capabilities in different instances and studies (Farrokhvar et al., 2018; Barzanti et al., 2017; Korolov et al., 2016) reported in the literature. However, prescriptive design knowledge, which would guide researchers and practitioners to systematically implement them for DSS in NPOs to analyse donor behaviour is lacking.

Two informal interviews were conducted with two experts from NPOs during this phase to expand awareness of the research problem outside the literature. The selection of experts was based on reviewing their profiles on social networks such as Linkedin and their experience working in NPOs. During interviews, we asked the experts (1) to describe their process of donor behaviour analysis, (2) state the challenges they face in designing a DSS that helps describe and predict donor behaviour, and (3) explicate the potential to create a design theory to guide the process of designing AI-enabled DSS. Table 3.1 summarises these interviews and states the process of analysing donor behaviour, the challenges faced by some NPOs, and suggestions for creating an artefact that analyses donor behaviour.

The interviews provided valuable insights, all of which were noted. For example, experts mentioned that descriptive and predictive analytics help NPOs make better decisions that increase efficiency and performance and understand factors that influence donations to NPOs. Furthermore, these analytics can be generated and visualised through a DSS. At this stage, the interviews helped identify the problem and increase awareness of creating a design theory of an artefact to analyse donor behaviour.

Expert category	**Process of analysing donor behaviour**	**Major challenges**	**Suggestions**
Data scientist in an NPO (5 years)	Occasionally analysing data of donors to create such analysis temporarily. No usage of DSS for analysing donor behaviour.	No collection of data of donor regularly. Different needs for creating such analysis, depending on NPOs' needs.	Building a DSS that analyse donor behaviour using ML techniques. Also, creating a design for the theory would increase the awareness of scholars to consider such analytics solutions.
Manager of NPO (8 years)	Normally using Excel sheets for creating tables and graphs about donors	Lack of human and technical resources. Low budget to afford such effective solutions. Spending a long time to make a decision.	Using an efficient system helps understand donors more comprehensively through such analysis.
Consultant of NPOs	Using Google analysis for analysing data of donors to help in making-decision.	Lack of human and technical resources. Lack of knowledge on designing a DSS for analysing donor behaviour.	Relying on ML capabilities to benefit more in creating visualisations that lead to understanding donors and volunteers and enhancing decision-making.

Table 3.1: Summary of Informal Interviews with NPO's Experts

3.2.3 Phase 2: Objective of a Solution

In this phase, the objectives and the requirements of the intended artefact are elicited to determine the main functionalities of the AI-enabled DSS. The initial requirements for creating an artefact are based on meta requirements of (Meth et al., 2015) and the decision theory by Silver, Mark S (1991). Also, the guidelines for developing an artefact from (Hevner et al., 2004) are followed. These guidelines are intended to help researchers, reviewers, authors, and readers understand what is required for effectual research (Hevner et al., 2004).

A design scientist must understand the artefact's objectives, which can be defined through design requirements (DRs). Table 3.2 introduces the initial design requirement derived from (Meth et al., 2015). Existing research in decision support theory typically describes two primary goals of decision-makers: ensuring maximum decision quality and reducing effort (Meth et al., 2015; Walls et al., 1992). However, a DSS may offer the user only limited selections of strategies (Silver, Mark S, 1991), which requires designers of a DSS to minimise the restrictions (Silver, 1988). System restrictiveness is the degree to which the DSS pre-selects decision techniques and, as a result, provides a limited variety of strategies, which may not include their preferred ones (Silver, 1988).

Ultimately, perceived advice quality, perceived cognitive effort, and perceived restrictiveness (Meth et al., 2015) are the most crucial characteristics of any DSS. Therefore, the three DRs borrowed from (Meth et al., 2015) offered a basement for our conceptual design and provided the potential for constructing our AI-enabled DSS design for analysing donor behaviour. Table 4.2 presents the DRs borrowed from (Meth et al., 2015) with an explanation and a justification for each DR.

Design Requirements	**Explanations**	**Justification**
Increase the decision quality by providing high-quality advice.	Providing advice with quality. Providing advice with quality. The process of analysing donor behaviour should be supported by a system that improves the quality of decisions.	Decision-makers have various objectives when making a decision (Gregor and Jones, 2007). Thus, they aim to achieve the maximum of good advice (Gregor and Jones, 2007). The AI-enabled DSS should provide a high-quality decision to help NPOs make better decisions about donors and volunteers.
Reduce the decision-makers' effort.	The system should prepare the decision and offer it to the decision-maker with the relevant information. For example, the system should provide information (through visualisations) about donors. This type of information can decrease the cognitive effort needed by NPOs decision-makers.	Decision-makers strive to make the minimum efforts when making decisions (Gregor and Jones, 2007). When the DSS provides high-quality advice, the effort of decision-makers will be reduced (Gregor and Jones, 2007).
Minimise system restrictiveness.	The system should offer several pre-selects decision strategies and offer decision-makers more flexibility to choose appropriate analytics.	The AI-enabled DSS should provide control and not restrict users (Gregor and Jones, 2007). For example, DSS users in NPOs must choose the type of analytics (predictive or descriptive).

Table 3.2: Design Requirements based on Meth et al. (2015)

3.2.4 Phase 3: Design and Development

This phase creates definitions of DPs and DFs which interpret the DRs from the previous phase. DPs can state what the artefact should do (Rhyn and Blohm, 2017b). DFs are unique artefact capabilities to fulfil DPs, according to (Meth et al., 2015), who also described DPs as statements to help develop an artefact that meets the DRs. DPs are essential to design theory elements because they contain important design knowledge (Meth et al., 2015). Because one aim is to build an artefact (AI-enabled DSS for analysing donor behaviour), the DPs should be stated as "they should do., or the system should fulfil... " (Gregor and Jones, 2007). Table 3.3 presents six DPs with their explanations. DFs are specific capabilities that map or address the DPs and DRs (Meth et al., 2015) as well as specific artefact functionalities required to meet DPs ((Meth et al., 2015). The DFs are introduced in the last conceptual design phase and are created to interpret the DPs (Table 3.4). The DPs and DFs components were developed during the informal interviews in phase one, the expert session and are supported by the existing literature (Rhyn and Blohm, 2017b).

DPs	Explanation
DP1: The AI-enabled DSS should learn based on ML	The AI-enabled DSS should be designed as an adaptive system (Silver, Mark S, 1991). The AI-enabled DSS should have predefined models to train the datasets. Therefore, ML techniques can learn based on the generated data of donors entered by decision-makers in NPOs (who use the AI-enabled DSS) to create effective descriptive and predictive models.
DP2: The AI-enabled DSS should describe donor behaviour.	Describing donor behaviour using ML is a key element of the AI-enabled DSS. NPOs may benefit from the interpreted results by the DSS to explain certain factors and information about donors, such as the most gender donating, and so on. Most importantly, ML techniques can describe the relative information about donors and visualise it properly.
DP3: The AI-enabled DSS should predict donor behaviour.	The AI-enabled of DSS should be able to predict donor behaviour using ML algorithms. Different predictive models can generate useful insights for NPOs decision-makers and support decision-making about donors. For example, the AI-enabled DSS should create a model to predict which age of previous donors may donate more in the future.
DP4: The AI-enabled DSS should describe volunteers' behaviour.	Describing volunteers' behaviour using ML is a key element of the AI-enabled DSS. NPOs need to rely on interpreted results by the DSS to explain certain factors and information about donors. For example, the AI-enabled DSS should create a model to predict who is likely to volunteer in the future.
DP5: The AI-enabled DSS should predict volunteers' behaviour.	The AI-enabled DSS should be able to predict volunteers' behaviour using ML algorithms. Different predictive models can generate useful insights for NPOs decision-makers and support decision-making about volunteers. Thus, ML techniques can describe the relative information about volunteers and visualise it properly.
DP6: The AI-enabled DSS should support decision-making with control and flexibility.	The AI-enabled DSS should maintain the control level by allowing decision-makers in NPOs (who use this system) to choose the predictive or descriptive analysis. Another example is allowing the NPOs' decision-makers to print a report or start a new analysis.

Table 3.3: Initial DPs of AI-enabled DSS to Analyse Donor Behaviour in NPOs

DFs	**Explanation**
DF1: Data import	The AI-enabled DSS should allow data import of donors. This feature will allow the user of the AI-enabled DSS to import the data from a spreadsheet containing important features. Users of the AI-enabled DSS should be able to load the datasets of donors and volunteers in order to start preparing the data.
DF2: Data pre-processing	This feature is to pre-process the data to ensure the adequacy of attributes. Meth et al. (2015) described pre-processing features as important. The pre-processing feature uses data pre-possessing techniques such as cleaning the data and formatting the dates.
DF3: Applying ML techniques (e.g. classifications and regressions)	The AI-enabled DSS should analyse the imported data using MLs techniques. ML techniques provide the means to structure the data, organise patterns and extract useful hidden information. For example, a classification technique can be chosen to classify donors based on their donations (high or low) and provide recommendations (high potential to donate in the future) or low (unlikely to donate again).
DF4: Self-Modifying code	Software systems that can independently change in a certain way are referred to as having self-modifying code, programs, or software (Đurić et al., 2016). The AI-enabled DSS should provide control for the users to maintain the workflow of making decisions (Rhyn et al., 2020). For example, enabling the user to choose the type of analysis from a list menu or removing unnecessary tooltips.

Table 3.4: Initial Design Features of AI-enabled DSS to Analyse Donor Behaviour in NPOs

After defining the DRs, DPs, and DFs, a conceptual design of the AI-enabled DSS (the first milestone of developing the AI-enabled DSS) is demonstrated in Phase 4, and evaluated in the first iteration. Required changes are applied, after completing the evaluation and reporting the results, to enhance the development of the AI-enabled DSS, and ensure the validity of DRs, DPs and DFs. The second milestone was to develop descriptive and predictive models for analysing donor behaviour in NPOs. Iteration two enhances the functionality and effectiveness of all developed models and aims to evaluate the descriptive and predictive models, applying any required changes. A full-functioning AI-DSS is then developed to ensure all the DRs. Iteration three follows to ensure all the DRs and objectives are met.

3.2.5 Phase 4: Demonstration

The demonstration phase presents an instantiation of the AI-enabled DSS. The aim of this phase is to show that the artefact's usage can be used to solve the problem. The demonstration phase is divided into two items: a conceptual design of the AI-enabled DSS and an artefact (AI-enabled DSS) for analysing donor behaviour. The first part of the demonstration is to present a conceptual design of the AI-enabled DSS to ensure the validity of DRs, DPs, and DFs in solving the research problem. We combined DRs of the AI-enabled DSS to map the DPs (objectives of the AI-enabled DSS), and the DFs from the previous Design and Development phase. A preliminary conceptual design is developed and evaluated with NPOs stakeholders. Before iteration one, we met with experts to demonstrate this preliminary conceptual design and explain how the components had emerged. The second part of the demonstration phase is to develop an AI-enabled DSS, applying changes from iteration one. The descriptive and predictive analysis was applied and introduced to form components of the AI-enabled DSS. Then, the AI-enabled DSS was evaluated (iteration two) to ensure its functionality and meet all DRs. After validating the AI-enabled DSS models and finalising iteration two, we demonstrated the DSS to experts who explored it and tested its functionality and operational ability. Finally, the AI-enabled DSS was evaluated by experts through interviews to report any other changes for iteration three. Iteration three ensured that all the objectives of the AI-enabled DSS are met and no further required changes during the research process model.

3.2.6 Phase 5: Evaluation

The process of measuring the success of a DSS is known as DSS evaluation (Maynard et al., 2001). It is beneficial in building and evaluating crucial systems such as DSS to include as many people and reference groups as feasible (Maynard et al., 2001). Evaluation in the DSR process model examines and quantifies how well the artefact contributes to a solution to the problem (Peffers et al., 2007). Evaluation can take many forms, including a comparison of the artefact's functionality with the solution objectives defined in phase 2 of conceptual design, objective quantitative performance measures, satisfaction and client feedback results, or simulations (Peffers et al., 2007).

In this phase, the framework of evaluation developed by (Venables and Ripley, 2002) is used, which has two types of evaluations, formative and summative. The assessment is meant to evaluate the AI-enabled DSS and relevant DRs, DPs, and DFs. Formative evaluations are utilised to generate experimentally verified interpretations to serve as the foundation for effective action in enhancing the traits or performance of the evaluated artefact (Venables and Ripley, 2002). Summative evaluation provides a foundation to produce common meanings of the evaluation in a different context. The evaluation phase runs three iterations as presented in (Peffers et al., 2007)' framework (as shown in Table 3.5); after each iteration, some changes are applied to enhance the design and development of the AI-enabled DSS. Later in this thesis, details of each iteration will present the involved steps, data collection and analysis, and the reporting of the outcomes of each iteration.

Iteration No.	**Evaluation Goal**	**Methods**
Iteration one	Formative (To validate relevancy)	Conducting Interviews
Iteration two	Formative (To validate feasibility)	ML validation techniques (e.g. K-fold cross)
Iteration three	Summative (to validate usefulness and effectiveness)	conducting Interviews

Table 3.5: An Overview of Evaluations and Related Iterations

3.2.7 Phase 6: Communication

This phase communicates the research problem, proposed solution, and results to scholars through publications. This phase informs researchers and other relevant audiences, such as practising professionals, about the problem and its importance, the artefact, its utility and novelty, the rigour of its design, and its effectiveness (Peffers et al., 2007). There are several studies published, which can be identified as follows:

3.2.7.1 Publication 1: Analysing Donors Behaviours in Nonprofit Organisations: A Design Science Research Framework (Alsolbi et al., 2022a)

This publication aims to present the research problem, define the initial requirements and objectives of a solution, and present the research process framework. This paper was accepted at the 3rd International Conference on Machine Intelligence and Signal Processing, NIT Arunachal Pradesh, India, in September 2021. During the presentation of the paper, there was a fruitful discussion on the planned outcomes of the research process framework.

The framework's significance is to give designers of intelligent decision support systems a theoretical foundation for developing generalized design principles and design features. It also demonstrates the abilities of ML and data analytics techniques to comprehend donor behaviour by examining the outside influences to understand donors and volunteers.

3.2.7.2 Publication 2: Data Analytics Research in Nonprofit Organisations: A Bibliometric Analysis (Alsolbi et al., 2022c)

This publication aims to provide an overview of data analytics applications in NPOs, using bibliometric analysis. It was introduced as a basement for initiating the research gaps and outlines the status of applying data analytics in NPOs. This paper was accepted at the 3rd International Conference on Machine Intelligence and Signal Processing, NIT Arunachal Pradesh, India, in September 2021. During the presentation of the paper, there was a useful discussion on the planned outcomes of the research process framework.
This paper also highlighted the authors' research findings and citation analysis, demonstrating the researchers' primary areas of interest when applying data analytics to non-profit organizations. The study's significance can be used to improve the knowledge already available on data analytics applications in NPOs and guide future research.

3.2.7.3 Publication 3: Different approaches of bibliometric analysis for data analytics applications in nonprofit organisations (Alsolbi et al., 2022b)

This paper is an extension of Publication 2, where the methodology has been extended to include a framework for using the Bibliometrix R-tool. This paper also presented the highlights on citation analysis, and scientific production of authors, and shows the focus of researchers who applied data analytics in NPOs. This paper is published in the Journal of Smart Environments and Green Computing. The research findings can be used to inform future studies and improve the body of knowledge already available on data analytics applications in NPOs.

Three major conclusions that may be derived from the findings are as follows: 1) In the context of non-profit organizations, reliable and traditional statistical methods-based data analysis procedures are consistently prevalent; 2) Healthcare and public affairs are two important industries that use data analytics to support decision-making and problem-solving; 3) Artificial intelligence (AI) based data analytics is a recently emerging trend, especially in the healthcare-related sector; however, it is still in its infancy, and more efforts are required to develop it.

3.2.7.4 Publication 4: A Systematic Review and Taxonomy of Data Analytics in Nonprofit Organisations (Alsolbi et al., 2023)

This publication systematically reviews data analysis applications by applying two approaches to collecting, assessing, and reporting the literature. Then, a basement of research gaps and future directions are introduced in this paper to guide scholars for further research in NPOs. This paper has been published in the Asia Pacific Journal of Information Systems (APJIS).

The study's contribution to the literature is significant because it sheds light on the advantages and disadvantages NPOs can experience from various domains' current use of data analytics applications. Future directions for research are also presented in this study. The findings reveal that the utilization of data analytics applications by NPOs has not been thoroughly investigated, highlighting the requirement for more study.

This study thoroughly evaluates the literature on data analytics applications in NPOs. (1) What are the suggested frameworks and adoption strategies for data analytics in NPOs? (2) What features of data analytics are used for NPOs' missions and activities? (3) What obstacles and problems must NPOs overcome in order to accept and use data analytics for their missions? We gathered, analysed, and used the data to create a new taxonomy in order to respond to the three research objectives. The findings reveal that the use of data analytics apps by NPOs has not been thoroughly investigated, highlighting the need for more study.

3.2.7.5 Publication 5: A Conceptual Design of AI-enabled Decision Support System for Analysing Donor Behaviour in Nonprofit Organisation

This publication aims to introduce the conceptual design of the AI-enabled DSS proposed in Publication 1 (Alsolbi et al., 2022a). The paper mainly focuses on defining the AI-enabled DSS's DRs, DPs, and DFs. Following that is the evaluation of the conceptual design using semi-instructed interviews. The paper is under submission process in one high-ranked journal.
A conceptual design is developed using a Design Science Research methodology to assess an AI-enabled DSS's initial DPs and features to analyse donor behaviour in NPOs. Interviewing participants from NPOs served as a formative assessment in the conceptual design evaluation phase. The Appreciative Enquiry framework was used to conduct the interviews in order to streamline the process. The findings of evaluating the conceptual design suggest that the require-

ments of the AI-enabled DSS be as efficient, effective, flexible, and usable as possible. This study adds to the body of design expertise for AI-enabled DSS that analyses donor behaviour in nonprofit organisations. A practical AI-enabled DSS for analysing donor behaviour in NPOs will be introduced through future research that combines theoretical elements. This study is limited to an analysis of volunteers and donors.

3.3 Chapter Summary

Analysing donor behaviour can enhance decision-making, support the decisions, and offer future predictions of donations and volunteering events. Thus, this chapter presented a DSR process model framework to guide the researcher to build an AI-enabled DSS for analysing donor behaviour in NPOs. This chapter described the six phases of the DSR model process framework, explaining each phase's components, inputs, and outputs. The six phases helped achieve the research aims and answer the research questions. Each phase required research activities, which ranged from defining the research problem, collecting the DRs, initiating the DP's and design features, and developing the artefact, evaluations, and readying results for publication.

CHAPTER 4

DATA ANALYSIS AND ARTEFACT DEPLOYMENT

4.1 Introduction

This chapter aims to provide a step-by-step explanation of data analysis and deployment of AI-enabled DSS. The three iterations and their involved methods are also presented. This chapter presents the methods applied during each iteration, the preliminary analysis, data processing and modelling, and essentially shows the process of building the AI-enabled DSS for analysing donor behaviour. The three iterations are presented in accordance with the DSR process model framework (Alsolbi et al., 2022a). Each iteration followed evaluations (formative and summative), and clarified what methods of validations have been utilised. This chapter represents a core stage in this thesis, as it facilitates data analysis and deployment of the AI-enabled DSS for analysing donor behaviour.

4.2 Iteration One: Evaluating the Conceptual Design of AI-enabled DSS

Iteration one aims to evaluate the initial DRs, DPs, and DFs. Iteration one resulted after DRs, DPs, and DFs were evaluated using formative assessment to ensure their relevance to our research aims and objectives. For iteration one evaluation, semi-structured interviews were conducted with NPO decision-makers, data scientists, volunteering experts, systems designers and analysts, NPO experts, and NPO managers. Experts were selected based on their technical experience and engagement with different NPOs in different nations. Interviews, one of the qualitative research methods, are frequently concerned with obtaining a thorough grasp of a situation or determining a specific phenomenon (Dworkin, 2012). During interviews, those experts evaluate our conceptual design; their changes or suggestions are applied. These iteration results lead to on DRs, DPs, and DFs developed in Phase 3: Design and Development. This iteration represents the main milestone of Phase 3, representing an evaluation of the conceptual design. We met with experts to demonstrate this preliminary conceptual design and explained how the components emerged. Figure 4.1 shows three main components of the conceptual design: three DRs, six DPs, and four DFs.

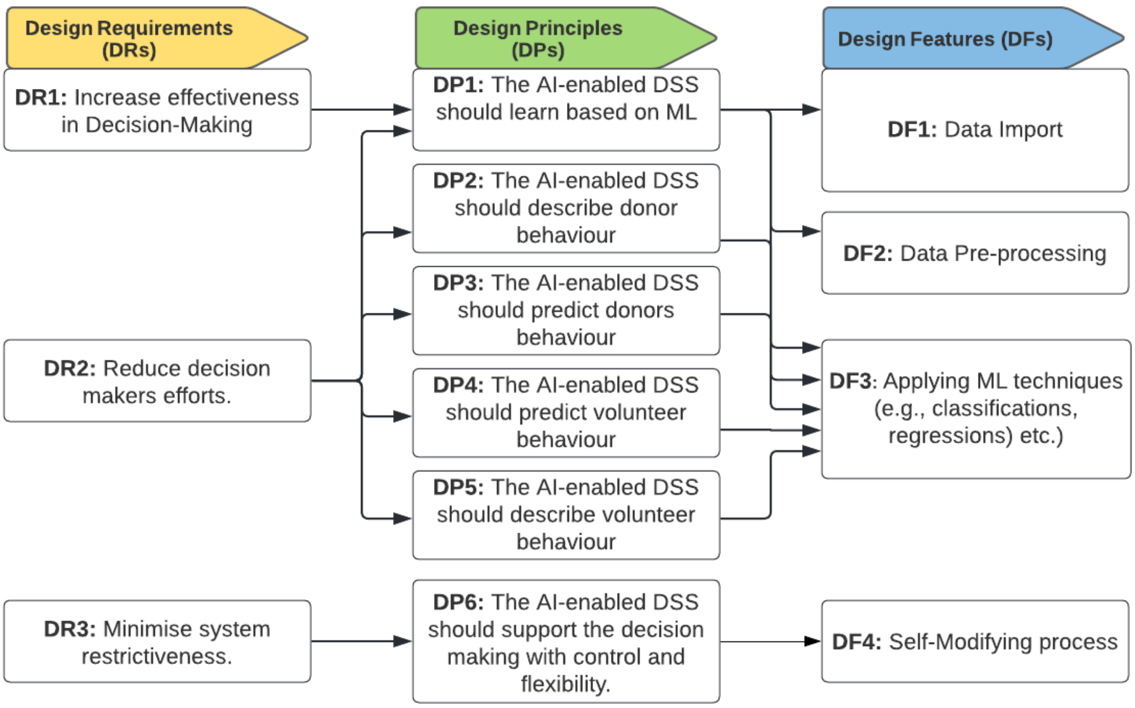


Figure 4.1: The Preliminary Conceptual Design of AI-enabled DSS for Analysing Donor and Volunteer Behaviour in NPOs

4.2.1 Data Collection

Iteration one was conducted using semi-structured interviews with a total of 16 interviewees from NPOs. In qualitative research methods, the sample number of interviews varies depending on the number of questions and the research objectives. The sample size is frequently smaller compared to quantitative research methods (Weston et al., 2001) because qualitative research methods are frequently concerned with gaining a thorough grasp of a phenomenon or determining the meaning (Downe-Wamboldt, 1992). Therefore, these 16 interviewees were invited to participate based on the variety of their experiences, their deep understanding of the research problem, and their availability to be interviewed within a certain period of the study.

Each interviewee was invited via email which included a brief introduction to the research problem and its proposed solution. They signed a consent form and agreed to be recorded to analyse the interviews. After an initial meeting, each interviewee views a short presentation which introduces the conceptual design. The 10-minute presentation duration included a brief introduction about the research problem, the research aims, the conceptual design, the expected output of the research, and an explanation of the interview process.

This is followed by introducing 11 questions (shown in Appendix C) distributed in five phases of Appreciative Inquiry Theory (Börjesson et al., 2007). Appreciative Inquiry is a method of focusing on what is excellent in a company in order to improve it and build a better future (Börjesson et al., 2007). Appreciative Inquiry has been selected because it can support a discussion of beliefs, emphasizing strengths and various levels of appreciation (Bushe, 2011). In Appreciative Inquiry, designing the questions should consider the need to provide the best guidance in obtaining the best answers from the stakeholders. Also, the questions were designed to be easily understandable by the participants so they could provide sufficiently detailed answers. Experts were asked several questions relative to each of the five phases, following the flow of the Appreciative Inquiry. The five phases were:

1- Participants: the questions are about experts' experience working in NPOs.

2- Discovery: the questions are to ask experts about their experience working on DSS, ML, and data analytics, either in NPOs or in profitable organisations.

3- Dream: the questions are to collect the experts' feedback on the conceptual design of AI-enabled DSS for analysing donor behaviour.

4- Design: the questions are to ask experts about any additional DRs, DPs and DFs that can be added to the conceptual design.

5- Destiny: the questions are to measure experts' expectations of the AI-enabled DSS for analysing donor behaviour in NPOs.

4.2.2 Data Analysis

All the records of the interviews were saved to the OneDrive of the research investigator from the University of Technology Sydney. Each interview was under an hour, which included an introduction to our research framework, an explanation of the conceptual design, and the questions.

Qualitative data analysis strategies vary widely, depending on the purpose of each collected piece of qualitative data (Saldaña, 2014). However, in this study, two strategies of qualitative data analysis, which are coding and categorising, are applied to the interview analysis. Coding is the process of understanding the meanings of different data sections; in some uses, coding entails giving a datum a symbolic meaning. On the other hand, categorising in qualitative data analysis means grouping similar or comparable codes for further analysis. In this research, four categories are provided to report the analysis results.

Interestingly, the four categories have various related codes, which are explained accordingly. Thus, some codes from different categories are linked to providing insightful information. We use MaxQDA software that specialises in analysing qualitative data to help categorise the data and support the coding process. The four categories of all interview responses are:

Category one: Working experience

This category groups the interviewees' relevant experience based on roles and the amount of time in those roles. This category summarises interviewees' answers during the first two phases of Appreciative Inquiry Theory: Participants and Discovery. Table 4.1 shows the code of working experience of participants in the interviews. The experts who were interviewed had experience in data science, software engineering, systems design and analysis, social science, management, and volunteering experience as consultants. The range of experiences provided richness in the answers and the evaluation. Conducting interviews also revealed categories of answers from different experts. These categories led to the discovery of hidden patterns among all the interviewees (Weston et al., 2001).

During the first part of our interviews, we asked the experts simple questions about their work experience. We found that most experts had some experience working or volunteering in different types of NPOs such as charities, religious centres, and youth centres. Following that, two software engineering experts, who had briefly volunteered in NPOs, provided relative answers during the interviews. Three NPO managers and one CEO, all from different NPO, also answered our questions, but they each comprehensively explained the challenges of analysing donor behaviour in NPOs.

Interestingly, one researcher in NPOs studies supports our claims that DSS are critical for NPOs to target more donors. He stated that NPOs require a clear vision of the donor behaviour of donations over a long time. The variety of involving experts in our interviews helped us raise awareness of the problem, understand some of the decision-making requirements of NPOs, and draw a path of opinions that assist us in designing the AI-enabled DSS in NPOs. These codes of category one are integrated to provide a meaningful analysis of the interviews.

Working experience code	Number of Experts	Number of experts on DSS	Length of experience	Number of Experts of donor behaviour	Experience length
NPO manager	4	1	4	1	10
Data Scientist	3	2	6 and 15	1	8
Consultant for NPO	2	0	0	2	4 and 8
Software engineer	2	1	7	1	2
Volunteering work experience	2	1	12	0	0
Researcher of NPO studies	1	0	0	1	13
Social expert in NPOs	1	0	0	1	5
System designer and analyst	1	1	18	1	13
Total of experts	16	6	-	8	-

Table 4.1: Category of Working Experience

The construction of each category construction attempts to group things that appear similar and appropriate (Saldaña, 2014). A categorisation is an interpretive act that may help interpret other categories and their codes. Thus, the category of working experience helps know different answers of different experts, with their variety of experience. Relevant and various experiences may include knowledge and skills in the evaluation process. Therefore, the codes of work experience are linked to the following categories and codes for maximizing the benefits of the evaluation and drawing useful conclusions from the interview analysis.

Category two: Evaluation of the conceptual design

This category collects the relevant answers to participants' conceptual design evaluations. This category summarises interviewees' answers during Dream, Phase 3 of the Appreciative Inquiry Theory. Experts were asked about their additional DRs, DPs, and DFs, and then they were asked about their opinions on mapping DRs, DPs, and DFs (the conceptual design). We then analysed each answer to assign it to a certain code to form the evaluation category. The evaluation category eventually had five codes of answers (Table 4.2), which were reported by experts evaluating our conceptual design.

This category collects the relevant answers of participant expectations about our AI-enabled DSS for analysing donor behaviour in NPOs, and groups them similarly. This category summarises interviewees' answers during Destiny, Phase 5 of the Appreciative Inquiry Theory.

Before concluding each interview, we asked the experts what they expected from the AI-enabled DSS for analysing donor behaviour. One question that was asked of all interviewees was: "What results/analysis do you expect when implementing the AI-enabled DSS to analyse donor behaviours?" Further, all answers were analysed and assigned under a code. As a result, four codes of expectations were obtained for the AI-enabled DSS. Table 4.4 shows the association between the work experience category and experts' expectations of AI-enabled DSS for analysing donor behaviour in NPOs. The work experience category codes combined all interviewees with the same role.

Most participants expected our artefact to predict and describe donor behaviour, which represented our main research objectives. Other experts expected the AI-enabled DSS to be a helpful tool to enhance decision-making in NPOs based on their understanding of the three conceptual components (DRs, DPs, and DFs). Essentially, one data scientist and a researcher of social studies in NPOs, expected that ML techniques are required to achieve the objectives of AI-enabled DSS for analysing donor behaviour.
The association of evaluation codes and work experience assist in providing useful feedback on users' different experiences. For example, when different experts agree on one code of evaluation, it indicates the importance of considering that code when applying such changes in the following iterations.

<table>
<tr><th>Code of experience</th><th>Number of experts</th><th>Codes of Evaluation</th></tr>
<tr><td rowspan="2">Data Scientist</td><td>2</td><td>Abstractive design: This code is means that the conceptual design is abstract but straightforward, providing designers and analysts with several ideas about the implementations.</td></tr>
<tr><td>1</td><td rowspan="2">Good design, but additions are required. There are additional requirements such as consideration of usability and quality of data.</td></tr>
<tr><td rowspan="2">NPO manager</td><td>1</td></tr>
<tr><td>2</td><td rowspan="4">Great design: This code means that experts stated that the conceptual design is great in its structure and mapping. There were no further comments from experts.</td></tr>
<tr><td>Consultant for NPO</td><td>2</td></tr>
<tr><td>Software engineer</td><td>2</td></tr>
<tr><td>Volunteering work experience</td><td>2</td></tr>
<tr><td>System designer and analyst</td><td>1</td><td>Systematic design: This code means that experts stated that the conceptual design is great in its structure and mapping. There were no further comments from experts.</td></tr>
<tr><td>CEO of NPOs</td><td>1</td><td>Great design: This code means that experts stated that the conceptual design is great in its structure and mapping. There were no further comments from experts.</td></tr>
<tr><td>Researcher of NPO studies</td><td>1</td><td>Abstractive design: This code is means that the conceptual design is abstract but straightforward, providing designers and analysts with several ideas about the implementations.</td></tr>
<tr><td>Social expert in NPOs</td><td>1</td><td>Great design: This code means that experts stated that the conceptual design is great in its structure and mapping. There were no further comments from experts.</td></tr>
</table>

Table 4.2: Category of Experts' Expectations

Category three: Additional DRs, DPs and DFs

This category is to combine the similarity of additional DRs, DPs, and DFs by participants in the interviews

This category maps the additional DRs, DPs, and DFs by participants in the interviews. This category summarises interviewee answers during the Design phase of the Appreciative Inquiry Theory. For example, three NPO managers required "useability," indicating usability as a key requirement for those lacking technical skills. Similarly, a "very friendly system" was required by one experienced NPO volunteer. Noticeably, "Quality of Data" is among the requirements because of its importance to data scientists. Any data analysis should be based on accurate and high-quality data (**?**). Wang and Strong (1996) grouped more than 100 quality data elements into four groups: relevance, accuracy, accessibility, and representation. However, the quality of data is not considered in this study due to data source limitations.

Data quality will be determined by checking these four categories of data quality during the step of data preparation prior to applying data analysis using ML techniques. Some experts requested unique additional requirements, such as "increasing efficiency" and "adaptive system". Increasing efficiency of decision-making is typical of our DR 2. However, an "adaptive system" is an interesting requirement for interactive systems (Weibelzahl et al., 2020). An adaptive system is typically used when all the necessary input characteristics are unknown or there are some slow variations in the input data (Narendra, 2016).

The "adaptive systems" requirement is out of our scope and research objective for this study and further studies of building an AI-enabled DSS for analysing donor behaviour in NPOs. In addition, a social expert in social science mentioned that more NPOs would benefit substantially when having a flexible system to install and edit contents of the AI-enabled DSS. This unique requirement was also considered when building the AI-enabled DSS's analytical models (iteration two) and (iteration three).

Interestingly, DPs are derived from the DRs. Therefore, we asked the interviewees to add DPs according to the additional DRs. For example, experts who asked for "Usability" as an additional DR, stated that "the AI-enabled DSS should be easy to use to describe and predict donor behaviour". Another social expert in NPOs claimed that a possible DP could be "the enabled DSS should be flexible to install and access by NPOs' stakeholders". This is to ensure that the "flexibility" of the additional DR can be achieved and save time and effort for decision-makers in NPOs. Consequently, experts who asked for additional DRs, were asked about any additional DFs. Coincidentally, experts who added "usability" as additional DRs, asked for "Tooltips", in addition to "easy to navigate" and "choice of colours" as other DFs. Table 4.3 represents the category of the additional DRs, DPs, and DFs, linked with the category of work experience.

Code of experience	**Number of experts**	**Additional DRs**	**Additional DPs**	**Additional DFs**
Data Scientist	3	Quality of data	-	-
		Increasing efficiency		
NPO manager	3	Adaptive system		
		Quality of data		
		Usability		Tooltips
				Choice of colours
Consultant in NPOs	2	Quality of data	-	-
		Security		Easy to navigate
Software engineer	2	Usability		-
Volunteering work experience	2		DSS should be usable to use to describe or predict donor behaviour	Tooltips Tooltips
System designer and analyst	1			
Technical committee in NPOs/ CEO of NPOs	1			
Researcher of NPO studies	1			
Social expert in NPOs	1	Flexibility to use		

Table 4.3: Category of the Additional DRs, DPs, and DFs

Category four: Expectations of the AI-enabled DSS

This category is to collect the relevant answers of participants' expectations about our AI-enabled DSS for analysing donor behaviour in NPOs, and group them similarly. This category summarises interviewees' answers during the destiny of the Appreciative Inquiry Theory phase.

Before concluding each interview, we asked the experts what they expect from the AI-enabled DSS to analyse donor behaviour. All interviewees asked one question: "What results/analysis do you expect when implementing the AI-enabled DSS to analyse donor behaviours?". Further, all answers were analysed and assigned under a code. As a result, four codes of expectations were obtained for the AI-enabled DSS. Table 5.4 shows the association between the work experience category and experts' expectations of AI-enabled DSS for analysing donor behaviour of AI-enabled DSS for analysing donor behaviour in NPOs. The work experience category codes combined all interviewees with the same role.

Most participants expected our artefact can predict and describe donor behaviour, representing our main research objectives. Other experts expected that the AI-enabled DSS would be a helpful solution to enhance decision-making in NPOs based on their understanding of the three conceptual components (DRs, DPs, and DFs). Essentially, one data scientist and a researcher in social studies in NPOs, expected that the ML techniques are required to achieve the objectives of AI-enabled DSS for analysing donor behaviour. The association of evaluation codes and work experience assist in providing useful feedback on users' different experience. For example, when different experts agree on one code of evaluation, it indicates the importance of considering it when applying such changes in the following iterations.

Code of experience	**Number of expert**	**Codes of Expectations**
Data Scientist	1	ML techniques are required: This code is to combine and summarise similar answers of experts who required a variety of ML techniques to be applied.
	2	Predicting donor behaviour: This code is to combine and summarise similar answers of experts who expected that the AI-enabled DSS can predict donor behaviour through predictive analysis.
NPO manager	1	Describing donor behaviour: This code is to combine and summarise similar answers of experts who expected that the AI-enabled DSS can describe donor behaviour through predictive analysis.
	2	Predicting donor behaviour: This code is to combine and summarise similar answers of experts who expected that the AI-enabled DSS can predict donor behaviour through predictive analysis.
Consultant for NPO	1	Helpful tool to enhance decision-making: This code is to combine and summarise similar answers of experts who expected that the AI-enabled DSS as a helpful tool to enhance the decision-making process in NPOs.
	1	Predicting donor behaviour: This code is to combine and summarise similar answers of experts who expected that the AI-enabled DSS can predict donor behaviour through predictive analysis.
Software engineer		
Volunteering experience	2	
System designer and analyst	1	Helpful tool to enhance decision-making: This code is to combine similar answers of experts who expected the AI-enabled DSS a helpful tool to enhance the decision-making process in NPOs.
CEO of NPOs	1	
Researcher of NPO studies	1	ML techniques are required: ML techniques are required: This code is to combine similar answers of experts who required a variety of ML techniques to be applied.
Social expert in NPOs	1	Helpful tool to enhance decision-making: This code is to combine similar answers of experts who expected the AI-enabled DSS a helpful tool to enhance the decision-making process in NPOs.

Table 4.4: Category of Experts' Expectations of the AI-enabled DSS

4.3 Iteration Two: Developing the AI-enabled DSS

After presenting the DRs, DPs, and DFs in the conceptual design, descriptive and predictive models are completed and fully functioning. The aim of this iteration is to ensure that the design and implementation of the AI-enabled DSS are fully completed. For the evaluation, k-fold cross-validation is applied to ensure these models' effectiveness and performance. K-fold cross-validation is a common technique to evaluate the models and estimates errors among practitioners (Anguita et al., 2012). This iteration requires a certain dataset to apply the validation techniques collected from a public source. This iteration represents the main milestone of Phase 4: Demonstration.

4.3.1 Data Sources

4.3.1.1 Donors' Data

A public dataset[1] was used, which included some features of donors such as age, state, gender, previous history of donations and amount of donations. This dataset was used in The Fourth International Conference on Knowledge Discovery and Data Mining KDD-98 (Parsa, 19898). A nonprofit organisation gathered the dataset; their mission was to offer activities and services to United States military veterans who had suffered spinal cord injuries or diseases. This NPO raised funds through direct mailing campaigns (Parsa, 19898). The available dataset includes a record of each donor who received the 1997 mailing but had not made a donation during the previous 12 months. The dataset consists of 19,1779 records. This iteration's outcome was to

[1]Metadata and descriptions of the data are available on appendix D

build a predictive analysis of donors and volunteers in NPOs using various data analysis and ML techniques, which will be validated to ensure the accuracy of analysis performance.

4.3.1.2 Volunteers' Data

The volunteer dataset [2] used for this research is part of a survey conducted by the Bureau of the Census for Corporation for National and Community Service (CNCS). Data were provided on volunteer activity involvement over the course of a year, from 2016 to 2017, as well as the kinds of organisations and voluntary activities that people participated in. Additionally, the survey is the only current comprehensive source of data on workers' occupations and the industries in which they work. Information is also available regarding individuals' desire for employment at the time of the survey, prior work history, and intended approach to the job search (Census Bureau, 2017).

4.3.2 Selecting the Development Platforms

For developing an AI-enabled DSS, R and Dataiku were selected as platforms for the execution due to some features on both platforms. R is a powerful programming language and software environment used for statistical computing and graphics, offering a wide range of data manipulation and analysis tools, such as linear and nonlinear modelling, classical statistical tests, time-series analysis, and classification (Liaw and Wiener, 2002; Venables and Ripley, 2002). Additionally, R is extensible, allowing users to write their functions and packages (Matloff, 2011), and has powerful graphical capabilities

[2] Metadata and descriptions of the data are available in Appendix D

for data exploration and understanding complex relationships (Venables and Ripley, 2002). Furthermore, R is platform-independent, meaning it can be used on different operating systems (Matloff, 2011), and has a large community of users and developers, making it easy to find help and support when needed (Matloff, 2011). Dataiku is a platform for complex analytics and data science that offers a full range of tools to speed up the entire data science process (Chuang, 2020). Dataiku's model monitoring feature allows users to track their models' performance and accuracy over time. With this feature, users can quickly identify potential problems or areas of improvement and optimise their models accordingly.

4.3.3 R Studio: Building a Probability Model of Donors and Volunteers

The statistical computing and graphics environment R is available for free. An R and Python development environment includes a console, a syntax-highlighted editor that enables direct code execution, and tools for graphing, history, debugging, and workspace management. R Language has many packages, however, in this research, the Shiny package was selected to build an interactive predictive model and then visualise it as a DSS. Creating interactive web applications using R is relatively simple, thanks to Shiny, which features a large collection of pre-built widgets with automatic "reactive" binding between outputs and inputs. Shiny simplifies interactions between the user and the code to create dynamic web applications directly from R.

4.3.3.1 Data Preparation

One dataset used for this project was constructed from the KDD-CUP098 dataset, which contained the donor data for paralysed veterans of America. This dataset included socio-demographics, which should be gathered in order to establish a knowledge of features (Pallant, 2020). Among the variables, we were interested in 1) donor attributes, including the state of the donor location, gender, age, education, socio-economic status, and income level, and 2) donor donation history, including total donor amount and the total number of past donations. We also created a variable called (donated within last year) for whether a donor donated after February 1996. As the data ended in February 1997, we used this variable to describe whether a donor donated again after she/he had donated before.

Another dataset is the volunteer dataset which is part of a survey conducted in 2017 by the Bureau of the Census for Corporation for National and Community Service, Washington. Similarly to the donor dataset, We were particularly interested in two of the variables: 1) volunteer features, such as the state where the volunteer locations or state, their gender, age, level of education, socio-economic status, and income; and 2) volunteering history, such as the total hours of volunteering divided by and the total number of previous volunteering. Also, we generated a variable called (volunteered during last year) to determine whether the volunteer has volunteered recently or not.

As the dataset was constructed to demonstrate DSS, we consolidated the values of some variables so that each has a sufficient number of samples for modelling and analysis. The gender includes male and female. The education level is grouped into a bachelor's degree or above and below a bachelor's degree. Socio-economic status is divided into high, middle, and low. The income level has seven brackets to divide the incomes accordingly.

4.3.3.2 Missing Data and Outliers

Most importantly, the present study validates the data to check for missing values or outliers. Missing data can bring biases into the prediction model or loss of information when creating descriptive analysis. It may lead to inappropriate statistical methods, and difficult to apply them for useful analysis (Peng et al., 2002). To solve the problem of the missing data, we used imputation techniques to replace the missing data with estimated values. This technique helps estimate the missing data using the mean imputation of the valid values (Tabachnick et al., 2007).

On the other hand, outliers are observations with distinct characteristics that distinguish them from other observations (Hair et al., 2010). There are significant problems with outliers, such as error variance increase, which may affect the prediction models. According to (Kline and Dolamore, 2020), deleting outliers in a large-scale dataset should not be a big concern. Therefore, we decided not to remove any outliers as both donor and volunteer datasets were large-scale. In a later step of the analysis, we found only a few outliers, but they carried useful explanations. For example, there was an extreme outlier in the number of donations from donors who were in their elder years, compared to the average amount. The interpretation revealed several factors that may encourage people to donate such as retiring and having a high amount of money for religious and authentic purposes.

4.3.3.3 Selecting Packages: glmnet package and logistic regression model

Glmnet is a package that uses penalised maximum likelihood to fit generalised linear and related models. For the regularisation parameter lambda, the regularisation route is calculated for the lasso or elastic penalty at a grid of values (on the log scale).

Glmnet is a package that uses penalised maximum likelihood to fit generalised linear and related models. For the regularisation parameter lambda, the regularisation route is calculated for the lasso or elastic penalty at a grid of values (on the log scale). Regression ML algorithms are used to predict a target value based on subjective variables (Naser and Alavi, 2021). The aim of the logistic regression model is to understand a binary answer or dependent variables (Hilbe, 2009). Logistic regression is chosen for the probability model of donors and volunteers because it can predict a binary outcome. For example, when predicting a donor's probability, there is a binary answer, (yes: a donor aged over 80 is likely to donate) or (no: a donor aged under 70 is unlikely to donate).

Logistic regression assumes that

$$y = \frac{1}{1 + e^{-}(\beta_0 + \beta_{1x1} + \beta_{2x2 \cdot\cdot} + \beta_{nxn})} \tag{4.1}$$

(4.2)

$$-\left[\frac{1}{N}\sum_{i=1}^{N} y_i(\beta_0 + x^T\beta) - log(1 + e^{(\beta_0 + x_i^T\beta)}) + \right] + \lambda[(1-\alpha)||\beta||_2^2/2 + \alpha||\beta||_1]$$

where λ and α are two tuning parameters

Tidymodel workflow

We chose R tidymodel package to build the glmnet model, and followed the standard workflow. The workflow starts with pre-processing the input data and ends with selecting the best model. It allows easy model deployment and prediction. The workflow includes the following steps:

- Splitting data into train and test stratified on the outcome variable. The training data is used to train the model, and the test data is used to report the performance of the trained model.
- Data pre-processing involves (1) replacing rare elements in the categorical variable state with "other"; (2) creating dummy variables from categorical variables; (3) log transforming a skewed numeric variable number of past donations; and (4) normalising numerical variables.
- Validating the model (more detail in subsection 4.4.2).

4.3.3.4 Probability Model of Donors and Volunteers: Building the Interfaces

The Shiny package is used to develop a functional front-end, back-end, and web-based DSS using the Shiny library in R, involving it in a platform of AI and ML, which can analyse, deploy models, and visualise the analysis through a dashboard. Shiny R is one of the most effective and interactive tools that help data scientists build a web-based application (Malasi, 2021). Shiny app is built using a combination of HTML, CSS, and JavaScript. It utilises the R package

Shiny to communicate with the R environment. Users can interact with the app using their web browser, allowing them to filter data, make selections, and view graphs. Shiny apps are hosted on a web server and can be shared with anyone with an internet connection (Malasi, 2021). After loading the package by running the command library (Shiny) in the R console, users must create the UI and server components of the application. The UI component defines the layout and appearance of the app, while the server component defines the behaviour (Chang et al., 2020).

To create a Shiny app, a directory should be created, with an app.R script placed within. This script contains a user interface object, a server function, and a call to the shinyApp function. The user interface determines the appearance and layout of the app, the server function stores the instructions necessary to construct the app, and the shinyApp function creates the app object. To it, use the command runApp("directoryName").

In this model, we have created three main interfaces which are:

- **Predictive Analysis of Donors:** This interface creates a probability percentage of a donor who is likely to donate in the future or not, based on critical variables such as state, age, number of donations, social economy level, income level, gender, and education level. A system user can play around with these options to see if there is any likelihood of people donating based on the provided features.

- **Predictive Analysis of Volunteers:** This interface creates a probability percentage of a person who is likely to volunteer in

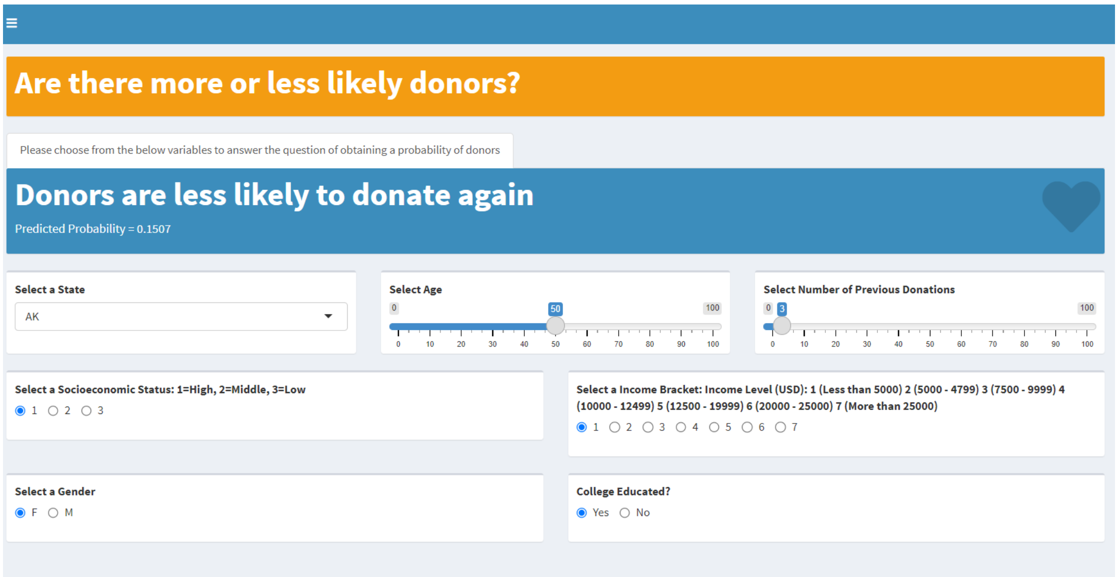


Figure 4.2: A Probability Model of Donors

the future or not based on critical variables such as state, age, number of donations, social economy level, income level, gender, and education level. These options can be manipulated by a system user to see if there is any likelihood of people volunteering based on the provided features.

- **About:** This page has essential information regarding the selected packages and usage of the model.

4.3.3.5 Probability Model of Donors and Volunteers: Operationality

The user of this model would basically have the answer to the question: Are donors/volunteers more or less likely to donate/volunteer? Then, several variables play critical roles in donor and volunteer intentions towards donations and volunteering.

1. State: Indicates whether the donor or volunteer lives in a state.

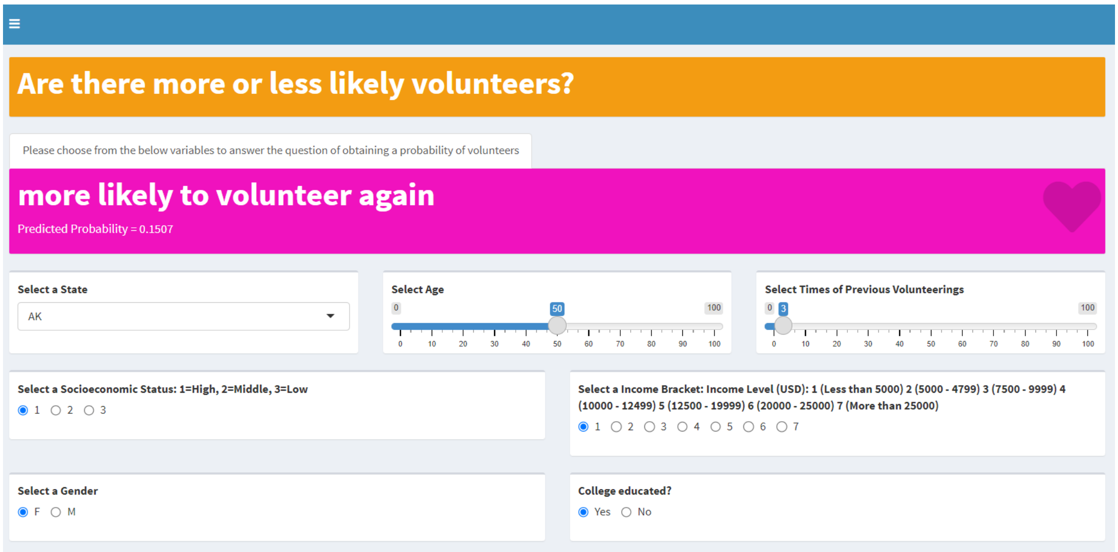


Figure 4.3: A Probability Model of Volunteers

2. Age: Overlay Age of a donor or a volunteer.

3. Several Previous Donations/volunteering: Indicating the number of donations or volunteering.

4. Socio-economic Status: Code indicating which cluster group the donor or the volunteer falls into. Indicating a cluster of socio-economic status, high, middle and low.

5. Income: Indicating the code of the income level: 1 (Less than \$5000) 2 (\$5000 - \$4799) 3 (\$7500 - \$9999) 4 (\$10000 - \$12499) 5 (\$12500 - \$19999) 6 (\$20000 - \$25000) 7 (More than \$25000).

6. Gender: Indicating whether the donor or the volunteer is a male or a female.

7. Level of Education: Indicating whether the donor or volunteer has a college certificate. This is also applied to whom obtained

higher education.

4.3.4 Dataiku Platform: Building the AI-enabled DSS

The dataset, which is a series of records, was imported into Dataiku to be prepared for processing. Processing the data includes creating samples, removing empty cells, and replacing values with other values. We visually and interactively developed data cleansing, normalisation, and enrichment scripts in DSS using visual data preparation. Then, a visual analysis can be created to the flow as a prepared recipe. The recipe is a function in Dataiku that provides cleaning and split data processing. The flow comprises both the datasets and the recipes, and the model of analysis can be created based on the cleaned or prepared data. Users can easily comprehend a data pipeline's flow. For the AI-enabled DSS, we created two flows of donors and volunteers which have recipes and prepared datasets, ML models, codes, and plugins (Dataiku.com, 2022b). An example of the flow of the AI-enabled DSS is shown in Figures 5.3 and 5.4.

4.3.4.1 Features Handling

Feature selection is a fundamental concept in ML that significantly impacts the model's performance. Both datasets of donors and volunteers have provided useful hints on selecting the features that have the potential to predict donating and volunteering-related matters. The activities of the feature selection process were done using the interactive approach of Dataiku for the descriptive analysis and predictive analysis. For instance, Dataiku DSS allows for specifying the variables or features and offers several activities of feature selection,

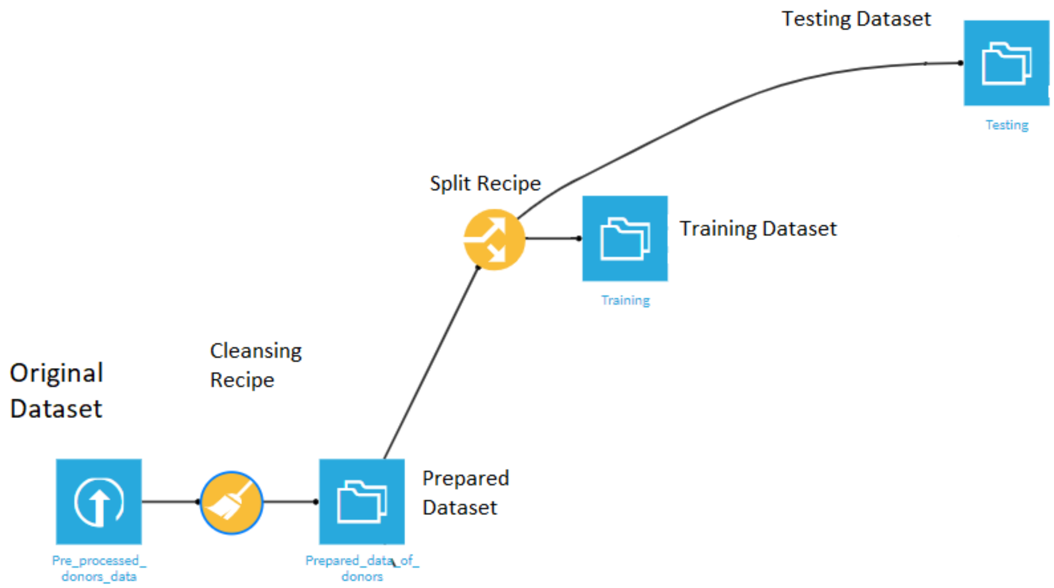


Figure 4.4: A Flow of Preparing Donors' Dataset

Figure 4.5: A Flow of Preparing Volunteers' Dataset

such as categorical variables and numerical variables. A categorical variable's treatment is defined by category handling, missing values, and associated controls. The missing values and numerical handling methods, as well as their associated controls, define how a numerical variable is handled (Dataiku.com, 2022a). Following the completion of the required data processing steps, descriptive analysis and predictive analysis were created to provide answers to the RQs.

4.3.4.2 Descriptive Analysis

Dataiku DSS provides an interactive statistics worksheet for performing an exploration of data analysis ((Dataiku.com, 2022a). We used this feature to create two descriptive analysis slides of donors and volunteers; both slides summarise data samples or features, draw conclusions from the sample data, and visualise the structure of the dataset. The types of analysis provided vary, including charts, correlation matrices, distribution and curve fitting, bivariate analysis (as shown in Figure 4.6), and univariate analysis (Dataiku.com, 2022a). To add a descriptive analysis of Dataiku, the "visual analysis" tab is used to explore and visualize the data and create descriptive plots. In addition, the "statistics" tab is used to generate a summary of the data, including measures of central tendency and dispersion (Dataiku.com, 2022a). Examples from the descriptive analysis of donors are shown in Figures 4.6 and 4.7, which show various graphs representing important statistics about donor behaviour. A tooltip is provided to help users understand the graphs clearly.

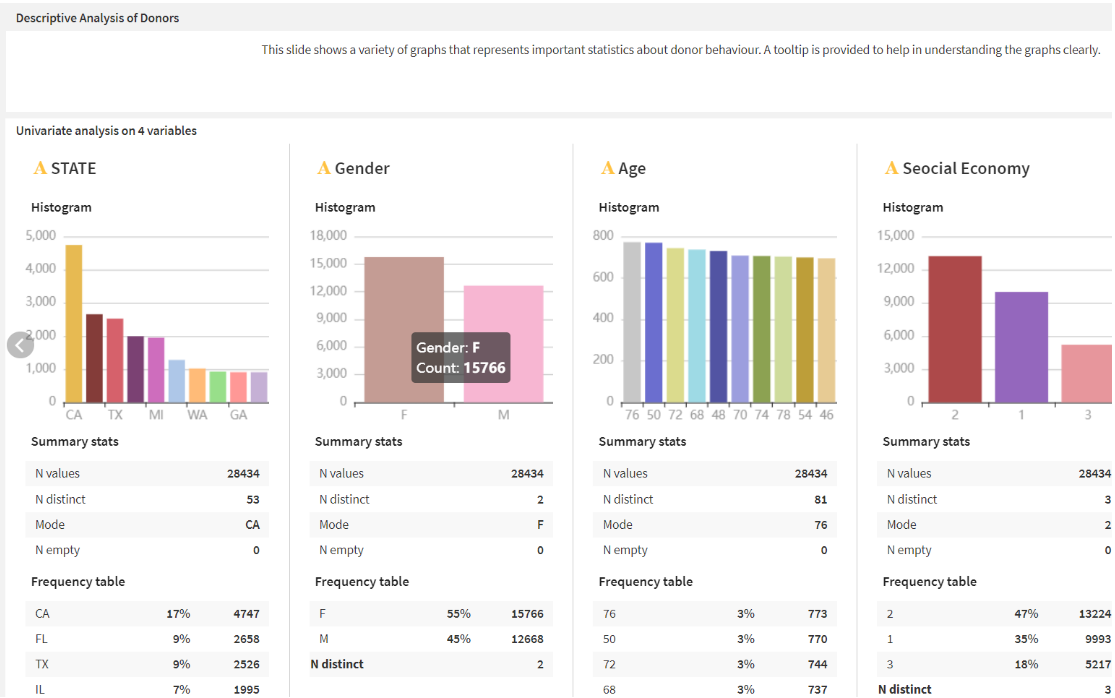


Figure 4.6: An Example of Age and Number of Donations Aggregation, Using Hexagon

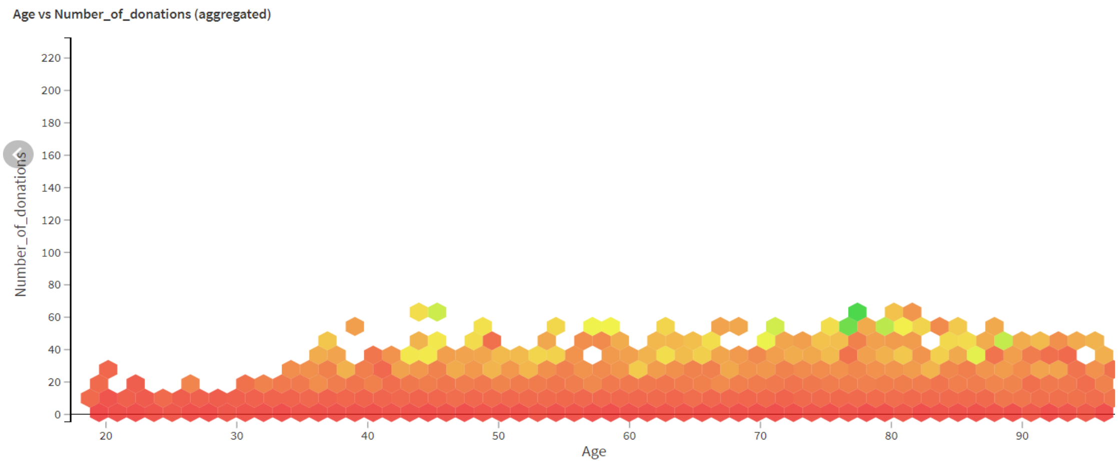


Figure 4.7: An example of Univariate Analysis from the AI-enabled DSS

4.3.4.3 Predictive Analysis

When a target variable needs to be forecasted, prediction is used. For instance, the decision maker might want to predict the number of donations or volunteer hours, using the total amount in dollars of donations, age, and education level. The amount of donations in this instance is the target, and the features utilised to make a prediction are the total amount of donations, age, and educational level.
Creating a predictive model on Dataiku requires several steps. First, the data must be imported and prepared by loading it into the platform, cleansing it, and combining it with other datasets if necessary. After the data is ready, an appropriate algorithm must be chosen to create the model. The parameters must be specified, such as the training and test data, the number of iterations, and the type of regularization (Dataiku.com, 2022a)). Once the model is run, the results can be evaluated with various tools, including visual inspections, accuracy scores, and feature importance plots (Dataiku.com, 2022a). Predictive analyses vary in their purposes. However, to narrow the scope of this study, random forest was selected because Dataiku DSS provided it and can prove the concept for this research.

Random Forest

Random forest is a computationally efficient ML technique that can work with large datasets (Oshiro et al., 2012). An example of an ensemble approach, random forest builds a number of decision trees that are used to classify a new instance based on popular vote. Each decision tree node employs a subset of attributes chosen randomly from the entire original collection of attributes. In addition, each tree, like bagging, employs a separate bootstrap sample data (Oshiro et al., 2012). Random forest can be defined formally as:

$$h_k, (X, T_k), k = 1, 2, ..., L, \tag{4.3}$$

where T_k are independent distributed random samples, and for the most well-liked class at input x, each tree casts a unit vote (Oshiro et al., 2012).

In general, random forests generate effective outcomes at the expense of the model's "explainability." (Dataiku.com, 2022a). In Dataiku DSS, when training the forest, a random sample of the training set is used for each tree, and a random subset of the input features is examined for each decision point in the tree (Dataiku.com, 2022a). Using the (prepared data of donors) dataset, we split the data into two sets (Training= 80%, and Testing =20%). After that, a random forest model flow is created to create predictions of donors (Figure 4.8). The same procedure was applied to volunteers' dataset datasets (Figure 4.9).

After creating the random models, a number of charts were deployed into the dashboard of the Dataiku to sort the slides of the AI-enabled DSS. The charts are lines, scatter plots, and curves, as shown in Figure 4.10 and Figure 4.11. Charts are graphical representations of data that allow us to observe the connections between various pieces of information and help us understand the analysis obtained from the model.

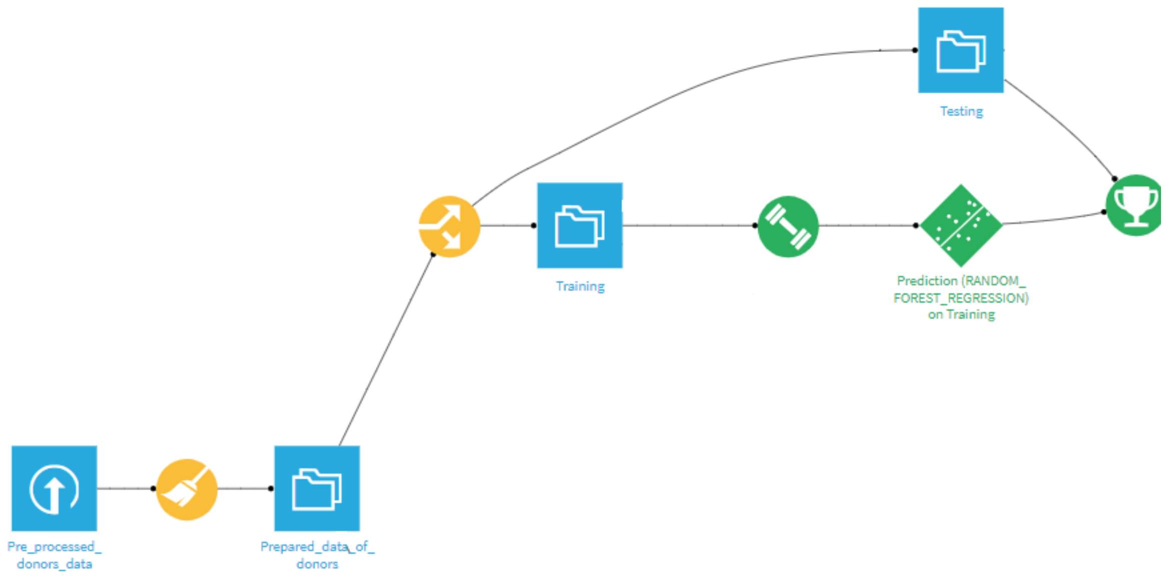


Figure 4.8: A Flow of Donors' Predictive Model Generation

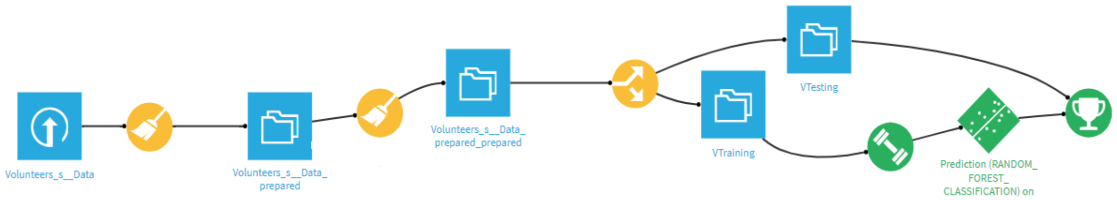


Figure 4.9: A Flow of Volunteers' Predictive Model Generation

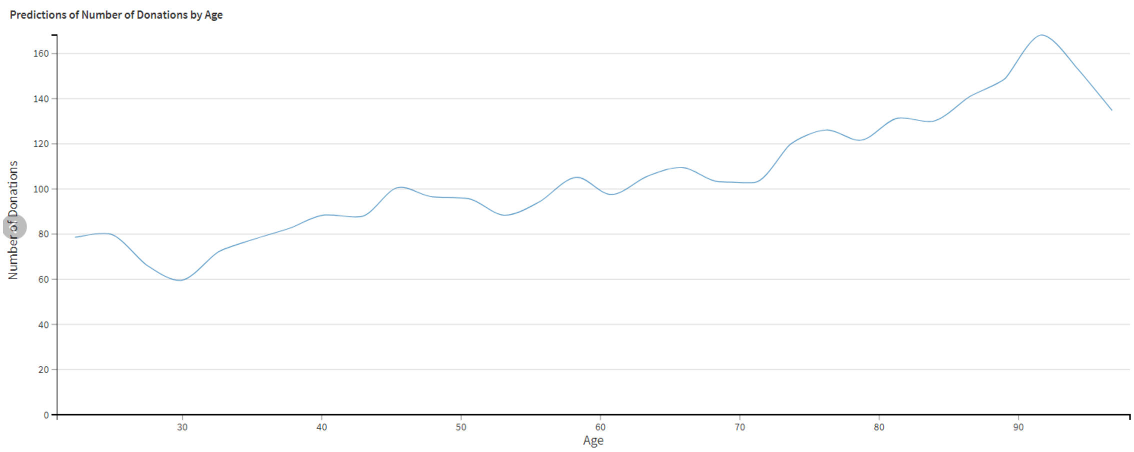


Figure 4.10: An Example of Predicting the Number of Donations by Age

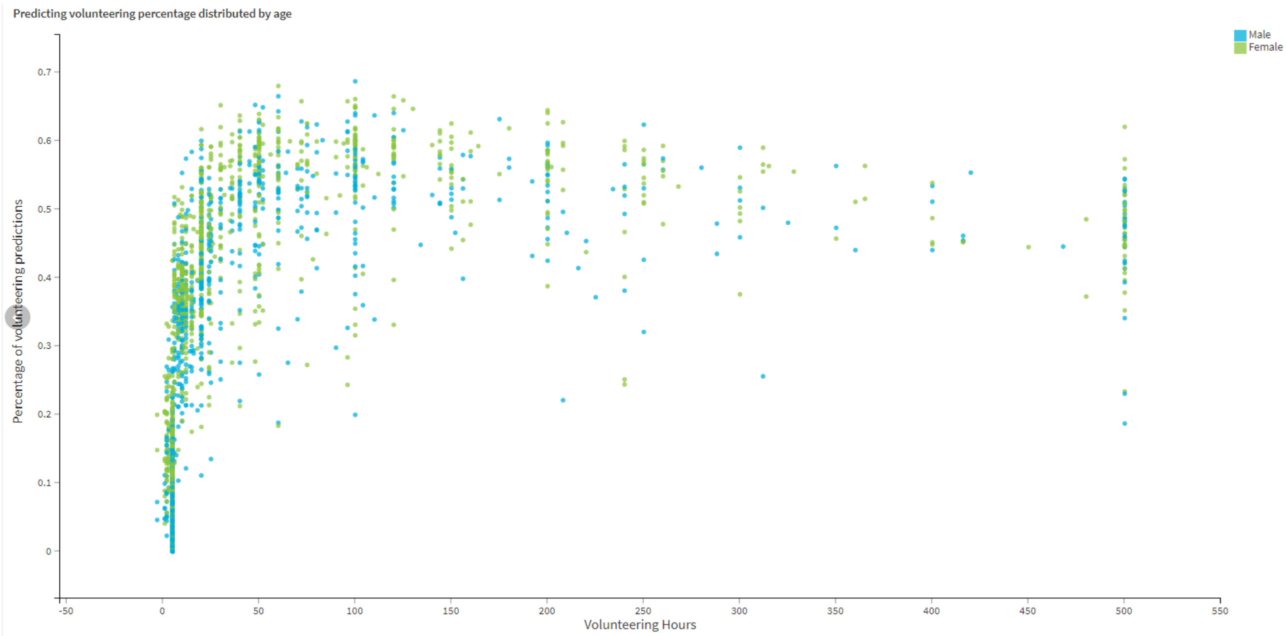


Figure 4.11: An Example of the Percentage of the Probability of Volunteering

4.3.5 Dataiku DSS Dashboard

Dashboards in Dataiku provide sharing of different elements of the created projects such as charts plots, models information, insights of models, and wiki articles. (Dataiku.com, 2022a). Dashboards in Dataiku share different elements of the created projects such as charts plots, models information, insights of models, and wiki articles. (Dataiku.com, 2022a). Dashboards can also be managed in terms of permission: who can access, edit, and read the contents, which allows an efficient workplace of a group who share the work analysis on Dataiku. Therefore, a decision-making team in an NPO can create a similar dashboard which allows them to manage, perform. and report the preferred analysis, and make better decisions. Interestingly, a project designer can export the selected elements to the dashboard and organise them in DSS. Titles, text boxes, and different options for changing colours and positions, and adding a frame can be added to the dashboard. For this research, we created the main dashboard and named it “AI-enabled DSS”. Then, we deployed the selected charts and plots in the dashboard [3] [4]. After that, we split the dashboard into seven slides, and each slide presents the relevant analysis of donors and volunteers [5].

1. **Introduction**: This page is a wiki article which simply has some information about the contents of the dashboard and its

[3] A demo is shown on: https://drive.google.com/file/d/1RlGZIMd3YALGF71e2iUlqDnsayDmAKCu/view

[4] Due to the multiple slides of the DSS dashboard, and for an overview of the dashboard, we provide investigators and interested scholars reading access to the AI-enabled DSS through the below link: `sydneyuniversityoftech.academics.dataiku-dss.io/projects/DONORSANALYSISTRIAL/`

[5] A user manual is attached in Appendix E

main aims.

2. **Descriptive Analysis of Donors**: Using the random forest model, different prediction plots have been created to give an overview of which donor features can be estimated, such as total donations, who will donate based on age, and who will donate more in the future. The contents of these slides help answer RQ.2.2.

3. **Predictive Analysis of Donors**: Using the random forest model, different prediction plots have been created to give an overview of what some features of donors can be estimated, such as total donations and what ages will donate, more in future. The contents of these slides help answer RQ.2.2.

4. **Descriptive Analysis of Volunteers**: The slide contains some statistics and data sample visualisations. Most importantly, we concentrate on the key traits that aid in the prediction and description of the features of donors and volunteers. The contents of these slides help answer the RQ.2.1.

5. **Predictive Analysis of Volunteers**: Using the random forest model, multiple prediction plots were generated to provide an overview of some contributor properties, such as total donations and the donor ages that will likely donate more in the future, which can be anticipated. The contents of these slides help answer RQ.2.2.

6. **About**: The slide shows the biographies of the team project.

4.4 Iteration Two: Validation Techniques

Verifying an ML model's validity is crucial for ensuring the model's dependability. It also is critical to comprehend a model and get familiar with its advantages and disadvantages. Understanding your model through and out will make it easier for you to analyse and spot mistakes in future output. Understanding how your model operates will also make it easier for you to see any confounding variables or drift.

4.4.1 k-fold Cross-Validation

K-fold cross-validation is derived from dividing the dataset into a k number of folds, which assesses the model's performance when presented with fresh data. K is the number of distinct groups into which the data sample is divided (Wong, 2015). The goal of the k-fold cross-validation approach is to minimise overfitting as the data is separated into folds 1, 2, 3, and 4. The first fold is a test set, and the following three folds are contained in the training data so we could use them for training our model. The first fold forms the test fold, the second fold likewise forms the test set and the other fold forms training data in the second instance. This suggests that each dataset and each fold were initially used for training purposes before gradually changing to test purposes. With the overfitting reduced, we can now determine the average variance.

In applied ML, cross-validation is commonly used to gauge how well an ML model performs on untrained data; that is, to employ a small data sample to assess how well the model will perform relatively

when employed to generate predictions on data that was not utilised during the model's training. It is a well-liked technique since it is easy to comprehend and typically yields a less-skewed or extremely naive assessment of the model performance compared to other techniques, including a straightforward train/test split (Wong, 2015).
The benefit of the k-fold cross approach is that the manner in which the data are separated is less important. Each dataset appears exactly once in the test set and k occurrences in the training set. The resultant estimate's variance is decreased as the ask is raised. The drawback of this approach is that the evaluation process requires k times as much computing because the training algorithm must be run k times from the beginning; this method can be modified by randomly dividing the data k times into a testing and training set (Jiang and Wang, 2017).

4.4.2 Validating the Probability Model of Donors and Volunteers

The training subsets and assessment subsets produced in k-fold cross-validation are demonstrated in the illustration below.

1. Tuning the model on a grid of λ (0, 0.01, 0.1, 0.05) and α (0, 0.5, 1) using 10-fold cross-validation. The best model parameters are selected which has the highest area under the curve. Report the performance of the best model on the test data.
   For example, when having 10000 samples, which were divided into 10 groups, each had 1000 samples. For given λ and α, we built a model using nine groups, groups 2-10. To check how good

this model is, we applied this model to the remaining group 1 and saw how good the prediction is. It is to prevent overfitting as group 1 is unknown to the trained model.

Next, we repeated the process but build a model using groups 1 and 3-10 to check the quality using group 2. By repeating these steps, we will create 10 models and know the quality of each model. The average quality of all models is the quality of models created with that given tuning parameter (or hyperparameter) λ and α.

By repeating the above for different λ and α, we got different models. We picked the pair of λ and α that generate the models with the best average prediction quality. Note that a model is described by tuning parameters λ and α, as well as model parameters. The model parameters if you know linear regression, are b0, b1, b2, .. in y = b0 + b1*x1 + b2 * x2 + ...

2. Training the final model using the best parameters on the whole dataset. The final model is saved for deployment. The above model tuning is to find the best λ and α. We still need model parameters. In the tuning, all the model parameters are obtained using 9 of 10 groups. In the last step, we trained the model using all ten groups to get the model parameters to avoid any data. λ and α are both zero after the tuning.

4.4.3 Validating the models in Dataiku

An ML expert uses the hyperparameters of an algorithm as levers to regulate how a model is educated by the algorithm. When training a decision tree, one of these settings, or hyperparameters, is called max depth. The max depth hyperparameter regulates how far into the future the model can go (Dataiku.com, 2022a). The ML practitioner is responsible for selecting the best hyperparameters to guide the algorithm. Determining values for several hyperparameters that an algorithm may have is a necessity. A grid search will thoroughly examine the values of all possible hyperparameter combinations.

The initial stage in model validation is to divide the dataset with known results into a train and a test set (Dataiku.com, 2022a). With the two datasets of donors and volunteers, in each flow, each dataset was prepared and split into training (80%) and testing (20%). In Dataiku, we divide our training set into "k" portions or folds, where "k" indicates a number, such as three. K-fold works by shuffling each fold so that each fold has an equal probability of being in both the training and validation sets (Dataiku.com, 2022a). Then, the model is trained on the training folds, tested, and the error is calculated on the validation fold. The folds are then shuffled in round-robin form until the error on all k-folds has been determined (Dataiku.com, 2022a). The cross-validated error for each set of hyperparameters is the average of these errors. Then, after training the model on the entire training set and computing the error on the test set (the ones that have not been trained yet), we select the combination of hyperparameters with the best cross-validated error. These test errors can then be used to compare against different algorithms.

For the predictive analysis in Dataiku using the random forest model, the mode of hyperparameters was a 5-fold cross-test for both donor and volunteer datasets (Figure 4.12). The validation process on Dataiku is done automatically in the user's back end, using the policy (split the dataset). In the example shown in Figure 4.12, the first iteration uses the first 20% of the total data for evaluation, the remaining 80% for training between 1 and 5 for testing, and between 5 and 25 for training. The second iteration uses the second 20% of the data for evaluation, the remaining 3 sets of data ([5-10] testing and [1-5 and 10-25] training), and so forth.

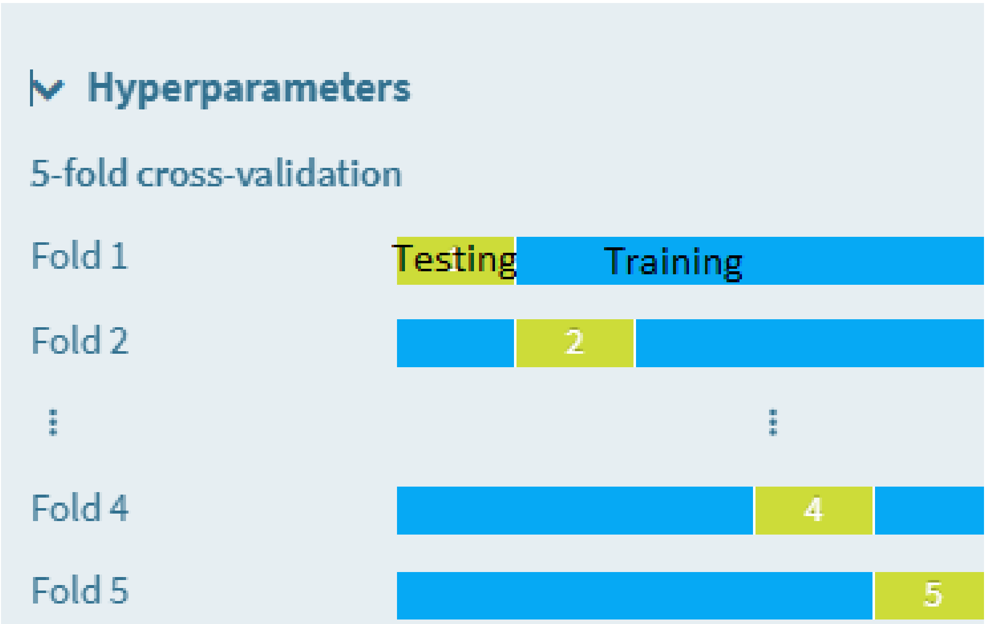


Figure 4.12: An Example of the Percentage of the Probability of Volunteering

The average of the numbers calculated in the iteration is the performance indicator provided by k-fold cross-validation. Although this method can be computationally costly, it does not discard a lot of data (unlike establishing an arbitrary validation set), which is a significant benefit in applications like inverse inference, where there are few samples.

However, the problem of overfitting may occur during the training of the data. Overfitting occurs when the model fits the training data too well (Dataiku.com, 2022a). It means the model has focused on memorising the dataset, rather than learning the input and output relationship. Thus, the resulting predictions of the model may be useless. During the process of creating predictive analysis in the AI-enabled DSS, we did not face any overfitting problems because we trained the datasets more than once, selected the best features that we expected to provide strong predictions, and used a cross-validation strategy.

4.5 Iteration Three: Evaluating the AI-enabled DSS

The use of summative evaluation ensures the success of AI-enabled DSS functionalities and meets the DRs in the conceptual design in iteration one. Summative evaluation is a process of gathering, combining, and interpreting data to decide on an artefact or a product (Dick, 1977). A functional front-end, back-end, and web-based DSS is developed using the Shiny library in R and Dataiku to present a predictive and descriptive analysis of donors and volunteers. For the iteration three evaluation, semi-structured interviews were conducted with NPO decision-makers, data scientists, volunteers, systems designers and analysts, experts in NPOs, and managers of NPOs. Interviews, one of the qualitative research methods, are frequently concerned with obtaining a thorough grasp of a situation or determining the specific phenomenon (Dworkin, 2012). These iteration results led to the application of any changes or suggestions on the outcomes and functionalities of the AI-enabled DSS.

4.5.1 Iteration Three: Data Collection

Iteration three was conducted using semi-structured interviews with a total of 22 interviewees whose professions included system designers, social experts, academics, and NPOs managers; all had different experiences with ML, DSS, and data analysis. The aim of this iteration was to provide feedback on the successful implementation of the design method and to support the initial presentation of "proof-of-concept" (Peffers et al., 2003; Peffers and Tuunanen, 2005).

In qualitative research methods, the sample number of interviews varies depending on the number of questions and the research objectives. Compared to quantitative research methods, the sample size in qualitative research methods is frequently used (Weston et al., 2001) because acquiring a complete understanding of a phenomenon or identifying its significance is typically the focus of qualitative research methods (Dworkin, 2012). Due to the range of experiences sought in this iteration, 22 interviewees were chosen to participate. This number of experts was chosen because of their in-depth knowledge of the research problem and their availability to conduct the interviews within a specific time frame of the study.

Each interviewee was invited via email, which included a consent form to sign and a brief introduction about the research problem and proposed solution. Participants agreed to be recorded, which would help analyse the interviews. Before meeting with each interviewee, the link to the AI-enabled DSS was to them sent for exploration. Furthermore, during each interview, a brief is introduced to the interviewees to give an overview of the research problem, research objectives, conceptual design, expected research output, and interview process. The questions are designed to measure each criterion and ensure that the relevant items of a particular design are met. The aim of iteration three is to observe and measure how well the artefact provides solutions to the defined problem (Peffers et al., 2007). Most importantly, comparing the objectives of the solutions (DRs) to the actual solutions or the developed artefact is important during the evaluation (Peffers et al., 2007). Therefore, the interview questions are distributed in five design categories to represent the DRs of the conceptual design in iteration one. In other words, we designed the interview questions in iteration three to validate the objectives of the solutions (DRs) that were stated in iteration one's evaluation.

4.5.2 Iteration Three: Data Analysis

In an approach similar to iteration one, iteration three used two qualitative data analysis strategies: coding and categorising. Four categories are provided in this paper to report the analysis results. Each category has different codes, which are discussed in detail. As a result, some codes from various categories are linked to useful information. The software used in the analysis is Microsoft Power BI. Microsoft Power BI is an acronym for Business Intelligence (BI). The majority of profitable and agile businesses now use AI to help them make decisions using its provided capabilities. The visual representation of the dashboard allows users to quickly see the relevant data for the decisions they must make. Power BI is user-friendly and has several applications.

4.5.2.1 Work Experience

This category grouped the roles and years of relevant experience held by interviewees, which helped link to additional categories in the interview analysis. The Work Experience category is essential, as it defines the required changes to the artefact. Table 4.5 shows the variety of work experience codes for whom were interviewed. The majority of interviewees are experts in ML and NPO management. We relied on academics from the universities, NPO consultants, and data scientists, who have been involved in similar projects or events at NPOs.

We discovered that the majority of experts have worked or participated in some NPO events such as at charities, religious centres, and youth centres. Consequently, data scientists with limited volunteer experience in NPOs presented relative answers during the interviews. Four NPO managers responded to our questions by expressing their impression about the capabilities of the AI-enabled DSS for analysing donor behaviour. ML and data science experts answered the interview questions with valuable answers and critical points on the functionality, operationality, and effectiveness of the AI-enabled DSS. The range of specialists we interviewed assisted us by providing useful information for evaluating the design and the implementation of the AI-enabled DSS, which assists in achieving our DRs.

Work Experience Code	Total of Experts
ML	4
Manager of NPO	4
Consultant of NPOs	3
University Academic	3
Data Scientist	2
Usability expert	2
DSS and Visualisations	1
Managing Donations Events	1
Managing Volunteering Events	1
Social Expert	1

Table 4.5: Category of Work Experience

4.5.2.2 Usability

This category grouped similar answers to questions related to the usability design requirement and linked them to the category of Work Experience. Table 4.6 presents an association of Work Experience codes and Usability codes. Most experts described the usage of AI-enabled interfaces as "easy". They explained the structure of the analysis, and how they organised it, applying the provided information to help make it easy to use. However, one NPO manager and an academic described it as only partially easy to use for two reasons. The first reason was the high number of plots and slides, where narrowing the number of analyses is suggested. The second reason was that not enough information was introduced to explain how to use the AI-enabled DSS. One university academic suggested building step-by-step guidelines, which could assist first-time users in making better decisions. Generally, the experts' answers, based on their various experiences, provided us with critical feedback on the AI-enabled DSS's ease of use.

Number of Experts	**Work Experience Code**	**How easy to use the AI-enabled DSS?**
4	ML	Easy to use
2	Consultant of NPOs	
2	University Academic	
2	Usability expert	
1	Data Scientist	
1	DSS and Visualisations	
1	Manager of NPO	
1	Managing Donations in NPOs	
1	Managing Volunteering events in NPOs	
1	Social Expert	
1	Consultant of NPOs	
1	Data Scientist	
3	Manager of NPO	Partially easy to use
1	University Academic	

Table 4.6: Codes of Work Experience Linked with Usability's Codes

In terms of navigation, 17 experts described moving between the slides of the AI-enabled DSS as “easy” (Figure 4.13). Similarly, four experts mentioned that the slides of the DSS had good navigation, but felt a slight modification was needed on the menu’s position on the slides. One expert described the navigation as “hard to navigate” because of the numerous slides. However, for the purpose of this research, the variety of different slides represents unique research objectives.

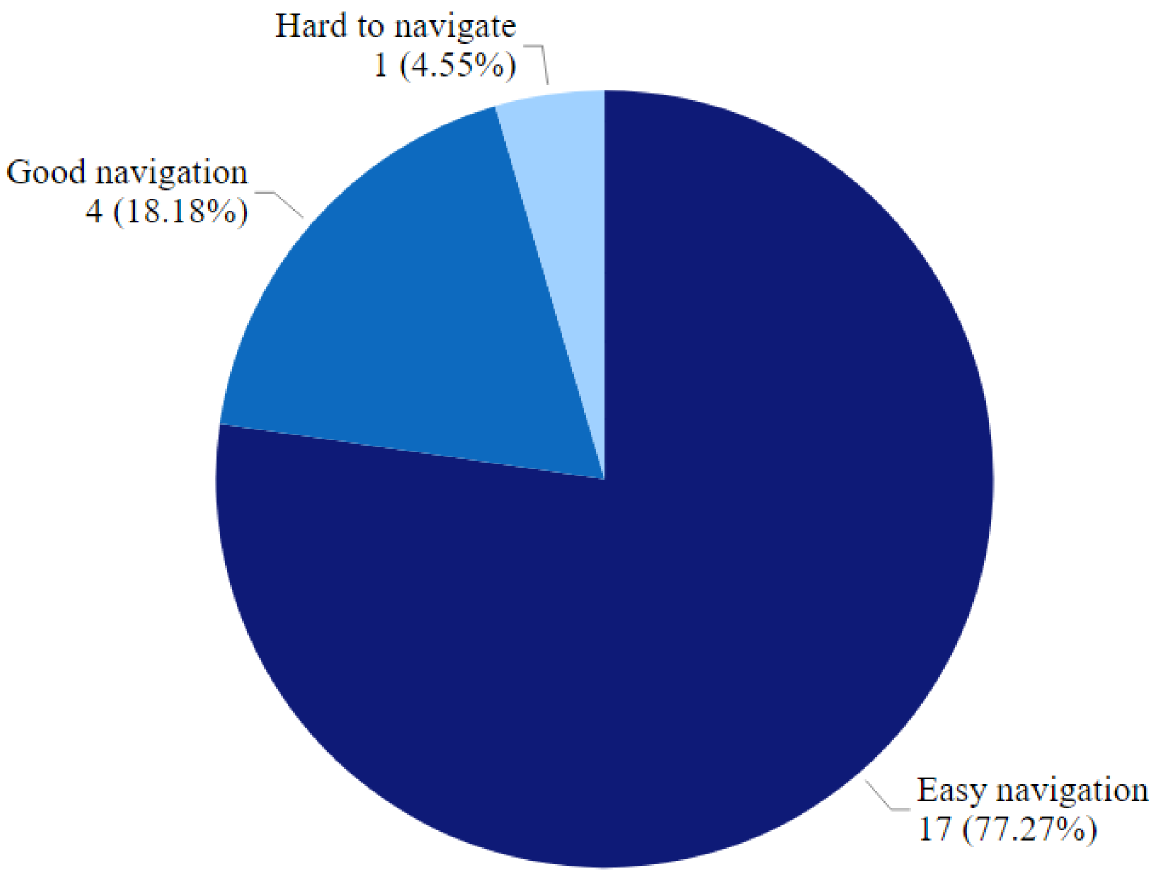


Figure 4.13: Navigation Measurement of the AI-enabled DSS

Interestingly, most of the experts were engaged with the AI-enabled DSS while using it, according to their valued answers (Figure 4.14). Most importantly, the variety of plots, visualisations, and the organisations which provided information influenced expert engagement. While experts vary in their experience, finding that the AI-enabled DSS engaged the experts is an indicator of achieving one element of usability, which is user engagement.

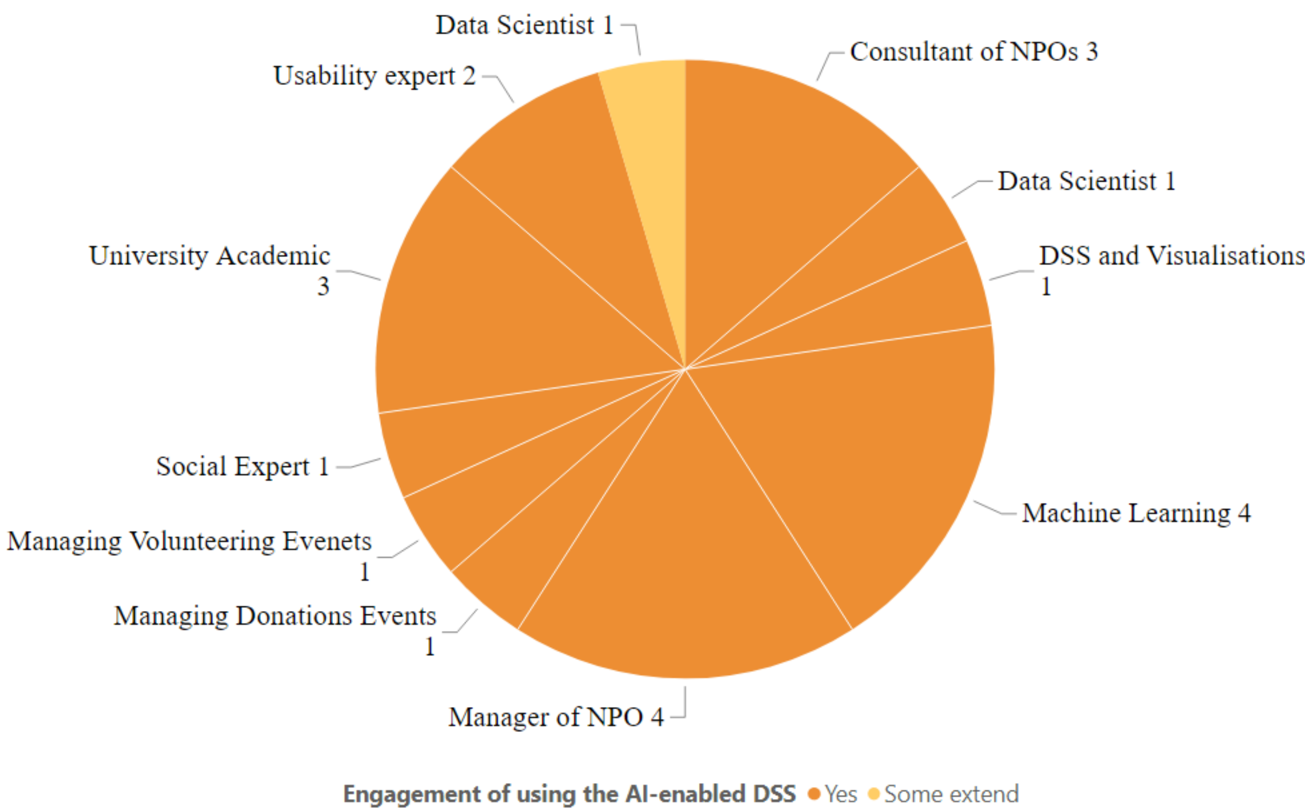


Figure 4.14: Codes of Work Experience Associated with Codes of Experts' Engagement of Using the AI-enabled DSS

Experts were asked for suggestions that could improve the usability of the AI-enabled DSS. As shown in Table 4.7, five suggested improvement codes that are shared with different work experience codes. The most suggested code, by three different experts, involved adding details about the figures, including a definition, the type, and the aim of the aim. One data scientist said, "Users with a lack of technical experience should obtain the information easily, to make a better decision".

Work Experience Codes	Suggested Improvements Codes
Data Scientist	Addign details about the figures
Machine Learning	
University Academic	
DSS and Visualisations	
Consultant of NPOs	Adding a user guideline
Usability expert	
Social Expert	Adding different analysis of donors
University Academic	Adding the blood type
Other experts	No additional requirements

Table 4.7: Suggested Improvements on the AI-enabled DSS

The second most suggested improvement was to create a useful guideline to help new users explore the functionality and features of the AI-enabled DSS. Following that was a suggestion for adding more analysis, suggested by a social expert. However, this suggestion is meant to add more figures and plots, which may cause overlapping with other visualised data features. Interestingly, one university academic suggested adding some analysis about blood donations, which is out of this research scope. The other experts did not suggest any additional requirements as they claimed that the current usability features meet the usability design requirement.

4.5.2.3 Effectiveness

Effectiveness assessments should concentrate on the most important issues, such as threats and opportunities affecting or potentially affecting the achievement of the system's objectives (Hockings et al., 2000). Thus, to assess the effectiveness of AI-enabled DSS, experts were asked questions related to any discovered issues, the outcomes of the AI-enabled DSS, and how the objectives of the AI-enabled DSS were achieved. Therefore, the category of effectiveness comes from grouping similar answers from experts who evaluated the effectiveness in accordance with the questions during the interviews.
Figure 4.15 shows the three codes of the effectiveness category, which are: large extent, some extent, and not sure. Most importantly, the codes of Work Experience were linked accordingly with the codes of effectiveness to provide useful meanings. Surprisingly, 18 experts found the AI-enabled DSS was effective to a large extent, due to the variety of involved ML and data analysis techniques. There were also three more experts who thought the AI-enabled DSS was effective to some extent, whereas one expert was not sure about the effectiveness of the AI-enabled DSS.

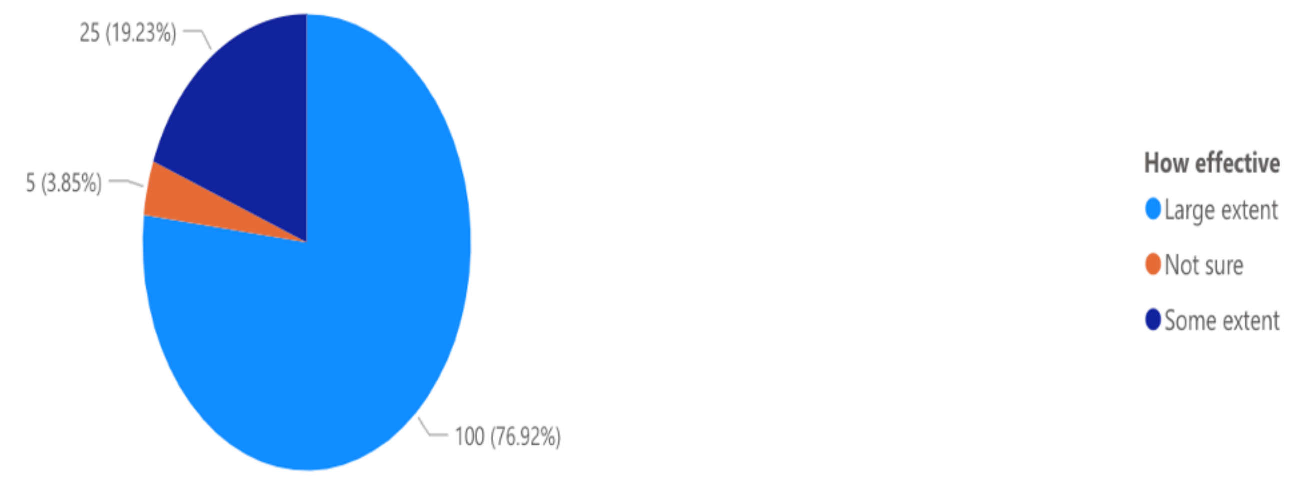


Figure 4.15: Effectiveness Measurement of the AI-enabled DSS

The AI-enabled DSS outcomes and the achievements of the objectives are presented in Table 4.8; the codes for objective achievement are: excellent achieved, very well achieved, and well achieved. The codes of outcomes for assessment achievement have been associated with the codes of Work Experience for obtaining insightful information. The majority of experts stated that the outcomes of the AI-enabled DSS are "very well achieved", or "well achieved", followed by four managers of NPOs who described the outcomes as "excellent achieved".

Work Experience Codes	**Total of Experts**	**The outcomes of AI-enabled DSS**
Manager of NPO	4	Excellent achieved
Consultant of NPOs	2	Very well achieved
Usability expert	1	
ML	3	
University Academic	3	
Consultant of NPOs	1	
Data Scientist	1	
DSS and Visualisations	1	
Managing Donations Events	1	
Social Expert	1	
Usability expert	1	
ML	1	Well achieved
Managing Volunteering Events	1	
Data Scientist	1	

Table 4.8: Codes of Work Experience and Codes of the Outcomes of the AI-enabled DSs

System effectiveness measures a system's ability to meet a set of specific mission requirements. However, measuring effectiveness may vary in the methods and techniques, whereas this research only focused on experts' answers towards the ability of a system to meet

DRs, find issues, and how well the objectives of the AI-enabled DSS were achieved via the outcomes.

4.5.2.4 Reducing effort

The reducing effort category combines codes of expert answers. By providing decision support, users can reduce the cognitive effort of making decisions (Meth et al., 2015). Table 4.9 shows Codes of Work Experience linked to Codes of Reducing the Effort of Making Decisions. Experts were asked two questions to assess their decision-making ability and obtain useful information about donors and behaviours. Hence, all experts were able to make a decision about donors and volunteers, where the examples vary between making decisions on donor frequency and some donor behaviour. For example, four ML experts were able to make decisions based on the ages when people donate most often, and the time and frequency at which people volunteer. Similarly, four managers of different NPOs used the predictions of donors and donations to decide which specific donors to appeal to.

Work Experience Codes	Total of Experts	Ability to make decision	Examples of decisions
Consultant of NPOs	1	Yes	Understanding donor and volunteer frequency. Experts were able to understand donor and volunteer frequency in a short time. Thus, their effort was reduced while making a decision.
Data Scientist	1		
DSS and Visualisations	1		
ML	4		
Managing Donations Events	1		
Managing Volunteering Events	1		
Social Expert	1		
University Academic	3		
Usability expert	2		
Data Scientist	1		Predicting donors and volunteers. Experts were able to understand some predictions of donors and volunteers in a short period of time. Therefore, their efforts were reduced.
Consultant of NPOs	2		
Manager of NPO	4		

Table 4.9: Codes of Work Experience and Reducing Effort Codes

4.5.2.5 Flexibility and Control

Providing flexibility to give users control is one important DR of an effective control system. Flexibility is an important attribute of AI-enabled systems, allowing them to adapt to changing conditions, take advantage of data, and respond rapidly to unpredictable scenarios. These systems are also capable of learning and adjusting their behaviour to achieve optimal outcomes (Walters, 2020). Furthermore, AI-enabled systems can carry out multiple tasks simultaneously and be used to automate tedious procedures (Yadav and Shukla, 2016). This flexibility of AI-enabled systems can lead to faster decision-making, improved resource utilisation, and enhanced customer service (Mohanty and Pradhan, 2019).

The flexibility of the AI-enabled DSS allows a manager or decision-maker who uses the system to apply changes. This means the control system itself must be adaptable to change. Another aim of the AI-enabled DSS's flexibility is to allow the users to move between the slides and explore the variety of details in the analysis. When explaining the flexibility and control of the AI-enabled DSS, 19 experts noted that the AI-enabled DSS is flexible for any applied changes or modifications, and can be controlled. In other words, the DSS is required to minimise system restrictiveness by enabling users to have control of the selection (Meth et al., 2015). System restrictiveness means a DSS pre-selected the decision strategies, which means users cannot have a variety of decision strategies (Silver, 1988).

Interestingly, most experts, while exploring the dashboard slides, found that the AI-enabled DSS was flexible and provided control (as shown in Figure 4.16). Meanwhile, three experts were unsure about the flexibility, as they were not allowed to edit the contents of the AI-enabled DSS accordingly. However, we clarified that the AI-enabled DSS was under evaluation, which means that any changes in the contents by experts may cause some errors.

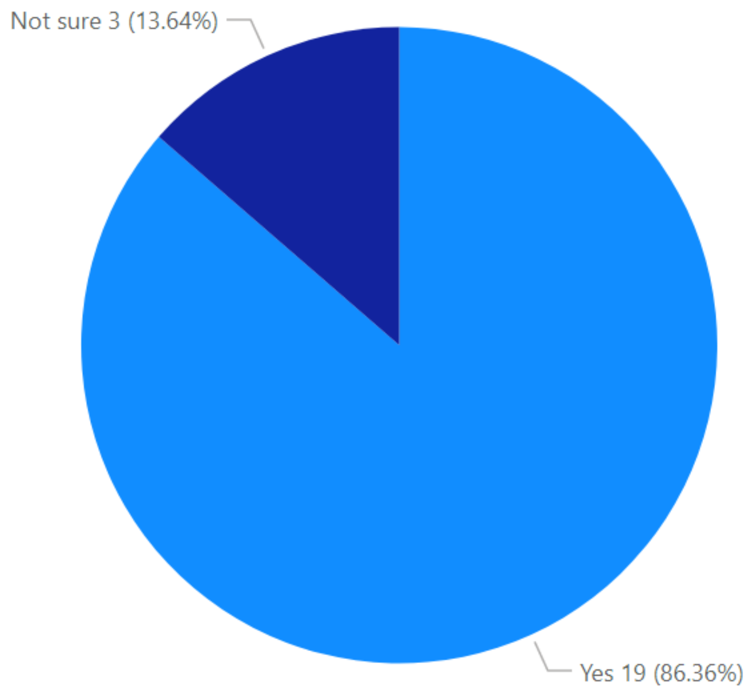


Figure 4.16: Flexibility and Control

4.6 Chapter Summary

This chapter described how data is extracted and prepared, along with methods of deployment for data modelling and validation techniques. It explained the techniques used to analyse the given data. and explained how the AI-enabled DSS dashboard is developed using the platform of R and Dataiku. In simple means, this chapter presented an analysis of the gathered data, including descriptive statistics, data analysis, and initial results in each iteration. This chapter represented the main work and steps of the DSR process model (Phase 3: Design and Development, Phase 4: Demonstration, and Phase 5: Evaluation).

CHAPTER 5

Research Results and Discussion

5.1 Introduction

This chapter presents the results of the three iterations and the outcomes of the DSR process model framework of Alsolbi et al. (2022a). It is divided into four main sections: the results of each of the three iterations, and the design theory. Following that, the responses to each research question meet the research aims and void the RGs. Iterations one and three took similar approaches (interviews), and the methods that were applied. However, the results varied due to the nature of the interview questions. Meanwhile, iteration two provided results of the models applied in the AI-enabled DSS, and relied on some performance metrics. This chapter presents the metrics' findings for validating the prediction algorithms' performance. Meanwhile, iteration two provides results of the models applied in the AI-enabled DSS, relying on some performance metrics. The metrics' findings for validating the prediction algorithms' performance are presented in this chapter.

The design theory was introduced after finalising the outcomes of the three iterations. The design theory provided functional justifications with a clear and attractive design that explained general solution elements in terms of their corresponding general requirements (Baskerville and Pries-Heje, 2010). In this research context, the design theory made the design problem (e.g., analysing donor behaviour for decision-making support) more controlled for professionals. Also, the design theory provides researchers with a theoretical foundation to predict and evaluate the patterns and lessons learned when designing the artefact.

The results ensure that each phase and iteration in the research process model was conducted to answer the RQs and achieve the research aims and objectives. The results in this chapter will provide a comprehensive examination of the research problem and how the designed artefact assists in solving the problem. The evolution in the DSR process model framework in (Alsolbi et al., 2022a) are three steps of evaluation. Therefore, the evaluation of the artefact offered feedback and a better understanding of the research problem in order to improve the quality and process of the design (Hevner et al., 2004). Thus, in this research context, each iteration presented the results and information from related evaluations to enhance the research process model execution. For example, the results of iteration one assisted the researcher in developing the predictive and descriptive analysis because the design requirements, principles, and features were well-defined and evaluated.

5.2 Results of Iteration One: Evaluating the Conceptual Design

The results of the analysis of the interviews provided insightful information about our conceptual design and what was required to analyse donor behaviour in NPOs using the AI-enabled DSS. The results of iteration one ensured the relevance of the design DPs and DFs to our research aims. A key insight from iteration one was that a traditional DSS did not meet NPOs decision-makers' requirements because it lacked efficiency and performance. However, DR1 supported the claim that a DSS should be designed to be effective and efficient. Thus, it was stated that decision-makers need to spend less time during the process of making decisions (Meth et al., 2015), which supports our DR2.
Most importantly, the interviews showed that decision-makers desired control and wanted to monitor the analysis while using the system. During the interviews, a software engineering expert said, "DR3 is an important requirement for any software designer". The first iteration's evaluation led to learning about the problem (analysing donor behaviour), the solution (designing the AI-enabled DSS), and adding an essential DR to the conceptual design, which was missed during the initial conceptual design stage. This experiment reflected how different stakeholders involved in the evaluation, all with rich experience working and volunteering in NPOs, gained different insights from literature and interviews with two experts. After finalising the

analysis, the research team looked at the results; they considered the variety of experts interviewed and the resources cited from the literature, and felt it was necessary to modify the initial conceptual design.

Most importantly, a minor change to the conceptual design was required based on the analysis of the interviews and the reflections gained during this experiment. From looking at the additional DRs, we found that usability was an additional requirement because the main target of the AI-enabled DSS was to help the main end users from NPOs make better decisions on donors. Despite the range of experience among interviewees, feedback on terms and considerations of usability were mostly repeated in the interviews. We added usability as a fourth main requirement in the DRs in the conceptual design. Usability is the second level of user experience, according to (Nielsen, 2010), a leader in the field. Once it is shown that the product can solve users' concerns, its usability is considered. The usability of a design is determined by how well its features suit users' demands and surroundings (Nielsen, 2010). Furthermore, some key elements of usability should be applied when considering "usability" during the design and development phase. Usability should include the following elements (Nielsen, 2010):

1. Effectiveness: it assists users in correctly performing actions.

2. Efficiency: users may do jobs quickly by following the simplest approach.

3. User engagement: Users find it enjoyable to use and relevant to

the industry/topic.

4. Error Tolerance: it covers a wide variety of user operations and only displays an error when something is truly wrong.

5. Ease of Learning: new users will have no trouble achieving their objectives and will have even more success on subsequent visits.

Usability is an important element of the design process of any system to ensure that the users of that system do not desert the system (Nielsen, 2010). The outcomes of any DSS are strongly affected by usability (Mt, 2017). A well-designed DSS is an interactive software-based system that assists decision-makers in compiling relevant information from assorted raw data, documents, personal knowledge, and business models to identify and solve problems, and make decisions (Mt, 2017).

Considering the additional DPs, DP7 was added, which states that the DSS should be easily usable by NPO stakeholders. Generally, most of the experts who required "usability" as an additional DR claimed that the AI-enabled DSS should be usable to predict and describe donor behaviour in NPOs. Thus, this additional DP would reflect on the additional DR and lead us to considerably add a corresponding DF that interprets how the DP7 will be achieved.

The DF5 of usability was to add a tooltip feature to the contents of the AI-enabled DSS. Tooltips give information that appears when a user presses a button (Isaksen et al., 2017). Seven different experts reported this feature in the interviews. For instance, a system designer and analyst stated, "When I move the cursor on a graph, I would like to know what numbers are, find useful information, and act like I do not know about data analysis. Tooltips can provide this type of advice". Tooltips help the user effectively navigate the system, which decreases the usage of help commands (Christie et al., 1980). Therefore, it was concluded that the tooltip feature would achieve the DP7 reflected on DR4. Essentially, other DFs reported by other experts during the interviews, such as "choice of colour" and "easy to navigate" would be considered as fundamentals of designing the AI-enabled DSS for analysing donor behaviour.

The other elements of usability, such as quality of information, easy navigation, error tolerance, and effective and efficient performance, will be considered when building the interfaces of AI-enabled DSS in a future study. The results of iteration one (shown in Figure 5.1) provided a general overview of requirements for building the AI-enabled DSS for analysing donor behaviour in NPOs. DRs, DPs, and DFs are well-defined and evaluated by experts. The results of iteration one are initials in Phase 3: Design and Development.

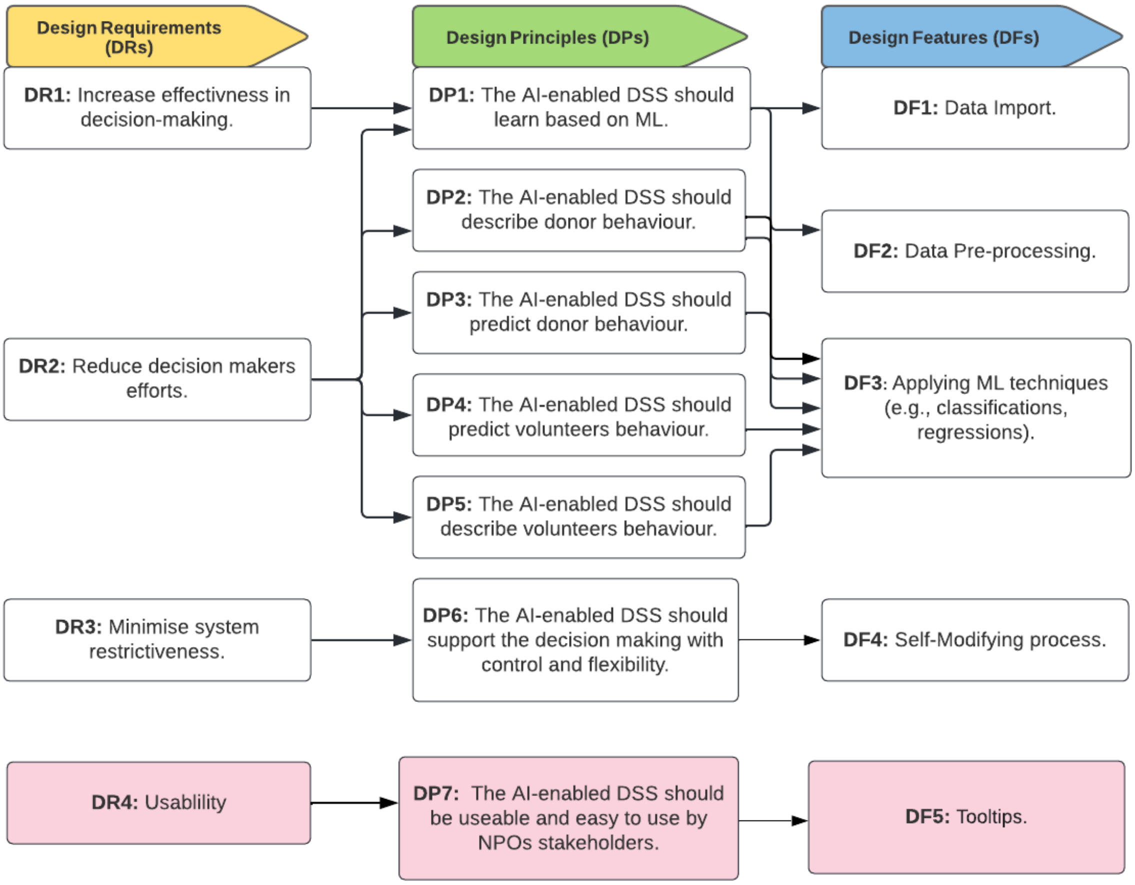


Figure 5.1: The Updated Conceptual Design of the AI-enabled DSS for Analysing Donor Behaviour in NPOs.

5.3 Results of Iteration Two: Evaluating the AI-enabled DSS Functionality and Performance

After presenting the DRs, DPs, and DFs in the conceptual design, descriptive and predictive models were completed and fully functional. Validation techniques associated with ML techniques, such as k-fold cross-validation are applied for the evaluation to ensure these models' effectiveness. Therefore, this iteration's results presented the performance of the models applied in the AI-enabled DSS.

5.3.1 Models Performance

Model performance metrics are essential parts of the ML pipeline. They indicate how well the models performed and report error percentages for deriving useful insights. Performance metrics compare the trained model and the actual dataset used for testing (Botchkarev, 2019).

5.3.1.1 Probability Model of Donors and Volunteers

The application of regression analysis predicts the dependent variables by fitting a model to a dataset of independent factors. A regression model is used to build the probability model of donors and volunteers in R. For the regression model (probability model of donors and volunteers), we used the following metrics:

1- **Accuracy** measures how many predictions made by a logistic regression model are correct. It is determined by dividing the number of correct predictions by the total number of predictions. The prediction-accuracy table assesses the effectiveness of a logistic regression model. It looks at the model's ability to accurately predict outcomes by comparing the number of correct predictions to the total number of predictions made. It is also sometimes referred to as the classification table, hit-miss table, or confusion matrix (Naser and Alavi, 2021).

2- **The receiving operating characteristic (ROC)** is a measure of classifier performance that involves plotting the proportion of correctly classified positive data points with the proportion of incorrectly classified negative data points. This creates a graph that illustrates the trade-off between accuracy and incorrect predictions.

Table 5.1 shows two metrics of donor and volunteer performance. Both models performed excellently as they obtained a percentage of more than 80%. These results initiate the answers to RQ.2, to be discussed further in this chapter. The results of the probability model of donors and volunteers showed a quality performance.

Metric	Donors' probability model	Volunteers' probability model
Accuracy	0.8669196	0.8756341
ROC	0.8883114	0.8978331

Table 5.1: Performance Metrics of the Probability Models

5.3.1.2 Random Forest Model

A classification ML typically generates a chance of belonging to one of two groups (Dataiku.com, 2022b). The actual predicted value is determined by the cut-off threshold chosen earlier. Dataiku trains a classifier (random forest model) which estimates the quality of donors and volunteers and filters them automatically. In Dataiku, the cross-validation results show that our random forest model had an average accuracy of 93.8% for the donor dataset and 91.8% for the volunteer dataset. Table 5.2 shows several classification metrics

Classification Metrics	Definition	Percentage in donors' model	Percentage in volunteers' model
Accuracy	Proportion of correctly-classified observations	93.5	91.8
Precision	The ratio of true positives and total positives predicted	71.1	72
Recall	Proportion of actual class X found by the classifier	70.5	69.7
ROC	Area under the ROC; from 0.5 (random model) to 1 (perfect model)	96.3	95.2

Table 5.2: Classification Metrics of Random Forest Models of Donors and Volunteers

generated by Dataiku for measuring the performance of predictive models of donors and volunteers.

- **Accuracy**: classification accuracy is the simplest metric which defines the number of correct predictions divided by the total number of predictions multiplied by 100. From Table 5.2, both models performed excellently in accuracy.

- **Precision**: A precision score of one indicates that the model did not miss any true positives and can distinguish between proper and incorrect labelling of donors and volunteers. As shown in Table 5.2, the precision of both models is not <50, indicating that the classifiers have fewer false positives.

- **Recall**: Recall towards 100 indicates that the model did not miss any true positives and is capable of accurately and wrongly labelling. Both models of donors and volunteers have percentages towards 100, indicating high performance.

- **ROC curve**: ROC is a figure generated after the model has been trained to show a classification model's performance. It plots the true positive rate against the false positive rate. The closer the ROC curve is to the top left corner of the graph, the higher the test's accuracy because in the upper left corner, sensitivity = 1 and the false positive rate = 0 (specificity= 1). Figures 5.2 and 5.3 show the ROC curves obtained from Dataiku.

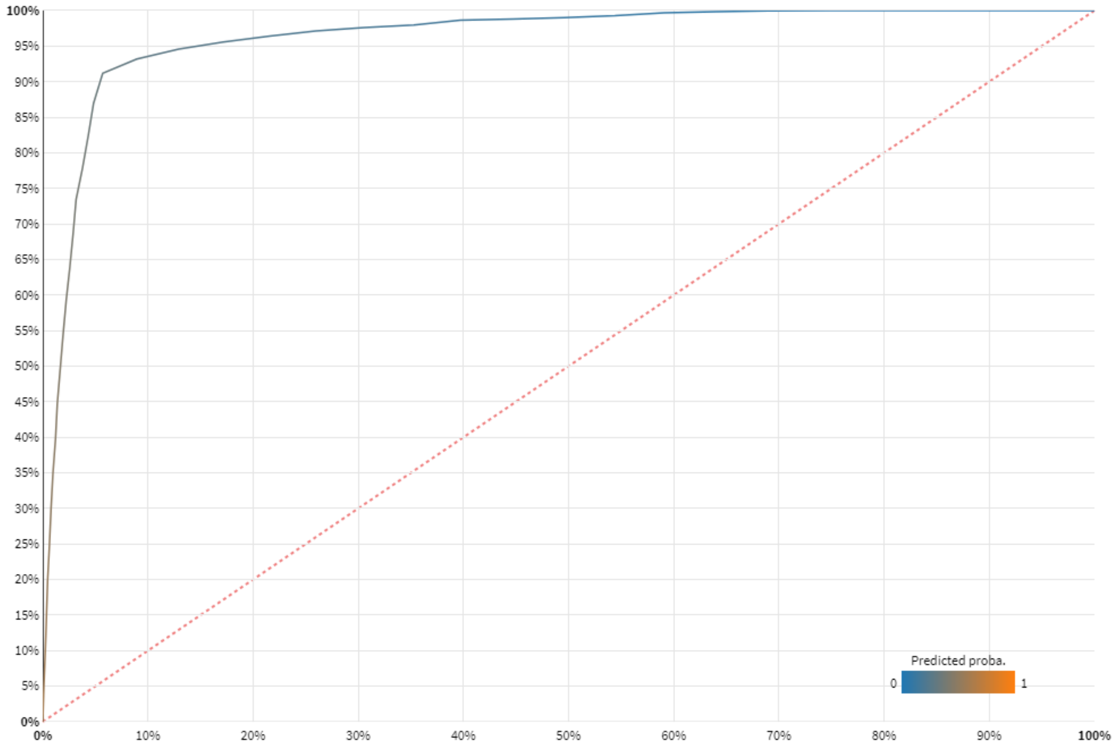


Figure 5.2: ROC Curve of Donors Random Forest Model

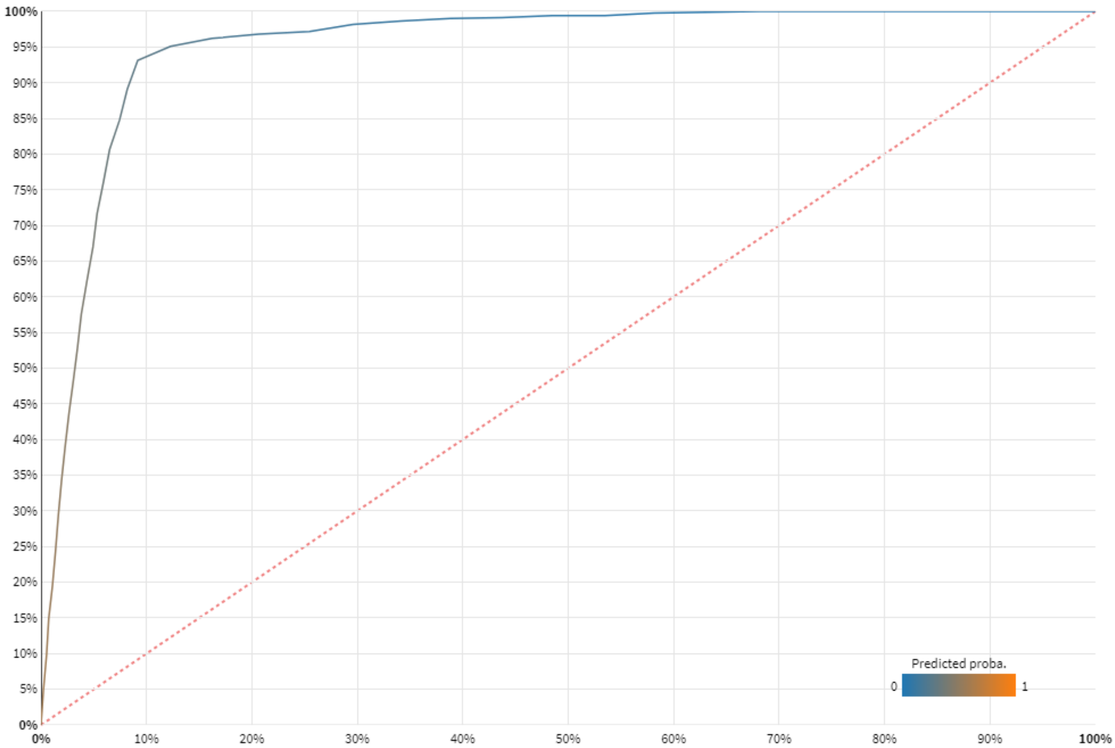


Figure 5.3: ROC Curve of Volunteers Random Forest Model

There are different measures for evaluating ML models depending on whether a regression model or classification model is under development. The above metrics provided useful information on how effective and well the models of donors and volunteers performed. Iteration two did not affect the functionality of the AI-enabled DSS because the models performed well. Therefore, iteration two enhances the organisation of slides in the dashboard of the AI-enabled DSS and provides some information on the slides. There were no major changes in the models or the analysis choices. After finalising iteration two, the cycle went back to Phase 2: Development and Design, to enhance the additional information on the AI-enabled DSS slides.

5.4 Results of Iteration Three

The results of iteration three offered useful insights for using the AI-enabled DSS to analyse donor behaviour in NPOs. A key insight from this iteration showed all the experts who tried and explored the slides of the AI-enabled DSS were satisfied that they achieved the research objectives. For instance, experts reported their satisfaction with the effectiveness, usability, efficiency, and control of the AI-enabled DSS, which reflects the design requirement of the artefact. However, there were some critical points to help users utilise the AI-enabled DSS effectively:

1- Providing guidelines or a step-by-step manual to assist less-technical users in obtaining the maximum benefits from the AI-enabled DSS. A user manual is a guide that aids in the seamless usage of a specific system, product, or service., An effective user manual is brief but communicates clearly and provides all the required information. A sample manual was developed and attached in Appendix D.

2- Comprehensively explaining each figure on the dashboard of the AI-enabled DSS. We found some experts with technical experience who could interpret the results. However, two NPO managers had difficulty interpreting two figures from descriptive donor analysis. One social expert suggested, "when you design, write the information as if you do not know the meaning of the figure". This suggestion indicated the importance of ensuring all analysis figures have adequate and helpful guidance to help new users obtain useful information and support for their decisions.

Generally, the results of iteration three did not reveal major changes in the structure, functionality, or operation. Instead, this iteration provided adequate information for each figure on the dashboard to support user decisions. By finalising the changes required for iteration three, a design theory for developing the AI-enabled DSS is created as part of the design science process model framework (Alsolbi et al., 2022a). Design theory describes how an artefact should be constructed to achieve the desired interventions and consequences. Thus, a design theory is created to enable instantiating the AI-enabled DSS for future development.

5.5 Design Theory

The design theory is one main component of the DSR process model (Peffers et al., 2007). It is a perspective of statements on how to design a solution to achieve certain goals for solving a known problem (Vaishnavi et al., 2019) and represents a knowledge contribution from DSR (Vaishnavi et al., 2019). In this research, the design theory follows the profile of the design theory adapted from (Gregor and Hevner, 2013). The design theory initially formed a profile of designing AI-enabled DSS for analysing donor behaviour in NPOs. Table 5.3 summarises the work's conceptual integration into an AI-enabled DSS design theory.

The first component of design theory states the purpose and scope (Gregor and Jones, 2007). This component defines the system and the stakeholders who use the artefact. The design of the AI-enabled DSS helps NPOs that have difficulty making decisions about donors and volunteers. Therefore, the AI-enabled DSS provides useful information on donors and volunteers through descriptive and predictive analysis. Experts who used the AI-enabled DSS during the third evaluation provided indications of its usefulness and effectiveness.

Constructs of the AI-enabled DSS are essential, as they incorporate different conceptual entities around the artefact. Most importantly, to collect the best advice on the key constructs, interactions between the designer and stakeholders or users of the artefact are required to capture the independent constructs (Meth et al., 2015). In this research, we strived to derive the best possible constructs of the AI-enabled through interviews with experts with long experience

working in NPOs and awareness of the research problem.
The knowledge or the principles form a design theory component and help provide a blueprint for building the AI-enabled DSS (Meth et al., 2015). Thus, we formed DPs generally to be implemented using DFs. However, the artefact mutability (the fourth component) may reflect on the knowledge base. The contents of the knowledge base used for the AI-enabled DSS and the underlying scheme for defining the requirements are context-dependent. They will evolve as more predictive and descriptive techniques are processed in a given context. Furthermore, stakeholders' requirements are likely to evolve when the technology is used more frequently, reducing the mining efforts even further. However, changes in context will have an impact on the artefact's design mutability in both circumstances.
For evaluating the AI-enabled DSS, we assumed that there are two types of evaluation (formative and summative) that may provide helpful indications for testing the functionality of the AI-enabled DSS. Furthermore, our design integrates the two types of evaluation through different testable methods to provide reassurance of effective testing propositions. One criterion for evaluating any system is testing the system against the objectives and requirements. Through the process of conducting the two types of evaluations, we obtained different measures of effective DRs, DPs, and DFs. The results of the two types of evaluation indicate that the objectives and the requirements of the AI-enabled DSS are met significantly.
The sixth component of design theory is justification knowledge, used for designing the AI-enabled DSS. The justificatory knowledge is

derived from the literature on DSS and conceptual abstracts of stakeholders and experts who support and evaluate the objectives and the constructs of the AI-enabled DSS.

Component	Description
1) Purpose and Scope	The literature shows a need for building an AI-enabled DSS for analysing donor behaviour to enhance decision-making in NPOs. The AI-enabled DSS assists NPOs in the process of making decisions.
2) Constructs	Reducing efforts of decision-makers, effective decision-making support, DSS perceived restriction, and usable DSS.
3) Knowledge of Form and Function	We derive DPs (DP1 to DP7) to support the DRs of developing AI-enabled DSS for analysing donor behaviour in NPOs. Meanwhile, DFs are suggested to correspond to the DFs.
4) Artefact mutability	The contents of the knowledge base used for the AI-enabled DSS, as well as the underlying scheme for defining the requirements, are context-dependent and will evolve as more predictive and descriptive techniques are processed in a given context. Furthermore, stakeholders' requirements are likely to evolve when the technology is used more frequently, reducing the mining efforts even further. However, changes in context will have an impact on the artefact's design mutability in both circumstances.
5) Evaluation and Validation Propositions	Two types of evaluations (formative and summative) have been used to evaluate the conceptual design and the practical components of the AI-enabled DSS. The two types of evaluations result in conducting three iterations to enhance the process design and development.
6) Justificatory Knowledge	DRs, DPs and DFs are derived from decision-making theory and prescriptive information from the literature.

Table 5.3: The Design Theory of the AI-enabled DSS for Analysing Donor Behaviour

This research intended to provide practical and theoretical components that aim to help NPOs analyse donor behaviour and find a foundation of analytical solutions from our final results. An overview of the phases of the DSR process model, its iterations, data analysis, and outcomes are represented in Figure 5.4.

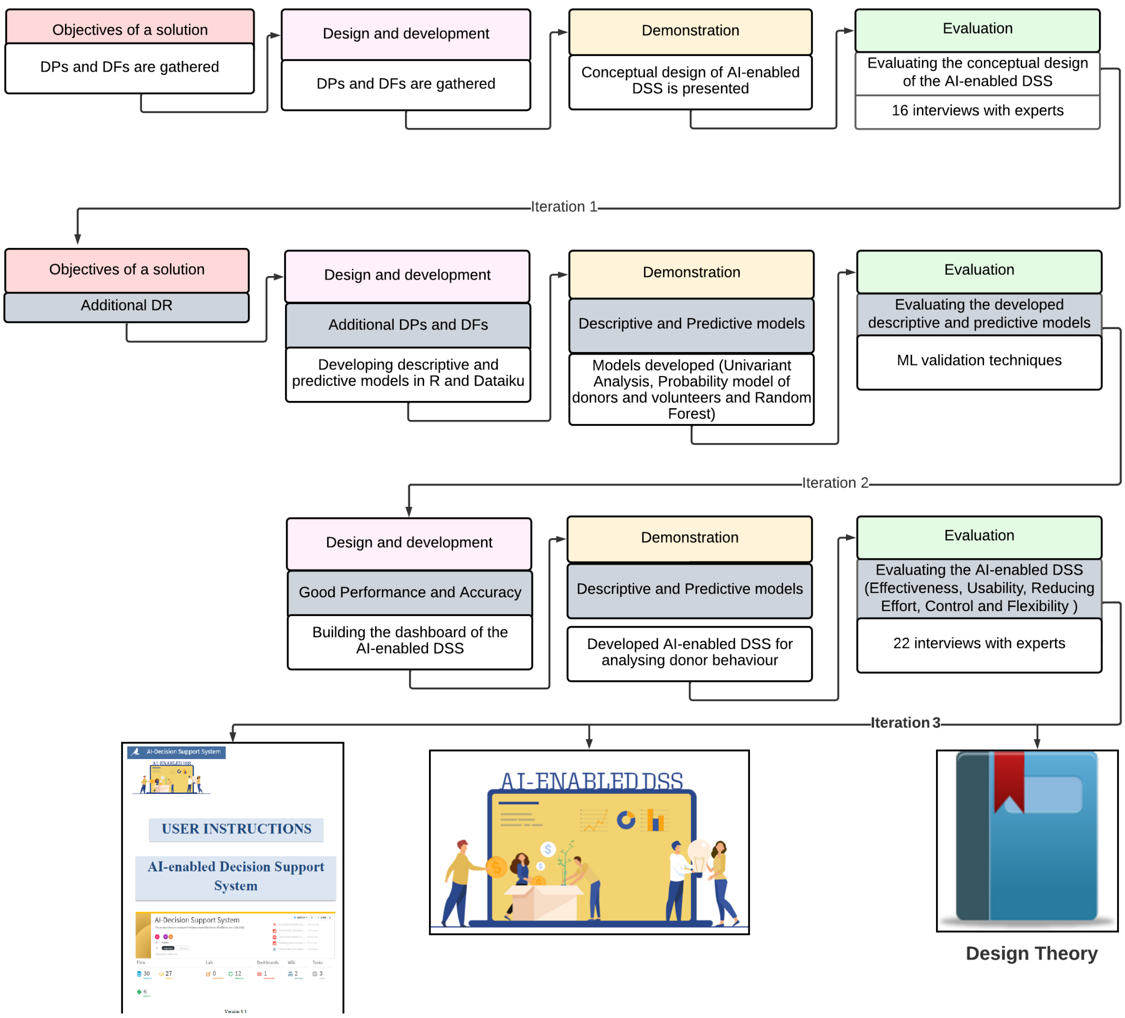


Figure 5.4: An Overview of Phases of the DSR Process Model Applied in the Research Methodology and its Outcomes.

5.6 Discussion on Research Questions' Answers

We discuss the research results to meet the research questions defined in **Chapter 1**.

5.6.1 RQ.1: Corresponding to RG1, what is the design theory that guides the development of an AI-enabled DSS to analyse donors' behaviours in NPOs?

The design of the AI-enabled DSS is synthesised into the design theory shown in Table 5.3 to consciously convey a conceptual, abstract solution rather than just one specific artefact. The aim of presenting the design theory is to contribute to the literature on DSS in NPOs by providing DRs, DPs, and DFs for creating the AI-enabled DSS to analyse donor behaviour in NPOs. To the best of our knowledge, this unique distinction between DPs and DFs is a novel addition to the relevant literature.

Consequently, further implementations of our design can act as both inventive modifications and extensions of the design we presented, as well as testing the underlying DPs and DFs. The AI-enabled DSS should provide added value and enhance the decision-making process for NPOs, as it is mainly built for analysing donor behaviour. Extracting essential information from the AI-enabled DSS about donors and volunteers enables decision-makers on their decisions.

The results of the three iterations show the potential benefits of integrating design requirements, DPs, and DFs. The proposed and evaluated DPs and DFs have proven that the DRs of the AI-enabled DSS are achieved. Therefore, our design theory is formed to be generalisable to allow the possible application in different contexts of building DSS. Our main answer to RQ1 is presented in Section 6.5 in this chapter. The following two sub-questions specifically discuss the main components of DPs and DFs.

5.6.1.1 RQ.1.1: What are DPs should an AI-enabled DSS follow to analyse donor behaviour in NPOs?

DPs are statements to prescribe how an artefact instantiates to meet the objectives and the design requirements (Meth et al., 2015; Chandra et al., 2015). The results of iteration one indicate that the DPs stated in the conceptual design are well-defined and presented. Thus, such interpretations are required to meet the design requirement. Seven DPs are formed in this research to facilitate the process of building the AI-enabled DSS for analysing donor behaviour:

1- The AI-enabled DSS should learn based on ML.

2- The AI-enabled DSS should describe donor behaviour.

3- The AI-enabled DSS should predict donor behaviour.

4- The AI-enabled DSS should describe volunteers' behaviour.

5- The AI-enabled DSS should predict volunteers' behaviour.

6- The AI-enabled DSS should support decision-making with control and flexibility.

7- The AI-enabled DSS should be usable and easy to use by NPOs stakeholders.

Our aim in creating these DPs is to offer a generalisation of these principles, which may help practitioners to implement highly effective solutions for analysing donor behaviour in NPOs. However, if any changes are made in the scope of the research problem or the objectives of the AI-enabled DSS, these principles may not be generalised. The defined DPs in this research may still benefit professionals who desire to analyse donor behaviour in NPOs.

5.6.1.2 RQ.1.2: What are DFs should an AI-enabled DSS follow to analyse donors' behaviours in NPOs?

DFs specify ways of implementing the DPs (Meth et al., 2015). In other words, DFs are important because they tell the designer the required features and functions of the system. In this research context, we created five DFs to interpret the DPs, and help achieve the requirements and the objectives of the AI-enabled DSS for analysing donor behaviour in NPOs.

1- Data import

2- Data pre-processing

3- Applying ML techniques (e.g., classifications, regressions, etc.)

4- Self-Modifying code

5- Tooltips

During the three evaluations, the DFs indicated useful measures of their ability to achieve the objectives of the AI-enabled DSS. For example, applying ML techniques, such as regression results, effectively predicts single donors using certain variables. Another significance of the DFs is the addition of tooltips, which saved decision-makers time and effort when reviewing their options about donors and volunteers. We strive to provide generic DFs that enable practitioners to deploy similar approaches for creating artefacts to analyse donor behaviour. However, these DFs may evolve in future research, depending on the nature of the research problem. We recommend these DFs because they have been evaluated by experts and have produced insightful indications of their capabilities for analysing donor behaviour. Changing some DFs for an artefact that analyses donor behaviour may require scholars to considerably change the design requirements or the DPs.

5.6.2 RQ.2: Corresponding to RG2, what are the main functionalities of a DSS for analysing donors' behaviours in NPOs?

Clearly, the objectives of the AI-enabled DSS stated the main requirements and the functionalities of the AI-enabled DSS. The design requirements, principles, and features are gathered to reach a stable solution for analysing donor behaviour in NPOs. Essentially, we strive to offer effective and usable functionalities of the AI-enabled DSS for analysing donor behaviour. Therefore, we created a descriptive and predictive analysis of donors and volunteers using big datasets to prove the concepts of analysing donor behaviour using the capabilities of ML techniques.
The main functionalities after implementing the AI-enabled DSS:

1- Exploring general statistics about donors and behaviours such as the average age, income, number of donations, and total of donations.

2- Obtaining predictions of donors and volunteers to know future contributions of donations and volunteering.

3- Knowing the probability of a single donor and volunteer considering variables such as age, income, state, level of education, and the number of donations and volunteering.

4- The ability to interpret the analysis with tooltips, colours, and useful information provided. As stated by some experts and managers of NPOs in iteration one, system users should be able to interpret the results. The interpretations of the results help in the process of making decisions.

For developing the above functionalities, different analyses have been applied using R environment and Dataiku. Both platforms enabled the researcher to develop an interactive DSS dashboard in AI-enabled DSS. The variety of analyses facilitates the process of big datasets preparations, processing, and modelling. The two sub-questions below explain how the AI-enabled DSS's main functionalities are combined and structured. In each sub-question, we explain what data analytics and models were performed to obtain such effective results.

5.6.2.1 RQ2.1: Upon our standing of different descriptive data models, what is the best model that tells the influential factors on donors' intentions to donate or volunteer?

Discovering the influential factors of donations and volunteering is an important task for NPOs to support decision-making. Determining these factors may vary with ML techniques and data analytics applications. In this research context, the probability model of donors and volunteers indicated the most important and influential factors on donors and volunteers. For example, when testing the single prediction of a donor on the probability model, we found that people with a college degree are more likely to donate. Therefore, certain factors have been discovered that influence donors and volunteers using the probability model of donors and volunteers. These factors are mainly age, state, income, socio-economic status and number of donations. The probability model (logistic regression) predicts the likelihood of donating or volunteering, considering the defined influential factors on donors and volunteers.

Univariate is another model for exploring the variables and the influential factors on donors and volunteers. Univariate analysis is a feature in Dataiku which allows a selection of different variables from the dataset to determine the distributions of the variables. The goal of univariate analysis is to describe and summarise the imported datasets. The summary of the dataset can lead to initiating some understanding of variables and their correlations for the process of creating useful models.

5.6.2.2 RQ.2.2: Who among the previous donors or volunteers is likely to donate or volunteer in the future?

Predictive modelling can be better employed to understand donor behaviour, preferences, and motivations. By leveraging predictive models, organisations have the potential to identify new donors and target segment populations that are more likely to donate. Moreover, predictive models can be deployed to explore relationships between donors and specific causes or geographic regions, as well as to uncover opportunities for increased donor engagement, optimise portfolios, and more accurately forecast revenue projections. Predicting the potential donors and volunteers requires under- standing of the relative dataset's variables. The logistic regressions built throughout this study point to several significant findings among potential donors to NPOs. The probability model of donors and volunteers indicated that NPOs could understand the strong positive likelihood of donations and volunteering events.

Moreover, they created two random forest models to obtain such predictions of donors and volunteers. random forest is used in our research context for predicting the probability of donors and volunteers. The random forest model combines and grows several trees to build a "forest". It starts with a single point and then goes down into two or more branches, and each branch offers potential outcomes of predictions. Therefore, the two random forest models of donors and volunteers provided different predictions because of the variety of features automatically selected by Dataiku. The random forest models indicated that most people over age 50 are more likely to donate money and participate in volunteering activities for NPOs. Income level also plays an important role in encouraging people to donate and volunteer. The frequency of donations or volunteer time are indicators of defining people who donate and volunteer. As an example of donor frequency, we found that people who donate more than ten times are likely to donate regardless of income level or age. Similarly, people are likely to volunteer multiple times as the frequency of volunteering increases. Meanwhile, we discovered that people who have full-time jobs are less likely to volunteer than people who work part-time or are searching for a job.

Donating and volunteering predictions may vary due to different variables in each related dataset. However, in this research, we relied on the reported variables in the literature that influence donors and volunteers and can indicate some insights about their behaviour. We found that age, state or location, income and education level, frequency of donations and volunteering, and social economy help obtain such useful predictions. Table 5.4 summarises the most important variables of donors and volunteers, which helped predict donor behaviour.

Variables of donors	**Variables of volunteers**
Age	Age
State	State
Income	Income
College degree	College degree
Social economy	Social economy
Number of donations	Frequency of volunteering
Total amount of donations	Type of work/job

Table 5.4: A summary of the Most Important Variables that Influence Donors and Volunteers

5.7 Chapter Summary

Chapter 5 presented the findings of the DSR process model framework for developing an AI-enabled DSS for analysing donor behaviour in NPOs. Mainly, three iterations' findings are shown to ensure the validity and functionality of the AI-enabled DSS. Also, the three iterations provided helpful feedback, which led to improved outcomes and requirements and assisted in achieving the objectives of the AI-enabled DSS. Finally, each RQ was discussed based on the findings of the DSR process model framework. The answers to each RQ met the research aims and objectives, which are crucial for their validity.

CHAPTER 6

CONCLUSION

6.1 Summary of the Research

Data analytics has the potential to revolutionise the nature of non-profit organisations. The analytical models and empirical studies that result from data analytics help the sector understand donors and volunteers. NPOs use data analytics and corresponding visualisations to discover and interpret donations and donor behaviour patterns, predict future funds, and analyse time series to undertake decisions and resolve issues.

Therefore, we conducted a systematic literature review which confirmed a dearth of literature on developing an intelligent DSS to analyse donor behaviour towards donating and volunteering. The literature points out some attempts to analyse donor behaviour using ML and big data capabilities; however, there is still no fine artefact or model to serve this purpose. NPOs typically lack the technological, financial, and human resources to develop a DSS for analysing donor behaviour. Also, NPOs lack the knowledge to design and develop technical solutions to analyse donor behaviour; they lack skills for managing data resources as well. Several studies showed that NPOs have massive data resources (Mayer, 2019; Hou and Wang, 2017). There are useful data analytics techniques that can be done using big datasets to provide insights and predictions and assist in understanding donor behaviour in NPOs.

Donor behaviour differs, depending on various factors such as income, education level, gender, previous history of donating, and frequency of donations and volunteering. Knowing and understanding these behaviours and the influential aspects of donations and volunteering is important for NPOs. However, analysing donor behaviour remains crucial yet difficult for many NPOs, as evidenced by the scarcity of literature on the subject. As a result, this research represents the first step towards using ML to analyse donor and volunteer behaviour in NPOs.

This research adopted a DSR process model developed by (Peffers et al., 2007) to form the AI-enabled DSS research framework. This framework consists of six phases, starting from initiating the research problem, defining the objectives of the solution, design and development, demonstration, evaluation, and finally, communication with scholars via publication. Essentially, there are three iterations involved in the evaluation to validate (1) the conceptual design of the design requirements, the design principles, and design features of the AI-enabled DSS, (2) the descriptive and predictive models developed in Dataiku and R, using the k-fold cross technique, (3) functionality, usability, and operationality, and benefits obtained from using the AI-enabled DSS. Experts with a range of experience have been involved during those interviews to provide valuable actions in iterations one and three, where self-validation was done for iteration two.

The current study provides new insights into the otherwise under-researched area of donor behaviour by comprehensively analysing various variables and factors that influence donors and volunteers. Descriptive and predictive analysis of big datasets of donors and volunteers have been developed and visualised via an AI-enabled DSS. A probability model is also developed using R and associated with the AI-enabled DSS for predicting the probability of donating and volunteering frequency. Another part of this thesis emphasised creating a design theory to help summarise the involved steps, methods, and validation methods used for developing the AI-enabled DSS. The design theory would be a basis for action as defined by (Gregor and Hevner, 2013), explicating prescriptions on how something is done.

The results of the three iterations conclude:

1- Usability is an essential design requirement which should be considered when developing the AI-enabled DSS. In addition, to consider the usability angles during the development (e.g., error tolerance and easy navigation), tooltips as a new design feature provided helpful information for some analysis on the AI-enabled DSS. Validating the predictive models, using the k-fold cross technique, presented high performance and accuracy achieved during self-validation in iteration two. One may argue that their other validation techniques may serve the purpose better. However, one beneficial feature of the Dataiku platform is to help perform k-fold cross in the back-end and show the results without interfering with users. This would help those with less technical experience in data analytics and modelling. Developing a probability model (through logistic regression) in R helped determine the probability of a donor and a volunteer setting their intentions to donate and volunteer in the future based on important variables.

2- Additional variables may be used to create a similar approach, which was not applied because it did not exist in the datasets. Certain variables can indicate donor behaviour and donor traits, such as age, income levels, gender, social economic level, education level, number of donations, volunteering frequency, and type of job. When using the applied analysis, these variables provided useful indications and presented meaningful figures for experts during interviews of iteration three. Interestingly, in previous theoretical studies of donor behaviour, these variables were found to be influencing donors and volunteers, where our findings confirm their influence from the perspective of using ML techniques.

3- Dataiku is a powerful platform that can be easy to learn and build useful analysis on a sample of data in NPOs; it is ideal for those without a solid statistical and analytics background. The platform can construct automatic models with minimal human intervention, but it also allows the user great flexibility in variable exclusion and model development. There are useful documents available that can guide scholars and professionals to develop useful DSSs and data analytics models and leverage the capabilities of ML and data analytics capabilities in NPOs. Experts found enough benefits of Dataiku's functionality to make a difference for NPOs with limited time and resources by allowing them to conveniently access data, gain insights quickly, and automate work. Dataiku is free for up to three users with limited features. NPOs can be granted a nonprofit licence under some qualifications (Dataiku.com, 2022b).

6.2 Overview of the Research Contributions

This thesis's contribution can be divided into two categories: theoretical contributions and practical contributions. First, we offer a theoretical foundation for creating an AI-enabled DSS for potential future directions and impacts. Second, the research contributes in a practical manner to the academic literature by employing ML approaches and utilising sizable datasets of donors and volunteers to conduct meaningful descriptive and predictive analyses. These evaluations of decision-making mechanisms may offer suggestions for NPOs' missions that add value or ways to increase the effectiveness and efficiency of internal data processing. The main contributions of this thesis are:

1- Introducing a design theory of AI-enabled DSS for analysing donor behaviour in NPOs. This design theory aims to guide future work by forming the required process and the involved methods. Scholars interested in extending our work would have a knowledge base to look at the research problem, the developed solution, and the applied methods from different perspectives. This knowledge base may expand the scope of this research, such as building more descriptive models or involving other validation methods in the research process model. Therefore, the design theory would be a guide for future work.

2- Developing descriptive and predictive models of donors and volunteers. To the best of our knowledge, this contributes to the literature as the first study that combines donor and volunteer behaviour analysis. These analytical models are not the only examples of highly effective tools that provide insightful information about donors and volunteers. However, presenting the analysis of donor and volunteer behaviour in this study proves the concept that NPOs would be wise to involve ML and data analytics techniques.

3- Developing the AI-enabled DSS for analysing donor behaviour in NPOs. Developing AI-enabled DSS is not complex, thanks to capable technologies in data analytics and ML, such as Dataiku. This thesis contributes to the literature by showing how different descriptive and predictive models can be combined, visualised, and presented in the form of a DSS dashboard which will help NPOs stakeholders make better decisions.

4- Involving experts in validating and evaluating the work. Not many studies involve experts from the same scope of the research problem to validate the proposed solutions. During the phases of this study, we have involved certain steps of evaluation and validation in the problem initiations phase, the collecting requirements phase, and the design and development phase.

6.3 Research Limitations

Like any other study, the current study has limitations. This study is limited to analysing donor and volunteer data, with considerations of money and time donations, which restricts the generalisability of the findings. The current study acknowledges that participants donate to charities in a variety of ways, which include responding to a disaster or donating for emergency relief, and responding to a charity's outreach, which includes direct mail appeals and charity events. They choose between various charities to support, and they support a friend's fundraising activity. Other types of donations, such as blood, gifts, services, and humanitarian aid, can be considered for a future study.

Firstly, the data was gathered only from individuals in the United States, leaving out donors from other areas of the world. Datasets from other countries may contain other features of donors and volunteers which may have affected the results if they were analysed. As a result, the current study's findings are contextual and time-bound. One should take caution when extrapolating conclusions beyond individuals who participated in this study, as there are likely donors and volunteers whose perspectives differ from those described in this study. During the recruitment process, however, the researcher made significant efforts to recruit datasets to build generalised results, which called for various sources and demographic backgrounds, including people of both genders, with diverse ages, levels of income, educational backgrounds, and occupations. Based on these efforts, the current work opens a new door and advances the study of analysing donor behaviour in NPOs using ML techniques.

Secondly, the current study was carried out from an interpretive standpoint, and the researcher is fully aware of the constraints imposed by this method. It is crucial to note that the researcher's beliefs may affect how the findings were interpreted and how they turned out. Thus, the thoughts and opinions formed may differ from those of another researcher working in the same area. The researcher's presence might have influenced the participation of experts during interviews.

Thirdly, due to some barriers to obtaining the relative datasets, the researcher did not consider cultural and religious impacts on donors and volunteers. Privacy, ethical, and religious concerns were the most common barriers to expanding the scope of this study beyond donations of money and time. However, the datasets used in this study are demographic and served the purpose of creating useful models and analytics. The present findings are nonetheless significant within the acknowledged constraints.

Fourth, the explorative methodology in the systematic review makes the study prone to subjective bias, although it ensures a comprehensive analysis of the research topic. There is a massive volume of secondary data from numerous databases that can be examined to achieve additional research objectives and establish the foundation for future studies. The systematic review used only three popular databases: Scopus, Web of Science, and ProQuest. However, some papers might have been skipped if the source was not in these databases; also, sources were restricted to publications appearing from 2010 to 2021. If the chosen period had been longer, more results could have been obtained. For example, other terms might appear with research focused on NPOs, such as "decision making" or "empowered data". However, irrelevant literature might appear in this research scope. Aiming to achieve specific results matching the keywords lists did not allow for extending the period.

Fifth, limitations were found in the bibliometric analysis used in this research. From a methodological perspective, co-term and SEP analyses are based on the correlation between semantic similarities of terms. Yet they neglect the causality or sentiment associations (positive or negative) between terms, limiting us from interpreting the deeper reasons for the term relationships and evolutionary patterns. Moreover, relevant research papers allocated beyond the scope of this study were excluded, although they had a high rate of keyword occurrence. For example, several studies related to medical research containing medical and health research keywords seem irrelevant to NPOs. The keyword selection was changed during the research phase several times to ensure that all documents found fell within the scope of the application of data analytics in nonprofit organisations.

Sixth, other tools and technologies can provide useful analysis of the collected data. Other tools may also help design the AI-enabled DSS. For example, Tableau is another software product for interactive data visualisation. However, due to the scope of this research and the time frame, we only used Dataiku and involved a probability model built in R. We believe that the capabilities of Dataiku saved substantial time in developing an application of data analytics, as it provided more documents and lessons.

Seventh, the researcher did not consider applying other ML techniques, such as clustering. Clustering is a type of unsupervised ML technique which is meant to group similarities between data samples. If there was scalability in the datasets, and unlabelled data, we would have applied such clusters to donors and volunteers.

Finally, the researcher also acknowledges the study's limitations when they were framed within a traditional donor behaviour analysis, which limited the types of questions that could be asked, particularly about emotions and emotional dimensions. For instance, the motivations behind giving in the wake of a catastrophe or a disaster could differ from those behind giving to an annual event. As a result, the researcher fully accepted the analysis applied to the selected datasets' limits in generalising the findings because of the magnitude of each type of dimension.

6.4 Future Directions

Although the current study makes several significant contributions, future researchers might further broaden its scope. Future research can therefore examine other types of donations such as blood, gifts, and goods. Despite being a suitable sample for this study, more research is needed before the results can be applied to other types of donation groups in different regions. This requires appropriate datasets from relevant NPOs. Researchers might find it intriguing to conduct a comparison study that leads them to distinguish between each instrument, look at their similarities and differences, and determine whether or not the variables in the current study behave similarly. Analysing other types of donations would help to understand other variables influencing donors towards their donations.

A key future direction is to apply ML techniques that consider cultural and religious aspects when analysing donor behaviour. Therefore, it would be beneficial to confirm the present findings in other contexts (religious and cultural variables), by contrasting the results in this study with future results. However, more datasets with adequate features are required to achieve this direction. A collaboration between NPOs and for-profit organisations may serve the purpose of overcoming the sensitive issues in NPOs, such as privacy and discrimination. Conducting this will result in more results to be generalised which may benefit the maximal number of NPOs across different nations.

Applying other data analytics models represents another future direction; some examples include clustering donors and volunteers based on their financial and time contributions. The current study only provides descriptive and predictive analysis of donor and volunteer datasets. For example, future researchers may find our AI-enabled DSS interesting and try to expand its functionality by adding clustering models and visualising the results accordingly. Most importantly, comprehensive databases will be added to ensure coverage of the literature. In an effort to identify people who have a high possibility of giving and the financial capacity to do so, a future study may benefit from including the prospective wealth of individuals within the donor and volunteer database. While the study included demographic data which provides information about the characteristics of a population, such as age, gender, ethnicity, income, education level, marital status, occupation, frequency of volunteering, and number of donations, it offered little insight into the specifics of the constituent's current life situation, such as marital status, job title, or family size. The variables within this analysis were focused on the characteristics of constituents as they pertain to NPOs. With the addition of this information, donor behaviour to target in the future would have been depicted more accurately. This research offers a place to start when examining different questioning strategies and a way to think about spectatorship at sporting events.

Most importantly, NPOs may face critical obstacles when adopting data, like technical concerns and insufficient resources. Challenges have been discovered when applying data analytics in NPOs; developing theoretical frameworks on how NPOs can adopt data analytics may resolve their critical issues. These frameworks could address privacy concerns and the NPO requirements that data analytics tools need to meet. Understanding and predicting users' perceptions (e.g., NPO managers and scholars) can overcome some of these challenges. For example, TAM, the technology acceptance model (Davis, 1985) can provide insights and indications concerning the perception of data in NPOs. Case studies conducted in several countries could validate the framework to ensure its effectiveness, accuracy, and generalisability. It will also reveal how aware NPOs scholars are of data concepts and applying data analytics.

Currently, there is no apparent study attempting to determine user adoption of data analytics by NPOs. Investigating the influence of quantitative or qualitative research methods to apply data analytics in NPOs offers another future research direction. These research methods will answer a range of questions about the characteristics of NPOs, including size, service types, scope, and intensity, which will help develop theories that provide adequate information for investing data analytics in NPOs (Johnson, 2015). Also, these methods could help data analytics practitioners be aware of post-data analytics needs. However, these studies should be conducted in the literature under a collaboration between NPOs and industrial sectors who may volunteer their experience to assist in applying data analytics.

Some innovative and remarkable ideas or procedures could be presented and used for some purposes based on their efficacy and practicability. ML and automatic algorithms, for example, are being used in different disciplines to resolve complicated problems, which may lead to improvements in the decision-making process. In the future, human-computer interactions in various sectors will be possible (Wang et al., 2010). Furthermore, there are numerous other issues and directions for applying data.

Another interesting future direction is political donation analysis and knowing the associated behaviours. Political donations can be analysed in various ways. This may involve looking at the total sum of money given to a particular party or politician, as well as the sources of the money, such as individuals or companies. Furthermore, the types of issues supported by the donors and the amount spent on political activities can be analysed too. Additionally, the timing of the donations can be assessed to see how it is linked to certain policies or events.

6.5 Bibliography

AHMED, M. H., A. AL-DHELAAN, A. AL-NUAIM, S. BANU, H. AL-OTAIBI, AND A. AL-AJLAN (2019): "Data analytics: An overview of techniques and technologies," *International Journal of Advanced Computer Science and Applications*, 10, 51–60.

AJZEN, I. (1991): "The theory of planned behaviour," *Organizational behaviour and human decision processes*, 50, 179–211.

ALSOLBI, I., R. AGARWAL, G. BHARATHY, M. SAMARAWICKRAMA, B. UNHELKAR, AND M. PRASAD (2023): "A Systematic Review and Taxonomy of Data Analytics in Nonprofit Organisations," *Asia Pacific Journal of Information Systems*, 33, 39–68.

ALSOLBI, I., R. AGARWARL, B. NARAYAN, G. BHARATHY, M. SAMARAWICKRAMA, S. TAFAVOGH, AND M. PRASAD (2022a): "Analyzing donors behaviours in nonprofit organizations: A design science research framework," in *Pattern Recognition and Data Analysis with*

Applications, ed. by D. Gupta, R. S. Goswami, S. Banerjee, M. Tanveer, and R. B. Pachori, Springer Nature Singapore, 765–775.

ALSOLBI, I., M. WU, Y. ZHANG, S. JOSHI, M. SHARMA, S. TAFAVOGH, A. SINHA, AND M. PRASAD (2022b): "Different approaches of bibliometric analysis for data analytics applications in non-profit organisations," *Journal of Smart Environments and Green Computing*, 2, 90–104.

ALSOLBI, I., M. WU, Y. ZHANG, S. TAFAVOGH, A. SINHA, AND M. PRASAD (2022c): "Data analytics research in nonprofit organisations: a bibliometric analysis," in *Pattern Recognition and Data Analysis with Applications*, ed. by D. Gupta, R. S. Goswami, S. Banerjee, M. Tanveer, and R. B. Pachori, Springer Nature Singapore, 751–763.

ANAF, J., F. E. BAUM, M. FISHER, E. HARRIS, AND S. FRIEL (2017): "Assessing the health impact of transnational corporations: a case study on McDonald's Australia," *Globalization and Health*, 13, 7.

ANGUITA, D., L. GHELARDONI, A. GHIO, L. ONETO, AND S. RIDELLA (2012): *The 'K' in K-fold cross validation*, Bruges, Belgium, 441–446.

ANHEIER, H. K. (2005): *Nonprofit organizations theory, management, policy*, London, United Kingdom: Routledge Taylor Francis Group.

ANITHA, P. AND M. M. PATIL (2018): "A review on data analytics for supply chain management: A Case study," *International Journal of Information Engineering and Electronic Business*, 11, 30.

ARIA, M. AND C. CUCCURULLO (2017): "bibliometrix : An R-tool for comprehensive science mapping analysis," *Journal of Informetrics*, 11, 959–975.

ARNOTT, D. AND G. PERVAN (2012): "Design Science in Decision Support Systems Research: An Assessment using the Hevner, March, Park, and Ram Guidelines," *Journal of the Association of Information Systems*, 13, 923–949.

BAMBHA, K., A. SHINGINA, J. L. DODGE, K. O'CONNOR, S. DUNN, J. PRINZ, M. PABST, K. NILLES, L. SIBULESKY, AND S. W. BIGGINS (2020): "Solid organ donation after death in the United States: Data-driven messaging to encourage potential donors," *American Journal of Transplantation*, 20, 1642–1649.

BARKER, E. (2019): "Optimizing individual membership dues for non-Profits," *The Non-profit Journal*, 23.

BARZANTI, L., S. GIOVE, AND A. PEZZI (2017): *A decision support system for non-profit organisations*, Springer Verlag, vol. 10147 LNAI, 270–280.

BASKERVILLE, R. AND J. PRIES-HEJE (2010): "Explanatory design theory," *Business Information Systems Engineering*, 2, 271–282.

BOPP, C., C. HARMON, AND A. VOIDA (2017): "Disempowered by data: Nonprofits, social enterprises, and the consequences of data-driven work," in *ACM SIGCHI Conference on Human Factors in Computing Systems*, Association for Computing Machinery, 3608–3619.

BÖRJESSON, A., L. HOLMBERG, H. HOLMSTRÖM, AND A. NILSSON (2007): “Use of appreciative inquiry in successful process improvement,” in *Organizational Dynamics of Technology-Based Innovation: Diversifying the Research Agenda*, ed. by T. McMaster, D. Wastell, E. Ferneley, and J. I. DeGross, Boston, United States: Springer, 181–196.

BOTCHKAREV, A. (2019): “A new typology design of performance metrics to measure errors in Machine Learning regression algorithms,” *Interdisciplinary Journal of Information, Knowledge, and Management*, 14, 045–076.

BUSHE, G. R. (2011): “Appreciative inquiry: Theory and critique,” *The Routledge companion to organizational change*, 87, 103.

CALLON, M., J.-P. COURTIAL, W. A. TURNER, AND S. BAUIN (1983): “From translations to problematic networks: An introduction to co-word analysis,” *Social Science Information*, 22, 191–235.

CENSUS BUREAU, U. S. (2017): “Current Population Survey,” www2.census.gov/programs-surveys/cps/techdocs/cpssep17.pdf.

CHADHA, Y., A. MEHRA, S. GREGOR, AND A. RICHARDSON (2015): “A Framework for Techniques for Information Technology Enabled Innovation,” .

CHANDRA, L., S. SEIDEL, AND S. GREGOR (2015): “Prescriptive knowledge in IS research: Conceptualizing design principles in terms of materiality, action, and boundary conditions,” in *2015 48th* *hawaii international conference on system sciences*, IEEE, 4039–4048.

Chang, W., J. C. Cheng, J. Allaire, Y. Xie, and J. McPherson (2020): "Shiny: Web application framework for R," *The R Journal*, 12, 495–517.

Charities, A. and N. for-profits Commission (2021): "Australian Charities Report 8th Edition," acnc.gov.au.

Chianese, A., F. Marulli, F. Piccialli, P. Benedusi, and J. E. Jung (2017): "An associative engines based approach supporting collaborative analytics in the Internet of cultural things," *Future Generation Computer Systems-the International Journal of Escience*, 66, 187–198.

Christie, D. M., D. D. Wayne, and E. U. Joseph (1980): "A methodology for the objective evaluation of the user/system interfaces of the MADAM system using software engineering principles," in *Proceedings of the 18th annual Southeast regional conference*, Association for Computing Machinery, 103–109.

Chu, P., J. Hernandez, and D. Lee (2019): "Using machine learning for non-Profit organizations," *Journal of Non-Profit Organizations*, 19, 214–229.

Chuang, A. (2020): "Dataiku: An advanced analytics and data science platform," *Applied Artificial Intelligence*, 34, 541–556.

Commission, P. (2010): "Contribution of the not-for-profit sector," Report, Australian Government.

COSTA, C. AND M. Y. SANTOS (2017): "Big Data: State-of-the-art concepts, techniques, technologies, modelling approaches and research challenges," *IAENG International Journal of Computer Science*, 44, 285–301.

DAG, A., A. OZTEKIN, A. YUCEL, S. BULUR, AND F. M. MEGAHED (2017): "Predicting heart transplantation outcomes through data analytics," *Decision Support Systems*, 94, 42–52.

DATAIKU.COM (2022a): "Dataiku DSS," https://doc.dataiku.com/dss/.

——— (2022b): "The Dataiku Story," www.dataiku.com/stories/the-dataiku-story/.

DAVIS, F. (1985): "A technology acceptance model for empirically testing new end-user Information Systems," Thesis.

DE VRIES, N. J., R. REIS, AND P. MOSCATO (2015): "Clustering consumers based on trust, confidence and giving behaviour: Data-driven model building for charitable involvement in the Australian not-for-profit sector," *PLoS ONE*, 10.

DENG, Q., Z.-H. YANG, T. FU, AND Y. YU (2014): "Data Mining for the online retail industry: A case study of RFM model-based customer segmentation using data mining," *Journal of Computer Science and Technology*.

DICK, W. (1977): "Summative evaluation," *Instructional design: Principles and applications*, 337–348.

DIETZ, R. AND B. KELLER (2016): "A deep dive into donor behaviours and attitudes," www.thenonprofittimes.com/wp-content/uploads/2016/04/donor-loyalty-study.pdf, Abila.

DONTHU, N., S. KUMAR, D. MUKHERJEE, N. PANDEY, AND W. M. LIM (2021): "How to conduct a bibliometric analysis: An overview and guidelines," *Journal of Business Research*, 133, 285–296.

DOWNE-WAMBOLDT, B. (1992): "Content analysis: method, applications, and issues," *Health Care for Women International*, 13, 313–321.

DUNFORD, L. (2016): *To give or not to give: using an extended theory of planned behaviour to predict charitable giving intent to international aid charities*, Minnesota, United states: University of Minnesota.

DWORKIN, S. L. (2012): "Sample size policy for qualitative studies using in-depth interviews," *Archives of Sexual Behavior*, 41, 1319–1320.

EIGNER, I., A. HAMPER, N. WICKRAMASINGHE, AND F. BODENDORF (2017): "Decision makers and criteria for patient discharge - A qualitative study," Association for Information Systems Electronic Library, AISeL.

ESFAHANI, H. J., K. TAVASOLI, AND A. JABBARZADEH (2019): "Big data and social media: A scientometrics analysis," *International Journal of Data and Network Science*, 3, 145–164.

ETZIONI, A. (1968): "The active society: A theory of societal and political processes," *New York: Free Press*.

EVERYACTION, T. (2018): "5 Things you need to know about nonprofits and big data," .

FARROKHVAR, L., A. ANSARI, AND B. KAMALI (2018): "Predictive models for charitable giving using machine learning techniques," *PLOS ONE*, 13, 1–14.

FISHER, J. W. (2016): "Design science research: A primer for Information Systems research," *Communications of the Association for Information Systems*, 38, 917–938.

FLEMING, L. (2001): "Recombinant uncertainty in technological search," *Management Science*, 47, 117–132.

FLEMING, L. AND O. SORENSON (2004): "Science as a map in technological search," *Strategic Management Journal*, 25, 909–928.

FREDRIKSSON, C. (2018): "Big data creating new knowledge as support in decision-making: practical examples of big data use and consequences of using big data as decision support," *Journal of Decision Systems*, 27, 1–18.

GAMAGE, P. (2016): "New development: Leveraging 'big data' analytics in the public sector," *Public Money and Management*, 36, 385–390.

GREGOR, S. AND A. HEVNER (2013): "Positioning and presenting design science research for maximum impact," *MIS Quarterly*, 37, 337–356.

GREGOR, S. AND D. JONES (2007): "The anatomy of a design theory," *Journal of the Association for Information Systems*, 8, 312–335.

GUPTA, S., N. ALTAY, AND Z. LUO (2019): "Big data in humanitarian supply chain management: a review and further research directions," *Annals of Operations Research*, 283, 1153–1173.

GUTIERREZ, C. (2021): "Diversifying philanthropy: Increasing diversity in private foundations," *The Nonprofit Quarterly*.

HACKLER, D. AND G. SAXTON (2007): "The strategic use of information technology by nonprofit organizations: Increasing capacity and untapped potential," *Public Administration Review*, 67, 474–487.

HAIR, J. F., D. J. ORTINAU, AND D. E. HARRISON (2010): *Essentials of marketing research*, vol. 2, New York, United States: McGraw-Hill.

HENRIKSEN-BULMER, J., S. FAILY, AND S. JEARY (2019): *Implementing GDPR in the charity sector: A case study*, International Summer School, Vienna, Austria: Springer International Publishin, 173–188.

HEVNER, R. A., S. MARCH, J. PARK, AND S. RAM (2004): "Design science in information systems research," *Management Information Systems Quarterly*, 28, 1–75.

HILBE, J. M. (2009): *Logistic Regression Models*, London, United Kingdom: CRC Press LLC.

HOCKINGS, M., S. STOLTON, AND N. DUDLEY (2000): "Evaluating effectiveness," *A Framework for Assessing the Management of Pro-*

tected Areas. Reihe: World Commission on Protected Areas (WCPA): Best Practice Protected Area Guidelines Series, 6.

HORVÁTH, I. (2000): "Conceptual design: inside and outside," in *Proceedings of the 2nd International Seminar and Workshop on Engineering Design in Integrated Product*, Citeseer, 63–72.

HOU, Y. AND D. WANG (2017): "Hacking with NPOs: Collaborative analytics and broker roles in civic data hackathons," *Proceedings of the ACM on Human-Computer Interaction*, 1, 1–16.

ISAKSEN, H., M. IVERSEN, J. KAASBLL, AND C. KANJO (2017): "Design of Tooltips for Data Fields," in *Design, user experience, and usability: understanding users and contexts*, ed. by A. Marcus and W. Wang, Springer International Publishing, 63–78.

JIANG, G. AND W. WANG (2017): "Error estimation based on variance analysis of k-fold cross-validation," *Pattern Recognition*, 69, 94–106.

JOHANNESSON, P. AND E. PERJONS (2014): *An introduction to design science*, Switzerland: Springer.

JOHNSON, M. P. (2015): "Data, analytics and community-based organizations: Transforming data to decisions for community development," *I/S: A Journal of Law and Policy for the Information Society*, 11, 49.

KALEMA, B. M. AND M. MOKGADI (2017): "Developing countries organisations' readiness for Big Data analytics," *Problems and Perspectives in Management*, 15, 260–270.

KASHIF, M. AND A. ZARKADA (2015): "Value co-destruction between customers and frontline employees: A social system perspective," *International Journal of Bank Marketing*, 33, 672–691.

KASSEN, M. (2018): "Adopting and managing open data: Stakeholder perspectives, challenges and policy recommendations," *Aslib Journal of Information Management*, 70, 518–537.

KAUR, S., D. MALIK, AND V. AGGARWAL (2017): "Big data analytics for non-profit organizations," *International Journal of Big Data Intelligence*, 4, 215–226.

KHOO-LATTIMORE, C., P. MURA, AND R. YUNG (2019): "The time has come: a systematic literature review of mixed methods research in tourism," *Current Issues in Tourism*, 22, 1531–1550.

KLIEVINK, B., B.-J. ROMIJN, S. CUNNINGHAM, AND H. DE BRUIJN (2017): "Big data in the public sector: Uncertainties and readiness," *Information Systems Frontiers*, 19, 267–283.

KLINE, A. AND S. DOLAMORE (2020): "Understanding data-driven organizational culture: A case study of family league of Baltimore," *Journal of Technology in Human Services*, 38, 247–270.

KOROLOV, R., J. PEABODY, A. LAVOIE, S. DAS, M. MAGDON-ISMAIL, AND W. WALLACE (2016): "Predicting charitable donations using social media," *Social Network Analysis and Mining*, 6.

LAWRENCE, C., T. TUUNANEN, AND M. MYERS (2010): *Extending design science research methodology for a multicultural world*, vol. 318, Berlin, Germany.

LEWIS, M. (2021): "What Are Nonprofit Organizations (NPOs)?" *Investopedia*.

LI, C. AND Y. WU (2019): "Understanding voluntary intentions within the theories of self-determination and planned behaviour," *Journal of Nonprofit Public Sector Marketing*, 31, 378–389.

LI, W., J. HE, Z. CHEN, AND Z. ZHANG (2018): "Customer service automation with machine learning: Survey and future directions," *IEEE Access*.

LIAW, A. AND M. WIENER (2002): "Classification and regression by Random Forest," *R News*, 2, 18–22.

LITOFCENKO, J., D. KARNER, AND F. MAIER (2020): "Methods for Classifying Nonprofit Organizations According to their Field of Activity: A Report on Semi-automated Methods Based on Text," *Voluntas*, 31, 227–237.

LNĚNIČKA, M. AND J. KOMÁRKOVÁ (2015): "The performance efficiency of the virtual Hadoop using open big data," *Scientific Papers of the University of Pardubice, Series D: Faculty of Economics and Administration*, 22, 88–100.

LOTT, R. AND D. BRINKERHOFF (2008): "The impact of donor recognition and appreciation on donor behaviour," *Nonprofit and Voluntary Sector Quarterly*, 37, 314–333.

MAHMOUD, M. A. AND B. YUSIF (2012): "Market orientation, learning orientation, and the performance of nonprofit organisations

(NPOs)," *International Journal of Productivity and Performance Management*, 61, 624–652.

MALASI, A. (2021): "All you need to know to build your first shiny app," https://towardsdatascience.com/all-you-need-to-know-to-build-your-first-shiny-app-653603fd80d9.

MATHIEU, B., H. SEBASTIEN, AND J. MATHIEU (2009): "Gephi: An open source software for exploring and manipulating networks," in *Third International AAAI Conference on Weblogs and Social Media*, Association for the Advancement of Artificial Intelligence, 361–362.

MATLOFF, N. (2011): *The art of R programming*, San Francisco, United States: No Starch Press.

MAXWELL, N. L., D. ROTZ, AND C. GARCIA (2016): "Data and decision making: same organization, different perceptions; different organizations, different perceptions," *American Journal of Evaluation*, 37, 463–485.

MAYER, L. H. (2019): "The promises and perils of using big data to regulate nonprofits," *Washington Law Review*, 94, 1281–1336.

MAYNARD, S., F. BURSTEIN, AND D. ARNOTT (2001): "A Multifaceted decision support system evaluation approach," *Journal of Decision Systems*, 10, 395–428.

MCKINNEY, M. AND S. CHATTERJEE (2008): "Optimizing fundraising campaigns using donor segmentation," *Nonprofit Management & Leadership*, 18, 327–344.

METH, H., B. MUELLER, AND A. MAEDCHE (2015): "Designing a requirement mining system," *Journal of the Association for Information Systems*, 16, 799–837.

MISHRA, D., A. GUNASEKARAN, T. PAPADOPOULOS, AND S. J. CHILDE (2018): "Big Data and supply chain management: a review and bibliometric analysis," *Annals of Operations Research*, 270, 313–336.

MOHAMED, A., M. K. NAJAFABADI, Y. B. WAH, E. A. K. ZAMAN, AND R. MASKAT (2020): "The state of the art and taxonomy of big data analytics: view from new big data framework," *Artificial Intelligence Review*, 53, 989–1037.

MOHANTY, S. AND B. PRADHAN (2019): "AI-Enabled Systems: A comprehensive review," *IEEE Access*, 7, 116051–116082.

MONGODB (n.d.): "Build the next big thing," www.mongodb.com/cloud/atlas.

MONTALVO-GARCIA, J., J. B. QUINTERO, AND B. MANRIQUE-LOSADA (2020): *Crisp-dm/smes: A data analytics methodology for non-profit smes*, Springer, vol. 1041, 449–457.

MT, A. (2017): "The Effect of Usability and Information Quality on Decision Support Information System (DSS)," *Arts and Social Sciences Journal*, 8.

NABAR, R. (2021): "A comparative study of machine learning models for fundraising success," Thesis.

NARENDRA, K. S. (2016): *Chapter two - hierarchical adaptive control of rapidly time-varying systems using multiple models*, Butterworth-Heinemann, 33–66.

NASER, M. Z. AND A. H. ALAVI (2021): "Error metrics and performance fitness indicators for Artificial Intelligence and Machine Learning in engineering and sciences," *Architecture, Structures and Construction*.

NIELSEN, J. (2010): "Usability 101: Introduction to usability," *Nielsen Norman Group*.

OSHIRO, T. M., P. S. PEREZ, AND J. A. BARANAUSKAS (2012): "How many trees in a Random Forest?" in *Machine Learning and Data Mining in Pattern Recognition*, ed. by P. Perner, Berlin, Heidelberg: Springer Berlin Heidelberg, 154–168.

PAGE, M. J., J. E. MCKENZIE, P. M. BOSSUYT, I. BOUTRON, T. C. HOFFMANN, C. D. MULROW, L. SHAMSEER, J. M. TETZLAFF, E. A. AKL, S. E. BRENNAN, R. CHOU, J. GLANVILLE, J. M. GRIMSHAW, A. HRBJARTSSON, M. M. LALU, T. LI, E. W. LODER, E. MAYO-WILSON, S. MCDONALD, L. A. MCGUINNESS, L. A. STEWART, J. THOMAS, A. C. TRICCO, V. A. WELCH, P. WHITING, AND D. MOHER (2021): "The PRISMA 2020 statement: An updated guideline for reporting systematic reviews," *International Journal of Surgery*, 88, 105906.

PALLANT, J. (2020): *SPSS survival manual: A step-by-step guide to data analysis using IBM SPSS*, London, United Kingdom: Routledge.

PARSA, I. (19898): "KDD Cup 1998 Data," American Association for Artificial Intelligence.

PATEL, B., S. ROY, D. BHATTACHARYYA, AND T.-H. KIM (2017): "Necessity of big data and analytics for good e-governance," *International Journal of Grid and Distributed Computing*, 10, 11–20.

PEFFERS, K., C. E. GENGLER, AND T. TUUNANEN (2003): "Extending critical success factors methodology to facilitate broadly participative Information Systems planning," *Journal of Management Information Systems*, 20, 51–85.

PEFFERS, K. AND T. TUUNANEN (2005): "Planning for IS applications: a practical, information theoretical method and case study in mobile financial services," *Information Management*, 42, 483–501.

PEFFERS, K., T. TUUNANEN, M. ROTHENBERGER, AND S. CHATTERJEE (2007): "A design science research methodology for information systems research," *Journal of Management Information Systems*, 24, 45–77.

PENG, C.-Y. J., K. L. LEE, AND G. M. INGERSOLL (2002): "An introduction to logistic regression analysis and reporting," *The journal of educational research*, 96, 3–14.

PICKERING, C. AND J. BYRNE (2014): "The benefits of publishing systematic quantitative literature reviews for PhD candidates and other early-career researchers," *Higher Education Research and Development*, 33, 534–548.

POWELL, L. J. AND H. WITTMAN (2017): “Farm to school in British Columbia: mobilizing food literacy for food sovereignty,” *Agriculture and Human Values*, 35, 193–206.

PRAKASH, M. AND G. SINGARAVEL (2015): “An approach for prevention of privacy breach and information leakage in sensitive data mining,” *Computers and Electrical Engineering*, 45, 134–140.

PYNE, S., P. RAO, AND S. B. RAO (2016): *Big Data Analytics : Methods and Applications*, New Delhi, India: Springer (India) Private Limited.

RATHI, D. AND L. M. GIVEN (2017): “Non-profit organizations’ use of tools and technologies for knowledge management: a comparative study,” *Journal of Knowledge Management*, 21, 718–740.

RHYN, M. AND I. BLOHM (2017a): “Combining collective and Artificial Intelligence: Towards a design theory for decision support in crowdsfuncding,” .

——— (2017b): “A Machine Learning approach for classifying textual data in crowdsourcing,” in *13th International Conference on Wirtschaftsinformatik*, 1–15.

RHYN, M., N. LEICHT, I. BLOHM, AND J. M. LEIMEISTER (2020): “Opening the black box: How to design Intelligent decision support systems for crowdsourcing,” in *15th International Conference on Wirtschaftsinformatik*, 1–15.

RUTKIN, A. (2015): “Big data aims to boost New York blood donations,” *New Scientist*, 225, 19.

RYZHOV, I. O., B. HAN, AND J. BRADIĆ (2016): "Cultivating disaster donors using data analytics," *Management Science*, 62, 849–866.

SALAMON, L. M. AND H. K. ANHEIER (1996): *The international classification of nonprofit organizations: ICNPO-Revision 1, 1996*, Baltimore, Maryland, United States: The Johns Hopkins Institute for Policy Studies.

SALDAÑA, J. (2014): "Coding and analysis strategies," *The Oxford handbook of qualitative research*, 581–605.

SARGEANT, A. AND E. JAY (2014): *Fundraising management: analysis, planning and practice*, London and New York: Routledge Taylor Francis Group, third ed.

SHAH, N., Z. IRANI, AND A. M. SHARIF (2017): "Big data in an HR context: Exploring organisational change readiness, employee attitudes and behaviours," *Journal of Business Research*, 70, 366–378.

SHAPIRO, S. J. AND V. OYSTRICK (2018): "Three steps toward sustainability: Spreadsheets as a data-analysis system for non-profit organizations," *Canadian Journal of Program Evaluation*, 33.

SHEHU, E., A. C. LANGMAACK, E. FELCHLE, AND M. CLEMENT (2015): "Profiling donors of blood, money, and time: A simultaneous comparison of the German population," *Nonprofit Management and Leadership*, 25, 269–295.

SHUKLA, A. K., P. K. MUHURI, AND A. ABRAHAM (2020): "A bibliometric analysis and cutting-edge overview on fuzzy techniques in

Big Data," *Engineering Applications of Artificial Intelligence*, 92, 103625.

SILVER, M. S. (1988): "User perceptions of DSS restrictiveness: An experiment," in *Proceedings of the Twenty-First Annual Hawaii International Conference on System Sciences.*, IEEE, vol. 3, 116–124.

SILVER, MARK S (1991): "Decisional guidance for computer-based decision support," *MIS quarterly*, 105–122.

SMITH, M. (2021): "Organizing effective fundraising events for non-Profits," *Charity Times*, 19, 17–22.

TABACHNICK, B. G., L. S. FIDELL, AND J. B. ULLMAN (2007): *Using multivariate statistics*, vol. 5, Boston, USA: Pearson.

TE, N. (2019): "Study helps you better understand donor behaviours,"
.

TECHPEDIA (2020): "Demographic Data," www.techopedia.com/definition/30326/demographic-data.

TEJADA, C. (2020): "The benefits of corporate sponsorships for non-Profits," *Philanthropy Today*, 25, 12–18.

THOMAS, A. R. (2020): *Data Analytics: Models and algorithms for Intelligent Data Analysis*, vol. Third edition of *Lehrbuch*, Wiesbaden, Germany: Springer Vieweg.

TRIVEDI, G. (2019): “Visualization and scientometric mapping of global agriculture big data research,” *Library Philosophy and Practice*.

TRIVERS, R. L. (1971): “The evolution of reciprocal altruism,” *The Quarterly Review of Biology*, 46, 35–57.

VAISHNAVI, V., B. KUECHLER, AND S. PETTER (2019): “Design science research in Information Systems,” *Association for Information Systems*, 1–62.

VENABLES, W. N. AND B. D. RIPLEY (2002): *Modern applied statistics with S*, United States: Springer.

VERNON, J., I. RAMSAY, AND G. STAPLEDON (2001): *Federal Law Review*, 29, 177–218.

WALLS, J. G., G. R. WIDMEYER, AND O. A. EL SAWY (1992): “Building an information system design theory for vigilant EIS,” *Information systems research*, 3, 36–59.

WALTERS, R. (2020): “A Guide to AI-Enabled Systems,” *IEEE Industrial Electronics Magazine*, 14, 10–18.

WANG, C. S., S. L. LIN, T. H. CHOU, AND B. Y. LI (2019): “An integrated data analytics process to optimize data governance of non-profit organization,” *Computers in Human Behavior*, 101, 495–505.

WANG, D. (2020): “Intellectual property: The benefits of royalties and other income streams,” *The Balance*.

WANG, R. Y. AND D. M. STRONG (1996): "Beyond Accuracy: what data quality means to data consumers," *Journal of Management Information Systems*, 12, 5–33.

WANG, Z., R. YAN, Q. CHEN, AND R. XING (2010): "Data Mining in nonprofit organizations, government agencies, and other institutions," *International Journal of Information Systems in the Service Sector*, 2, 42–52.

WATSON, R. T. (2015): "Beyond being Systematic in Literature Reviews in IS," *Journal of Information Technology*, 30, 185–187.

WEIBELZAHL, S., A. PARAMYTHIS, AND J. MASTHOFF (2020): "Evaluation of adaptive systems," Association for Computing Machinery.

WEINGER, A. (2019): "The importance of donor data and how to use it effectively," www.philanthropynewsdigest.org/features/the-sustainable-nonprofit/the-importance-of-donor-data-and-how-to-use-it-effectively.

WEISMAN, R. (2020): "How to generate income for non-Profits," *Nonprofit Quarterly*, 22, 30–37.

WESTON, C., T. GANDELL, J. BEAUCHAMP, L. MCALPINE, C. WISEMAN, AND C. BEAUCHAMP (2001): "Analyzing interview data: The development and evolution of a coding system," *Qualitative Sociology*, 24, 381–400.

WHITE, K., R. HABIB, AND D. W. DAHL (2020): "A review and framework for thinking about the drivers of prosocial consumer behavior," *Journal of the Association for Consumer Research*, 5, 2–18.

WITJAS-PAALBERENDS, E. R., L. P. M. VAN LAARHOVEN, L. H. M. VAN DE BURGWAL, J. FEILZER, J. DE SWART, E. CLAASSEN, AND W. T. M. JANSEN (2018): "Challenges and best practices for big data-driven healthcare innovations conducted by profit–non-profit partnerships–a quantitative prioritization," *International Journal of Healthcare Management*, 11, 171–181.

WONG, T.-T. (2015): "Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation," *Pattern Recognition*, 48, 2839–2846.

YADAV, S. AND S. SHUKLA (2016): "Analysis of k-fold cross-validation over hold-out validation on colossal datasets for quality classification," in *2016 IEEE 6th International conference on advanced computing (IACC)*, IEEE, 78–83.

YANG, E. C. L., C. KHOO-LATTIMORE, AND C. ARCODIA (2017): "A systematic literature review of risk and gender research in tourism," *Tourism Management*, 58, 89–100.

ZAREI, E. AND A. JABBARZADEH (2019): "Knowledge management and social media: A scientometrics survey," *International Journal of Data and Network Science*, 3, 359–378.

ZHANG, Y., A. L. PORTER, Z. HU, Y. GUO, AND N. C. NEWMAN (2014): ""Term clumping" for technical intelligence: A case study on dye-sensitized solar cells." *Technological Forecasting and Social Change*, 85, 26–39.

ZHANG, Y., G. ZHANG, D. ZHU, AND J. LU (2017): "Scientific evolutionary pathways: Identifying and visualizing relationships for scientific topics," *Journal of the Association for Information Science and Technology*, 68, 1925–1939.

ZHOU, X., T. WILDSCHUT, C. SEDIKIDES, K. SHI, AND C. FENG (2012): "Nostalgia: The gift that keeps on giving," *Journal of Consumer Research*, 39, 39–50.

ĐURIĆ, G., MITROVIĆ, G. VOROTOVIĆ, I. BLAGOJEVIC, AND M. VASIĆ (2016): "Developing self-modifying code model," *Istrazivanja i projektovanja za privredu*, 14, 239–247.

APPENDIX A

APPENDICES

A.1 Keywords Search

A.1.1 Keywords Search String Used in the Systematic Review

This appendix shows the keywords search string used in the three databases (Scopus, Web of Science and ProQuest) to collect the relative articles to be reviewed for the systematic review. Noticeably, when using these string, the results of search may vary from our results as more documents were published till the time of writing this thesis. This string would help researchers whom interested in conducting further related research.

Database	Search string
Scopus	((TITLE-ABS-KEY (“Big Data”) OR TITLE-ABS-KEY (“Data-driven”) OR TITLE-ABS-KEY (“Data analy*”) OR TITLE-ABS-KEY (“Predictive analy”))) AND ((TITLE-ABS-KEY (“Nonprofi*”) OR TITLE-ABS-KEY (“Non-Profi*”) OR TITLE-ABS-KEY (“non for profit*”) OR TITLE-ABS-KEY (“Charit*”) OR TITLE-ABS-KEY (“Fundraising”) OR TITLE-ABS-KEY (“Donations”) OR TITLE-ABS-KEY (“Donors”)))
ProQuest: It does not offer a searching field in the keywords	ab(((“Big Data”) OR (“Data-driven”) OR (“Data analy*”) OR (“Predictive analy”)) AND ((“Nonprofi*”) OR (“Non-Profi*”) OR (“non for profit*”) OR (“Charit*”) OR (“Fundraising”) OR (“Donations”) OR (“Donors”))) OR ti(((“Big Data”) OR (“Data-driven”) OR (“Data analy*”) OR (“Predictive analy”)) AND ((“Nonprofi*”) OR (“Non-Profi*”) OR (“non for profit*”) OR (“Charit*”) OR (“Fundraising”) OR (“Donations”) OR (“Donors”)))
Web of Science	ti=(((“Big data”) OR (“Data-driven”) OR (“Data analy*”) OR (“Predictive analy*”))) AND ti=(((“Nonprofi*”) OR (“Non-Profi*”) OR (“Non-for-Profi *”) OR (“Charit*”) OR (“Fundraising”) OR (“Donations”) OR (“Donors”))) OR ab=(((“Big data”) OR (“Data-driven”) OR (“Data analy*”) OR (“Predictive analy*”))) AND ab=(((“Nonprofi*”) OR (“Non-Profi*”) OR (“Non-for-Profi *”) OR (“Charit*”) OR (“Fundraising”) OR (“Donations”) OR (“Donors”))) OR ak=(((“Big data”) OR (“Data-driven”) OR (“Data analy*”) OR (“Predictive analy*”))) AND ak=(((“Nonprofi*”) OR (“Non-Profi*”) OR (“Non-for-Profi *”) OR (“Charit*”) OR (“Fundraising”) OR (“Donations”) OR (“Donors”)))

A.1.2 Keywords Search String Used in the Research Survey

This appendix lists the search terms that were used to locate relevant papers for the research survey across the three databases (Scopus and Web of Science). Notably, when using these strings, the search results may differ from our results because more documents have been released at the time this thesis was being written. Researchers who are interested in undertaking additional related study could benefit from using this string.

Field Tags	Search String
Scopus: TITLE-ABS-KEY	TITLE-ABS-KEY ("Data analy*" OR "Data-driven" OR "Predictive Analy*" OR "Analy*" OR "Big Data" OR "large*scale data" OR "Open Data" OR "natur* language process*" OR "NLP" OR "Machine Transla*" OR "lexical analys*" OR "Information extract*" OR "knowledge Graph" OR "Feature Select*" OR "Natur* language generat*" OR "NLG" OR "Natur* language interact*" OR "mode identif*" OR "virtual personal assistant" OR "Text to Speech" OR "sentiment analys*" OR "data mine*" OR "text mine*" OR "document mine*" OR "linguistic mine*" OR "data analys*" OR "text analys*" OR "document analys*" OR "linguistic analys*" OR "data Process*" OR "text Process*" OR "document Process*" OR "linguistic Process*" OR "Text Classif*" OR "Text Cluster*" OR "SYNTACTIC ANALYS*" OR "automatic summarize" OR "information filter*" OR "Expert System" OR "Decision Support System" OR "Model-based" OR "intelligen* system" OR "multi-agent system" OR "knowledge Management" OR "knowledge Represent*" OR "Semantic Net*" OR "Predicate logic" OR "knowledge engineer*" OR "Decision Tree" OR "Linear Regression" OR "BP Neural Network" OR "neural comput*" OR "Artificial Neural Network" OR "Bayesian Classification" OR "Support Vector Machine" OR "Logistic Regression" OR "Spectral Clustering" OR "Dimensionality Reduction" OR "Classification Accuracy" OR "Fuzzy Clustering" OR "Association Rules" OR "Combined Training" OR "Deep Learning" OR "Machine Learning" OR "Reinforcement Learning" OR "depth learning") AND TITLE-ABS-KEY ("Nonprofi*" OR "Non*Profi*" OR "no*for*profit*" OR "Charit*" OR "third sector" OR "giving dector" OR "voluntary sector*" OR "voluntary sector*" OR "civil society giving" OR "nongovernmental" OR "organization philanthropy" OR "social capital")
WoS: the search field changed three times using different tags which are: (TI= Title, AB = Abstract AK = Author Keywords)	TI=("Data analy*" OR "Data-driven" OR "Predictive Analy*" OR "Analy*" OR "Big Data" OR "large*scale data" OR "Open Data" OR "natur* language process*" OR "NLP" OR "Machine Transla*" OR "lexical analys*" OR "Information extract*" OR "knowledge Graph" OR "Feature Select*" OR "Natur* language generat*" OR "NLG" OR "Natur* language interact*" OR "mode identif*" OR "virtual personal assistant" OR "Text to Speech" OR "sentiment analys*" OR "data mine*" OR "text mine*" OR "document mine*" OR "linguistic mine*" OR "data analys*" OR "text analys*" OR "document analys*" OR "linguistic analys*" OR "data Process*" OR "text Process*" OR "document Process*" OR "linguistic Process*" OR "Text Classif*" OR "Text Cluster*" OR "SYNTACTIC ANALYS*" OR "automatic summarize" OR "information filter*" OR "Expert System" OR "Decision Support System" OR "Model-based" OR "intelligen* system" OR "multi-agent system" OR "knowledge Management" OR "knowledge Represent*" OR "Semantic Net*" OR "Predicate logic" OR "knowledge engineer*" OR "Decision Tree" OR "Linear Regression" OR "BP Neural Network" OR "neural comput*" OR "Artificial Neural Network" OR "Bayesian Classification" OR "Support Vector Machine" OR "Logistic Regression" OR "Spectral Clustering" OR "Dimensionality Reduction" OR "Classification Accuracy" OR "Fuzzy Clustering" OR "Association Rules" OR "Combined Training" OR "Deep Learning" OR "Machine Learning" OR "Reinforcement Learning" OR "depth learning") AND TI = ("Nonprofi*" OR "Non*Profi*" OR "no*for*profit*" OR "Charit*" OR "third sector" OR "giving dector" OR "voluntary sector*" OR "voluntary sector*" OR "civil society giving" OR "nongovernmental" OR "organization philanthropy" OR "social capital")

A.1.3 Relevancy of the Collected Article from the Literature

The aim of this appendix is to present each relevant paper collected in the literature. This appendix is an essential for the research to obtain such inspirations and lessons from the expriments done in the literature.

No.	Article Name	Authors	Relevancy to Data Analytics Applications in NPOs
1.	Adopting and managing open data: Stakeholder perspectives, challenges and policy recommendations.	(Kassen 2018)	The study helps to expand awareness of open data in NPOs. Analysing donor behaviour could be one several applications in NPOs to use the data analytics applications.
2.	An approach for prevention of privacy breach and information leakage in sensitive data mining.	(Prakash & Singaravel 2015)	The study helps in supporting proposed solutions over using data analytics in NPOs. Therefore, when analysing donor behaviour, some concerns of privacy may rise by NPOs' stakeholders.
3.	An associative engines based approach supporting collaborative analytics in the Internet of cultural things	(Chianese et al. 2017)	This study can benefit in showing an example of different technical terms used in organisations to build a system of analysing large contents of data in NPOs.
4.	An integrated data analytics process to optimize data governance of non-profit organization	(Wang et al. 2019)	The study helps by presenting a usage of the Knowledge Discovery in Databases (KDD) process in scientific decision-making methods and use the four main pre-processing, collection, data mining and assessment processes. Similarly, analysing donor behaviour can a similar approach of using a decision support system.
5.	Big data creating new knowledge as support in decision-making: practical examples of big data use and consequences of using big data as decision support	(Fredriksson 2018)	The provided examples in this paper can benefit to learn more lessons of applying big data analytics in NPOs, such as analysing donor behaviour.
6.	Challenges and best practices for big data-driven healthcare innovations conducted by profit–non-profit partnerships–a quantitative prioritization.	(Witjas-Paalberends et al. 2018)	The interviews and surveys with key opinion leaders would also provide more lessons about challenges of applying big data analytics in NPOs.
7.	Clustering consumers based on trust, confidence and giving behaviour: Data-driven model building for charitable involvement in the Australian not-for-profit sector	(De Vries, Reis & Moscato 2015)	This study presents a novel unsupervised clustering technique (MST-kNN) which can be a beneficial lesson when analysing donor behaviour in NPOs.
8.	Crisp-dm/smes: A data analytics methodology for non-profit smes	(Montalvo-Garcia, Quintero & Manrique-Losada 2020)	This paper presents an example of showing an example of managing information sources in NPOS. One benefit is to manage donors' information or data through the approach used in this study.
9.	Cultivating disaster donors using data analytics	(Ryzhov, Han & Bradić 2016)	The study is useful to identify designs that exert a significant impact on the outcome of a fundraising campaign by analysing large scale datasets. So, it is another tangle of analysing donor behaviour in NPOs.

10.	Data analytics and community-based organizations: Transforming data to decisions for community development	(Johnson 2015)	From a survey of existing technologies data and decision science, this paper can benefit in refencing to different applications of analysing donor behaviour in NPOs.
11.	Data and Decision Making: Same Organization, Different Perceptions; Different Organizations, Different Perceptions	(Maxwell, Rotz & Garcia 2016)	The study provides insights of applying decision support systems in NPOs. One benefit is to manage donors' information or data through the approach used in this study.
12.	Data Mining in Non-profit Organizations, Government Agencies, and Other Institutions	(Wang et al. 2010)	A summary of the current and exist data mining technologies with presenting examples. One benefit is to analyse donors' information or data through the examples provided in this study.
13.	Decision makers and criteria for patient discharge - A qualitative study	(Eigner et al. 2017)	A lesson of this paper is to understand patients using a decision-making approach.
14.	Disempowered by data: Non-profits, social enterprises, and the consequences of data-driven work	(Bopp, Harmon & Voida 2017)	The paper is beneficial in identifying the critical impacts of data practices in NPOs. This study supports some relevant arguments about applying data analytics in NPOs.
15.	Hacking with NPOs: Collaborative analytics and broker roles in civic data hackathons	(Hou & Wang 2017)	The paper is beneficial in identifying the critical impacts of data practices in NPOs. This study supports some relevant arguments about applying data analytics in NPOs.
16.	Implementing GDPR in the charity sector: A case study	(Henriksen-Bulmer, Faily & Jeary 2019)	The study helps in supporting proposed solutions over using data analytics in NPOs. Therefore, when analysing donor behaviour, some concerns of privacy may rise by NPOs' stakeholders.
17.	Methods for Classifying Non-profit Organizations According to their Field of Activity: A Report on Semi-automated Methods Based on Text.	(Litofcenko, Karner & Maier 2020)	The paper presents an important application of classifying NPOs, however; it can help in classifying donor behaviour based on the nature of donations.
18.	Necessity of big data and analytics for good e-governance.	(Patel et al. 2017)	The paper is beneficial in identifying the critical impacts of data practices in NPOs. This study supports some relevant arguments about applying data analytics in NPOs.
19.	Non-profit organizations' use of tools and technologies for knowledge management: a comparative study.	(Rathi & Given 2017)	The study provides insights of applying decision support systems in NPOs. One benefit is to manage donors' information or data through the approach used in this study.
20.	Predicting heart transplantation outcomes through data analytics	(Dag et al. 2017)	This paper is relevant to analysing donor behaviour in predicting organs donations.
21.	Solid organ donation after death in the United States:	(Bambha et al. 2020)	This paper is relevant to analysing donor behaviour in predicting organs donations.

	Data? Driven messaging to encourage potential donors.		
22.	The performance efficiency of the virtual Hadoop using open big data.	(Lněnička & Komárková 2015)	This paper provides some highlights on the importance of big data analytics in NPOs.
23.	The promises and perils of using big data to regulate non-profits.	(Mayer 2019)	The paper is beneficial in identifying the critical impacts of data practices in NPOs. This study supports some relevant arguments about applying data analytics in NPOs.
24.	Three Steps Toward Sustainability: Spreadsheets as a Data-Analysis System for Non-Profit Organizations	(Shapiro & Oystrick 2018)	The study helps in supporting proposed solutions over using data analytics in NPOs. Therefore, when analysing donor behaviour, some concerns of privacy may rise by NPOs' stakeholders.
25.	Understanding Data-Driven Organizational Culture: A Case Study of Family League of Baltimore	(Kline & Dolamore 2020)	The paper is beneficial in identifying the critical impacts of data practices in NPOs. This study supports some relevant arguments about applying data analytics in NPOs.

APPENDIX

B

APPENDICES

B.1 Ethics Approvals

B.1.1 "Approval of Ethics for a research project: "AI-enabled decision support system for analysing donor behaviour"

This ethical approval was obtained as a research requirement of collecting data such as interviews. The the interviews are done for the purpose of evaluations by involving experts in data science, decision-makers, big data analysts, NPOs managers and social science researchers to evaluate the conceptual design of the AI-enabled DSS in Iteration one and, and the artefact of AI-enabled DSS in Iteration three. The ethical approval was classified as low risk by under the local research office in Faculty of Engineering and Information Technology, University of Technology Sydney.

From: research.ethics@uts.edu.au
To: Mukesh Prasad; Idrees Alsolbi
Subject: Your ethics application has been approved as low risk - ETH21-5802
Date: Monday, 22 March 2021 8:24:38 AM
Attachments: Ethics Application.pdf

Dear Applicant,

Re: ETH21-5802 - "Ethics Application for a research project "AI-enabled decision support system for analysing donors behaviours""

Your local research office has reviewed your application and agreed that it now meets the requirements of the National Statement on Ethical Conduct in Human Research (2007) and has been approved on that basis. You are therefore authorised to commence activities as outlined in your application, subject to any conditions detailed in this document.
You are reminded that this letter constitutes ethics approval only. This research project must also be undertaken in accordance with all UTS policies and guidelines including the Research Management Policy.
Your approval number is UTS HREC REF NO. ETH21-5802
Approval will be for a period of five (5) years from the date of this correspondence subject to the submission of annual progress reports.
The following standard conditions apply to your approval:

- Your approval number must be included in all participant material and advertisements.
- Any advertisements on Staff Connect without an approval number will be removed.
- The Principal Investigator will immediately report anything that might warrant review of ethical approval of the project to the Ethics Secretariat (Research.Ethics@uts.edu.au).
- The Principal Investigator will notify the UTS HREC of any event that requires a modification to the protocol or other project documents, and submit any required amendments prior to implementation. Instructions on how to submit an amendment application can be found here.
- The Principal Investigator will promptly report adverse events to the Ethics Secretariat. An adverse event is any event (anticipated or otherwise) that has a negative impact on participants, researchers or the reputation of the University. Adverse events can also include privacy breaches, loss of data and damage to property.
- The Principal Investigator will report to the UTS HREC annually and notify the HREC when the project is completed at all sites.
- The Principal Investigator will notify the UTS HREC of any plan to extend the duration of the project past the approval period listed above through the progress report.
- The Principal Investigator will obtain any additional approvals or authorisations as required (e.g. from other ethics committees, collaborating institutions, supporting organisations).
- The Principal Investigator will notify the UTS HREC of his or her inability to continue as Principal Investigator including the name of and contact information for a replacement.

This research must be undertaken in compliance with the Australian Code for the Responsible Conduct of Research and National Statement on Ethical Conduct in Human Research.

You should consider this your official letter of approval.

If you have any queries about this approval, or require any amendments to your approval in future, please do not hesitate to contact your local research office or the Ethics Secretariat.

Ref: 12a

B.1.2 Participant Information Sheet and Consent Form

This sheet is inform participants essential information about the process of interviews, and how their answers will be collected, analysed, and stored.

Ethics Approval Number: UTS HREC REF NO. ETH21-5802

PARTICIPANT INFORMATION SHEET

WHO IS DOING THE RESEARCH?

My name is **Idrees Alsolbi** and I am a **student** at UTS. My supervisors are **Dr. Mukesh Prasad, Prof. Renu Agarwal and Prof. Bhuvan Unhelkar**

WHAT IS THIS RESEARCH ABOUT?

This research is to find out about analysing donor behaviour in nonprofit organisations (NPOs). The researcher presents a design science research project concerned with designing of AI-enabled decision support systems of donor behaviour analysis in NPOs. The research project applies the underlying machine learning techniques, guidance from Information Systems (IS) by applying a design science research methodology, to analyse donors' attitudes towards donations and volunteering activities for more effective and efficient decision making in the nonprofit sector.

FUNDING

There is no fund received for this project.

WHY HAVE I BEEN ASKED?

You have been invited to participate in this study because you are one of the chosen data scientists/decision-making teams. Your contact details were obtained by email from the online site of your organisation.

IF I SAY YES, WHAT WILL IT INVOLVE?

If you decide to participate, I will invite you to

- **I will ask you to be involved in a 30-minutes semi-structured interview to test the functionality of our AI-enabled decision support system for analysing donorbehaviour.**
- **Participate in a 30-minutes semi-structured interviews that will be video recorded and transcribed to test our final version of our Decision Support System for analysing donors' behaviours.**

ARE THERE ANY RISKS/INCONVENIENCE?

Yes, there one risks/inconvenience. The risk might be loosing of interview data which we might record the session again. However, this risk is too low, because we will use back-up solutions, of the interview records.

DO I HAVE TO SAY YES?

Participation in this study is voluntary. It is completely up to you whether or not you decide to take part.

WHAT WILL HAPPEN IF I SAY NO?

If you decide not to participate, it will not affect your relationship with the researchers or the University of Technology Sydney. If you wish to withdraw from the study once it has started, you can do so at any time without having to give a reason, by contacting **[Idrees Alsolbi – via email Idrees.alsolbi@student.uts.edu.au].**

If you withdraw from the study, your samples will be destroyed; the study tapes will be erased; the transcripts will be destroyed.

However, it may not be possible to withdraw your data from the study results if these have already had your identifying details removed.

If you decide to leave the research project, we will not collect additional personal information from you, although personal information already collected will be retained to ensure that the results of the research project can be measured properly and to comply with law. You should be aware that data collected up to the time you withdraw will form part of the research project results. If you do not want them to do this, you must tell them before you join the research project.

CONFIDENTIALITY

By signing the consent form you consent to the research team collecting and using personal information about you for the research project. All this information will be treated confidentially. The data and interviews records will be stored on Stach, a data management system in UTS. Only the researcher and his supervisors will have access to it. Your information will only be used for the purpose of this research project and it will only be disclosed with your permission, except as required by law.

We plan to discuss and publish the results of the research via publications in conferences and journals. In any publication, information will be provided in such a way that you cannot be identified.

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 me on my email: Idrees.alsolbi@student.uts.edu.au or my principle supervisor: Mukesh.Prasad@uts.edu.au

You will be given a copy of this form to keep.

NOTE:
This study has been approved in line with the University of Technology Sydney Human Research Ethics Committee [UTS HREC] guidelines. If you have any concerns or complaints about any aspect of the conduct of this research, please contact the Ethics Secretariat on ph.: +61 2 9514 2478 or email: Research.Ethics@uts.edu.au], and quote the UTS HREC reference number. Any matter raised will be treated confidentially, investigated and you will be informed of the outcome.

UTS UNIVERSITY OF TECHNOLOGY SYDNEY

CONSENT FORM

I ____________________ agree to participate in the **research project** [AI-enabled Decision Support System for Analysing Donors' Behaviours in Nonprofit Organisations], [**UTS HREC approval reference number**: ETH21-5802] being conducted by **[Idrees Alsolbi, Idrees.alsolbi@student.uts.edu.au,]**.

I have read the Participant Information Sheet or someone has read it to me in a language that I understand.

I understand the purposes, procedures and risks of the research as described in the Participant Information Sheet.

I have had an opportunity to ask questions and I am satisfied with the answers I have received.

I freely agree to participate in this research project as described and understand that I am free to withdraw at any time without affecting my relationship with the researchers or the University of Technology Sydney.

I understand that I will be given a signed copy of this document to keep.

I agree to be:
☒ Audio recorded
☐ Video recorded

I agree that the research data gathered from this project may be published in a form that:
☐ Identifies me

I am aware that I can contact **[Idrees Alsolbi]** if I have any concerns about the research.

____ ____/____/____
Name and Signature [participant] Date

_______________*Idrees Alsolbi* _________________________ ____/____/____
Name and Signature [researcher or delegate] Date

APPENDIX C

APPENDICES

C.1 Interviews' Questionnaire

C.1.1 Questionnaires of First Interviews to Evaluate the Conceptual Design of the AI-enabled DSS

The questionnaires in the first interviews are shown in this appendix. These questionnaires are to evaluate the conceptual design of the AI-enabled DSS.

UTS HREC REF NO. ETH21-5802

UTS UNIVERSITY OF TECHNOLOGY SYDNEY

Project Name	AI-enabled decision support system (DSS) for analysing donors' behaviours in non-profit organisations (NPOs).	
Project Team	Idrees Alsolbi	PhD Student
	Dr. Mukesh Prasad	Principle Supervisor
	Prof. Renu Agarwal	Co-supervisor
	Prof. Bhuvan Unhelkar	Co-supervisor
Location	School of Computer Science - University of Technology Sydney	

Instructions:

1. This is a list of questions for participants during our first interview to evaluate our AI-enabled decision support system.
2. The researcher has obtained the required human research ethics approval (**Eth21-5802**) from the University of Technology Sydney.
3. <u>**The participant should sign on the consent form before answering the questions in this interview.**</u>
4. This interview will be recorded for purposes of analysing the interview data. The recording will not be shared with anyone outside of the research team. The results should drive useful insights on the evaluation of the conceptual design.
5. The below list contains formal questions for participants who individuals from NPOs are (data scientists, decision-making systems experts, software engineers, and NPOs managers).
6. The questions will go through 5 stages (Participation, Discovery, Dream, Design and Destiny). The interview should not take more than 20 minutes.

- # Questionnaires

Stage	Script/ Questions
Introduction (2 minute)	• *Thanks for meeting with me. My name is Idrees Alsolbi, and I am supervised by Dr. Mukesh Prasad and Dr. Renu Agarwal, from School of Computer Science, University of Technology Sydney.* • *I'd like to briefly summarize why we're having this interview today. We're trying to evaluate our initial conceptual design of AI-enabled DSS to analyse donors behaviours in NPOs. The conceptual design is a part of our design science framework for designing an AI-enabled DSS for analysing donors behaviours in NPOs. As it is stated in the presentation, the results of this interview should reflect on the conceptual design.* • *Before we start, do you have any questions regarding the introduction of the research (video)?* • *Let us start the questions now.*
Participation (5 minutes)	1. Can you tell me about your experience working in NPOs? (Go to Q.5 if the interviewee has not worked in NPOs). 2. How long have you been working in NPOs? 3. What are the main challenges that face data scientist/Decision makers in NPOs to analyse donors' behaviours? 4. What are the main tasks for data scientists/decision makers in NPOs to analyse donors/volunteers data?
Discovery (5 minutes)	5. Have you ever been involved in designing DSS to analyse donors'? If yes, please explain. 6. Have you ever been involved in analysing donors/volunteers using Machine Learning techniques? If yes, please explain. 7. How would you describe the conceptual design of our AI-enabled DSS to analyse donors behaviours?
Dream (3 minutes)	8. Do you see the mapping of these Design Requirements, Design Principles, and Design Features can achieve our objectives? If yes, please explain.
Design (3 minutes)	9. What Design Requirements/Features/Functions are critical to analyse donors behaviours? And why?
Destiny (3 minutes)	10. What results/analysis do you expect when implementing the AI-enabled DSS to analyse donor behaviours?
Conclusion	Thank you for your collaboration and participation in this interview. I hope we can speak to you in the future for our second interview of the evaluation.

C.1.2 Questionnaires of Second Interviews to Evaluate the Artefact of the AI-enabled DSS

The questionnaires and their according stages in the second interviews are shown in this appendix. These questionnaires are to evaluate the operationality and the functionality of the AI-enabled DSS.

Project Name	AI-enabled decision support system (DSS) for analysing donor behaviour in non-profit organisations (NPOs).	
Project Team	Idrees Alsolbi	PhD Student
	Dr Mukesh Prasad	Principle Supervisor
	Prof. Renu Agarwal	Co-supervisor
	Prof. Bhuvan Unhelkar	Co-supervisor
Location	School of Computer Science - University of Technology Sydney	

Instructions:

1. This is a list of questions for participants during the first interview to evaluate the proposed AI-enabled decision support system.
2. The researcher has obtained the required human research ethics approval (**ETH22-7329**) from the University of Technology Sydney.
3. **<u>The participant should sign the consent form before answering the questions in this interview.</u>**
4. This interview will be recorded for purposes of analysing the interview data. The recording will not be shared with anyone outside of the research team. The results should drive useful insights into evaluating the instantiation of the AI-enabled DSS.
5. The below list contains formal questions for participants from NPOs (data scientists, decision-making systems experts, software engineers, and NPOs managers).
6. The questions will be covered in 5 stages (Participation, Discovery, Dream, Design and Destiny). The interview will take around 20 minutes.

• Questionnaires

Design Requirements	Measurement	Questions
Usability	Engagement, Easy to learn, Efficient, Error tolerant	1. What is your impression of the "interface" of the AI-enabled DSS?
		2. Have you engaged while using the AI-enabled DSS?
		3. How easy was it to use/navigate the components of the AI-enabled DSS?
		4. Were there any errors? Lost in translations? If yes, please specify
		5. What could be improved to enhance the usability of the AI-enabled DSS?
Effectiveness	The ability of a system to meet its specified needs	6. To what extent the AI-enabled DSS is effective for analysing donor behaviour?
		7. Do you find the variety of Machine Learning (ML) techniques effective analysis?
		8. Will you consider the AI-enabled DSS frequently to analyse donor behaviour if you are a decision-maker in an NPO?
		9. Were you able to obtain a result and useful information in a short time?
Operationality	How well is the system operating	10. How well do the components of the AI-enabled DSS operate?
Reducing efforts	Reducing efforts of stakeholders	11. Were you able to understand some of the donor/volunteer behaviour? Can you give an example?
Control	Providing control for stakeholders	12. Were you able to have control and flexibility when using the AI-enabled DSS for analysing donor behaviour?

APPENDIX

D

APPENDICES

D.1 Descriptions of datasets used in this research

D.1.1 Donors Datasets

This appendix is to present some useful information about donors' datasets and some basic statistics of the selected features.

PARALYZED VETERANS OF AMERICA (PVA) DATA DICTIONARY TO ACCOMPANY

KDD-CUP-98

The Second International Knowledge Discovery and
Data Mining Tools Competition

Held in Conjunction with KDD-98

The Fourth International Conference on Knowledge
Discovery and Data Mining
[www.kdnuggets.com] or
[www-aig.jpl.nasa.gov/kdd98] or
[www.aaai.org/Conferences/KDD/1998]

Sponsored by the

American Association for Artificial Intelligence (AAAI)
Epsilon Data Mining Laboratory
Paralyzed Veterans of America (PVA)

The original dataset is available on:
https://www.inf.ed.ac.uk/teaching/courses/irds/miniproject-datasets.html

Background and Objectives

The data set for this year's Cup has been generously provided by the Paralyzed Veterans of America (PVA). PVA is a not-for-profit organization that provides programs and services for US veterans with spinal cord injuries or disease. With an in-house database of over 13 million donors, PVA is also one of the largest direct mail fund raisers in the country. Participants in the '98 CUP will demonstrate the performance of their tool by analyzing the results of one of PVA's recent fund raising appeals. This mailing was sent to a total of 3.5 million PVA donors who were on the PVA database as of June 1997. Everyone included in this mailing had made at least one prior donation to PVA.

The analysis file includes all 191,779 Lapsed donors who received the mailing, with responders to the mailing marked with a flag in the TARGET_B field. The total dollar amount of each responder's gift is in the TARGET_D field.

The average donation amount (in $) among the responses is:
Learning Data Set
Target Variable: Donation Amount (in $) to 97NK
Mailing

```
   N      Mean     Minimum     Maximum
  -----------------------------------------------
```

4843	15.6243444	1.0000000	200.0000000

Validation Data Set
Target Variable: Donation Amount (in $) to 97NK Mailing

N	Mean	Minimum	Maximum
4873	15.6145372	0.3200000	500.0000000

The selected features in this research:

TCODE Donor title code
ODATEDW Origin Date. Date of donor's first gift to PVA YYMM format (Year/Month).
OSOURCE Origin Source
TCODE Donor title code
STATE State abbreviation (a nominal/symbolic field)
ZIP Zipcode (a nominal/symbolic field)
MAILCODE Mail Code " "= Address is OK
B = Bad Address
PVASTATE EPVA State or PVA State
Indicates whether the donor lives in a state served by the organization's EPVA chapter
P = PVA State
E = EPVA State (Northeastern US) Summary statistics are provided for the numeric variables only.
DOB Date of birth (YYMM, Year/Month format.)
MDMAUD The Major Donor Matrix code
The codes describe frequency and amount of giving for donors who have given a $100+ gift at any time in their giving history.
An RFA (recency/frequency/monetary) field.

The (current) concatenated version is a nominal or symbolic field. The individual bytes could separately be used as fields and refer to the following:

First byte: Recency of Giving
C=Current Donor
L=Lapsed Donor
I=Inactive Donor
D=Dormant Donor

2nd byte: Frequency of Giving
1=One gift in the period of recency
2=Two-Four gifts in the period of recency
5=Five+ gifts in the period of recency
HOMEOWNR Home Owner Flag

H = Home owner
U = Unknown
NUMCHLD NUMBER OF CHILDREN
INCOME (we did this categories for easy understanding and reading during the analysis stage)
1= (less than 5000), 2= (5000 - 4799), 3= (7500 - 9999), 4= (10000 - 12499), 5= (12500 - 19999), 6= (20000 - 25000), 7= (More than 25000)

GENDER Gender
M = Male
F = Female
U = Unknown
J = Joint Account, unknown gender
WEALTH1 Wealth Rating (we used this feature as social economy, and has three categories: Low, Medium and High).
HIT MOR Flag # HIT (Mail Order Response)
Indicates total number of known times the donor has
CARDPROM Lifetime number of card promotions received to date. Card promotions are promotion type FS, GK, TK, SK, NK, XK, UF, UU.
MAXADATE Date of the most recent promotion received (in YYMM, Year/Month format)
NUMPROM Lifetime number of promotions received to date
CARDPM12 Number of card promotions received in the last 12 months (in terms of calendar months translates into 9603-9702)
NUMPRM12 Number of promotions received in the last 12 months (in terms of calendar months translates into 9603-9702)

RAMNTALL Dollar amount of lifetime gifts to date
NGIFTALL Number of lifetime gifts to date
CARDGIFT Number of lifetime gifts to card promotions to date
MINRAMNT Dollar amount of smallest gift to date
MINRDATE Date associated with the smallest gift to date
MAXRAMNT Dollar amount of largest gift to date
MAXRDATE Date associated with the largest gift to date
LASTGIFT Dollar amount of most recent gift
LASTDATE Date associated with the most recent gift
FISTDATE Date of first gift
NEXTDATE Date of second gift
TIMELAG Number of months between first and second gift
AVGGIFT Average dollar amount of gifts to date
EC5 Percent Adults 25+ has a college degree or not (Yes= 1, No= 0)
Some Statistics about the datasets of donors.

Variable	Minimum	Maximum
ODATEDW	8306.00	9701.00
TCODE	0	72002.00
DOB	0	9710.00

AGE	1.0000000	98.0000000
NUMCHLD	1.0000000	7.0000000
INCOME	1.0000000	7.0000000
WEALTH1	0	9.0000000
HIT	0	241.0000000
CARDPROM	1.0000000	61.0000000
MAXADATE	9608.00	9702.00
NUMPROM	4.0000000	195.0000000
CARDPM12	0	19.0000000
NUMPRM12	1.0000000	78.0000000
NGIFTALL	1.0000000	237.0000000
CARDGIFT	0	41.0000000
MINRAMNT	0	1000.00
MINRDATE	7506.00	9702.00
MAXRAMNT	5.0000000	5000.00
MAXRDATE	7510.00	9702.00
LASTGIFT	0	1000.00
LASTDATE	9503.00	9702.00
FISTDATE	0	9603.00
NEXTDATE	7211.00	9702.00
TIMELAG	0	1088.00
AVGGIFT	1.2857143	1000.00

D.1.2 Volunteers Dataset

This appendix is to present some useful information about volunteers' datasets and some basic statistics of the selected features.

CURRENT POPULATION SURVEY, September 2017 Volunteering and Civic Life FILE

Current Population Survey, September 2017: Volunteering and Civic Life Supplement [machine-readable data file] / conducted by the Bureau of the Census for Corporation for National and Community Service (CNCS). Washington: Bureau of the Census [producer and distributor], 2017.

Description:

Data are provided on labour force activity for the week prior to the survey. Comprehensive data are available on the employment status, occupation, and industry of persons 15 years old and over. Also shown are personal characteristics such as age, sex, race, marital status, veteran status, household relationship, educational background, and Hispanic origin. The Volunteering and Civic Life Supplement questions were asked of persons aged 16 years old or older. Data are provided on participation in volunteer activities during a one-year period from September 1, 2016, to the date of the interview, types of organizations volunteered for, and the types of activities for volunteer service. Data are also provided on whether donations of money, assets, or property valued at 25 dollars, or more were made in the past year to charitable or religious organizations. Data are also provided on interactions with family, friends, neighbours, and people from other backgrounds; political engagement; group membership and participation.

Reference Materials:

Current Population Survey, September 2017: Volunteering and Civic Life Supplement Technical Documentation. Additional copies are available from Marketing Services Office, Customer Services Centre, Bureau of the Census, Washington, DC 20233. Bureau of the Census. The Current Population Survey Design and Methodology (Technical Paper 66) describes in detail the sample design and survey procedures used as well as the accuracy of estimates and sampling errors. Reference copies should be available from most public libraries or Federal Depository Libraries. For information about the Current Population Survey and other Census Bureau data products, be sure to visit our online Question & Answer Centre on the Census Bureau's home page (http://www.census.gov/) where you can search our knowledge base and submit questions. File Availability: You can order the file on disc from the Customer Services Centre at (301) 763-INFO (4636) or through our online sales catalog (click "Catalog" on the Census Bureau's home page).

Descriptions of the selected features:

The selected features were selected and customised during the process of cleaning and preparing the data. The customise is for the purpose of meeting the research objectives and answering the research questions, with considerations of ML techniques in pre-processing the data. For more information and details about all features of the dataset, please visit the original database attachments can be found at:

https://www2.census.gov/programs-surveys/cps/techdocs/cpssep17.pdf

Variable	Description	Some statistics
Volunteer City	City of a volunteer	N values 319489 N distinct 11009
Volunteer State	State of a volunteer	N values 319489 N distinct 51
House_Type	Type of House (own, rent, shared)	N values 319489 N distinct 11
Tele_Home	Home Telephone	N values 319489 N distinct 7
Income_Level	Income level of a volunteer ((we did this categories for easy understanding and reading during the analysis stage) 1= (less than 5000), 2= (5000 - 4799), 3= (7500 - 9999), 4= (10000 - 12499), 5= (12500 - 19999), 6= (20000 - 25000), 7= (More than 25000))	N values 319489 N distinct 2
Owing_Business	Does the volunteer own a business or no? (Yes, No)	N values 319489 N distinct 2
Martial_Status	(single, Married, Divorce, Engaged, never been in a relationship)	N values 319489 N distinct 4
Gender	(Female or Male)	N values 319489 N distinct 2
Volunteering_Armed_Force	Has a volunteer joined the armed forced (in the USA)?	N values 319489 N distinct 3
Working_Armed_Force	Has a volunteer worked or worked for the armed forced (in the USA)?	N values 319489 N distinct 3
Level_of_Education	Does a volunteer has college degree or not (Yes= 1, No= 0)	N values 319489 N distinct 2

	Some Statistics about the datasets of donors.	
Birth_Place	Place of birth of a volunteer	N values 319489 N distinct 161
Citizenship	Citizenship of a volunteer	N values 319489 N distinct 6
Searching_for_job	Is the volunteer looking for a job?	N values 319489 N distinct 4
Having_more_than_job	Does the volunteer have more than a job?	N values 319489 N distinct 3
Working_hours	Working hours of a volunteer in total	N values 319489 N distinct 51
Weekly_working_hours	Working hours of a volunteer per week	N values 319489 N distinct 11
Reasons_for_part-tim	Any reasons of volunteering a part-time job?	N values 319489 N distinct 11
Take_off_hours	How many off-work hours per week?	N values 319489 N distinct 63
Working_last_year	Did the volunteer work last year	N values 319489 N distinct 3
Job_industry	What is job industry a volunteer belongs?	N values 319489 N distinct 53
Volunteering_last_year	Did the volunteer work last year?	N values 319489 N distinct 2
Volunteering_frequency	How many times per year a volunteer joins a volunteering opportunity?	N values 319489 N distinct 8
Volunteering_hours_last_year	How many hours a volunteer did as a volunteer last year?	N values 319489 N distinct 2
Social economy * this feature is a combination of different feature such as wealth, owing a business, number of volunteering, etc. We customised the original features to obtain a useful feature that can help in the probability model and predictive analysis.	Low, Medium, and High	N values 319489 N distinct 2

APPENDIX

E

APPENDICES

E.1 User Manual

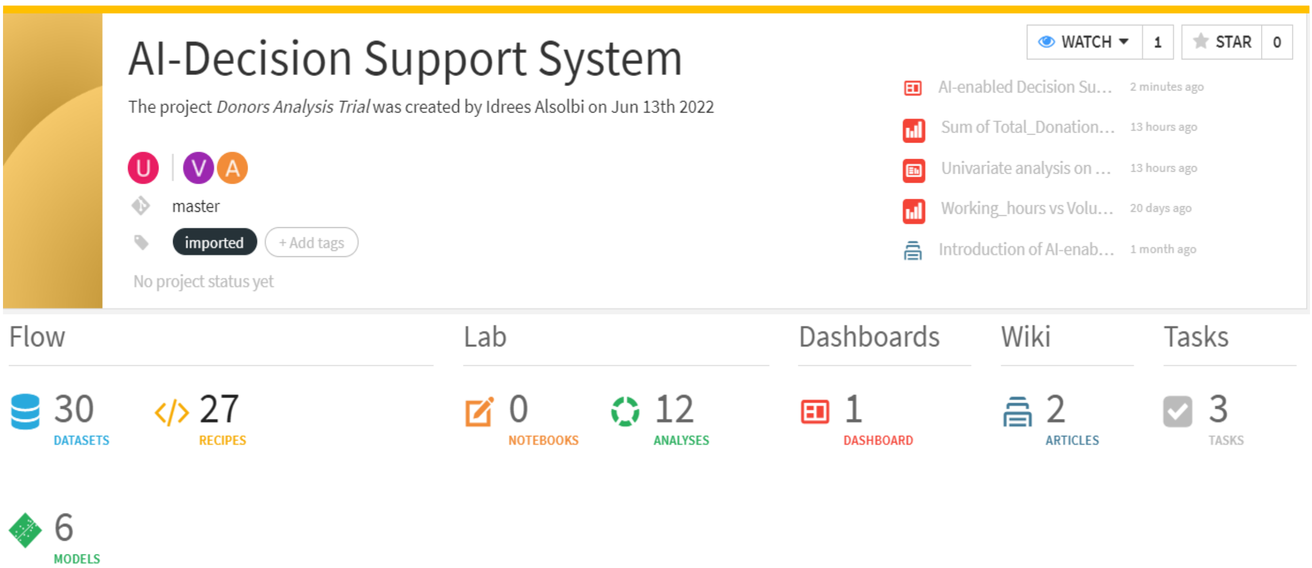
E.1.1 User Guidelines of the AI-enabled DSS for analysing donor behaviour in NPOs

This appendix is to present our user guidelines for the AI-enabled DSS in NPOs for analysing donor behaviour.

AI-Decision Support System

USER INSTRUCTIONS

AI-enabled Decision Support System



Version 1.3

DISCLAIMER

This prototype is a research output at the University of Technology Sydney. The AI-enabled Decision Support System represents an outcome of the Design Science Research process model framework to identify the research problem, collect the design requirements, build and design, execute the instantiation, evaluate the results, and report the results. The AI-enabled Decision Support System aims to apply such machine learning techniques on donors and volunteers in nonprofit Organisations to enhance the decision-making process.

[Australian Artificial Intelligence Institution]

[Faculty of Information Technology, School of Computer Science]

[University of Technology Sydney]

[15 Broadway, Ultimo NSW 2007]
[+6403428488]

Table of contents

PREFACE

1.1 Description of the user

The AI-enabled Decision Support System is intended to help decision-makers at nonprofit organisations for analysing donor behaviour. There are many types of potential users may use the system for:

A nonprofit organisation decision-maker
A nonprofit organisation manager/Chief Executive/Head or Principle.
A data scientist
A technical support manager
A social studies consultant
A nonprofit organisation decision maker

1.2 Explanation of safety warnings

WARNING! A change in the information provided or the types of figures, trying to reach out to the admin's authenticity, may turn the system down.

CAUTION! Some slides and components of the system may take some time to be loaded.

1.3 Accessing the System

1.3.1 Internet

The latest version of the system is available on:

https://sydneyuniversityoftech.academics.dataiku-dss.io/projects/DONORSANALYSISTRIAL/

Username: user1

Password: 101010

For any difficulties of access, please email the researcher: Idrees.alsolbi@student.uts.edu.au.

Description of the System

1.4 Purpose of the System

The system is to analyse donor behaviour in nonprofit organisations. Different types of analysis help decision-makers to understand and analyse donor behaviour intensively. Donors include those who have donated money and time over the past many years (donors and volunteers). This system is excluded from analysing blood donations, gifts, or goods.

1.5 Technical data

Icon	Explanation
30 DATASETS	**The Number of databases processed and used for applying such analysis of donors and volunteers.**
27 RECIPES	**The Number of recipes which are used to create new datasets by performing transformations on existing datasets.**
12 ANALYSES	**The Number of analysis or different types of analyses used to analyse donor behaviour.**
1 DASHBOARD	**The Number of dashboards that present visualisations of the 12 different analysis.**
2 ARTICLES	**The Number of the articles that used to introduce the system and provide such valuable information about the usage of the systems.**
6 MODELS	**The Number of predictive and descriptive models that have been applied to analyse donor behaviour.**

Table 1 System elements

1.6 System Compliance

This product is performed on Dataiku, which is a Data-Driven Decision-Making platform.

- Dataiku can:
- Generate deeper intelligence faster from seamless data access, smart data preparation and repeatable, transparent data transformation.
- Uplevel machine learning (ML) skills and experiment with AutoML.
- Effectively communicate insights with dashboards and customisable apps.
- Collaborate with technical gurus to bring your business insights to life.

1.7 Dashboard elements

1. Once you click on the Dashboard 1 DASHBOARD icon
2. There are different slides at the bottom of the Dashboard.
3. You can click on each slide to see its contents of it.

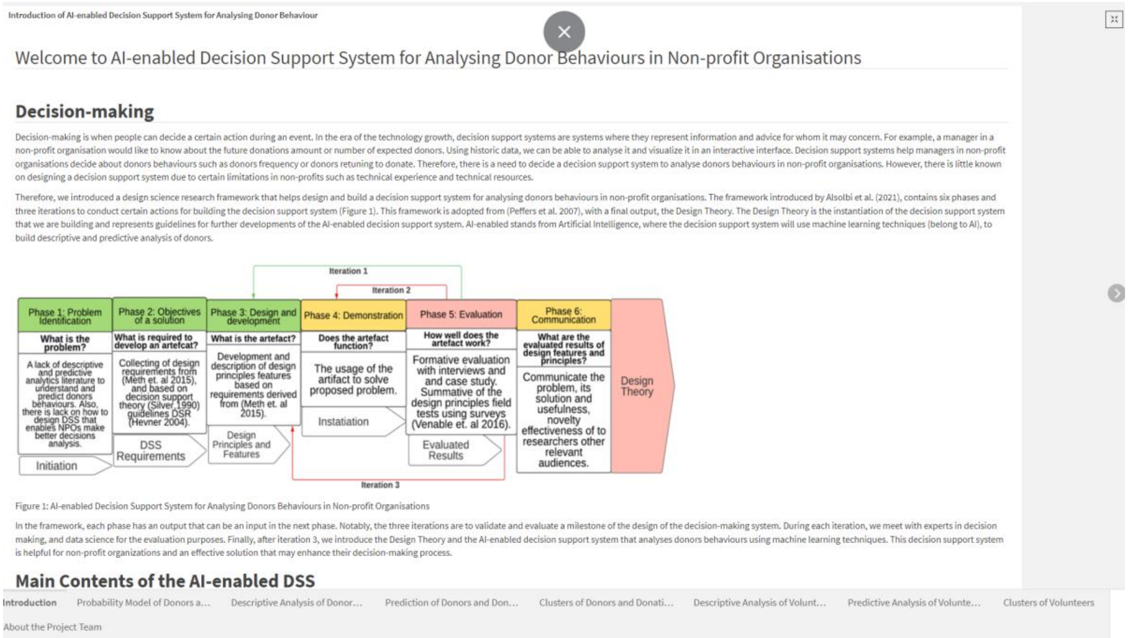
Slide	Explanation
Introduction	This slide presents some content to introduce the research process model framework that was used to design and build the AI-enabled Decision Support System.
Probability Model of Donors and Volunteers	This probability model has been built using R (with the shiny App) tool. A probability model is a mathematical representation of a random phenomenon. It is defined by its sample space, events within the sample space, and probabilities associated with each event.

Descriptive Analysis of Donors	This slide shows a variety of graphs that represents important statistics about donors. A tooltip is provided to help in understanding the graphs clearly.
Predictive Analysis of Donors	This slide shows some plots of predictions of different variables of donors. Also, the slide visualises the decision tree for predicting the highest state of donations, using 19 features from the data.
Descriptive Analysis of Volunteers	This slide shows a variety of graphs that represents important statistics about volunteers. A tooltip is provided to help in understanding the graphs clearly.
Predictive Analysis of Volunteers	This slide shows some plots of predictions of different variables of volunteers. Also, the slide visualises the decision-tree for predicting the highest state of donations, using 19 features from the data.
About the project Team	This slide provides information about the members of this research outcome.

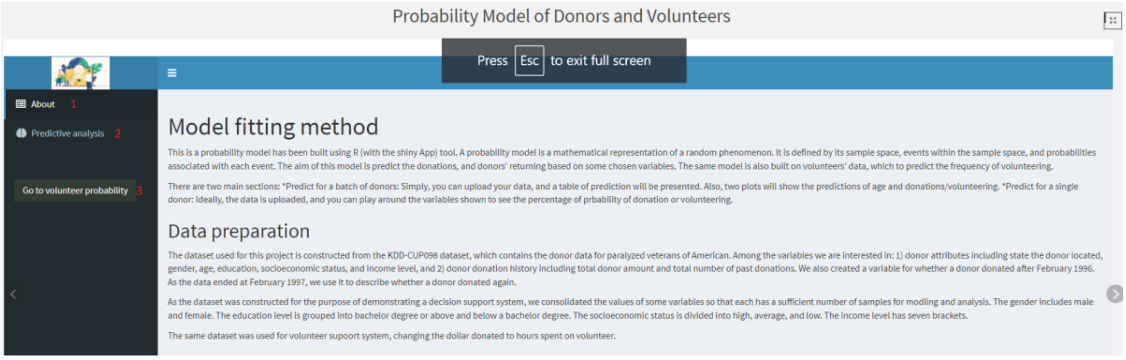
Table 2 AI-enabled DSS Dashboard Slides

1.8 Understanding the user interface: Introduction

- Read the contents of this page to obtain better information about the research framework and the main contents of the AI-enabled Decision Support System.



1.9 Understanding the user interface: Probability Model of Donors and Volunteers

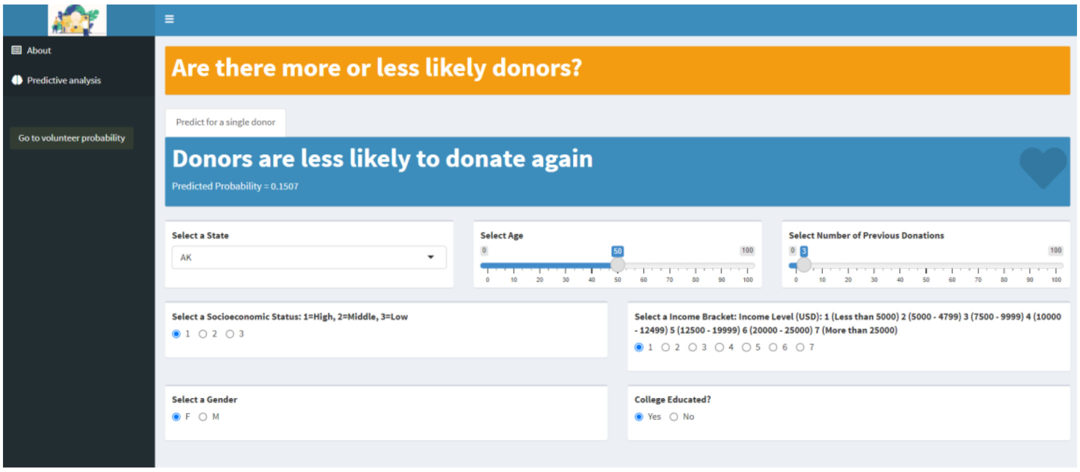


This model aims to predict the donations and donors' returns based on some chosen variables. The same model is also built on volunteers' data to predict the frequency of volunteering.

This interface has three icons:

1. About: it contains some information about the model and the environment that was built in.
2. Predictive analysis: this function performs a single analysis on donors based on some chosen variables. This analysis answers the question: Are there more or less likely donors?

How to obtain such a result of the question: Are there more or less likely donors?



The answer to the predicted probability of donors' percentage is highlighted in blue. It also means the probability percentage of certain donors will likely be donating based on the below variables:

1= Sate: the user can change state accordingly to see the difference in the probability percentage.

2= Age: the user can change Age accordingly to see the difference in the probability percentage.

3= Number of donations: the user can change the number of donations accordingly to see the difference in the probability percentage.

4= Social status: the user can change social status accordingly to see the difference in the probability percentage.

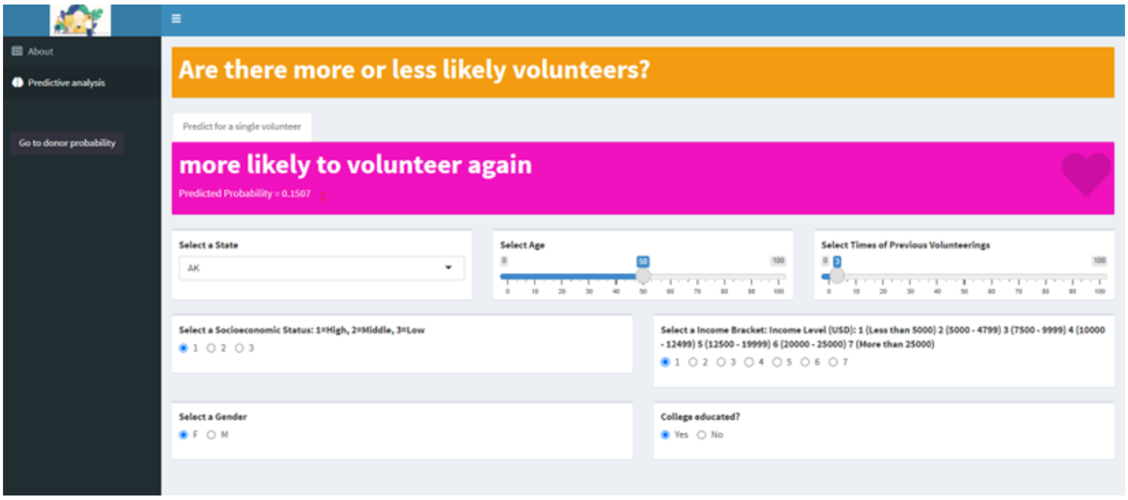
5=Income bracket or level: the user can change income level accordingly to see the difference in the probability percentage.

6= Gender: the user can change Gender accordingly to see the difference in the probability percentage.

7= Education: the user can change the Education status according to the difference in the probability percentage.

3. Go to volunteer probability: to move for predicting the probability of volunteers based on some variables.

How to obtain such a result of the question: Are there more or less likely donors?



The answer to the predicted probability of volunteers' percentage is highlighted in pink. It also means the probability percentage of certain donors will likely be donating based on the below variables:

1= Sate: the user can change state accordingly to see the difference in the probability percentage.

2= Age: the user can change Age accordingly to see the difference in the probability percentage.

3= Number of donations: the user can change the number of donations accordingly to see the difference in the probability percentage.

4= Social status: the user can change social status accordingly to see the difference in the probability percentage.

5=Income bracket or level: the user can change income level accordingly to see the difference in the probability percentage.

6= Gender: the user can change Gender accordingly to see the difference in the probability percentage.

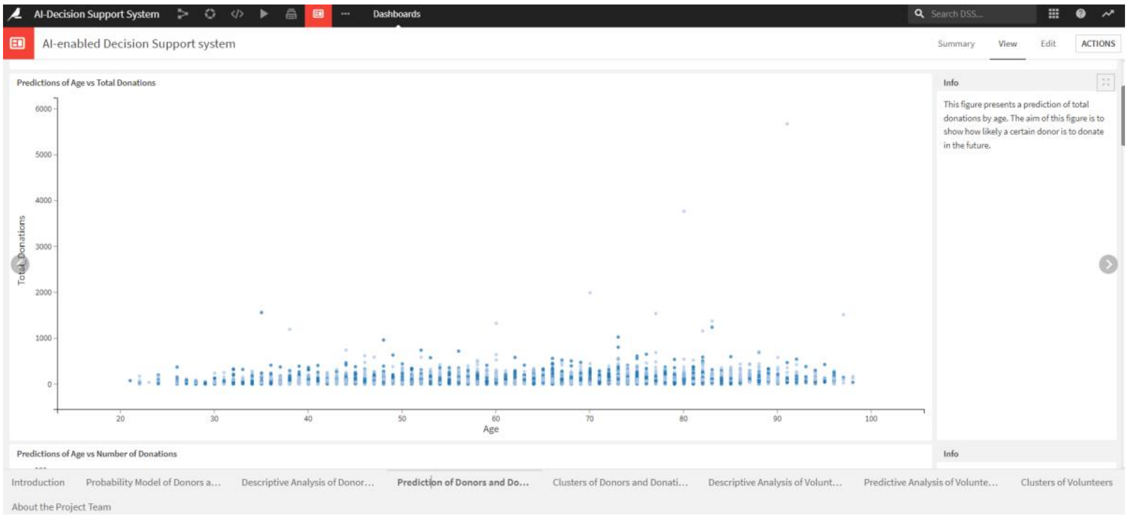
7= Education: the user can change statue of Education according to see the difference in the probability percentage.

1.10 Understanding the user interface: Descriptive Analysis of Donors

Descriptive Analysis of Donors shows various graphs representing important statistics about donor behaviour. A tooltip is provided to help in understanding the graphs clearly. Each figure has an info box on the right side of the Dashboard to give some useful information about the figure.

- You can simply scroll down to explore more of the statistics and figures of donors' analysis.

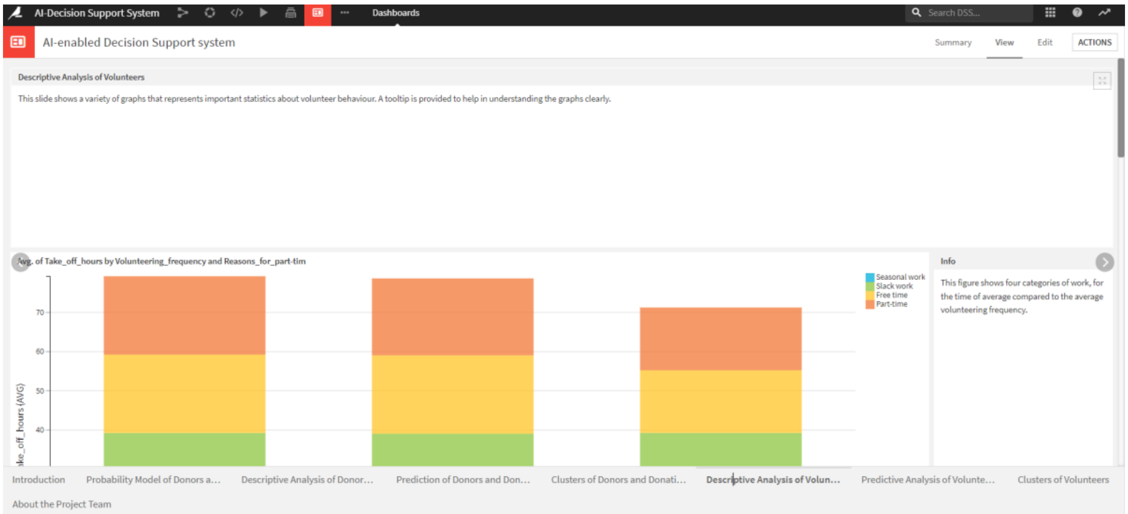
1.11 Understanding the user interface: Predictive Analysis of Donors



This slide shows some plots of predictions of different variables of donor behaviour. Also, the slide visualises the decision tree for predicting the highest state of donations, using 19 features from the data.

- You can scroll down to explore more of the statistics and figures of donors and donations predictions.

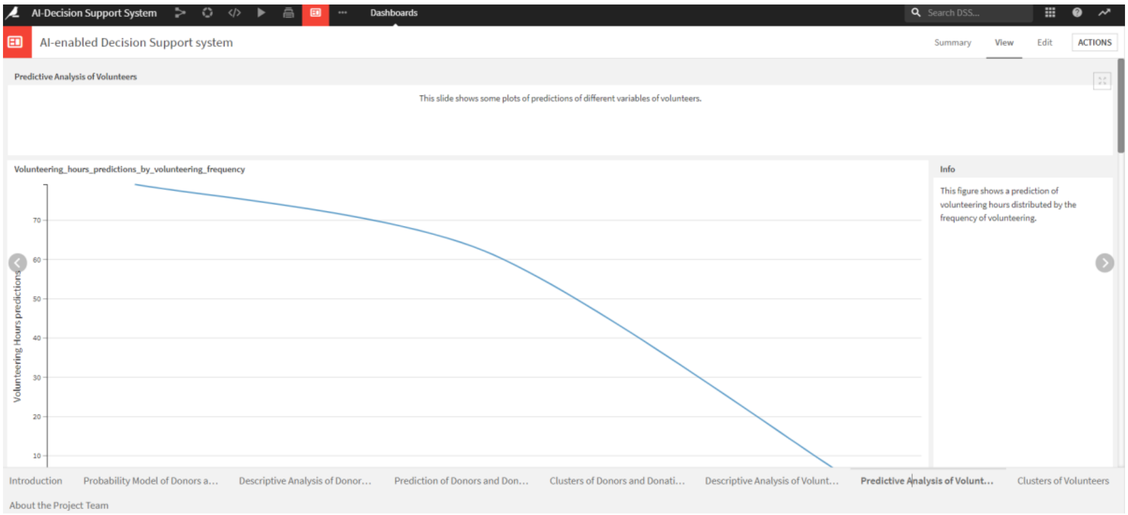
1.12 Understanding the user interface: Descriptive Analysis of Volunteers



The descriptive Analysis of Volunteers slides shows various graphs representing important statistics about volunteer behaviour. A tooltip is provided to help in understanding the graphs clearly.

- You can scroll down to explore more of the statistics and figures of volunteers' analysis.

1.13 Understanding the user interface: Predictive Analysis of Volunteers



The predictive Analysis of Volunteers slide shows some plots of predictions of different variables of volunteers. This slide aims to show some useful predictions of volunteers and related variables that may assist in attracting more volunteers.

- You can explore more of the statistics and figures of volunteers' analysis.

1.14 Understanding the user interface: About the project Team

This slide is to provide some information about the project team of the AI-enabled Decision Support System.

USER INSTRUCTIONS

AI-enabled Decision Support System