Data-driven Adaptive Personalized Property Investment Risk Analysis: Frameworks, Methods and System

by Nur Atiqah Rochin Demong

Thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

under the supervision of Professor Jie Lu and Associate Professor Dr. Farookh Khadeer Hussain

University of Technology Sydney
Faculty of Engineering and Information Technology

October 2020

TABLE OF CONTENTS

CERTIFICATE OF AUTHORSHIP/ORIGINALITY

I, Nur Atiqah Rochin Demong declared that this thesis is submitted in fulfillment 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: Production Note:
Signature removed prior to publication.

Date: 27 October 2020

DEDICATION

I would like to dedicate this thesis to the Allah SWT. Every challenging work needs self-efforts as well as guidance from elders especially those who are very close to my heart. My humble effort I dedicate to my sweet, loving, and amazing husband, Mohd Norhedhir Yaakub, for sharing the pain, sorrow, and depression during the hard time, and above all for the unconditional love, patience, and encouragement. Whose affection, love, encouragement and prays of day and night make me able to get such success and honor.

To my wonderful kids, Athirah, Akram, Aqilah, Amirah and Akmal, for being the best support and encouraging me to complete my thesis and going through this challenging and exciting journey.

Along with all hardworking and respected supervisors, this thesis is dedicated to Prof. Jie Lu and Dr. Farookh Khadeer Hussain, for without their early coaching and enthusiasm none of this would have happened.

ACKNOWLEDGMENTS

One of the great pleasures of writing this thesis is acknowledging the efforts of many peoples who were involved directly or indirectly to complete my PhD study.

I would like to express my sincere gratitude to my principal supervisor, Professor Jie Lu for accepting me as one of her PhD students and guiding me to complete my study. Thank you for your continuous support, guidance, comments, suggestions, patience, encouragement and teaching me to be analytical and help me to improve in my writing skills. Moreover, her commitment in supervising me has shaped me into a good researcher and will be a great benefit to me in my future research work and life. I also would like to address my sincere thanks to my co-supervisor, Associate Professor Dr. Farookh Khadeer Hussain for his ideas, suggestions, and invaluable comments to strengthen my research throughout my study.

I also wish to express my appreciation to all my friends and the members of the Decision System & e-Service Intelligence (DeSI) Lab in Faculty of Engineering and IT (FEIT), University of Technology Sydney, for their help, participation and invaluable comments in every presentation I made during my study. Support and encouragement developed during my study will be good memories in my future research life. I wish all the very best in our future research works.

I also would like to thank Jemima McDonald, Barbara Munday and Sue Felix for being proofreaders of my published papers and thesis. I sincerely appreciate the efforts you have given for my paper to be accepted.

Finally, I would like to thank the Universiti Teknologi MARA for funding and providing support during my PhD study. I am grateful to have been able to study and have experiences in Australia.

LIST OF FIGURES

LIST OF TABLES

Abstract

The risk analysis for real estate property investment which incurred a high cost and high risk has been qualitatively and quantitatively assessed by various techniques. These techniques consider the heuristic risk factors mainly based on the expert survey, weigh and rank the factors using algorithms and mathematical formulas and decide the best investment based on the performance index of alternatives given. Currently, identifying, weighing, and ranking of the risk factors in investment decisions largely depend on expert judgment using traditional approaches. Additionally, current research is lacking in considering investors' different preferences and requirements. The motivation of this study is how to identify, weigh and rank the risk analysis factors when experts have a different point of view or judgments' and expert subjective justifications.

This thesis describes a new personalized multidimensional process (PMP) framework based on knowledge discovery to overcome the weaknesses of existing risk identification and measurement techniques. This framework comprises two new methods namely personalized association mapping (PAM) method and personalized multidimensional sensitivity analysis (PM-SA) method. This thesis also proposes an adaptive personalized property investment risk analysis (APPIRA) method to identify the property investment determinants. This APPIRA method adopts a data-driven and personalization technique to weigh risk factors and ranking using a multi-criteria decision-making model by applying the TOPSIS method for optimal solutions. The existing real estate analytic systems, which only serve as search tools and do not benefit homebuyers in terms of search time, flexibility, and intuitive results. Thus, a prototype of adaptive personalized property investment risk analysis system (APPIRAS) developed to validate the framework and methods proposed to overcome the limitations.

The innovations of this research were the justification of risk factor identification as the determinant for different objectives; weighing which is based on historical data-driven to decision support using the knowledge discovery approach (analytical decision-making) and the investor's adaptive personalization of the risk factors which fulfil their requirements; and ranking using multi-criteria decision-making model. The methodology in this research incorporated literature review, frameworks development, methods development, prototype software system development and evaluation. A case study has been developed to show the applicability of the developed framework, methods, and system. The outcomes of this study have a significant impact in helping investors to achieve their objectives and APPIRAS as the decision support tools to achieve optimal decisions. These methods can be applied to practice and benefit the property industry directly.

CHAPTER 1

INTRODUCTION

This chapter discussed in detail the background and issues related to risk analysis for investment in the real estate industry.

1.1 Background

Decision-making is a process of gathering input and processing the data collected for analysis to produce a list of outcomes based on given sources and limitations. All decisions made will involve low, medium, or high-level risk based on the uncertainty factors affecting the analysis. For instance, the higher the uncertainty of factors related to decision-making, the higher the level of risk and vice-versa. If decision makers are familiar with the sources and factors that will affect the decision-making, however, and know how to monitor and control the uncertainty factors, the risk will be lower. Property investment risk analysis involves modeling that considers the uncertainty of risk factors.

Risk-based decision-making (RBDM) concepts and applications have been discovered by many researchers who applied different techniques or methods to support the decision-making process in different fields. For example, practices on RBDM for investment in the real estate industry study have been conducted to investigate the risk related issues and it was found that many decisions made based on investigating and analysing factors by giving weight, calculate and select the best option based on a high performance index (Liow & Kwame.Addae-Dapaah, 2010; Peng, Zhang, Tang, & Li, 2011; N. Piyatrapoomi, A. Kumar, & S. Setunge, 2004). There is a gap found in the literature review where there is a need of the justification of risk factor identification, weighting and ranking. Thus, a new approach which is based on historical data-driven to decision support using knowledge discovery and investors personalization for real estate investment risk analysis proposed.

The real estate industry is one type of business form which is related to tangible assets such as houses, units, or buildings. The investment in the real estate industry is characterized as a discrete asset value category where the value of these assets still retains even if it is removed from the business operation. However, the value of these assets is depending on many factors and the changes of the value will be affected immediately and the percentage of change can be drastically increased or decreased.

Real estate industry business processes include managing buying and selling properties, rental services for properties, and advertising properties to sell or rent. Property investment can be complicated with layers of legal and administrative challenges that need to be considered carefully before doing any transactions. Real estate has been regarded as a high profit industry and currently develops quickly in China (S. Hui, Zhi Qing, & Ye, 2009). Even though the investment in the real estate industry incurred high cost and slow liquidity, however, it gives more value and high rate return of investment in a short period of time. Thus, that makes the investors who have enough budgets and capital will go for it. Some investors gain their capital for investment through mortgage either from a bank or organization. The prices and value of the property in the real estate industry normally increased all the time especially when the location of the property is surrounded by facilities and amenities even though it is still in the development phase. For example, if there is public transport or other amenities such as children's playground or park, community centres', schools which are still in development progress, the prices or values of the property will increase once these facilities are complete and ready to be used.

Real estate has been playing an important role in the national economy and has become a pillar industry and being able to stimulate gross domestic product (GDP) rapidly (Yu & Xuan, 2010b). However, the real estate projects are characterized by high risk, high return and long cycle, which needs real estate investors carefully research each project so as to maximize the return and minimize the risk (Juhong & Zihan, 2009; Ren & Yang, 2009; Shiwang, Jianping, & Na, 2009; Yu & Xuan, 2010a, 2010b). The real estate business is very risky due to off-scale and long-term, and other factors such as natural, social, economic, regulatory, psychological factors. Once the decision is at fault, the investor will suffer huge or even destructive damages. Therefore, risk analysis is necessary using scientific methods and tools to understand the risk situation thoroughly when deciding on investment (Yu & Xuan, 2010a).

Property investment transactions in the real estate industry involve multi-criteria decision-making (MCDM) analysis that cause a challenging decision for investors. Property investment is high-risk, high-cost, dynamic, and involves uncertainty (Chong, Guo, & Wang, 2008; E. C. M. Hui, Yu, & Ip, 2010; Xiaobing. Zhang & Li, 2009). Risk in property investment is typically defined as the uncertain elements that are affected by qualitative or quantitative factors in which the higher the uncertainty, the higher the risk (L. Liu, Zhao, & Liu, 2007; S. J. Zhou, Li, & Zang, 2008). Currently, weighing and ranking the uncertain factors in investment decisions largely depends on expert judgment. Hence, investors have different preferences and requirements, risk analysis and modelling support for purchase decisions should include a personalization mechanism. Existing studies do not address these issues well.

Risk with high uncertainty leads to a higher potential for failure, and the consequences of failure will have a greater impact. Uncertainty, in dynamic risk prediction, is dependent on many factors throughout the life of an investment, from initial purchase to the final stages of development. Moreover, another factor that affects uncertainty is the investor's and the experts' understanding on how to interpret the identified risk factors.

Real estate investment risk analysis is performed to achieve the investor's goals and requirements. Real estate investment normally relates to capital risk and liquidity risk. According to Liu (2007), real estate portfolio model is stated as the following: each project is totally competitive, each investor cannot monopolize the market or control return rate (Cao 2005 cited in Liu et.al., (2007)); Return rate of each project conforms to normal distribution, and risk can be expressed by variance of return rate, and return can be expressed by the expected return rate; Each investor knows about the expected return rate and variance; Each project is related with each other, in other words, the return rate of each project is connected, and conforms to joint state distribution. Correlation coefficient and covariance can show the related degree; each investor expects to maximize the return on the fixed risk or minimized the risk with the fixed return; each investor can divide his investment in as many projects as he wants.

Risk occurs at different stages of the investment process. The risk arises because of possible consequences and associated uncertainties, and there are several risk factors or variables and risk sources that will affect the level of risk for the given alternatives. The evaluation of risk will be affected by different types of risk factors and risk sources for investment in the real estate industry.

Most risk analysis techniques used in real estate such as analytical hierarchy process (AHP) and analytic network process (ANP) assign a weighing to each risk factor, based on expert experience. There are many limitations and disadvantages in using expert judgment to help property investment decision-making, but the main concern with this approach is that an investor's goals may not be similar to the expert's goals. Additionally, experts may be familiar with the factors that affect decision-making and know how to monitor or control uncertainty factors, but they may have different perceptions or methods of analysis (L. Chen, 2010; Lafleur, 2011), which could result in a misrepresentation of the weight and rank of those risks from the investor's perspective. The need for a comprehensive and interactive system as an automated tool that investors can use to model and predict risk in their decision-making process is clear.

The risk analysis and prediction methods currently available do not provide interactive tools that allow investors to personalize their goals and objectives, set limitations, or the levels of risk they are willing to assume. The online real estate and property investment application systems that are available only provide investors with basic information and current reports on the properties for rent or purchase (Y.-M. Fang, Lin, Huang, & Chou, 2009; Yuan, Lee, Kim, & Kim, 2013). However, these systems have three weaknesses: 1) they do not include modelling or risk prediction; 2) they provide a tremendous amount of information to the end-user, often resulting in information overload; and 3) the investor's goals and acceptable levels of risk cannot be personalized.

Hence, this study proposed a PMP framework, methods and system for adaptive personalized property investment risk analysis in the real estate industry to overcome the following issues.

Issue 1: High cost and high risk

Investment in the real estate industry involved with high cost and high capital. Moreover, the properties are not easily sold which make the risk level of the real estate investment will be very high. Moreover, property features such as house price are affected by many factors, for example, interest rate, land supply and inflation rate (E. C. M. Hui et al., 2010). There will be a list of projects for investment that needs to be analyzed and priority will be given to the most beneficial project with the limited or available budget and time. Real estate investment is speculative, and its return and risk are influenced by many factors, such as natural environment, socioeconomic environment, market and enterprise purchasing capability (L. Liu et al., 2007; S. J. Zhou et al., 2008). Moreover, the risk factors change dynamically over a period of time and this will affect the risk level of real estate portfolio. Therefore, there is a need to measure and analyze the risk effectively and efficiently to achieve best results and gain profits tangible or intangible. It is imperative to know the determinants of risk factors contributing to the specific goals or objectives in different geographical areas.

Risk analysis is the process of identifying the security risks to a system and determining their probability of occurrence, their impact, and the safeguards that would mitigate that impact.

Issue 2: Uncertainty of risk factors

Risk measurement for decision-making is dealing with the measurement of uncertainty and probability or consequences for the choices given for the investment. The uncertainty of variables or factors that will affect the risk management processes such as risk evaluation results will give impacts to the successes of the project investment in the real estate industry. The investors must have knowledge, understand, and know about how to manage the factors that will affect the investment in the real estate industry. Thus, the investors need to make wise decisions for investment in the real estate industry. The risk analysis is part of the process involved when deciding on which project should be given priority for investment with the limited budget and meets the investor's goals and objectives. In order to control risk and enhance return, investors of real estate should not exclusively choose those projects of "both high" (high risk and high return) or "both low" (low risk and low return), instead they need to rationally utilize capital to develop different types of projects (L. Liu et al., 2007).

Issue 3: Insufficient comprehensive risk analysis

There are many risk analysis models and methodologies applied to real estate and other industries. Each of these models or methods has its features or characteristics, advantages, disadvantages, and limitations. Risk analysis consists of three stages: risk identification, risk estimation and risk assessment (Yu & Xuan, 2010a).

Several examples of risk identification methods at present include mainly Delphi, brainstorming, Fault Tree Analysis, SWOT analysis and expert survey. The expert survey method is widely used to identify the risk factors based on the heuristic approach which is not sufficient for comprehensive risk analysis.

According to (S. J. Zhou et al., 2008), real estate investment risk evaluation is a complex decision-making problem with multiple factors and multiple targets. The most existing real estate investment risk evaluations give priority to the single-goal decision-making, using the single indexes, such as the maximum expectation, the largest variance, and the minimum standard deviation rate to evaluate the real estate investment. These evaluating methods are easy to understand, but they cannot comprehensively evaluate the quality of an overall program. Also, some use Multi-element Analysis Model (MAM) for real estate investment risk evaluation. The traditional MAM assumes that the whole is subject to the normal distribution. Yet, the whole distribution of real estate investment programs is uncertain, thus, it is inadequately precise to use MAM for real estate investment risk analysis. Furthermore, many evaluation programs or models involve many evaluation indexes that the dimensions are different, and the weights are difficult to determine, there are some difficulties in the practical application.

Issue 4: Misinterpretation and different judgments of risk factors

From the review of literature, studies done by other researchers categorize the risk sources and risk factors differently. Some researchers classified the risk factors according to the stages of real estate investment and some other group it according to the sources. This will lead to misinterpretation of factors that will affect the risk analysis result. Thus, this research will apply the combination of personalization technique and data-driven approach (analytical decision-making) to gather information from literature reviews of journals, reports, domain databases of Australian Property Monitor and other secondary and published documents to identify the risk sources and its factors that will affect the property investment risk analysis. Existing techniques determine the risk evaluation index based on expert surveys to determine the weight of the factors that will influence the decision-making process. Experts in the field have different opinions and judgment about the factors and this will affect the result of decisions made. Hence, this study identifies the determinants of risk factors for different goals, estimates and weighs the risk factors based on the personalization techniques and maps it with investor's requirements to achieve the investor's goals and objectives.

Issue 5: Complexity on practical implementation of risk management analysis technique

The risk analysis of investment in real estate projects refers to the overall consideration of the risk attributes, the target of risk management and the risk bearing capability of risk subjects on the basis of investment risk identification and estimation, thus determining the degree of influence of investment risks in the system (X.-L. Tang & Liu, 2009). There are several models proposed by other researchers to evaluate, analyze, assess, or predict the risk for several alternatives or options given. Some of the methods or models include a Monte Carlo method, fuzzy set theory (Y. Sun, Huang, Chen, & Li, 2008), Markowitz, fuzzy-analytical hierarchy process (F-AHP), a real option method (Rocha, Salles, Garcia, Sardinha, & Teixeira, 2007), and a hidden Markov model. These methods or models have different characteristics, advantages and limitations to be applied in different fields (Y. Sun et al., 2008); Lander & Pinches 1998, cited in (Rocha et al., 2007). Some of the problems include the lack of mathematical skills, restrictive modelling assumptions, increasing complexity and limited power to predict investment in competitive markets (Lander & Pinches 1998 cited in (Rocha et al., 2007).

Issue 6: Unavailability of high-quality data for strategic information

Zeng, An and Smith (2007) believes that high quality data is a prerequisite for an effective application of sophisticated quantitative techniques. They therefore agree and suggest that it is essential to develop new risk analysis methods to identify major factors, and to assess the associated risks acceptably in some various environments.

The correct methodology is important to ensure the right decision is made and will be beneficial to the investors, users or agents and it is necessary in decision-making processes (O. K. Hussain, Chang, Hussain, & Dillon, 2007; Zeng et al., 2007). Many risk analysis techniques currently used in the UK construction industry are comparatively mature, such as Fault Tree Analysis, Event Tree Analysis, Monte Carlo Analysis, Scenario Planning, Sensitivity Analysis, Failure Mode and Effects Analysis, Program Evaluation and Review Technique (Zeng et al., 2007). However, in many circumstances, the application of these tools may not give satisfactory results due to the lack of data available.

There are many substantial studies related to the application of risk analysis in different fields including analysis of investment in the real estate industry. However, none of the studies applied the personalized multidimensional process for risk analysis in the real estate industry based on knowledge discovery from data-driven approaches such as data mining processes. Thus, this research was undertaken to propose a new personalized multidimensional process framework that meets the investor's requirements to achieve their goals and objectives. Hence, a new personalized association mapping method and personalized multidimensional – sensitivity analysis method embedded in adaptive personalized property investment risk analysis (APPIRA) method proposed.

Issue 7: Requirement of historical data and information using computer systems for risk factors identified for specific objectives

Yu & Xuan (2010a) agreed that it is feasible to develop a practical computer system for the real estate risk-based investment decision-making that could effectively predict the level of financial risk of real estate development projects and could help the investor control risk effectively. However, the real estate investment decision-making has a strong feature of valuing practice and experience. The expertise practical experiences will have a strong impact on decision-making and so do the investors' risk preferences. Moreover, the estimation of the interval values of items in the cash flow statement should be based on adequate market research and historical data and information accumulation. The proposed model should establish a strict risk management system, emphasize the accumulation of historical data and information, and try to make a market research reflect market conditions as much as possible, so the data in the cash flow statement could be of strong accuracy and the calculation result of computer system could scientifically predict the project financial risk. Hence, a prototype of an online application for the real estate property analysis developed to ease and help the investors to choose the best real estate portfolio that meets investor's requirements. Moreover, the application of a data-driven approach using data mining techniques or knowledge discovery used to determine, weigh and rank the risk factors that will affect the decision-making process in the real estate industry which applied the deterministic approach will be more reasonable.

1.2 Research Questions

To overcome the limitation of the existing ranking method for investment risk analysis in the real estate industry as described above, this study aims to answer the following specific research questions.

Question 1: How to support the investors analyze risk for the real estate property investment automatically?

The decision-making process involved with certainty, uncertainty and risk options for the factors that will affect the result of the decision made. The factors discovered can affect the result on its own or combined with other factors either in pairs or multidimensional. Not all solutions provided by the technology meet the investor's goals and objectives. A PMP framework based on data-driven approach and knowledge discovery dealing with the risk analysis process was developed to deal with real estate property investment analysis that meets investor's requirements. The PMP framework helps the investors to achieve their goals and objectives using deterministic and personalization approaches using automatic analytic tools as compared to an expert's manual judgment.

Question 2: How to identify the risk factors that align with investor's requirements?

The determinants of risk factors subject to the investor's goals and limitations will be identified based on the pattern discovery from a data-driven approach using data mining processes. These weights and ranking of factors provided by the data mining technique might not meet the investor's requirement. Hence, the PAM method (algorithm) was developed to map the determined factors by knowledge discovery with investor's personalization that will affect the decision-making process. The investor will personalize the weights of the determinant and the system will rank the properties options that align with the investor's requirements.

Question 3: How to analyze system responses and generate personalized suggestions to meet the investor's requirements?

Since uncertainties are involved in the factor determination over a while, it is very important to understand how sensitive the factors are to the variation of the investor's personalization. Sensitivity analysis can be used to see the variation of the result if there is variation in the factors personalized by the investors that will affect the real estate property sold in a certain period. Hence, in this research, a PM-SA method (algorithm) was developed.

Question 4: How to support the investors who manage and deal with the real estate property analysis that fulfils their personalized requirements?

Since it involves a high volume of data extracted for knowledge discovery using data mining technique, Adaptive Personalized Property Investment Risk Analysis System (APPIRAS), a prototype system has been developed to support the real estate property analysis. APPIRAS helps an investor to identify the sales pattern of real estate property, determine the factors that influence the sales pattern and rank the factors and map the factors with their requirements. The APPIRAS will be used by investors who would like to search for the best real estate property that meets their requirements. The system will implement the framework and the methods, which have been developed in early stages.

1.3 Research Objectives

Concerning the issues or problem statement stated earlier, the area of this research is to develop a PMP framework of decision-making under risk, particularly real estate property analysis. Personalized multidimensional process framework and methods developed and discovered based on real estate property portfolio characteristics and data-driven analysis to support decision-making that fulfil investor’s requirements. In relation to the research question described in 1.2.1, the main objective is divided into four specific objectives.

Objective 1:

To develop a new personalized multidimensional process (PMP) framework for property dynamic risk analysis

Based on the model of a generic framework for knowledge discovery proposed by Fong and Hui (2010), an extended PMP framework developed based on knowledge discovery and investor’s personalization for effective decision-making to deal with the real estate property analysis that meets investor’s requirements. The data-driven approach of data mining category specifically the decision tree induction technique applied to discover the pattern of data-driven approach to decision support process. This framework provides the pattern of data, hence, determines the factors that contribute to the buying or selling of real estate property analysis which discover three main types of questions: what? why? and when? For example, the data-driven approach will identify and explain which factors act as determinants that contribute to the short time frame for the property sold. Is it because of the features of the property, location, price, type of property, type of sale, sale result, size of property for a certain period of time or which real estate agency that handles the transactions? This framework presents a data-driven system and a process from data to patterns and from patterns to applicable rules/methods for decision support act as analytical decision-making.

This PMP framework is described by a diagram as depicted in Figure 13 to show its new features which is extended from the proposed generic framework for knowledge discovery.

Objective 2:

To develop PAM method to map the weight and ranking of risk factors with investor's personalised requirements

A personalized association mapping method developed to map the weight and ranking factors between pattern discovery based on a data-driven approach from knowledge discovery with investor's requirements to achieve their goals and objectives. The data mining processes applied the SPSS, WEKA, data mining add-ins, Crystal Ball and TIBCO Spotfire end-user access data analysis tools to set up a decision tree induction. For example, in real estate property analysis, the pattern of data shows that the trend of property for sale is influenced by its features as the main contributors to the short time frame for the house to be sold as determinants. This main factor might not parallel with the investor's requirement as for them the most important factor is the location because they wanted to find a property for rental.

Objective 3:

To develop a PM-SA method to understand the sensitivity of the risk factors of real estate investment

A PM-SA method developed to analyze the variation of the multidimensional factors that meets the investor's requirements dynamically. This PM-SA method determines the effects of any pattern changes over a period of time with investor's personalization towards the analysis results. This PM-SA method describes both graphical and mathematical forms for analytical and descriptive decision-making processes. Another challenge of this research is to define both the most deterministic and descriptive factors for sensitivity analysis to obtain the best decision-making support. Sensitivity analysis explores the changes in weight for selected variables of property investment personalized by investors to see the variation of results. Controllable and uncontrollable risk factors for property investment listed as a guideline for the sensitivity analysis.

<u>Objective 4:</u>

To develop a prototype of Adaptive Personalized Property Investment Risk Analysis System (APPIRAS) by implementing the proposed framework and methods

A prototype of Adaptive Personalized Property Investment Risk Analysis System (APPIRAS) as a tool and to support dynamic risk analysis based on the personalized multidimensional process framework. This system developed based on the PMP framework developed in objective 1 and implemented the PAM method and PM-SA method developed in objectives 2 and 3, respectively. This prototype system developed according to several modules which includes introduction to APPIRAS, overall real estate property for sales or rent analysis module with several sub-modules: agent performance analysis module, property for the sales analysis module, property for rent analysis module, personalized property for sales or rent analysis module and potential property for sale/rent based on investor's goals and objectives module. DBMiner, WEKA, Crystal Ball, Risk Simulator and data mining add-ins using excel applied to develop the APPIRAS prototype system as these tools integrates major data mining functions with online analytical processing (OLAP) operations as well as data warehousing technologies, and discovers various kinds of knowledge at multiple conceptual levels of large relational databases.

The summary of relationship between research objectives and research questions is depicted in Figure. 1.1.

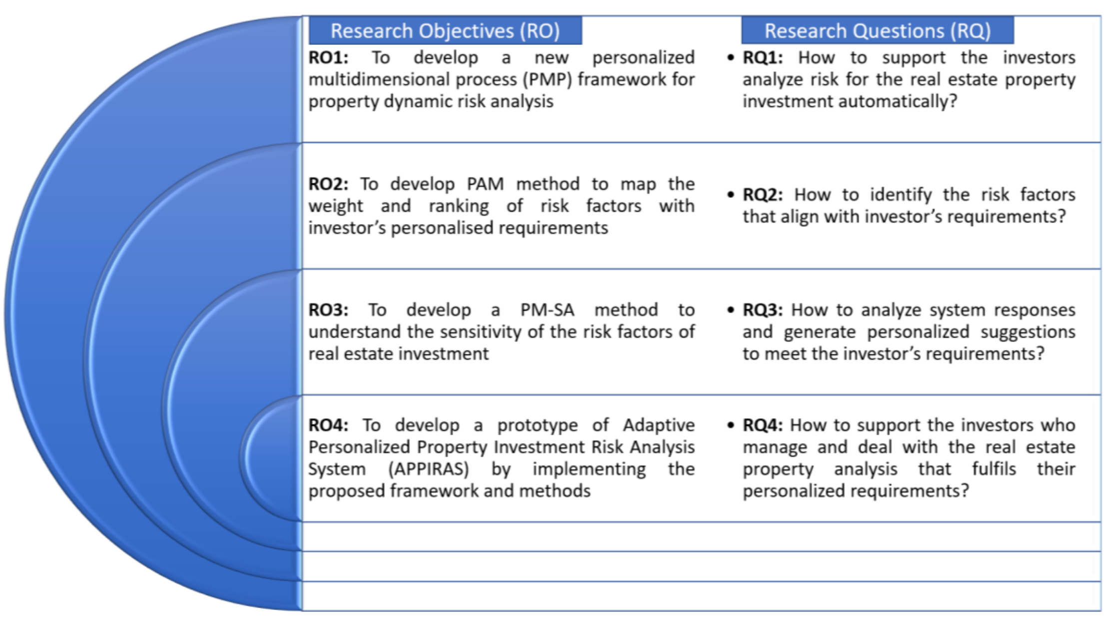


Figure 1.1 Relationship between Research Objectives and Research Questions.

1.4 Research Outcomes

In relation to the research objectives described in 1.3, the outcomes of this research listed as follows.

1): A new PMP framework to achieve Objective 1 which is described in Figure 3.1

A new PMP framework to deal with risk-based decision-making and risk analysis based on descriptive and data-driven or deterministic approach developed based on a generalized framework for knowledge discovery in business environments. This framework is new because it combines the concepts of personalization using multidimensional modelling to help investors analyzing risk and fulfil their requirements. It applied the data-driven approach of knowledge discovery as an input for the decision support that will be mapped with investor's requirements to achieve optimized decisions.

2): An algorithm to identify, weigh and rank the factors based on the pattern of data from the knowledge discovery approach for investors to evaluate and support them in the decision-making process to achieve Objective 1.

It provided an algorithm to identify, weigh and rank the factors based on historical data extracted from operational databases to the knowledge warehouse for data mining processes automatically as compared to existing technique which refers to expert's judgement manually.

3): A new personalized association mapping (PAM) method to completely achieve Objective 2.

A new PAM method developed to map the investor's requirement with pattern data from knowledge discovery. It is different from other methods because it will map the investor's requirements based on historical data provided using knowledge discovery techniques. This method is described as the mapping between investor's requirements with a data pattern from knowledge discovery using data mining technique and implemented in the model of outcome 1.

4): A new personalized multidimensional – sensitivity analysis (PM-SA) method to completely achieve Objective 3.

A new PM-SA method incorporated in the proposed framework to see the variation of the result if there is a change of pattern and investor's requirements as uncertainties are related to risk analysis for real estate risk factors. This sensitivity analysis is different because it applied the personalization concepts to investigate the variation of the risk factors based on investor's requirements. It is described as sensitivity analysis using multidimensional modelling and used to analyze the impacts of investor's personalization that meet their requirements. This method helps the investors understand the contributing factors for the optimum property investment. This PM-SA method was implemented in the model of contribution 1.

5): A new APPIRA method to completely achieve Objectives 3 and 4.

An adaptive personalization method for property investment risk analysis is needed to optimize and help to achieve investor's goals. Furthermore, it helps to fulfil the requirements of investors based on their limitations. Different investors will have different goals and will set their risk factors differently. Thus, the proposed method helps to fully consider the investors' constraints or limitations such as the amount of capital while achieving their goals. As shown in Figure 4.1, there are four main modules in the proposed adaptive personalization method, integrated to gather input from investors and provide output to investors for decision-making analysis. The input for APPIRA comes from adaptive personalization, risk factor determinants module and knowledge discovery module. Risk analysis module gathered and processed these inputs to produce the output for optimal decision-making, Main characteristics of APPIRA method refers to personalization, automatic processing, and user requirements diversification. However, the limitations of the APPIRA method are depending on the datasets available or stored in the system to produce the output and input from the investors. Timely and valid data are needed to achieve better results and contribute to optimal decision-making.

6): A prototype of an APPIRAS to achieve objective 4.

This APPIRAS system supports the real estate property analysis by PMP framework involving the personalized mapping method and PM-SA method. APPIRAS helps and supports the investors analyzing risks to achieve an optimal decision. This system integrates the framework with the two methods proposed. This system incorporates several modules that are implemented according to the proposed model of outcome 1. There are three main inputs for this system namely: pattern of data based on knowledge discovery in database using data mining technique, investor's personalization that meet their requirements and the investor's requirements sensitivity analysis using multidimensional modelling. Based on these three inputs, the output of the system will be the list of solutions based on investor's requirements.

The relationships between research questions and research outcomes depicted in Figure 1.2.

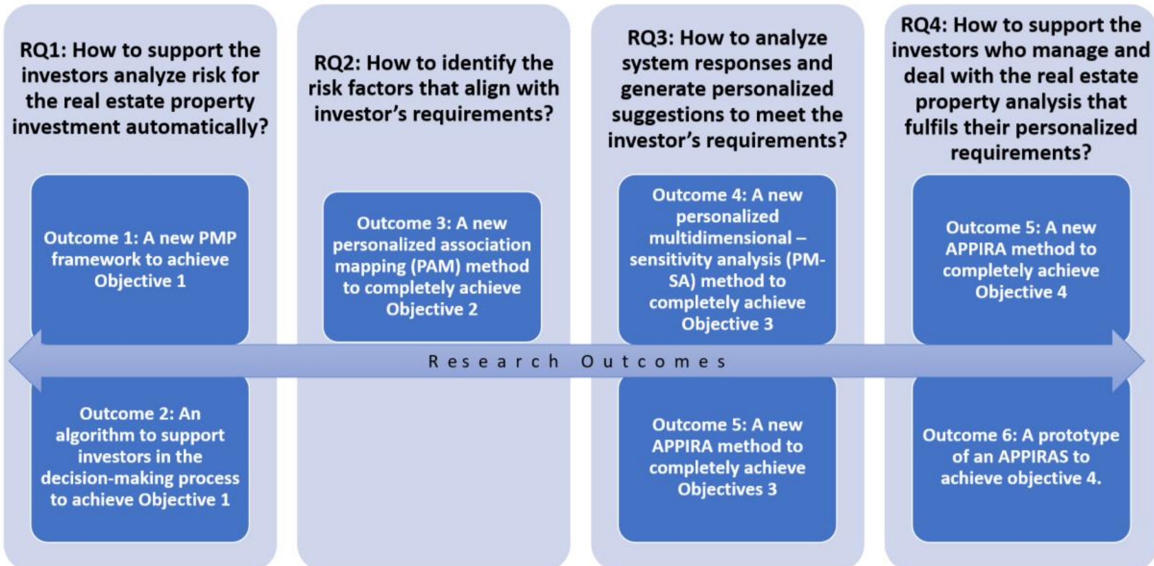


Figure 1.2 Relationship between Research Questions and Research Outcomes.

1.5 Significance and Innovations

This study contributed to both the new directions for real estate property risk analysis using the PMP framework and the application of APPIRAS that will ease and help the investors to choose the best real estate property for investment or other reasons. It is important to define the relationships among risk sources and its factors as determinant for specific objectives and weight the risk evaluation index based on knowledge discovery and map it with the investor's requirements to achieve the best result of dynamic risk analysis.

1.5.1 Significance

As mentioned in the statement of the problem, there is a need to study the techniques used and applied to measure the weight of risk factors that will affect the real estate investment dynamic risk analysis. This research filled the gap in the literature by contributing to a better understanding of the technique for identifying the determinants of risk factors based on different sets of goals or objectives, determining the weight of risk factors in the process of analyzing the dynamic risk of real estate investment. In addition, the PMP framework proposes in this research act as a guideline and a reference for the investors to be aware of when making the real estate investment dynamic risk analysis. Four significant improvements in the dynamic risk analysis of real estate investment can be gained from the output of this proposed approach.

(1) The PMP framework solved the weaknesses of existing dynamic risk analysis for investment dealing with measurement of risk factors weight and ranking that will affect the analysis result. The weight of the risk factors based on heuristic approach or expert survey applied in existing approaches has its limitations as it uses the single indexes, such as the maximum expectation, the largest variance, and the minimum standard deviation rate to evaluate the real estate investment. These methods do not consider the investor's requirements and the result might not fulfil the needs of investors. Thus, personalization is needed to fulfil the investor's requirements.

(2) The PAM method helps the investors to identify and set their justification and requirements based on the pattern of data provided by the analysis of historical data analyzed using a data mining approach. The combination of personalization approach (investor's personalization) and deterministic approach (data-driven approach) will meet the investor's requirements.

(3) The PM-SA method helps the investors to propose changes to the dynamic risk analysis for investment to improve the decision-making process. The PM-SA method will evaluate the impacts of the changes in the risk factors weight and ranking of the dynamic risk analysis results. If the result in the dynamic risk analysis does not change significantly, then there is no difference in risk factors that contribute to the result that has been personalized by the investors and vice versa. The investors will make the decision based on comparing the results from the analysis.

(4) The APPIRA method is a data-driven and personalized approach which personalizes the weight of risk factors using the TOPSIS model for optimal solutions. Since there are no existing approaches in real estate property analysis that integrate both a heuristic/analytic decision process and a deterministic approach to identify matters

dealing with risk identification, risk estimation, and risk assessment, this thesis explores and implements the APPIRA method. Moreover, it collects knowledge using multidimensional data models and data mining techniques and integrates it with the investor's personalization. The APPIRA method maps the investor's requirements against stored data using knowledge discovery to extract relevant patterns using knowledge discovery. These patterns indicate the highest contributing factors to selling a property in the shortest period.

(5) The APPIRAS prototype system supports the investors with the dynamic risk analysis for investment in the real estate industry in practice.

1.5.2 Innovations

(1) The proposed PMP framework overcomes the limitations of existing dynamic risk analysis techniques by implementing personalization association mapping and personalized multidimensional – sensitivity analysis. Existing risk analysis techniques do not deal with investor's personalization and requirements. Existing technique provides the risk evaluation index mostly based on expert surveys and many evaluation programs or models involving many evaluation indexes that the dimensions are different and the weights are difficult to determine, there are some difficulties in the practical application. Moreover, the existing approach does not include the personalized multidimensional sensitivity analysis which determines the impacts of different weights of risk factors towards the analysis results.

(2) The PAM method applied the investor's personalization to meet their requirements. Moreover, the investors weigh the risk factors based on data pattern from knowledge discovery which takes into account the historical data using data-driven or deterministic approach if compared to current approach which is normally focusing on heuristic approach (based on expert survey).

(3) The PM-SA method applied sensitivity analysis based on investor's personalization towards the weight of risk factors based on data patterns from the data mining technique.

(4) The APPIRA method provides a new way to deal with risk-based decision-making and risk analysis by combining a heuristic/analytic decision process and a deterministic approach. The key idea of the proposed method is to identify the determinant and map the weight and rank of relevant risk factors against investors' requirements (analytical approach), using data-driven techniques and multidimensional data modelling (deterministic approach) to identify relevant patterns in the data.

(5) The APPIRAS system overcomes the limitations of the existing real estate portfolio analysis for investment by integrating the data mining technique with online analytical processing and input from investor's personalization to deal with dynamic risk analysis. Existing system mostly provides risk analysis of real estate portfolio based on the heuristic approach and did not consider the investor's requirements. Thus, this research developed the prototype of APPIRAS that can deal with dynamic risk analysis for investment which includes the investor's personalization combined with the analytic and a deterministic approach.

This research has four main innovative results as shown in Figure 1.2.

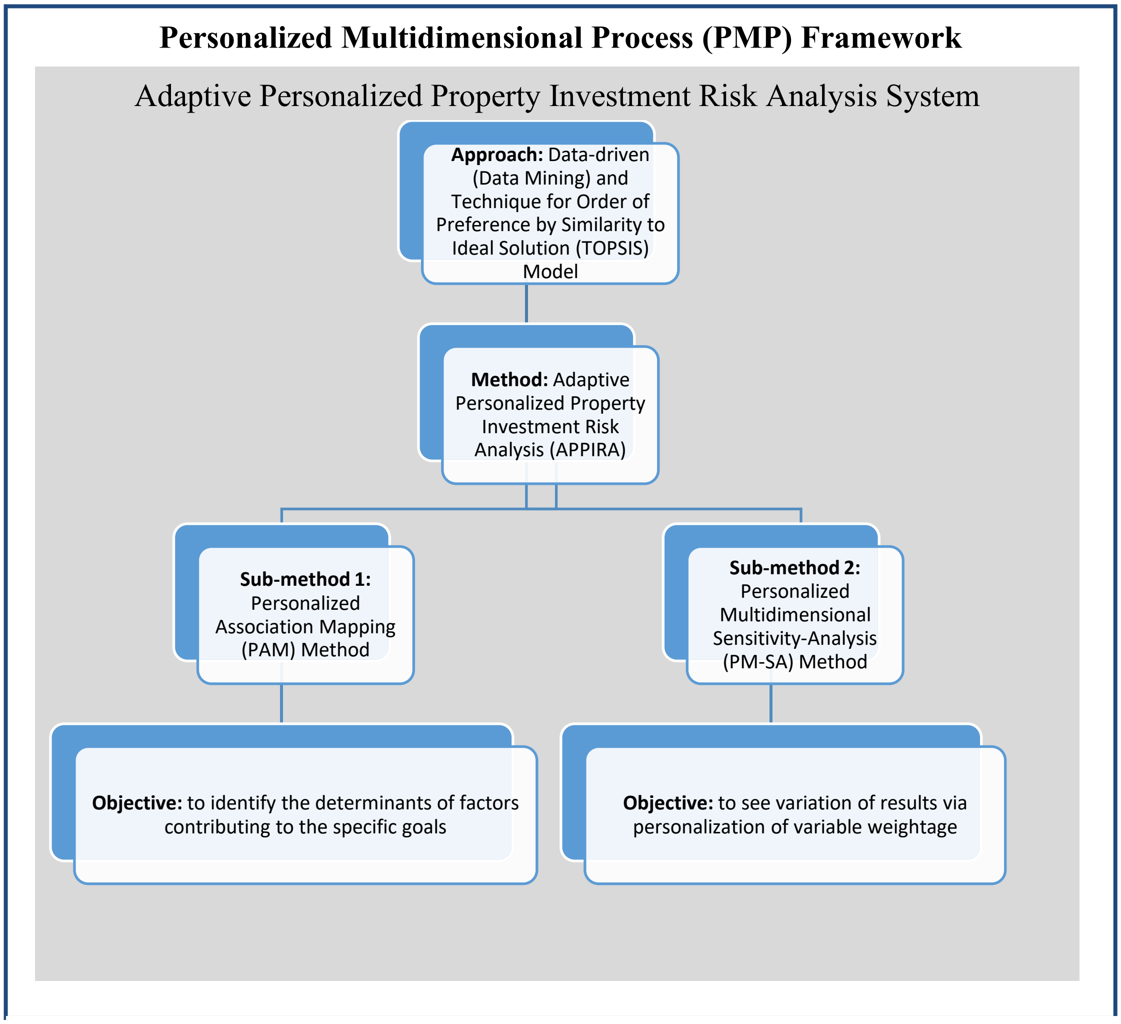


Figure 1.3 Four main innovative results in this study.

1.6 Research Methodology

This section describes how the research objectives given in Section 1.3 have been achieved in this study. Methodology in this research incorporated literature review, framework development, variable classification, variable ranking, data collection, data processing and analysis, mapping method development, sensitivity analysis method development, two new algorithm development and real estate case study prototype software system development and evaluation as depicted in Figure 1.2.

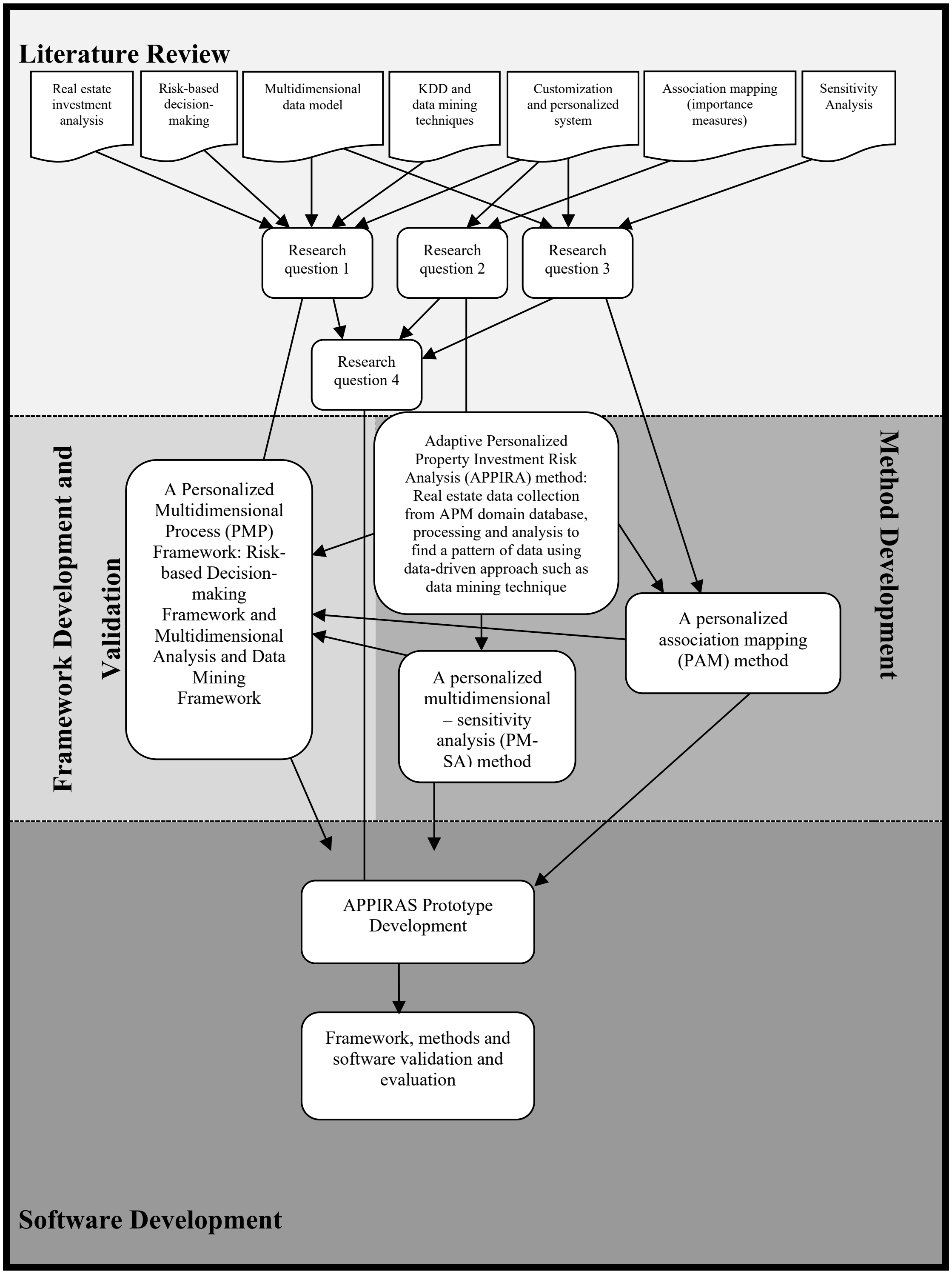


Figure 1.4: Research methodology for this study.

1.6.1 Problem Definition

At this stage, thorough and in-depth analysis of relative old and current issues related to risk-based decision-making and risk analysis for investment, explored and investigated. Existing methods, models, techniques, and approaches related to risk-based decision-making and risk analysis for investment identified and analyzed to discover its features, characteristics, limitations, advantages, and disadvantages. Moreover, issues and critical review of the models applied in different industries are identified to find out the gaps and limitations of the proposed model by other researchers. In line with this identification, existing methods, and approaches to improve property investment risk analysis were compared to uncover their common disadvantages and finally to reveal possible research gaps. At this stage, possible research gaps found in the literature review structured to be investigated and studied for this research project. Further investigations of existing approaches, techniques, framework, or model have been explored to compare and identify the issues and problems. The main problems identified and divided into several sub problems to define the specific research questions. Then, the research objectives and outcomes defined based on research questions listed earlier.

1.6.2 Research Planning

This stage involves several decisions and assessments. New framework, algorithm, method or model, techniques, approaches and software system needed to overcome the weaknesses of existing property investment risk analysis which refers to expert's judgment for ranking and weighting the property variables found in the previous stage are set. Resources and strategies needed for analysis and validation to see the feasibility of the proposed solution are also determined at this stage. Finally, a comprehensive road map to achieve the research objectives is then defined in a project management plan and ready for full-scale implementation.

1.6.3 Method Development

At this stage, the new framework, model, technique, approaches, and software system that have been defined in the previous stage are realized to gain the research objectives. Based on a deep understanding of existing risk-based decision-making models, a PMP framework developed to solve the issues and limitations of other approaches. The development of the PMP framework is based on the concepts of data model, an AHP and ANP that have been developed to deal with risk-based decision-making processes. After analyzing these two processes, a multidimensional data model is applied as an extension from these two processes integrated with a personalization approach to support decision-making with risk based on knowledge discovery of data mining techniques. The PMP framework applies the combination of heuristic and deterministic approach to support decision-making with risk. This new framework applies the concept of data-driven approach to support the decision-making process. Several data mining tools such as Crystal Ball, Risk Simulator, WEKA, TIBCO Spotfire, data mining add-ins using excel, and decision tree induction of data mining techniques applied in the proposed framework to discover the pattern of data or knowledge discovery. After analyzing the pattern of data, a PAM method developed to map the investor's requirements with the knowledge discovered to achieve their goals and objectives. The PMP framework integrates online analytical processing (OLAP) and data mining techniques which are referred to as the data warehouse end-user access tools. For PAM method, dimensional modelling of OLAP data model applied to analyze the multidimensional factors from the data mining process mapping with investors weight and ranking of factors customization. Then, a PM-SA conducted to discover the changes of factors that will affect the result of the decision-making process and the investors' requirements dynamically. Both PAM and PM-SA methods were tested using the APPIRA method and adopted in APPIRAS. The PMP framework validated by comparing the results of decision-making with risk based on case study to show its applicability for risk-based decision-making analysis.

1.6.4 Analysis and Validation

At this stage, illustrative case studies as well as real-world applications were conducted to analyze and validate the performance and the effectiveness of the developed property investment risk analysis system, technique, and approaches. These analyses and validation explore their capability and feasibility for identifying, ranking, and weighing the risk factors as well as existing property investment websites. Moreover, the results of the analysis and evaluation were used for further improvements and to find new directions for development. Furthermore, the inputs for the proposed framework are gathered and extracted for processing and producing the output. This stage involves data collection, data analysis and designing of multidimensional data models (dimensionality modelling).

Step 1 Data Collection

At this stage, real estate data from online transaction processing (OLTP) domain database of Australian Property Monitor (APM) subscribed by the University of Technology Sydney (UTS) Library and reports collected and stored in the end-user access tools for case study. Attributes of databases and reports provided from APM analyzed and the important attributes determined to be the use of risk analysis for decision-making. Data from reports provided by APM gathered with data from the domain database. A database created to store the data extracted from the APM domain database. More than 1000 rows of data extracted from the data analysis. Each property portfolio comprises 50 attributes available for the analysis of the comparative market appraisal report. For market search activity, there are 11 attributes analyzed for investment risk analysis. The data collected based on the time frame given in the database to keep the historical data and used to determine the pattern of data over a period of time. The data collected and stored in the database or end-user access tools used for a case study.

Step 2 Data Analysis

After related data have been collected and extracted, the data were analyzed using end-user access tools. The data extracted need to be in standard format, time-variant, subject-oriented and non-volatile. The denormalization technique used to group the category of data for subject-oriented purposes which are useful for data analysis.

Step 3 Data Design

At this stage, the data that have been analyzed and useful for data analysis mapped into a multidimensional data model of OLAP. A dimensional modelling, Unified Modeling Language (UML) class diagram of the selected attributes for analysis and a star schema designed based on denormalized data for analysis purposes. A multidimensional data model using dimensional modelling designed for the use of OLAP and data mining application. A dimensional modelling which contains dimension table and fact table for OLAP will calculate the aggregate tuples or rows of the data contained in the database. The multidimensional data model views the risk factors (determinants) as the dimensions in a star schema surrounding the fact table for the decision-making analysis.

Step 4 Personalized association mapping (PAM) method development (to achieve Objective 2)

After gathering of output from Step 3, which refers to the output from OLAP and data mining (OLAM), the pattern of data given to the investor for their judgment and investor will personalize the queries based on the results to meet their requirements. This will lead to the development of a personalized association mapping (PAM) method to map the pattern of data from multidimensional and data mining processes, OLAM with the investor's weight and ranking of factors to meet their goals and objectives. A case study to map the pattern of data from multidimensional and data mining processes, OLAM with the investor's weight and ranking of factors to meet their goals and objectives produced to show the applicability of the PAM method.

Step 5 Personalized multidimensional – sensitivity analysis (PM-SA) method development (to achieve Objective 3)

Since risk factors are related to uncertainty and investor's requirements are different from time to time, a personalized multidimensional – sensitivity analysis (PM-SA) developed to determine the sensitivity of the risk factors dynamically. The investors will send queries based on their personalization where the combination of output from Step 3 compared to the decision-making process.

Step 6 Adaptive Personalized Property Investment Risk Analysis System (APPIRAS) development (to achieve Objective 4)

A new personalized multidimensional process (PMP) framework developed in Step 4 used as the structure of the adaptive personalized property investment risk analysis system (APPIRAS), developed at this stage. The development of the software includes identifying the requirements of the system, designing the interface, coding, and testing the system developed. The collection, analysis, and design of data in Step 3 used as the input for the system according to the multidimensional data model in Step 4 and Step 5. The APPIRAS implemented the PAM method in Step 4 and PM-sensitivity analysis method in Step 5.

Step 6.1: Determining the requirements of the Adaptive Personalized Property Investment Risk Analysis System (APPIRAS)

At this stage, the flowchart of how the system works and the requirement of input and output of the system are identified and determined. A list of modules of the system produced and UML diagram produced to show on how the systems work.

Step 6.2: Designing of APPIRAS

At this stage, the user interface is designed including the screen layouts, pseudocode and rules using selected end-user access tools such as WEKA, DBMiner, Crystal Ball. Selected end-user access tools used to develop this system because these tools can support both SQL-like data mining query language, called DMQL, and a graphical user interface for interactive mining of multiple –level knowledge. Moreover, these tools provide multidimensional data visualization support and interact with standard data sources through an open database connectivity (ODBC) interface. Furthermore, these wide range of tools support multi-functional, online analytical mining (OLAM) in large databases.

Step 6.3: Coding the APPIRAS

At this stage, the coding of the system generated based on Step 3 until Step 5. This step ensures that the design must be translated into a machine-readable form.

Step 6.4: Testing the APPIRAS

At this stage, once the code is generated, the software program is tested. System testing and unit testing will be applied at this stage to make sure that the system runs smoothly, and each module fulfils the criteria of Step 3 until Step 5.

Step 6.5: PAM and PM-SA Method validation

The PAM method developed in Step 4 and PM-SA method in Step 5 validated by running these two methods within the developed APPIRAS at this stage. The validation mechanism is by using the output of Step 4 as the input mapping with investors personalization query based on their goals and objectives. The output of different combinations compared using a PM-SA method to support the decision-making process.

Step 6.6: APPIRAS evaluation

At this stage, the APPIRAS developed, tested, and explored to see its applicability for real estate dynamic risk analysis which consists of risk identification, risk estimation and risk analysis. User testing applied in which any respondent, who is new to the system will be asked to use the system. Potential users who would like to invest in real estate property asked to test the system and the result analyzed whether it meets their requirements and satisfaction. Their feedback was to improve the system’s modules, functionality, and its user interface design to produce an effective, efficient, and user-friendly system.

1.6.5 Evaluation and Revision

At this stage, data extracted and stored in the database is processed using DBMiner software tool and other end-user access tools to produce the output. End-user access tools used to integrate major data mining functions with OLAP operations as well as discovering various kinds of knowledge at multiple conceptual levels of large relational databases. End-user access tools provide a wide range of tools to support multi-functional, on-line analytical mining (OLAM) in large databases. The output of OLAP is used for decision-making processes for structured data types. Decision tree induction as a data mining technique used to find the pattern of data from the database.

1.7 Thesis Structure

The structure of this thesis is as follows: In Chapter 2, the literature review is discussed. In Chapter 3, the adaptive personalized property investment risk analysis (APPIRA) frameworks were illustrated and explained. In Chapter 4, this paper describes in detail the APPIRA method which integrates the PAM and PM-SA methods. The APPIRA system prototype screenshots are illustrated, and its main features elaborated in Chapter 5. In Chapter 6, case study inclusive of data analysis of an experiment and results discussed in detail to depict on how these new methods were implemented. The conclusions and future direction of this research are discussed in Chapter 7.

References list all sources that have been used to complete the study and to write this thesis. The relationships amongst chapters of this thesis are graphically described in Figure 1.2.

1.8 Publications Related to This Thesis

Below is the list of publications relating to this thesis from the beginning of the study to present.

Book Chapters

1. Demong, N. A. R. & Lu, J. 2011, 'Risk-based Decision-making Framework for Investment in the Real Estate Industry', In: Lu J., Jain L.C., Zhang G. (eds) Handbook on Decision-making. Intelligent Systems Reference Library, vol 33. Springer, Berlin, Heidelberg
2. Demong, N.A.R., Lu, J. & Hussain, F. K. 2013, 'Personalised Property Investment Risk Analysis Model in the Real Estate Industry', Human-Centric Decision - Making Models for Social Sciences, 7092, Springer-Verlag Berlin Heidelberg.

International Conferences

3. Demong, N. A. R., 'Gaining Tacit Knowledge based on Deterministic Approach for Risk Analysis in the Real Estate Industry', Proceeding of the Advances in Business Research International Conference (ABRIC 2015)
4. Demong, N. A. R. & Lu, J. 2011, 'Personalized Multidimensional Process Framework for dynamic risk analysis in the real estate industry', Sixth International Conference on Construction in the 21st Century, Kuala Lumpur - Malaysia July 5-7, 2011.
5. Demong, N.A.R., Lu, J. & Hussain, F. K. 2012, 'Multidimensional and Data Mining Analysis for Property Investment Risk Analysis', International Conference on Information Systems, December, 6 – 7, 2012, Penang, Malaysia World Academy of Science, Engineering and Technology, 72.

International Journals

6. Demong, N.A.R., Lu, J. & Hussain, F. K. (2019). An Adaptive Personalized Property Investment Risk Analysis Method Based on Deterministic Approach International Journal of Information Technology & Decision-making, (Status: Revisions being processed – under third review, Manuscript No: IJITDM-D-18-00132).

CHAPTER 2

LITERATURE REVIEW

Existing, current as well as relatively past ten years sources dedicated to the problems of this research were reviewed. Since this study involves property investment risk analysis, existing techniques, real estate analytic systems, data-driven decision-making approach, personalization and multi-criteria decision-making techniques, we review literature on these topics below.

2.1 Property Investment Risk Analysis

Property investment is high-risk, high-cost, dynamic, and involves uncertainty (E. C. M. Hui et al., 2010, Bergmann et al., 2020, Lu et al., 2020). The relationship between risk, cost, and rate of return is very complicated and each correlate in a positive way where high cost will lead to high risk but will get a high rate of return. In real situations, risk in property investment is typically defined as the uncertain elements that are affected by qualitative or quantitative factors – the higher the uncertainty, the higher the risk (Lieser & Groh, 2014).

Most risk analysis techniques used in real estate assign a weight to each risk factor, based on the expert's experience. Experts in the real estate industry refer to real estate agents, brokers, appraisers and property management teams. There are many limitations and disadvantages in using expert's judgment to help property investment decision-making, but the main concern with this approach is that incomplete preferences and loss of information. Moreover, an investor's goals may not be similar to the expert's goals. Additionally, experts in the field may be familiar with the factors that affect decision-making and know how to monitor or control uncertainty factors. However, they may have different perceptions or methods of analysis (L. Chen, 2010; Lafleur, 2011, Pires et al., 2018), which could result in a misrepresentation of the weight and rank of those risks from the investor's perspective. The need for a comprehensive and interactive system that investors can use to model and predict risk in their decision-making process is clear. This study examines the important role of risk allocation for property features selected by the investor to fulfil their requirements.

Property investment in the real estate industry is not a short-term business. As an investor, there are many criteria to look at when choosing the right investment for optimal decision-making. Understanding the types of investment risk in the market is crucial in achieving the investor's goals and the decision made relative to current investor's constraints or limitations. Real estate investment is a high-cost, high-risk, high-return activity (Beracha, Downs, & MacKinnon, 2017). Risk is high because the factors affecting investments are uncertain and involve decision-making based on multiple criteria (Almeida, Baştürk, & Golan, 2017). When the future is uncertain and investments are durable and illiquid, the decision to invest at a certain point in time and the correct assessment of risks are key issues. Investors need to know how to measure risks and identify the relationship between risks borne and risks premiums demanded (D'Alpaos & Canesi, 2014). Risk with high uncertainty leads to a higher potential for failure, and the consequences of failure will have a greater impact (Ebert, Wei & Zhou, 2020). Uncertainty, in dynamic risk prediction, is dependent on many factors throughout the life of an investment, from initial purchase to the final stages of development (Almeida et al., 2017). Another factor that affects uncertainty is the investor's and the experts' understanding of how to interpret identified risk factors. Since property investment risk analysis deals with dynamic risk factors, thus, the next section will review the dynamic risk-based decision-making concepts in detail.

2.1.1 Dynamic Risk-based Decision-making

This section discusses the concepts of risk-based decision-making which includes the definition of risk, types of risk and brief explanation about risk analysis. These three concepts are mainly related to the concepts of risk-based decision-making for investment applied in the real estate industry.

2.1.1.1 Definition of Risk

According to Aven (2007), "risk is defined as the combination of possible consequences and associated uncertainties (uncertainties of what will be the consequences), and a source is a situation or an event with a potential of a certain consequence". Hence, Aven (2007) defined risk as the combination of sources (including associated uncertainties and vulnerabilities) as shown in Figure 2.1. Three categories of sources: threats, hazards, and opportunities.

Exposure of a system to certain threats or hazards, can lead to various consequences or outcomes such as economic loss, number of fatalities, the number of attacks, the proportion of attacks being successful (Terje Aven, 2007).

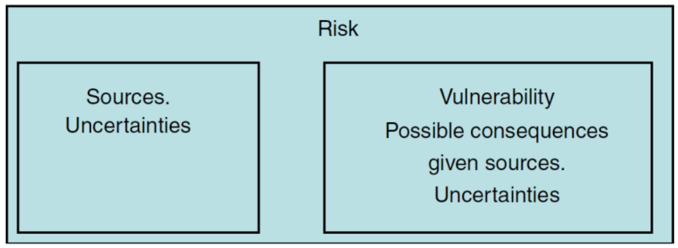


Figure 2.1: Risk viewed as a combination of sources and vulnerability (Terje Aven, 2007).

Risk-based decision-making is a process of making decisions based on risk related to the choices in which data and information about the list of choices need to be analyzed. The result of the analysis of the choices in the list will vary depending on the uncertainty factors that will affect the decision-making process. The risk arises because of possible consequences and associated uncertainties, and there are several risk factors or variables and risk sources that will affect the level of risk for given alternatives. The risk of each investment project is represented as probability and consequences or unwanted outcomes that should be minimized or eliminated to optimize the benefits of the investment. Risk factors and risk sources which are uncertain and hard to predict will give impact to the probability and consequences of each investment project in the list.

2.1.1.2 Types of Risk

There are several types of risk that affect the decision-making process for investment in the real estate industry. The two most common categories of risk in real estate industry investment are systematic risk (beta) and unsystematic risk. Unsystematic risk is also known as idiosyncratic risk. According to Chauveau and Gatfaoui (2002), systematic risk is a measure of how the asset covaries with the economy and unsystematic risk which is independent of the economy. According to the Capital Asset Pricing Model (CAPM), a company's total risk consists of two types of risk: unsystematic and systematic risk (Lintner, 1965; Sharpe, 1963, 1964 as cited in Lee and Jang (2007). Systematic risk referred to a type of risk that influences a large number of assets. Systematic risk cannot be avoided despite diversified stock portfolio diversification (Brealey & Myer, 2000 cited in Lee and Jang (2007). Unsystematic risk or idiosyncratic risk is sometimes referred to as specific risk which is sensitive to diversification if compared with undiversifiable systematic risk. Idiosyncratic risk matters for asset pricing because it inhibits the intergenerational sharing of aggregate risk (Storesletten, Telmer, & Yaron, 2007).

According to the CAPM (Lintner, 1965; Sharpe, 1963, 1964, cited in Lee and Jang (2007), the total risks are calculated as follows:

Total risk = Systematic risk + Unsystematic risk

Systematic Risk. Systematic risk refers to a type of risk that influences a large number of assets. It cannot be avoided despite stock portfolio diversification (Brealey & Myer, 2000, cited in Lee and Jang (2007). According to Lee and Jang (2007), systematic risk can differ from period to period. Managerial decisions about operations, investments, and financing will influence the performance of the company, consequently affecting how its returns vary with market returns. The CAPM suggests that the expected rate of return on a risk asset can be obtained by adding the risk-premium to the risk-free rate; the expected risk premium varies in direct proportion to beta in a competitive market (Chen, 2003; Gencay, Selcuk & Whitcher, 2003; Lintner, 1965; Sharpe, 1963, 1964; Sheel, 1995, cited in Lee and Jang (2007). Mathematically, the expected rate of return is described as

$$R_i = R_f + (R_m - R_f)\beta_i$$

where R_i is the expected return on the ith security R_f the risk-free rate; R_m the return on the market portfolio; β_i the estimated beta of the ith security; $(R_m - R_f)\beta_i$ the risk premium. Based on CAPM, systematic risk refers to a type of unavoidable risk on the stock market. Systematic risk is presented by beta which is calculated by linear analysis between the daily prices of stocks and the security index of the stock market (J. Zhou, Wu, & Xin, 2006).

Unsystematic or Idiosyncratic Risk. Unsystematic risk, or idiosyncratic risk, is sometimes referred to as a specific risk which is sensitive to diversification, contrasting with systematic risk, which is undiversifiable. Idiosyncratic risk is significant for asset pricing because it inhibits the intergenerational sharing of aggregate risk (Storesletten et al., 2007).

CAPM and Arbitrage Pricing Theory (APT) assert that idiosyncratic risk should not be priced in the expected asset returns, and the recent surge of interest in the idiosyncratic risk of common stocks has generated substantial evidence on the role of idiosyncratic risk in equity pricing (Liow & Kwame.Addae-Dapaah, 2010). The main reason for this interest is that most investors are under-diversified due to wealth constraints, transaction costs or specific investment objectives; as such, idiosyncratic risk may be important to less well-diversified real estate investors who wish to be compensated with additional risk premium. Such investors need to consider idiosyncratic risk (together with market risk) when estimating the required return and the cost of capital on assets or portfolios. Both systematic (market) and idiosyncratic volatility are relevant in stock asset pricing (Campbell et al., 2001, cited in Liow & Addae-Dapaah, (2010)).

Various intelligent techniques including the Real Option method, Multi-State approach, variable precision rough set (VPRS), Condition Value-at-Risk (CVaR), AHP, Support Vector Machine (SVM), Radial Basis Function Neural Network, Fuzzy Comprehensive Valuation Method and Projection Pursuit Model based on Particle Swarm Optimization (PSO) have been applied to deal with and support unsystematic or idiosyncratic risk-based decision-making.

2.1.1.3 Risk Analysis

The risk analysis concept has always been present in business transactions especially in the real estate industry which involve high cost and high capital (Chong et al., 2008). Risk analysis has proven its value in reducing risks in projects. Regardless of the type of risk analysis process, the application of risk management has a positive effect in finding and taking actions to avoid events that could cause negative consequences for the project and the organization (Olsson, 2007). Risk analysis is a vital process for project investment in the real estate industry with low liquidity and high cost. Risk analysis consists of three stages: risk identification, risk estimation and risk assessment (Yu & Xuan, 2010a). Risk analysis includes identification of risks, gathering information related to risk analysis or risk assessment and determining the risk level for each choice or alternatives given and last but not least deciding and selecting the best option for investment. There were several risk analysis techniques, tools, models and methodologies invented or created for analyzing risk applied in different industries.

Risk occurs at a different stage of the investment process. The risk assessment or risk analysis is part of the process involved when deciding on which project should be given priority for investment with the limited budget and time given. There are many models and methodologies available for risk-based decision-making. The investors or investors should understand and must know these models and methodologies for optimal decision-making.

Risk analysis is an important process that needs to be conducted to achieve optimal decision-making. The real estate franchisors can achieve their goals and objectives if they fully understand and can identify uncertain factors or variables that will affect the level of risk for the alternatives given. The uncertain factors or variables will lead to probability and consequences and can be reserved as a list of threats that will affect the risk level. Therefore, it is important to discover the framework of risk analysis for investors in the real estate industry.

Aven, Vinnem and Wiencke (2007) proposed a framework which comprises the basic elements of risk management. The elements of the framework are as follows: problem definition; (challenges, goals and alternatives), stakeholders, concerns that affect the consequence analyses; and the value judgments related to these consequences and analyses (frame conditions and constraints), identification of which consequence analyses to execute and the execution of these, managerial review and judgement, and the decision (T. Aven et al., 2007). Risk analysis involves decision-making in situations involving high risks and large uncertainties, and such decision-making is difficult as it is hard to predict what would be the consequences of the decisions (Deng & Ma, 2008; C. Lin, Meng, & Pan, 2001; H. Sun, Fan, & Shi, 2009; Xuefang. Zhang & Ji, 2010).

Another risk assessment framework proposed by Li et al. (2007) included the risk factor system, data standards for risk factors, weights of risk factors, and integrated assessment methods are used to quantitatively analyze the outbreak and spread of highly pathogenic avian influenza (HPAI) in mainland China. They used a Delphi method in determining the risk factors according to the predetermined principles. Moreover, an analytical hierarchy process (AHP) and integrated multicriteria analysis was used to assess the HPAI risk.

2.1.2 Types of Risk-based Decision-making Process

According to Koller (2005), smart decisions and choices are facilitated by the following: (1) Assuring the risk/uncertainty assessment process is holistic and includes such considerations from all pertinent aspects and (2) Assuring that the value of a project, however calculated, has been properly impacted by all risks and uncertainties. Problem solving is a critical activity for any business organization (Stair & Reynolds, 2008). In order to make the decision, it involves a series of steps to ensure the right decision is made.

The decision-making process can be divided into two main types: static decision-making process and dynamic decision-making process. These two main types of decision-making processes are related to different strategies of investment in the real estate industry either simultaneous strategy or sequential strategy. Simultaneous and sequential investments are common in the real estate market (Rocha et al., 2007).

2.1.2.1 Static Risk-based Decision-making Process.

Static decision-making normally corresponds to the simultaneous investment strategy. It is a now-or-never decision where all irreversible resources are compromised at once. Static decision-making is related to the simultaneous strategy for investment in the real estate industry. The simultaneous strategy is usually implemented during periods of increasing demand and implies lower construction costs but, in turn, carries more uncertain returns.

Bitter experiences with residential housing developments and mega-entertainment resorts started simultaneously having generated profits only after five or more years of construction (Rocha et al., 2007). Figure 2.2 shows the static decision-making process.

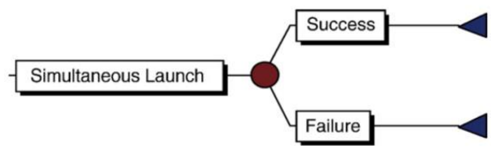


Figure 2.2: Static (simultaneous) decision-making process (Rocha et al., 2007).

2.1.2.2 Dynamic Risk-based Decision-making Process.

Dynamic decision-making process is related to sequential investment strategy, where risks are faced in sequence with relatively smaller increments at every phase of the project, but at the expense of relatively higher construction costs. In sequential investment strategy, the initial outflow is lower than in the simultaneous investment strategy and expected inflows of previous phase may finance subsequent ones (Rocha et al., 2007). Dynamic decisions arise in many applications including military, medical, management, sports, and emergencies. Dynamic risk prediction is an important process that needs to be focused on when dealing with investment with a limited budget, time and other constraints and to achieve the optimal decision-making. The dynamic decision-making process for the prediction of risk level generally related to three managerial flexibilities characteristics of real option methods as depicted in Figure 2.3.

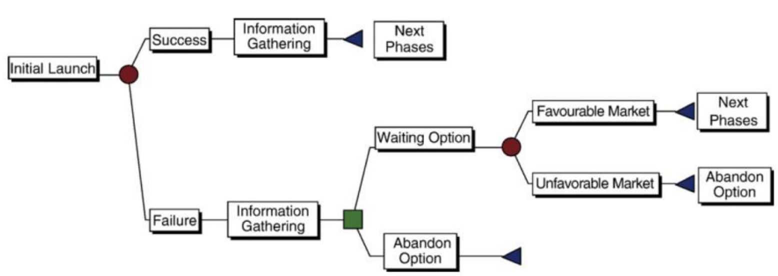


Figure 2.3: Dynamic (sequential) decision-making process (Rocha et al., 2007).

Due to controllable and uncontrollable property features, property investment risk analysis related to dynamic decision-making processes and need to carefully analyze to achieve optimal solutions.

The most important managerial flexibility characteristic related to dynamic decision-making is the waiting option. The challenges investors' face with bandit problems is that the characteristics of each option are often unknown, especially when dealing with higher uncertainty of risk factors or risk sources. Higher uncertainties of risk sources or risk factors will lead to higher probability or consequences of the decision made for the given options. Besides, owing to the increasing complexity of the decision, the uncertainty of evaluation also increases. Under this situation, investors would not be able to use precise numbers to express their evaluations, but they can still give approximate ranges of evaluations by their knowledge and cognition (Y.-H. Lin, Lee, & Ting, 2008). However, to achieve optimal decision-making, the knowledge of risk analysis is important for them to master and they need to analyze the risk sources and risk factors for the given options.

2.1.3 Decision under Uncertainty

The uncertainty of the risk factor in the real estate industry led to a high cost and a high risk for the risk analysis (Zhai, Chen & Chen, 2018). Most existing techniques for the risk analysis in the real estate industry refers to experts in the field to rank and weigh the risk factors, especially for the novice investor. This technique creates misinterpretation and different judgments from different experts in the field which led to inaccurate risk measurement.

Dynamic risk analysis in the real estate industry has always dealt with the uncertainty factors (Pomerol, 2001). The uncertainty factors create high risk as it involves the high cost of investment in the real estate industry (Kwon & Kim, 2012; Lafleur, 2011; Rocha et al., 2007). The uncertainty of the risk factor for investment risk analysis in the real estate industry includes financial risk, economic risk, location risk, scheduled risk, technical risk, policy risk, contractual risk and others. For example, the financial risk refers to the uncertainty of profits which originates from the process of financing, money allocation and transfer, interest payment as financial aspects of a project (A. Piyatrapoomi, A. Kumar, & S. Setunge, 2004). Another example, the economic risk includes economic risk, regional development risk, market supply and demand risk and inflation risk which is uncertain (L. Chen, 2010).

The investors need to have in-depth knowledge in the field to decide the best investment and to reduce the risk of loss (L. Liu et al., 2007; Niu, Lu, & Zhang, 2009). Hence, knowledge management of dynamic risk analysis is important to help investors make better decisions and be more reliable. Risk measurement to rank and weigh the risk factors is a major step to consider. Existing risk analysis techniques are still lacking sufficient and comprehensive evaluation for investors to make a good decision (Demong & Lu, 2011). Tools to rank and weigh the uncertain risk factors for dynamic risk analysis, such as AHP and Delphi methods have existed for some time. These techniques referred to expert's opinion in ranking and weighting the risk factors for risk analysis (J. Gao & Wang, 2009). Thus, it creates misinterpretation and different judgments from the experts or professionals in the fields.

Moreover, these tools have significant shortcomings for settings personalized by constraining the investor's goals and objectives that change dynamically. In addition, different users will have different requirements and goals of the investment. Some requirements are simple and seem straight forward, while for others, are complex and require more analysis to make the decisions. Therefore, there is a need for a new method that is more reliable and trusted as compared to referring to the experts in the field. Expert opinion was unable to incorporate more empirical evidence that contributed to project failure (Glorio, Mazón, Garrigós, & Trujillo, 2012).

Besides, other issues related to the application of expert opinion include lack of reliable reference (O. K. Hussain et al., 2007), expert opinion may change (Glorio et al., 2012) and it depends on the expert's level of experience in the field (Shiwang et al., 2009). As a result, a more systematic approach to accumulating and reporting evidence that can provide in-depth knowledge and can solve the problems from different user's requirements.

2.1.4 Risk Sources and Risk Factors in the Real Estate Industry

This section explains the risk sources and risk factors for investment in the real estate industry. The five main categories of risk sources and risk factors for investment in the real estate industry, namely financial risk, economic risk, scheduled risk, policy risk, and technical risk and others. Each of these risk sources has its own risk factors as a sub-element.

2.1.4.1 Financial Risk

Financial risk refers to the uncertainty of profits which originates from the process of financing, money allocation and transfer, interest payment as financial aspects of a project (Yu & Xuan, 2010a). According to Ma and Meng (2009), financial risk includes own fund risk, bank loan risk, shares risk and financial structure risk. Financial risk analysis is the core of the real estate investment risk analysis and will directly determine the decision-making of investment. The financial risk consists of three sub-category risk factors: policy, engineering and market (Yu & Xuan, 2010a). Even though the investment in the real estate industry incurred high cost and slow liquidity, however, the investment in the real estate industry gives more value and high rate return of investment in a short period of time (Shujing. Zhou, Wang, & Li, 2010). Moreover, property prices such as house price are affected by many factors, for example, interest rate, land supply and inflation rate (E. C. M. Hui et al., 2010). The real estate industry has high risks as it is heavily dependent on bank loans as the financial risk in real estate investment and those risks involved can trace back to the asset security of bank loans (Yang, 2008). Financial risk also can be mitigated by analyzing the real estate portfolio based on financial requirements for the real estate investment (Chong et al., 2008). According to Liu, Zhao and Liu (2007), real estate portfolio models are stated as the following: each project is totally competitive, each investor cannot monopolize the market or control return rate (Cao, 2005 as cited in Liu, (2007).

2.1.4.2 Economic Risk

Economic risk includes economic risk, regional development risk, market supply and demand risk and inflation risk (Ma & Meng, 2009). Li and Suo (2009) defined economic risk factors consist of sales cycle, industry competitiveness, economic operation, exchange and interest rate. Sun et al. (2008) proposed a model on general relationships among the significant elements for dynamic risk prediction for real estate franchises in the real estate industry using the real options method.

2.1.4.3 Scheduled Risk

Schedule risk will affect the degree of risks for alternatives given (Y. Sun et al., 2008). Real estate investment normally relates to capital risk and liquidity risk. Liquidity risk related to scheduled risk because both risk sources are dependent on time series. Yu and Xuan (2010a) define schedule risk as the delay of some part of the whole project process or even the entire project, which is often accompanied with the increase of cost; The Critical Path method based on the Work Breakdown Structure (WBS) is the mainly-used methodology to the schedule risk.

2.1.4.4 Policy Risk

Policy risk refers to objective existences, uncontrollable, of great harm and small likelihood of occurrence as the source of risk in real estate investment (Yu & Xuan, 2010a). Jin (2010) highlighted the policy environment risk as the lifecycle risk impact factors of the real estate project. Organization policy and industrial policy also an example of the variable or factor that will affect the result of risk analysis (Y. Sun et al., 2008). Ma and Meng (2009) defines policy risk which includes monetary policy risk, industrial policy risk, land policy risk, housing policy risk, tax policy risk and town planning risk. Li and Suo (2009) points out that the policy factors consist of environmental policy, tax policy, financial policy, and industrial policy.

2.1.4.5 Technical Risk and Others

Technically risk refers to the harm and danger caused by technical deficiencies or defects (Yu & Xuan, 2010a). Tendering management, design change and project construction as the elements of risk factors for technical risk sources (Y. Sun et al., 2008). Leifer et al. (2000) as cited in Strong et al. (2009), have defined three major dimensions of uncertainty that are relevant for all innovation development projects targeting new lines of business: technological, market organizational, and resource uncertainties.

There are several factors that may affect the result when making risk analysis or assessment. For example, construction risk analysis, especially at the early stages of the project, is intricate because the nature of risk is usually affected by numerous factors including human error and the data and information available. It may be extremely difficult to assess the risks associated with a project due to the great uncertainty involved (Zeng et al., 2007). Hussain et al. (2007) propose a criterion as the demand or the set of factors which show specifically what investors want in their interaction with the problems in the particular context. Within the project, various degrees of application of risk management exist by the functional risk management process applied (Olsson, 2007).

There is an element of risk inherent in all decisions made, as there is a degree of uncertainty associated with all decision outcomes. Perceptions of risk are an inherent part of the decision-making process (J. Williams & M. Noyes, 2007). The real estate business is very risky due to large-scale and long-term, and other factors such as natural, social, economic, regulatory, psychological factors. Once the decision is at fault, the investor will suffer huge or event destructive damages (Yu & Xuan, 2010a). Type of users who will make the decision, their goals and objectives to be achieved, and several feasibility study factors which include behavioral, organizational, organizational policy and contractual, sociological and political feasibility study are some of the other variables that will affect the degree of risks for alternatives given (Y. Sun et al., 2008).

According to Saleem and Vaihekoski (2008), currency risk can have very important implications for the portfolio management, the cost of capital of a firm, asset pricing as well as currency hedging strategies, as any source of risk which is not compensated in terms of expected returns should be hedged. Real estate investment is speculative and its return and risk are influenced by many factors, such as natural environment, socioeconomic environment, market (Wilhelmsson & Zhao, 2018), enterprise purchasing capability (L. Liu et al., 2007).

Ma and Meng (2009) describe risk factors and risk sources as a risk evaluation index system. The index system of evaluating the real estate project risk established in their article can comprehensively reflect the risk factors real estate projects face in the process of the development. The evaluation index (risk source) for the other index analysis is the construction risk includes nature condition risk, the risk of delays in the project, project quality risk, development cost risk and construction claim risk.

Li and Suo (2009) highlighted two other main risk factors for real estate investment as the following: location factor; and settlement risks: sales return sum, settlement ability, the settlement period. Yu and Xuan (2010a) suggest the other source of the risk is manageable risk: those originated from errors or changes in management, linked with how a project is organized, managed and implemented. Zhang and Li (2009) stated that the risk indicators for early stage of real estate projects include purchasing land risks, removing and resettlement risks, survey risks, design risks, financing risks, bidding risks, contract risks and approval risks. Based on the review of literature, Table 2.1 depicts the risk factors that will affect the dynamic risk analysis for investment in the real estate industry.

Table 2-1: A summary of risk factors that will affect the dynamic risk analysis for investment in the real estate industry.

Risk Analysis Factors	
Sociological risk	Technical risk
Organizational policy risk	Economic risk
Contractual risk	Behavioral risk
Types of user (investors)	Organizational risk
Goals and objectives of decision maker	Scheduled risk
Political risk	Currency risk
Social risk	Technological risk
Financial risk	Policy risk
Market organizational risk	Resource uncertainties risk
Construction risk	Psychological risk
Regulatory risk	Natural environment risk
Socioeconomic environmental risk	Market risk

Risk Analysis Factors	
Enterprise purchasing capability risk	Nature condition risk
Location factor risk	Settlement risk
Management risk	Purchasing land risk
Removing and resettlement risk	Survey risk
Design risk	Bidding risk
Contractual risk	Approval risk

Zhao et. al., (2009) divided risk factors into four stages of real estate project investment as the following: Risk factors during the investment decision process: development opportunity risk, risk of regional economic environment, risk of regional social environment and risk of project positioning; Risk factors during the land to obtain process: risk of market supply and demand, risk of development cost, risk of financing and risk of levy land and remove; Risk factors during the item construction process: risk of project quality, risk of project duration, risk of development cost, risk of contracting , risk of project technology, risk of construction claim, risk of natural conditions and risk of contract mode; Risk factors during the rent and sale management process: risk of marketing opportunity, risk of sales planning, risk of operating contract and risk of others such as natural disaster and contingency.

Jin (2010) draws the lifecycle risk impact factors of the project according to a different stage of real estate project as follows: The risk in investment decision stage includes: policy environment risk, investment opportunity choice risk and investment property type choice risk; The risk in land acquisition stage includes: land price change risk, land idle risk, levy land and dismantle risk and raise funds risk; The risk in construction stage includes: invite public bidding mode risk, contract way risk, contract bargain risk, quality risk, schedule delay risk, development cost risk, construction safe risk and construction claim risk; The risk in lease and sale stage includes: lease and sale opportunity risk, lease and sale contract risk; The risk in property operation stage includes: natural disaster risk and contingency risk.

2.1.5 Existing techniques/models for property investment risk analysis

Investment in the real estate industry in emerging economies demonstrates tight working capital, low liquidity, slow payback, high sunk cost, capital intensive outflows that are not immediately recovered, enduring uncertainties about demand, price/m^2, land costs, and short to medium construction times (Chen et al., 2020). It is very important for investors to have an approach or technique to analyze the real estate project investment to minimize the uncertainty or risks that will affect their profits and margins (Rocha et al., 2007; Y. Sun et al., 2008).

AHP allows better, easier, and more efficient identification of selection criteria, their weighting and analysis. AHP is a relatively effective evaluation method of multi-level analysis, strong network analysis, effective for comprehensive evaluation and trend prediction of multi-factor, multi-standard, and multi-program. It makes the qualitative analysis and quantitative analysis with the result of practical feasibility and objectivity Normally, by using AHP, it allows decision makers to find the weight of each criterion and requires them to provide judgments about the relative importance of each criterion and then to specify a preference for each decision alternative using each criterion. According to the nature of the problem and the overall objective, the users of AHP decompose it into different factors, and then aggregately combine them at different levels in the light of interaction and membership, forming a multilayer analysis structure model. It finally comes down to the weights of relative importance or the relative merits of ranking that the lowest level (alternatives, measures, indexes, etc.) is related to the top level (total goal). The hierarchical structure model generally includes three main levels: goal level is about the problem to be solved or the target to be achieved; criterion level (sometimes followed by index level) is an intermediate link involved to realize a predetermined target by certain measures and policies; alternative levels provide specific measures or policies. The relationship between previous level factors and the succeeding level factors is indicated by lines. This is based on the decision maker's (expert survey method) experience in the field and might lead to incorrect decision if there is no proper reference (Y. Sun et al., 2008).

Furthermore, there are many automated intelligence systems available to support decision-making, such as online analytical processing systems and data mining. Modeling risk factors and risk interaction is an essential part of the decision-making process and is characterized by growing complexity, higher uncertainty, and tighter constraints (Beracha et al., 2017). As such, decision support systems are an important part of these platforms. It is extremely difficult to assess the risks associated with an investment decision due to the uncertainty involved – such as fluctuations in property prices, incomplete risk data, or a lack of reliable, relevant data (Y.-M. Fang et al., 2009). Using the correct methodology for the decision-making process is crucial for ensuring users make the most beneficial decisions (Crosby & Henneberry, 2016; Shinzato, 2018). Risk analysis involves three main processes: risk identification, risk estimation, and risk assessment (C. H. Jin, 2010).

The most challenging part of risk analysis is risk identification – determining the risk factors to include in the decision-making process. Risk estimation quantifies identified risks by measuring and weighting each factor to be included in the risk analysis. There are many existing techniques or models applied in property investment risk analysis that can be grouped into two categories either done manually or automatically. Delphi technique (Park, D'Angelo, & Gunashekar, 2018) and both AHP and ANP determine the risk evaluation index and the weight of the risk factors based on a survey of experts (Angelou & Economides, 2009). The drawback of using an expert survey to estimate the weighting of risks is that investors may hold a different opinion to that of the experts (Bolger & Wright, 1994). Several existing techniques available for property investment risk analysis are shown in Table 2.2.

Table 2-2: Applications of existing techniques/models in property investment risk analysis

Methods/Techniques	Example study
Delphi method	• Define responsible property investing (RPI) by using the Delphi Method to prioritize criteria for the evaluation of property investments. An international panel from the real estate and social investing sectors evaluated 66 criteria in terms

	of materiality to investors and importance to the public interest. A moderate to strong level of consensus was achieved (Pivo, 2008).
Analytical Hierarchy Process (AHP)	• Based on the AHP model, help investors to manage risk exposure and opportunities in property investments. Numerical examples on urban development projects are presented in order to test the effectiveness of the AHP model in supporting decisions and adapting strategies to a permanently changing environment (D'Alpaos & Canesi, 2015) • Asked senior valuers in the UK profession to identify, and score, the investment quality risks of prime offices at the date of valuation. The focus is on the principal elements of investment quality risk comprising yield movement, lease length, rental movement and change in occupier demand. Analysis of the results enabled the development of a generic market model to be used to risk score individual property investments. (Hutchison, Adair, & Leheny, 2005). • Weighting has been performed through the application of the pairwise comparison approach utilized within the AHP a multi criteria evaluation method proposed by Saaty in the 1980s. The AHP assists with decision-making processes by providing decision-makers with a structure to organize and evaluate the importance of various objectives and the preferences of alternative solutions to a decision (Sdino, Rosasco, & Magoni, 2018). • Based on the AHP model, will help investors to manage risk exposure and opportunities in property investments. Numerical examples on urban development projects are presented in order

	to test the effectiveness of the AHP model in supporting decisions and adapting strategies to a permanently changing environment (D'Alpaos & Canesi, 2015).
Analytical Network Process (ANP)	• A case study of a residential and commercial mixed-use project in Liverpool City Centre was used to demonstrate the effectiveness of the ANP model. The result of an initial case study revealed that ANP is an effective tool to support developers in making decisions based on risks assessment. It was found through this study that the established ANP model is effective and can be adopted by real estate developers to suit for the business needs (Khumpaisal & Chen, 2010). • A study aims to assess the risk factors based on the ANP model and to prioritize the key risk factors to identify which risk factor is highly affected by the commercial development process. The results revealed that there are five major risk factors such as environmental, social, economic, technological and political risk, and 32 sub-risk factors (Thilini & Wickramaarachchi, 2019). • A paper aims to introduce a novel decision-making approach to risks assessment in commercial real estate development against social, economic, environmental, and technological (SEET) criteria. It therefore aims to describe a multiple criteria decision-making model based on ANP theory, and to use an experimental case study on an urban regeneration project in Liverpool to demonstrate the effectiveness of the ANP model. An ANP model is set up with 29 risk assessment criteria, and results from an experimental case study reveal that the ANP method is effective to support decision-making based on risk assessment to select the most appropriate development plan;

	and therefore it is applicable in the commercial area (Z. Chen & Khumpaisal, 2009). • An article examines the expectations of real estate practitioners regarding risk assessment techniques. The data collected from the interviews were analysed using an ANP application called 'Superdecision 1.6.0' , developed by Saaty (2005) as cited in (Khumpaisal, Ross, & Abdulai, 2010)
Brainstorm technique	• Each group of operators have selected the most important criteria (or characteristics) which must be evaluated during the design and planning phase (Sdino et al., 2018).
Real Option	• The paper thoroughly reviews the major dissimilarities in the suppositions underlying the discounted cash flow (DCF) and the real options approach and develops a conceptual framework of real options for the entire real estate development process. The findings provide the evidence needed to support the practical appeal of the method to practitioners in the industry. This will enable property practitioners to capture the upside potentials and limit downside losses for investment projects (Mintah, 2016). • The purpose of paper is to exhibit a summary of current literature surrounding real option-based valuation methodology for real estate property development projects besides applying this theory to an empirical case study of a yet-to-be-exercised residential real estate development project in Victoria, Australia, with a 12 step binomial process. The results suggest that in this case, the development should be delayed/deferred, since it is probabilistically expected that the economic environment may reveal to be more favorable in the future than at present (Vahdatmanesh & Firouzi, 2017).

	• The main aim of this case study is to present a practical application of the investment valuation and to construct an option pricing model for real estate investment which considers and integrates as many aspects of the investment and market environment as possible to describe the best situation of the real estate market and its development. The valuation of the investment is carried out using a universally applicable numerical method of binomial trees. The results obtained are subjected to the sensitivity analysis with respect to the discount rate, value of the most influential parameter of the volatility and the input option parameters. The results of the valuation of the project obtained using the real option approach are important mainly for the management of the company in the process of quantification of the present value of future investments (Durica, Guttenova, Pinda, & Svabova, 2018).

There are many approaches to confirm assigned weightings to the risk factors. These typically reflect experts' practical experiences, but they are influenced by personal views, and this will affect the investor's decision-making (Lafleur, 2011). Besides, other tools could be used by discovering the information from social networks and by means of sentiment analysis. Morente-Molinera, Kou, Pang, Cabrerizo, and Herrera-Viedma (2019) proposed a novel method that uses sentiment analysis procedures in order to automatically create fuzzy ontologies from free texts provided by users in social networks. Multi-granular fuzzy linguistic modeling methods are used in order to select the best representation to store the information in the fuzzy ontology. Information is transformed and presented in an organized way making it possible to properly work with it using the method proposed.

2.1.6 Existing real estate analytic systems

There are many real estate systems available on the internet, making it easy to find real estate properties to rent or buy. However, homebuyers usually have to access a multitude of different websites to gather useful information because so many website structures and formats differ (C. Fang & Marle, 2012). Additionally, many web-based real estate systems provide too much information (Yuan et al., 2013). For example, realestate.com.au, Australia's number one property site, lists every property that meets the user's search criteria, with no associated risk analysis, and the resulting information overload requires a significant amount of time to peruse in order to select the right property. Advanced search functions should be available that allow users to limit the number of results beyond basic property features and price specifically focusing on risk analysis for options selected. Furthermore, search criteria that allow the user to personalize their search with the attributes needed to fulfil their financial goals and objectives are needed. Personalization mechanism helps the investor to weigh the attributes of the property based on their requirements such as the higher percentage given to the number of bedrooms available. Furthermore, the investor also weighs the risk analysis factor based on their goals and limitations.

According to the American National Association of Realtors (NAR) On-line Technology Survey (2011), 88% of home buyers chose the internet as an information source when searching for a home – an increase of 14% since 2010 – but the median time home buyers spent on their search was 12 weeks, the same as in 2009 (Yuan et al., 2013). This indicates that the use of real estate websites did not improve the efficiency of the home buyers' search (D'Urso, 2002). Technological innovations have resulted in new applications that allow home buyers to select affordable price ranges, take virtual tours, obtain information about the neighborhood and the surrounding environment, access comparative data for different districts, and view property values over time, but, so far, none of these applications include a risk prediction for return on investment. Existing real estate websites have become a significant transaction platform. However, these platforms only serve as search tools and do not benefit homebuyers in terms of search time, flexibility, and intuitive results. Research has been conducted to investigate the impact of search tools on investors' property investment decisions (Crosby & Henneberry, 2016; Lieser & Groh, 2014; M. Zhang, Guo, & Chen, 2016). An ontological structure to improve information management efficiency that includes case-based reasoning to improve recommendation accuracy was developed and tested in (Yuan et al., 2013). The results demonstrate that the proposed system to be effective and validated the authors' goals. Nevertheless, existing models do not allow investors to rank and weigh their own risk factors. The APPIRA method proposed in this study provides these features to assist the decision-making process in dynamic risk analysis scenarios. An effective system must be able to extract anomalous patterns from search results and present this knowledge in a way that allows investors to evaluate risks and modify their investment criteria (Peng et al., 2011).

2.2 Data-driven Decision-making Approach

Data-driven decision-making involves making decisions based on data analysis and interpretation with references to information processed from datasets gathered online or offline sources. A wide range of enterprise tools available either proprietary or open source application software available for users specifically businesses to get data and analyze it to support decision-making and contribute to strategic planning. Some of the decision-making tools available are decision support system, data mining, OLAP and group decision support system. The review of literature indicates that there is much work to be done in developing an intelligent decision support system (IDSS) for handling risk-based decision-making in business operations or as a tool for businesses managing their business task especially when they deal with decision-making processes in their daily routine as a manager. This is perhaps the most important concern for the future of an information system related to risk analysis or risk aggregation for managers who deal with decision-making processes, since it will promote vital and useful technology that helps investors identify the risk involved in making certain decisions to meet the organization goals and objectives.

Managers need to integrate intelligent information systems that are capable of supporting them throughout the decision-making lifecycle, which starts with structuring a problem from a given set of symptoms and ends with providing the information needed to make the decision (Delen & B. Pratt, 2006). Research on how to design, build and implement intelligent decision-making support systems (i-DMSS) from a more structured and software engineering/systems engineering perspective are still missing in 1980 – 2004 period (Mora et al., 2006). The quality, speed and realization of the decision-making can be increased when the right information is available to the right persons, at the right time, and in the right form (Karacapilidis, Gupta, A. Forgionne, & Mora T, 2006).

2.2.1 Knowledge Discovery and Data Mining

Knowledge discovery and data mining (KDD) have become areas of growing significance because of the recent increasing demand for KDD techniques, including those used in machine learning, databases, statistics, knowledge acquisition, data visualization, and high-performance computing. KDD can be extremely beneficial for the field of artificial intelligence in many areas, such as industry, commerce, government, and education. Information can be converted into knowledge about historical patterns and future trends. Data mining allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified (Horeis & Sick, 2007; Kechadi & Savvas, 2010; Lobur, Stekh, Kernytskyy, & Sardieh, 2008; Qi, 2008). Data mining is taken as a process of transforming knowledge from data format into some other human understandable format like rule, formula, and theorem. A data mining process is not only mining knowledge from data, but also from humans (G. Wang, Hu, Zhang, Xianquan, & Zhou, 2008). It is necessary to find some interesting patterns and rules in data mining with accuracy, semantically correct, consistent and complete data (C. Gao & Wang, 2010). In quality measurement of data, integrity constraints play a significant role. Similarly, in data mining processes, use of constraints is useful for finding patterns and rules (Sumon Shahriar & Anam, 2008). Based on the knowledge discovered, the decision makers would match their requirements with measurement of risk factors generated by the decision support system in order to achieve their goals. Data mining is the core part of the KDD process shown in Figure 2.4.

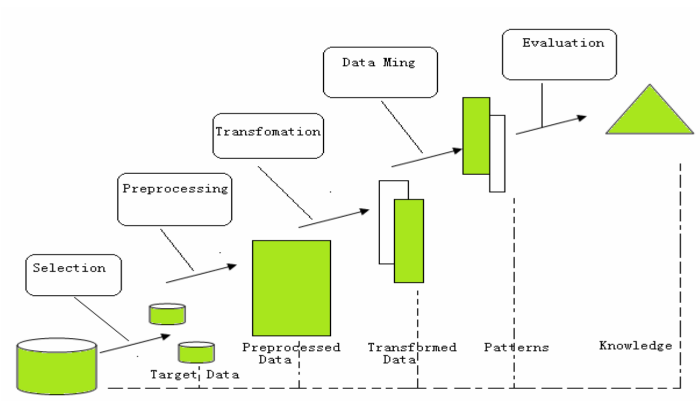


Figure 2.4: Knowledge Discovery in Database process (Qi, 2008).

Risk analysis concept has always been present in business transactions especially in the real estate industry which involve high cost and high capital (Chong et al., 2008). Risk analysis has proven its value in reducing risks in projects. This study focuses on the measurement of identifying, ranking, and weighing the risk factors for property investment risk analysis based on a data-driven approach using multidimensional analysis and data mining technique as a deterministic approach.

Property investment risk analysis involved with micro-level of analysis as detailed features of the property would be analysed using multidimensional analysis and data mining techniques. Some of the risk factors that might affect the property investment decision-making are as follows: location, minimum and maximum price, size of the property, number of bathroom, bedroom and car park, additional features such as balcony, walk-in wardrobe, sauna and zoning. The ultimate aim of any investor is to maximize his returns and minimize the risk (Gumparthi & Venkatachalam, 2010). Investors engage with the property market through the asset market and their determination of yields; the rates of return required to attract direct investment into property. There is evidence of significant pricing inaccuracies in the property market.

A set of factors that are specific to local property markets and that affect both the risk premium and expectation of rental growth includes: the structure and performance of the underlying local economy, location, the quality of local infrastructure and services, the balance between demand for and supply of business accommodation and local market stability and liquidity (Halbert, Henneberry, & Mouzakis, 2014). Hence, there is a need for multidimensional analysis in property investment risk analysis which needs the high dimensionality of data and accurate information. Property investment risk analysis is depending on investment type risk which is mainly divided into residential and commercial real estate. Each type has a different sensitive degree to location, policy, macroeconomics, and so does the risk-withstand ability (Ye, 2011). The deterministic approach which applies the concept of multidimensional data model and data mining technique involves decision-making tools and technologies to support the decision-making process.

2.2.1.1 Multidimensional Data Model

The multidimensional data model provides users with the flexibility to view data from different perspectives. This model is applied in data warehouse architecture which contains different types of end-user access tools such as data mining and OLAP. OLAP is one of the technologies that enable client applications to efficiently access the data multi-dimensionally. OLAP provides many benefits to analytical users, for example, an intuitive multi-dimensional data model makes it easy to select, navigate and explore the data, an analytical query language provides the power to explore complex business data relationships and pre-calculation of frequently queried data, enables very fast response time to ad-hoc queries (Huang, Tseng, Li, & R. Gung, 2006).

OLAP systems are a popular business intelligence technique in the field of enterprise information systems for business analysis and decision support (Shi & Zhu, 2009). OLAP not only integrates the management information systems (MIS), decision support systems (DSS) and executive information systems (EIS) functionality of the earlier generation of information systems but goes further and introduces spreadsheet-like multi-dimensional data views and graphical presentation capabilities. The core component of an OLAP is the data warehouse, which is a decision-support database that is periodically updated by extracting, transforming, and loading data from several OLTP databases (Caron & Daniels, 2008). OLAP implementations typically employ a star schema (dimensional modelling), which stores data denormalized in fact tables and dimension tables. The fact table contains mappings to each dimension table, along with the actual measured data.

In the multidimensional data model, data are organized into multiple dimensions, and each dimension contains multiple levels of abstraction defined by concept hierarchies. This organization provides users with the flexibility to view data from different perspectives (Huang et al., 2006). Whereas, data warehouses and data marts are the data stores for analytical data, OLAP is one of technologies that enable client applications to efficiently access this data multi-dimensionally (English, 1999; Thomsen, 1997; Inmon, 1996 as cited in Huang et. al, (2006)). OLAP provides many benefits to analytical users, for example, an intuitive multi-dimensional data model makes it easy to select, navigate and explore the data, an analytical query language provides the power to explore complex business data relationships and pre-calculation of frequently queried data, enables very fast response time to ad hoc queries (Thomsen, 1997; Jacobson, 2000 as cited in Huang et. al, (2006). A star model representing a multi-dimensional database is shown in Figure 2.5.

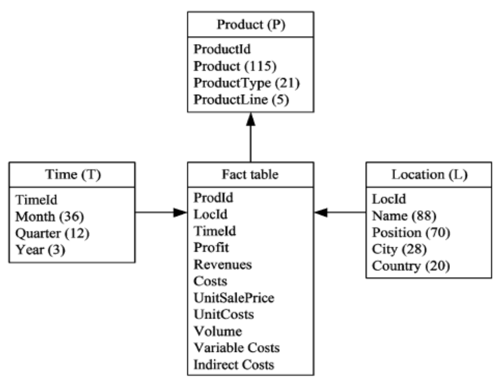


Figure 2.5: An example of a star schema (dimensional modeling) with three dimension tables and a central fact table representing the GoSales data set (Caron & Daniels, 2008).

Hans and Kamber (2000) as cited in Huang et. al. (2006) draws five OLAP operations in the multi-dimensional data model which includes: roll-up, drill down, slice and dice, and pivot (rotate) operations; which correspond to vertical analysis (e.g.. the “father” and “son” relationship) and parallel analysis (e.g., the “brother” relationship) in the qualitative analysis a shown in Figure 2.6.

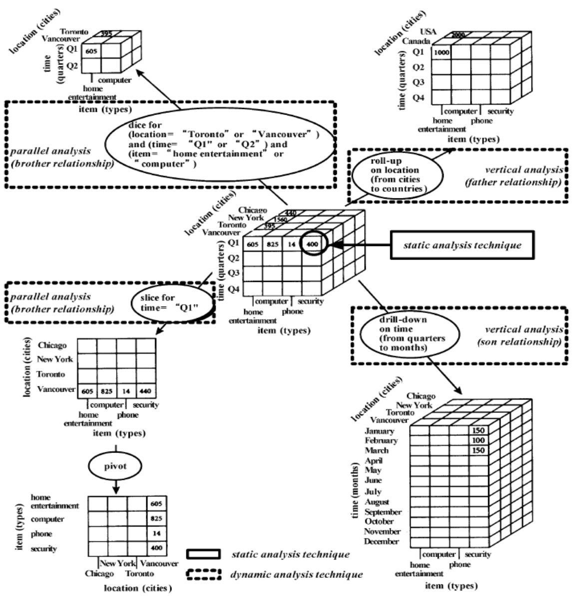


Figure 2.6: OLAP technical operation diagram (Huang et al., 2006).

Huang et. al. (2006) lists two different types of multi-dimensional analysis namely quantitative vs. qualitative and static vs. dynamic as shown in Figure 2.7. The quantitative multi-dimensional analysis is used to operate on the numerical type of data, while the qualitative multi-dimensional analysis is used to operate the non-numeric type of data.

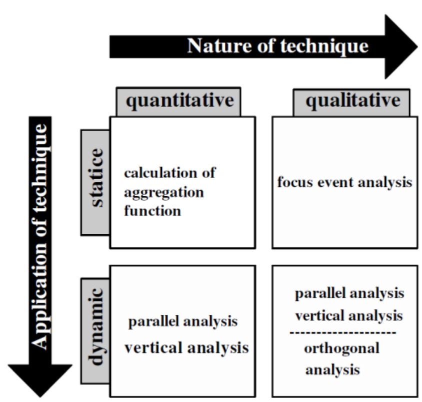


Figure 2.7: Structure of multi-dimensional analysis technique (Huang et al., 2006).

There is a small amount of literature published in the qualitative multi-dimensional analysis area (Huang et al., 2006). The differences of the concept and application between qualitative and quantitative multi-dimensional analysis are depicted in Table 2.3.

Table 2-3: Comparison of quantitative and qualitative multi-dimensional analysis.

Features	**Quantitative technique**	**Qualitative technique**
Data type of measure	Numerical data	Non-numerical data
Application objective	Information aggregation	Selection and analysis of events
Dimensional framework	Detailed information	Classification structure
Hierarchical framework	Upper level members summarize from lower level members; lower level members represent a component for each upper level member. Only each lower level member illustrated its individual definition (each upper level treated as a summary of its lower level members)	Upper level members do not summarize directly from lower level members; lower level members do not represent as the direct component for each upper level member. Both upper level members and lower level members illustrate their individual definition
System operation	Active and automatic operations	Passive and non-automatic operations

For static vs dynamic multi-dimensional analysis, the key difference is that "static" represents merely a fact without considering any variability, while "dynamic" reflects incorporating all changes. For example, in the time dimension analysis, emphasizing sales of only one specific month is in a "static" status. Modeling the sales changes among different months is a "dynamic" analysis. The differences of the static vs dynamic multi-dimensional analysis are depicted in Table 2.4.

Table 2-4: Comparisons of static vs dynamic multi-dimensional analysis.

Type of data	**Static multi-dimensional analysis**	**Dynamic multi-dimensional analysis**
Quantitative	The aggregation function embedded in the data cube can be generated. The results of the aggregation function are only facts of the situation	This approach should be capable of modelling any changes in the values of the aggregation function from data cubes through the selection of different dimensions and levels. Handle two types of analysis: vertical and parallel as shown in Figure 2.8.
Qualitative	Operates on non-numerical data which shows detailed facts in the data cube, can be observed	Handle three types of analysis: vertical, parallel and orthogonal as shown in Figure 2.9.

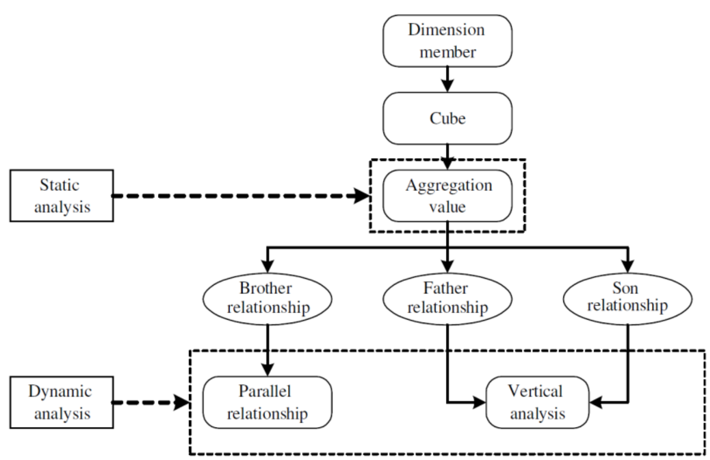


Figure 2.8: Models of multi-dimensional analysis for quantitative data (Huang et al., 2006).

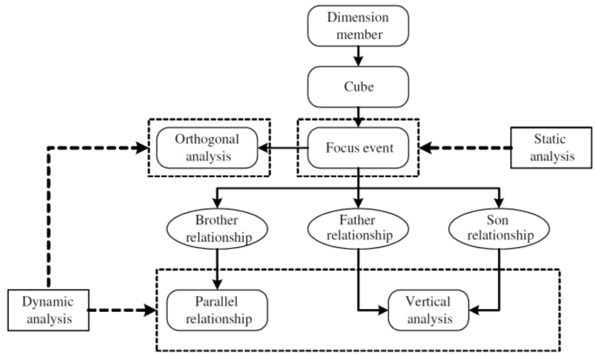


Figure 2.9: Models of multi-dimensional analysis for qualitative data (Huang et al., 2006).

2.2.2 Intelligent Decision Support System (IDSS) for Risk-based Decision-making

The review of literature indicates that there is much work to be done in developing an IDSS for handling risk-based decision-making in business operations or as a tool for businesses managing their business task.

Some of the decision-making tools available are decision support system, data mining, OLAP and group decision support system. Recent developments of decision support system tools have included the artificial intelligent (AI) techniques to provide an IDSS as a tool that helps investors make their decision more effective and efficient. AI techniques or mechanisms such as fuzzy logic, expert systems, neural networks, case-based reasoning, genetic algorithms and intelligent agents have been applied to the DSS application to make the decision-making support tools in organizations becoming more effective. An intelligent agent needs to be integrated with the decision-making support systems tool to provide a system that fully realized its promise for users, groups and organizations (Mora et al., 2006). The application of intelligent decision support systems for handling risk-based decision-making has been explored and becoming more popular since 2007 till now. Furthermore, IDSS is more reliable and has profited quite well after use because the factors in the decision support system are numerous and a large number of modules and methods are provided to help a policymaker analyze the problem (R.-S. Lin, Wang, Hu, Gao, & Lu, 1996).

Uncertainty and complexity are becoming common facts leading to the greater recognition of systematic and holistic approaches to problem solving (Cassaigne et al., 2006). Moreover, the review has also explored current developments in the field of decision support technology use for risk management or risk analysis. For instance, Delen and Pratt (2006) developed an integrated and intelligent DSS for manufacturing systems that are capable of providing independent model representation concepts. In their study, they believe managers need to integrate intelligent information systems that are capable of supporting them throughout the decision-making lifecycle, which starts with structuring a problem from a given set of symptoms and ends with providing the information needed to make the decision. They also report on a collaborative research effort whose aim has been to fill this need by developing novel concepts and to demonstrate the viability of these concepts within an advanced modeling environment. According to Mora et al. (2006), research on how to design, build and implement i-DMSS from a more structured and software engineering/systems engineering perspective are still missing in this entire research period (1980 – 2004). According to Karacapidilis et al., (2006), the quality, speed and realization of the decision-making can be increased when the right information is available to the right persons, at the right time, and in the right form.

2.2.3 A generalized framework for knowledge discovery in business environments

Web-based applications monopolize the business world and, as a result, web applications currently hold the largest repository of knowledge (Hussain & Asghar, 2013).

The deterministic approach proposed applied the concept of data mining in order to find hidden knowledge in the dataset stored in the system. Data mining is the construction of mining technology as an end-user access tool involving services, data mining techniques, user queries, and knowledge discovery. Data mining discovers patterns in web data collections and can be divided into three categories: content mining, structure mining, and usage mining (Chaoyang, 2013). This paper applies the concept of data mining, which comprises four stages: data collection, data preprocessing, pattern discovery, and pattern analysis in which extend the generalized framework by Fong and Hui, (2010) as shown in Figure 2.10.

Data mining technology can help to disseminate knowledge by producing analysis results based on user inquiries. It can solve problems such as discovering correlations, finding event sequence relationships, classifying and clustering items, and improving variable structures and paths among variables of analysis (Hussain & Asghar, 2013; Xiaoyan & Ping, 2011). Data mining has been used for exploring connectivity (link) structures, which can lead to predictions about future links among objects (Krishna, Jose, & Suri, 2014).

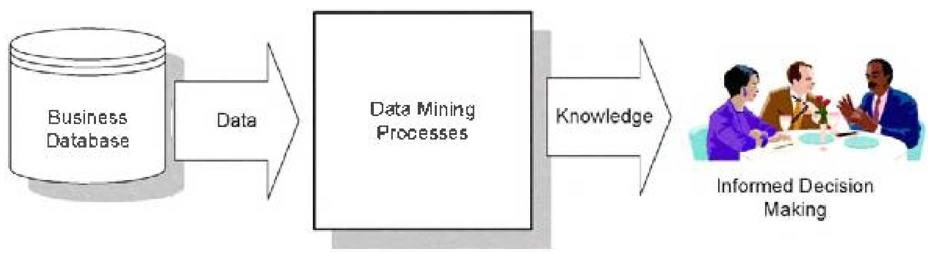


Figure 2.10: The essence of the proposed framework. Raw data from one or more business database(s) are selected and extracted, which are then subjected to the number of data mining processes (Fong & Hui, 2010).

There are many techniques and models available for risk analysis in the real estate industry. However, research on personalization with multi-criteria decision-making process is still limited and did not include the personalization mechanism to weigh the risk factors. Furthermore, available techniques do not include the weight of the property attributes that will fulfil the investors' requirements. Thus, the APPIRA method is proposed to address these issues. Deterministic approach guarantees that the probability of that the outcome will occur is 100%.

2.3 Personalization Technique

This section discusses the concepts of personalization and adaptive personalization model. These concepts relate mainly to property investment risk analysis as applied in the real estate industry.

It is universally agreed that real estate property investment creates big profits, compared to other types of investment such as cash and fixed interest investments, bonds and superannuation (C. H. Jin, 2010). However, the risk and cost factors involved are also very high, compared to other types of investment (Angelou & Economides, 2009; Bolger & Wright, 1994; E. C. M. Hui et al., 2010; Juhong & Zihan, 2009; Lu, Zhang, Ruan, & Wu, 2007; Ren & Yang, 2009). Scientific investment decision-making processes are needed to carefully analyse and ensure investment decisions are correct and effective. The feasibility studies of investment projects and property have been researched and various methods have been proposed and applied to measure risk analysis with the alternatives given (Vesanen, 2007; Zeng et al., 2007).

In this research project, the term personalization refers to the mapping and satisfying of a user's/business's goal in a specific context with a service's/business's goal in its respective context (Riecken, 2000 as cited in Zhang & Zeng, (2006). Personalization is motivated by the recognition that a user has needs and meeting them successfully is likely to lead to a satisfying relationship with what they require. Personalization involves a process of gathering user-information during interaction with the user, which is then used to provide appropriate assistance or services; tailor-made to the user's needs (Bonett, 2001 as cited in (P. Zhang & Zeng, 2006) (Zheng, Yao, & Niu, 2008) (Kao, Tseng, & Lee, 2010; Soujanya & Kumar, 2010). The technology of personalized service has been applied to many different fields. Modern personalized service can provide pertinent service for different users so that the specific demands from them can be met. In the Internet field, the technology of personalized service can improve the quality of web service and the efficiency of users' access (Liang, Lai, & Ku, 2007)(Zhen et. al, 2006; (Dong & Li, 2010; Y.-n. Xiong & Geng, 2010).

2.3.1 Adaptive personalization

Personalization is more refined than customization, in that personalization is undertaken by a marketer for their customers, whereas customization stems from customers' requests (Montgomery & Smith, 2009). Effective computer-based decision support systems closely pair their technology and applications with personalization systems (Montgomery & Smith, 2009).

Personalization is a strategic tool for differentiating a product or service and has become very popular in e-services and online systems (Arentze, 2013; Ding, Wang, Wu, & Olson, 2017; M. Zhang et al., 2016). Many personalized e-services systems, such as the recommender systems on Amazon.com (Kwon & Kim, 2012), have helped leverage a significant competitive advantage while, at the same time, allowing customers to set their own personal preferences and limitations to improve their decision-making process (Glorio et al., 2012). Real estate websites, however, do not yet accommodate personalized risk prediction for property investment.

According to Vesanen (2007), there are four different types of personalization: adaptive, cosmetic, transparent, and collaborative. Adaptive personalization, such as Yahoo.com, gives customers a choice of options. Cosmetic personalization offers customers a standardized product with a choice of packaging or presentation; for example, Google.com. Transparent personalization refers to different products marketed with the same presentation. In collaborative personalization, the organization and customer build the product together. Furthermore, Claus and Raubal (2004), suggest that personalization should be provided to support decision-making as user's capacity to evaluate decision alternatives are limited and to consider individual decision maker's preferences.

This paper applies the concept of adaptive personalization to risk analysis and decision support in real estate, because the investor chooses their investment from different property options to align with their goals. Investor preference and satisfaction can be optimized through adaptive personalization. A combination of personalization and adaptation is proven to be effective and efficient, with a significant impact, and is feasible. (Arentze, 2013; Ding et al., 2017; M. Zhang et al., 2016).

2.4 Multi-criteria Decision-making Techniques

Multi-criteria decision-making (MCDM) techniques related to property investment risk analysis as the decision made is based on several parameters of property features that need to be analyzed in order to meet the investor's requirements. Several of the most frequently used multi-criteria decision-making model includes MAXMIN (assumed that the overall performance of an alternative is determined by its weakest attribute) or MAXMAX (an alternative is selected by its best attribute value) algorithm, Simple Additive Weighting (SAW), AHP, Technique for Order Preference by Similarity to the Ideal Solution (TOPSIS), The Simple Multi Attribute Rating Technique (SMART), and Elimination and Choice Expressing Reality (ELECTRE) (Chen, Hwang, 1992 as cited in Roszkowska (2011). It is a useful tools to help the decision maker(s) to select options in the case of discrete problems especially with the help of computers, those methods have become easier for the users, thus they have great acceptance in many areas of decision-making processes in economy and management ((Roszkowska, 2011).

Multi-criteria analysis focuses mainly on three types of decision problems: choice − select the most appropriate (best) alternative, ranking − draw a complete order of the alternatives from the best to the worst, and sorting − select the best k alternatives from the list.

2.4.1 Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) Approach

According to Roszkowska (2011), in general, the process for the TOPSIS algorithm starts with forming the decision matrix representing the satisfaction value of each criterion with each alternative. Next, the matrix is normalized with a desired normalizing scheme, and the values are multiplied by the criteria weights. Subsequently, the positive-ideal and negative-ideal solutions are calculated, and the distance of each alternative to these solutions is calculated with a distance measure. Finally, the alternatives are ranked based on their relative closeness to the ideal solution. The TOPSIS technique is helpful for decision makers to structure the problems to be solved, conduct analyses, comparisons and ranking of the alternatives. The classical TOPSIS method solves problems in which all decision data are known and represented by crisp numbers. Most real-world problems, however, have a more complicated structure. Based on the original TOPSIS method, many other extensions have been proposed, providing support for interval or fuzzy criteria, interval, or fuzzy weights to modelled imprecision, uncertainty, lack of information or vagueness.

The main advantages of this method are the following (Hung, Cheng, 2009 as cited in Roszkowska (2011)): simple, rational, comprehensible concept; intuitive and clear logic that represent the rationale of human choice; ease of computation and good computational efficiency; a scalar value that accounts for both the best and worst alternatives ability to measure the relative performance for each alternative in a simple mathematical form and possibility for visualization.

The idea of classical TOPSIS procedure can be expressed in a series of following steps (Chen, Hwang, 1992; Jahanshahloo, Lofti, Izadikhah, 2006a as cited in Roszkowska (2011)).

Solving each multi-criteria problem begins with the construction of a decision-making matrix (or matrices). In such matrices, values of the criteria of alternatives may be real, intervals number, fuzzy numbers or qualitative labels. Parameter value for property investment ranges from different data types listed.

Let us denote by $D = \{1,2,\dots,K\}$ a set of decision makers or experts. The multi-criteria problem can be expressed in $k -$ matrix format in the following way:

$$\begin{array}{c|cccc} & C_1 & C_2 & \dots & C_n \\ \hline A_1 & x_{11}^k & x_{12}^k & \dots & x_{1n}^k \\ A_2 & x_{21}^k & x_{22}^k & \dots & x_{2n}^k \\ \dots & \dots & \dots & \dots & \dots \\ A_m & x_{m1}^k & x_{m2}^k & \dots & x_{mn}^k \end{array}$$

where:

- $A_1, A_2, \dots, A_m$ are possible alternatives that decision makers have to choose from,
- $C_1, C_2, \dots, C_n$ are the criteria for which the alternative performance is measured,
- x_{ij}^k is the $k -$ decision maker rating of alternative A_i which respect to the criterion C_j (x_{ij}^k is numerical, interval data or fuzzy number).'

In this way for m alternatives and n criteria we have matrix $X^k = (x_{ij}^k)$ where x_{ij}^k is valued of $i -$ alternative with respect to $j -$ criterion for $k -$ decision maker, $j = 1,2,\dots,n$, $k = 1,2,\dots,K$.

The relative importance of each criterion is given by a set of weights which are normalized to sum to one. Let us denote by $W^k = [w_1^k, w_2^k, \dots, w_n^k]$ a weight vector for $k -$ decision maker, where $w_j^k \in \Re$ is the $k -$ decision maker weight of criterion C_j and $w_1^k + w_2^k + \dots + w_n^k = 1$. In the case of one decision maker we write, x_{ij}, w_j, X, respectively.

Step 1. Construct the decision matrix and determine the weight of criteria.

Let $X = (x_{ij})$ be a decision matrix and $W = [w_1, w_2, \dots, w_n]$ a weight vector, where $x_{ij} \in \Re$, $w_j \in \Re$ and $w_1 + w_2 + \dots + w_n = 1$.

Criteria of the functions can be: benefit functions (more is better) or cost functions (less is better).

In this study, risk refers to cost so less is better.

Step 2: Calculate the normalized decision matrix.

This step transforms various attribute dimensions into non-dimensional attributes which allows comparisons across criteria. Because various criteria are usually measured in various units, the scores in the evaluation matrix X have to be transformed to a normalized scale. The normalization of values can be carried out by one of the several known standardized formulas. Some of the most frequently used methods of calculating the normalized value n_{ij} are the following:

$$n_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^{m} x_{ij}^2}}$$

$$n_{ij} = \frac{x_{ij}}{\max_i x_{ij}}$$

$$n_{ij} = \begin{cases} \dfrac{x_{ij} - \min_i x_{ij}}{\max_i x_{ij} - \min_i x_{ij}} & \text{if } C_i \text{ is a benefit criterion} \\ \dfrac{\max_i x_{ij} - x_{ij}}{\max_i x_{ij} - \min_i x_{ij}} & \text{if } C_i \text{ is a cost criterion} \end{cases}$$

for $i = 1, \dots, m; j = 1, \dots, n.$

Step 3. Calculate the weighted normalized decision matrix.

The weighted normalized value v_{ij} is calculated in the following way:

$$v_{ij} = w_j n_{ij} \text{ for } i = 1, \dots, n.$$

where w_j is the weight of the j-th criterion,

$$\sum_{j=1}^{n} w_j = 1 .$$

Step 4. Determine the positive ideal and negative ideal solutions.

Identify the positive ideal alternative (extreme performance on each criterion) and identify the negative ideal alternative (reverse extreme performance on each criterion). The ideal positive solution is the solution that maximizes the benefit criteria and minimizes the cost criteria whereas the negative ideal solution maximizes the cost criteria and minimizes the benefit criteria.

Positive ideal solution A^{+} has the form:

$$A^{+} = (v_1^{+}, v_2^{+}, \dots, v_n^{+}) = \left(\left(\max_i v_{ij} \,\middle|\, j \in I\right), \left(\min_i v_{ij} \,\middle|\, j \in J\right)\right)$$

Negative ideal solution A^{-} has the form:

$$A^{-} = (v_1^{-}, v_2^{-}, \dots, v_n^{-}) = \left(\left(\min_i v_{ij} \,\middle|\, j \in I\right), \left(\max_i v_{ij} \,\middle|\, j \in J\right)\right)$$

where I is associated with benefit criteria and J with the cost criteria, $i = 1, \dots, m; j = 1, \dots, n$.

Step 5. Calculate the separation measures from the positive ideal solution and the negative ideal solution.

In the TOPSIS method a number of distance metrics can be applied*. The separation of each alternative from the positive ideal solution is given as

$$d_i^{-} = \sqrt{\sum_{j=1}^{n} \left(v_{ij} - v_j^{-}\right)^2}, \; i = 1,2, \dots, m \tag{3}$$

Step 6. Calculate the relative closeness to the positive ideal solution.

The relative closeness of the i-th alternative A_j with respect to A^{+} is defined as

$$R_i = \frac{d_i^{-}}{d_i^{-} + d_i^{+}}$$

where $0 \leq R_i \leq 1, i = 1,2, \dots, m$.

Step 7. Rank the preference order or select the alternative closest to 1.

A set of alternatives now can be ranked by the descending order of the value of R_i.

2.5 Summary

From the review of literature, studies done by other researchers categorize the risk sources and risk factors differently. Some of the researchers classified the risk factors according to the stages of real estate investment and some other group it according to the sources. This will lead to misinterpretation of factors thereby affecting the risk analysis result. Thus, this study will apply the combination of heuristic and deterministic approach to identify and weigh the risk sources and its factors that will affect the property investment risk analysis.

Furthermore, there are a number of issues related to available methods or models for risk analysis in the real estate industry (Rocha et al., 2007; Y. Sun et al., 2008). For example, real option methodology has problems in the practical implementation of risk analysis, such as lack of mathematical skills, restrictive modeling assumptions, increasing complexity and limited power to predict investment in competitive markets (Zhang, 2011; Lander & Pinches, 1998 cited in Rocha et al., (2007)). Moreover, Zeng, An and Smith (2007) believe that high quality data are a prerequisite for the effective application of sophisticated quantitative techniques. It is essential to develop new risk analysis methods to identify major factors, and to assess the associated risks in an acceptable way in various environments.

The majority of existing real estate investment risk evaluations give priority to single-goal decision-making, use single indices such as the maximum expectation, the largest variance, the minimum standard deviation rate to evaluate the real estate investment. Besides, there is a need for an effective and efficient technique to identify major factors and to assess the associated risks in an acceptable way in various environments, as older tools cannot be effectively and efficiently applied (Rocha et al., 2007; Y. Sun et al., 2008; S. J. Zhou et al., 2008). Additionally, dynamic risk analysis concepts and applications have been discovered by many researchers which applied different techniques or methods to support the decision-making process in different fields. For example, practices on risk-based decision-making for investment in the real estate industry study has been conducted to investigate the risk related issues and it was found that many decisions made based on investigating and analyzing factors by giving weight, calculate and select the best option based on the highest performance index (A. Piyatrapoomi et al., 2004). There is a gap found in the literature review where there is a need for the justification of risk factor weight and ranking which is based on historical data-driven to decision support using knowledge discovery and investors personalization for real estate investment risk analysis.

Hence, this study has estimated and weighed the risk factors based on the pattern of data discovered using data mining techniques (deterministic approach) and maps it with investor's requirements (heuristic approach) to achieve the investor's goals and objectives. The deterministic approach will apply and use the concept of technology called decision support tools. Decision support tools help businesses to manage their business tasks efficiently and effectively, especially when managers deal with decision-making processes in their daily routine (Lu et al., 2007; Niu et al., 2009).

A literature review reveals that similar studies categorize risk sources and risk factors differently. Some researchers classify risk factors according to the stage of investment, while others classify risk factors according to their source. This issue leads to misinterpretation and affects the results of risk analysis. Consequently, this study uses an integrated heuristic and deterministic approach to identify and weight the sources and factors of risk affecting property investment.

Additionally, there are a number of issues related to existing methods or models for risk analysis in the real estate industry (Rocha et al., 2007; Y. Sun et al., 2008). Real option methodology lacks mathematical tools and has restrictive modeling assumptions, increased complexity, and limited power with which to predict investment in competitive markets, making practical implementation problematic (Rocha et al., 2007). Moreover, Zeng, An and Smith (2007) believes that high-quality data is a prerequisite for the effective application of sophisticated quantitative techniques. It is essential to develop new risk analysis methods to identify and assess key risk factors.

CHAPTER 3

AN ADAPTIVE PERSONALIZED PROPERTY INVESTMENT RISK ANALYSIS FRAMEWORKS

3.1 Introduction

The limitation of existing property investment risk analysis arises from the lack of sufficient personalization technique to identify the determinants of property investment align with investor's specific goals and constraints. This chapter presents three different frameworks of an adaptive personalized property investment risk analysis framework referring to personalized multidimensional process framework, risk-based decision-making framework and multidimensional analysis and data mining framework to overcome this limitation. The relationships between these three frameworks depicted in Figure 3.1.

The personalized multidimensional process framework introduces an adaptive personalization approach, which is integrated into the data-driven analysis, to deal with property investment determinants for specific goals that differ among investors. The introduction of this adaptive personalization approach will overcome the limitation of the property investment risk analysis by experts discussed earlier in Chapter 1.

The remainder of this chapter is structured as follows. The framework of the personalized multidimensional process is described in Section 3.2. In Section 3.3, the risk-based decision-making framework is explained in detail and a multidimensional analysis and data mining framework is given in Section 3.4. Finally, this chapter is summarized in Section 3.5. The work presented in this chapter has been reported in three of our publications listed in Section 1.8, i.e. publications 1, 4 and 5.

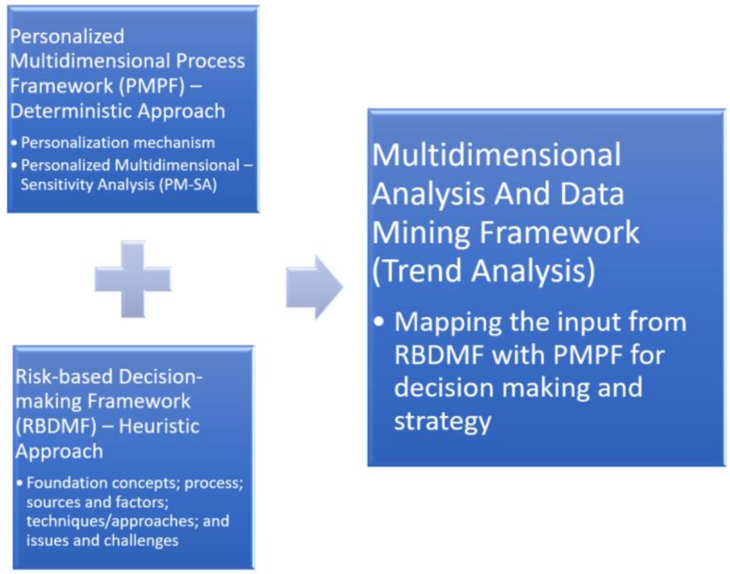


Figure 3.1: The relationships between three different frameworks of an adaptive personalized property investment risk analysis.

3.2 Personalized Multidimensional Process Framework

The PMP framework consists of nine major different steps. This framework applies the bottom up approach to support the decision-making process for dynamic risk analysis for investment in the real estate industry. The investors need to set up their goals and objectives before proceeding with the data analysis. Once the goals and objectives have been identified, the pattern of data using SAS tools and knowledge discovery in database or data mining techniques is explored.

A PMP framework for this research has been developed as shown in Figure 3.1.

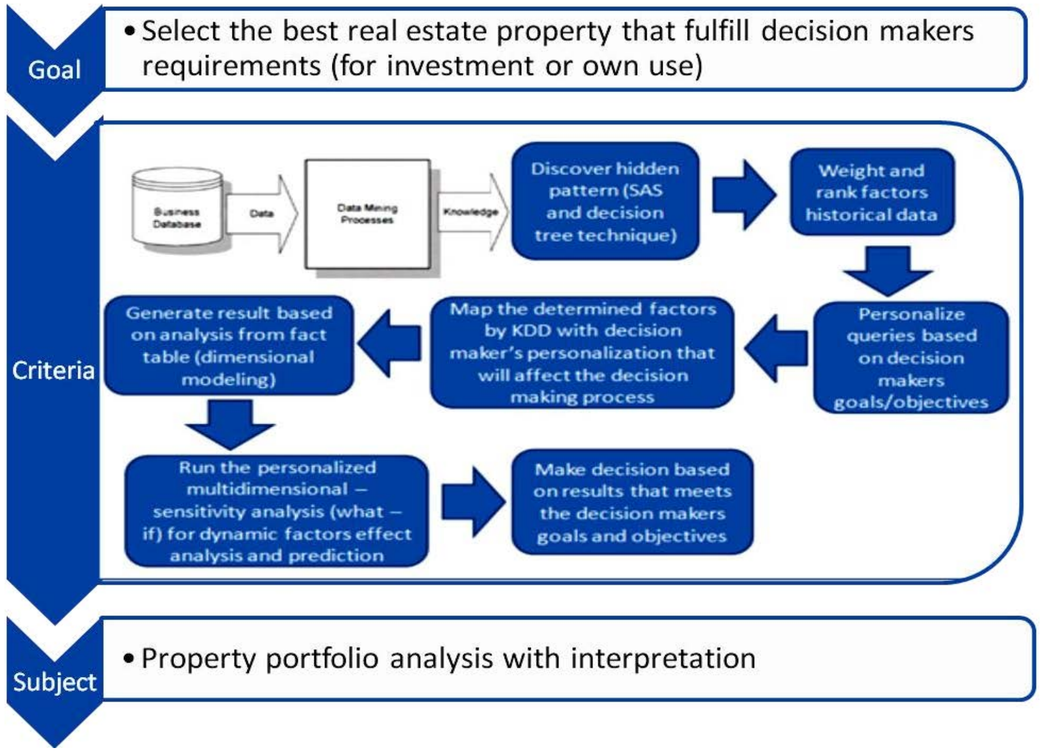


Figure 3.2: A personalized multidimensional process (PMP) framework.

Based on the pattern of data discovered, the system automatically gives the weight and ranking of the risk factors. A data mining technique known as decision tree induction used to identify the pattern of data and give the weight and ranking of the factors of the real estate property to be sold in a certain period of time. These weights and ranking of factors provided by the data mining technique might not meet the investor's requirement. Hence, the investors personalized their queries based on their requirements.

Then, the PAM method developed to map the determined factors by knowledge discovery with investor's personalization that affect the decision-making process. This method helps the investors to achieve their goals and objectives using deterministic and heuristic approaches. Deterministic approach is based on historical and pattern data while the heuristic approach is based on investor's personalization. The system generates the result based on the investor's personalization and pattern of data from the online analytical processing combined with data mining approach. A dimensional modeling using star schema used to generate the result. A star schema refers to a fact table which is surrounded by a dimension table. The time dimension is a mandatory dimension in a star schema to identify the changes of attributes over a period of time.

Next, sensitivity analysis is conducted to understand how sensitive the factors are to the variation of the investor's personalization since the uncertainties are involved in the factor determination over a period of time. The variation of the result of the factors personalized by the investors that will affect the real estate property sold in a certain period of time can be traced and identified using the PM-SA method. Based on the findings or results of the analysis, the investors will choose the best option that meets their requirements. As a summary, the PMP framework requires three main inputs for data analysis processing as shown in Figure 3.2.

Figure 3.3: Three main inputs for the PMP framework.

3.2.1 Nine Steps of Personalized Multidimensional Process Framework

The PMP framework presents comprehensive risk analysis with a 9 steps process which takes into account both qualitative and quantitative risk factors (Demong & Lu, 2011). It comprises two important methods, the PAM method and PM-SA method to minimize and mitigate the impact of risk in the decision being made as depicted in Figure 3.3.

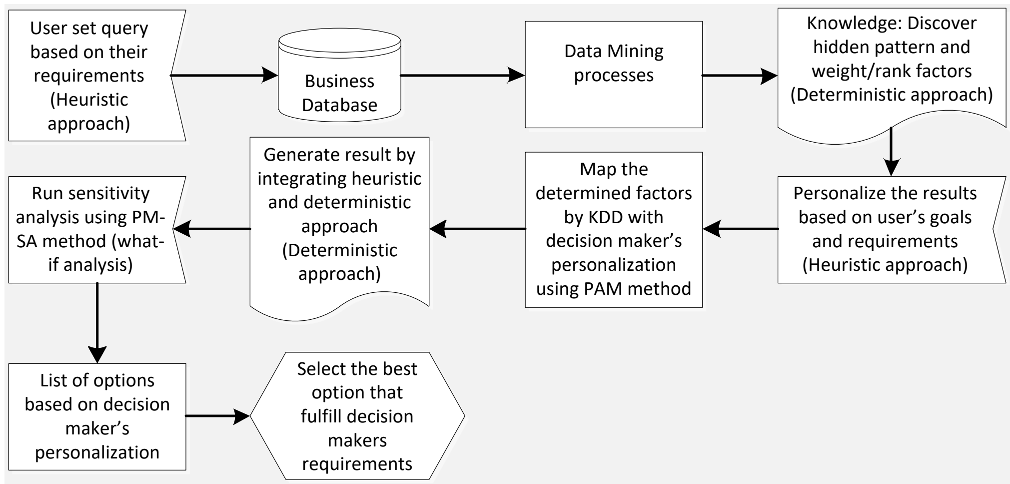


Figure 3.4: A personalized multidimensional process (PMP) framework.

This framework applies the bottom-up approach to support the decision-making process for dynamic risk analysis for investment in the real estate industry. The bottom-up approach in this paper refers to the information processing based on input from investors as a query for the system to analyze and provide hidden knowledge based on data available. Input of data for the data mining process is extracted from different operational databases, forms, reports and etc. The investor will send queries to the system and output will be displayed from the system for the investor to personalize the risk factors ranking and weighting. Then the result will be displayed based on the data stored in the system. Subsequently, sensitivity analysis will be conducted. The result of the variation of the factors personalized by the investors can be traced and identified by using the PM-SA method. For example, the investor might run sensitivity analysis of risk factors that affect the real estate property sold in a certain period of time. Based on the findings or results of the analysis, the investors will choose the best option that meets their requirements.

Figure 3.4 depicts the flow chart for the PMP framework which displays how the PAM method functions. This paper will focus more on the algorithm on how the PAM method functions. The process begins with the user or the investor will need to set their goals, objectives and limitations which are referring to the heuristic approach. The data set by the users will be sent to the system for the matching process and the system will display the matching criteria results with patterns of data or more valuable information which might be useful for users to be aware of. This process is known as a deterministic approach. Then, the users will personalize the rank and weight of the factors based on their requirements. Next, the system will map these criteria set by the user with available data stored in the database. The integration between heuristic and deterministic approach begins with this step to map the ranking and weight of the risk factors from the user's perspective and pattern of data stored in the database.

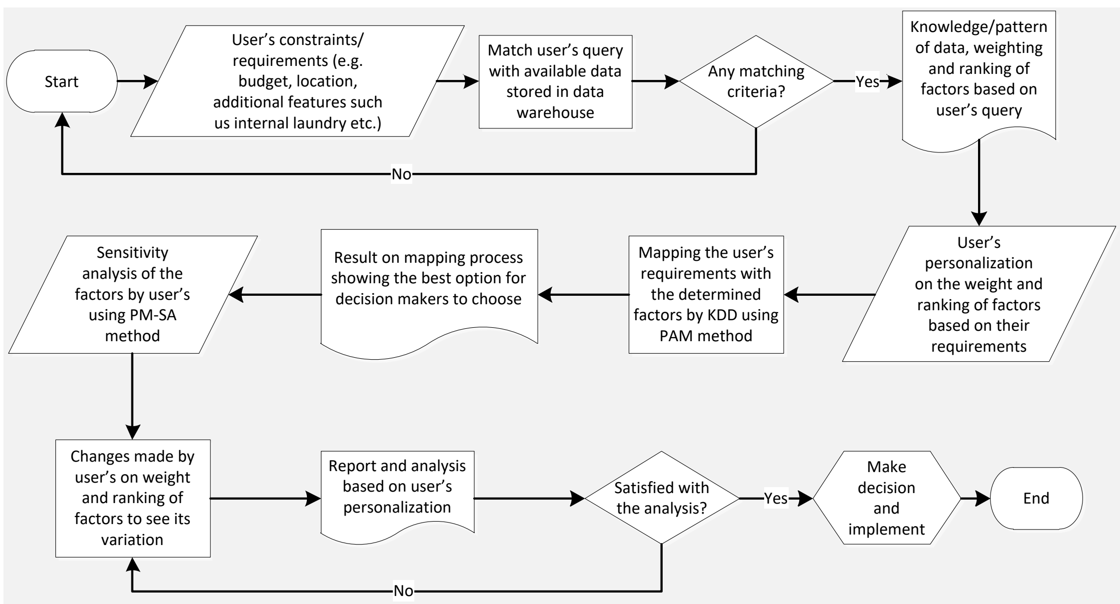


Figure 3.5: Flow chart for the PMP framework.

3.3 Risk-based Decision-making Framework

The limitation of existing risk analysis arises from the lack of comprehensive risk management and dealing with a high volume of data or input dispersed everywhere. This section presents a risk-based decision-making framework for investment in the real estate industry to overcome this limitation and which aims to reduce risk. The background and structure related to the risk-based decision-making framework proposed for investment in the real estate industry discussed in detail.

Risk analysis decision-making is an important tool because such investments normally yield a high return but at the same time pose a high risk to success (Ma & Meng, 2009; S. J. Zhou et al., 2008). There are many substantial studies related to risk analysis techniques and approaches for the real estate industry. In principle, risk-based decision-making techniques involve risk analysis and support the decision-making process. The literature shows that various risk-based decision-making techniques have been integrated with decision support tools and intelligent agents to enhance the usefulness of the technique. Predicting and controlling risk has become the key to the success or failure of a project (W. Li et al., 2009). Several techniques have been proposed and applied in e-service intelligence to evaluate, analyze, assess or predict risk, including the Analytic Hierarchy Process (AHP) and Monte Carlo Simulation, Markowitz Portfolio Optimization Theory, real options methodology (Rocha et al., 2007), and a Hidden Markov Model (Y. Sun et al., 2008) Lander & Pinches 1998, cited in Rocha et al., (2007)).

Risk analysis with uncertainty in the decision-making process deals with the measurement of uncertainty and probability and the likely consequences for the choices made for the investment. The uncertainty of variables or factors that will affect the risk analysis process will impact the success of a project investment in the real estate industry. Techniques such as fuzzy set theory (Y. Sun et al., 2008) and F-AHP have been proposed to solve such problems.

This section proposes a risk-based decision-making framework for investment pre-analysis in the real estate industry, as shown in Figure 3.5. The proposed risk-based decision-making framework comprises five main elements namely: foundation concepts; process; sources and factors; techniques/approaches; and issues and challenges. These five elements will be explained in subsequent sections. This framework can be applied to different problems or issues in various industries and can support decision makers to render their decision-making process more structured and manageable.

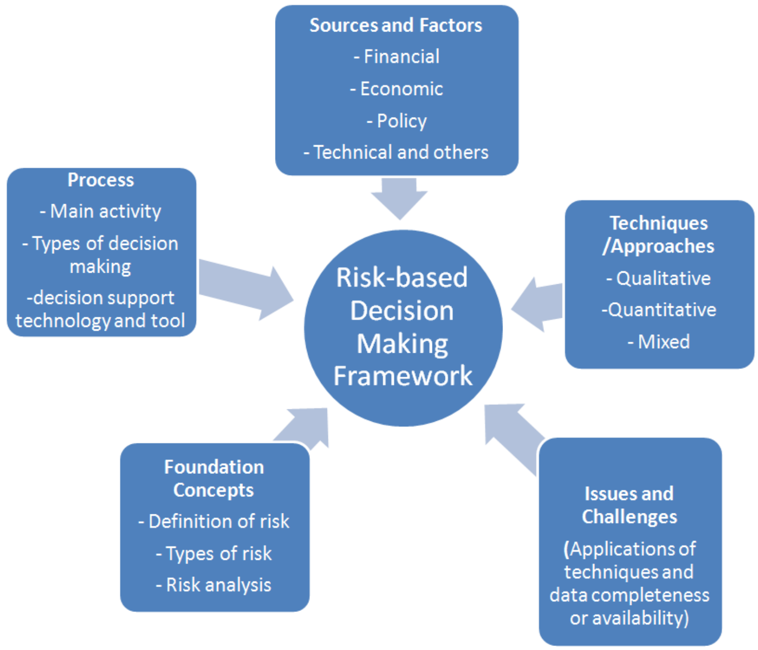


Figure 3.6: The structure of the proposed risk-based decision-making framework.

3.3.1 Risk-based Decision-making Concepts

The concepts of risk-based decision-making including the definition of risk, types of risk, and a brief explanation of risk analysis explained and elaborated in Chapter 2 Section 2.1.1. These concepts relate mainly to risk-based decision-making for investment as applied in the real estate industry.

Risk analysis is the process of identifying the security risks to a system and determining their probability of occurrence, their impact, and the safeguards that would mitigate that impact (Syalim, Hori, & Sakurai, 2009). The risk analysis concept is present in business transactions, especially in the real estate industry which involves high cost and high capital (Chong et al., 2008). Regardless of the types of risk, the application of risk analysis has a positive effect in identifying events that could cause negative consequences for a project or organization and taking actions to avoid them (Olsson, 2007). Risk analysis is a vital process for project investment in the real estate industry which has low liquidity and high cost. It mainly consists of three stages: risk identification, risk estimation and risk assessment (Yu & Xuan, 2010a).

Risk Identification. There are currently several risk identification techniques at present including the Delphi technique, brainstorming, Fault Tree Analysis, SWOT analysis and expert survey. Of these, Delphi is the most widely used.

Risk Estimation. Risk estimation quantifies the risks that exist in the process of investment in real estate projects. It uses risk identification, determines the possible degree of influence of such risks and objectively measures them to make evaluation decisions and subsequently choose the correct method to address the risks. The theoretical basis of risk estimation includes the Law of Large Numbers (LLN), the Analogy Principle, the Principle of Probabilistic Reasoning and the Principle of Inertia (X.-L. Tang & Liu, 2009).

Risk Assessment. The risk assessment or risk evaluation of investment in real estate projects refers to the overall consideration of the risk attributes, the target of risk analysis and the risk bearing capability of risk subjects on the basis of investment risk identification and estimation, thus determining the degree of influence of investment risks on the system (X.-L. Tang & Liu, 2009).

Risk occurs at different stages of the investment process. Several risk analysis techniques, tools, and methodologies have been developed to analyze risk in different industries. Some of these techniques, such as the even swaps method, have been integrated with decision support tools called Smart-Swaps to support multi-criteria decision analysis and assist decision makers, in particular, the project manager, to engage in optimal decision-making (Mustajoki & P. Hämäläinen, 2007).

Risk analysis is an important process that needs to be conducted to achieve optimal decision-making. Real estate franchisors can achieve their goals and objectives if they fully understand and can identify the uncertain factors and variability that will affect the level of risk for each given alternative. The uncertain factors or variables will lead to probability and consequences and can be retained as a list of threats that will affect the risk level. It is therefore important to propose a framework of risk analysis as a guideline for investors in the real estate industry.

3.3.2 Risk-based Decision-making Process

Chapter 2 Section 2.1.2 discusses the risk-based decision-making process which includes the types of decision-making process and the decision support technology for risk-based decision-making. The decision support technology discussed briefly as this section focuses on the five main elements of risk-based decision-making framework for investment in the real estate industry. Next section explains in detail the main risk-based decision-making activities.

3.3.2.1 Main Risk-based Decision-making Activity

According to Busemeyer and Pleskac (2009), decision-making processes become more complex, experience greater uncertainty, suffer increasing time pressure and more rapidly changing conditions, and have higher stakes. Thus, it is important to have guidelines or step-by-step activity that will ensure all the requirements and elements for the risk-based decision-making are clearly identified, defined and prioritized.

The main activity for the risk-based decision-making needs to be listed and perform accordingly as required to ensure the decision made is beneficial. Moreover, risk analysis needs to be performed carefully to ensure there are no undetected or potential problems on the horizon as the risk factors and its sources have the tendency to be uncertain. The uncertainty of risk factors will lead to probability and consequences to the outcome of the decision-making process with risk. Risk can be distinguished from other events due to the unwanted effects associated with it, and its ability to change the outcome of the interaction in a negative way or towards an unwanted direction. The consequences are the outcome of an event expressed qualitatively or quantitatively, being a loss, injury, disadvantage or gain. There may be a range of possible outcomes associated with an event (O. K. Hussain et al., 2007). Figure 3.6 shows the main activities of risk-based decision-making for investment in the real estate industry.

The process starts from defining and gathering information sources or data about a current, incoming or upcoming real estate project in the list. For example, the details about the properties or property portfolio need to be collected from valid sources such as real estate agencies. Information regarding the project available in the real estate industry should be gathered as much as possible for the risk-based decision-making. A checklist is used to determine all the elements of potential risk factors are identified and to ensure they are all covered. All the potential risk factors will be identified by asking questions such as what can happen, why it happens and how it happens. The risk identification processes include internal and external sources through brainstorming, inspections, professional consultations, case studies, audits and questionnaires.

The identified risk will be analyzed and the risk level for each project in the list will be determined using several related techniques. The analysis involves ranking and prioritizing the risks for each project based on the project's profile and the risk-consequences analysis. The risk analysis might use qualitative or quantitative techniques, or a combination of the two, to identify the risk level and the consequences of each project. Feasibility studies for all elements in the list will be checked and aim to uncover the strengths and weaknesses of the identified risk factors and determine whether or not the project will be successful. Once the analysis is done, the best project for investment will be selected and monitored. Feedbacks from the decision makers will be asked for each of the output of all of the risk analysis activities involved to ensure all elements or criteria have been met.

The output or decision point of the decision-making process can be categorized into three choices or alternatives: hold, proceed (go) or discard (stop) (Strong et al., 2009). The hold state stipulates waiting for a better time to continue the process; the proceed (go) decision point is to proceed with the potential or actual line of business; and discard (stop) decision point terminates the process. Feasibility studies on all the sources and factors or variables that will affect the project investment need to be carried out to eliminate, hedge or mitigate the risk.

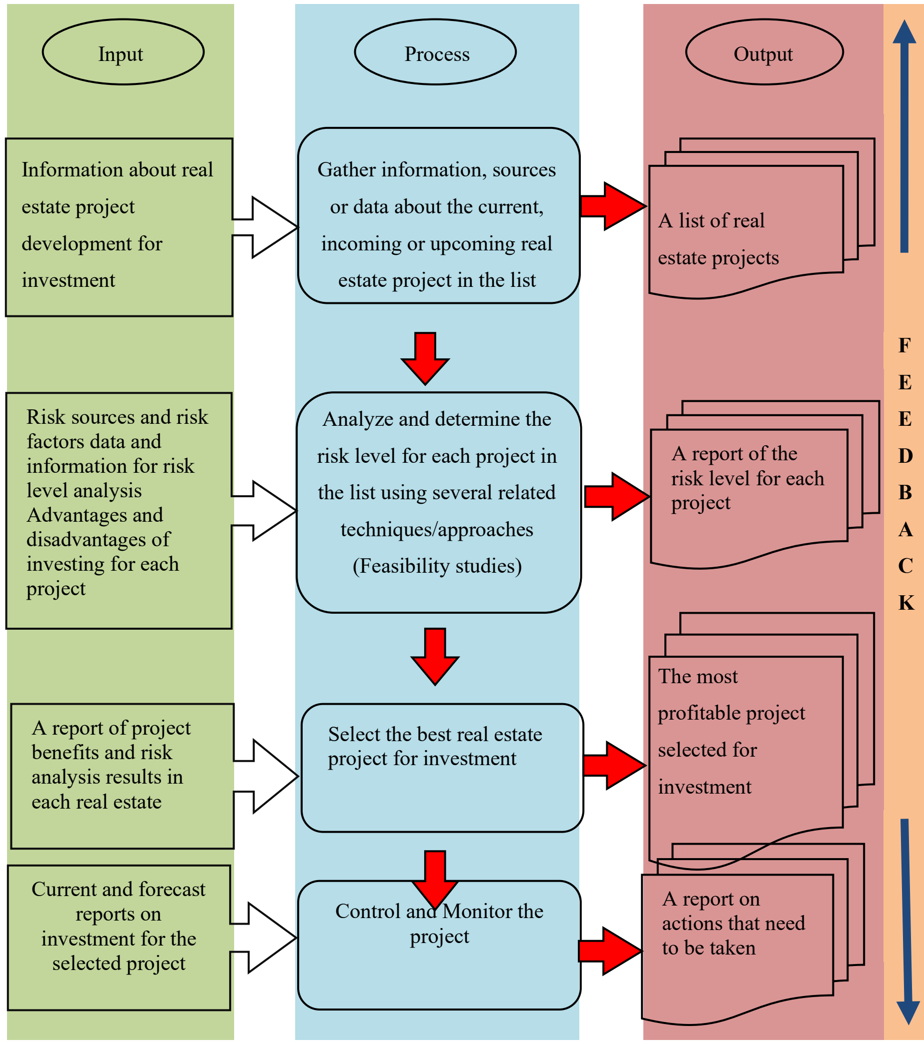


Figure 3.7: Risk-based decision-making main activity for investment in the real estate industry.

3.3.3 Risk Sources and Risk Factors in the Real Estate Industry

The five main categories of risk sources and risk factors for investment in the real estate industry, namely financial risk, economic risk, scheduled risk, policy risk, and technical risk and others. Each of these risk sources has its own risk factors as a sub element as discussed in Chapter 2 Section 2.1.4.

3.3.4 Risk-based Decision-making Techniques for Real Estate Project Investment

Investment in the real estate industry in emerging economies demonstrates tight working capital, low liquidity, slow payback, high sunk cost, capital intensive outflows that are not immediately recovered, enduring uncertainties about demand, price/m^2, land costs, and short to medium construction times. It is very important for investors to have an approach or technique to analyze the real estate project investment to minimize the uncertainty or risks that will affect their profits and margins (Rocha et al., 2007; Y. Sun et al., 2008).

There are many techniques or approaches available for risk-based decision-making, each of which has its own features or criteria and is used for quantitative or qualitative analysis, or both. Some researchers have combined or embedded these techniques to conduct both quantitative and qualitative analysis. For example, the real options method is strictly concerned with quantitative analysis. However, Information and Communication Technology (ICT) investments experience tangible and intangible factors, and the latter can be mainly treated by qualitative analysis. They have proposed a decision analysis technique which combines real options, game theory, and an analytic hierarchy process for analyzing ICT business alternatives under threat of competition (Angelou & Economides, 2009).

This section will explain some of the risk-based decision-making techniques for investment in the real estate industry, which are divided into three categories, namely, quantitative, qualitative and hybrid technique.

3.3.4.1 Quantitative RBDM Technique

Quantitative technique refers to the analysis of variables or elements that can be measured using either discrete or continuous numerical data involving statistical data analysis. Some examples of quantitative RBDM techniques for investment in the real estate industry include Beta, Projection Pursuit model based on Particle Swarm Optimization (PSO), condition value-at-risk (CVaR), Maximal Overlap Discrete Wavelet Transform (MODWT), Markowitz's Portfolio Analysis, Regression Analysis, Statistical Stepwise Regression Analysis and Neural Network Sensitivity Analysis.

i. **Beta.** Beta is a risk measurement for systematic risk (Yan & Lei, 2008). Beta measures the degree of co-movement between the asset's return and the return on the market portfolio. In other words, beta quantifies the systematic risk of an asset (X. Xiong, Zhang, Zhang, & Li, 2005). The systematic risk, as denoted by β_i, is a measure of the slope of a regression line between the expected return on the ith security (R_i) and the return on the market portfolio (R_m) such as Standard and Poor's 500 (S&P 500 cited in Lee and Jang (2007) and Stock Index and New York Stock Exchange (NYSE) Index. Mathematically, the beta (β_i) is expressed as

$$R_i = \beta_0 + \beta_i R_m + e_i. \tag{3}$$

Based on the formula given above, an asset with a higher beta should have a higher risk than an asset with a lower beta (G. Tang & Cheong Shum, 2003).

ii. **Projection Pursuit Model Based on Particle Swarm Optimization (PSO).** The Projection Pursuit model based on PSO is used to process and analyze high dimensional data, especially non-linear and non-state high-dimensional data. PSO can be used to solve a large number of non-linear, non-differentiable and complex multi-peak optimization problems and has been widely used in science and engineering. The Projection Pursuit model can make exploratory analysis and is also referred to as the deterministic analysis method (Shujing Zhou & Li, 2010). Modeling steps for the Projection Pursuit Model are as follows: Investment risk assessment program of the normalized values; Projection index structure function; Optimized projection target function; Scheme selection.

iii. **Condition Value-At-Risk (CVaR).** A dynamic condition value-at-risk (CVaR) technique is one of the new tools of risk measurement for studying investment in real estate. This technique was proposed by Rockafellar and Uryasev as cited in Meng et al., (2007), and has many good properties, such as being computable, convex and more efficient than the Markowitz value-at-risk technique for portfolio investment.

iv. **Maximal Overlap Discrete Wavelet Transform (MODWT).** Maximal Overlap Discrete Wavelet Transform (MODWT) provides a natural platform for investigating the beta or systematic risk behaviour at different time horizons without losing any information (X. Xiong et al., 2005). They proposed this method to decompose a time series of any length into different timescales and listed the advantages of MODWT over the Discrete Wavelet Transform (DWT) as follows: 1) The MODWT can handle any sample size, while the Jth order DWT restricts the sample size to a multiple of 2^J; 2) The detail and smooth coefficients of a MODWT multi-resolution analysis are associated with zero-phase filters; 3) The MODWT is invariant to circularly shifting the original time series; 4) The MODWT wavelet variance estimator is asymptotically more efficient than the same estimator based on the DWT.

v. **Markowitz's Portfolio Analysis and Regression Analysis.** Lin and Chen (2008) carried out a study on the identification of default risk as a systematic risk based on Chinese stock markets using two analyses, namely portfolio analysis and regression analysis to check whether default risk is systematic and to discover the relationship between the expected return and the default risk. Regression analysis was used to examine whether default risk is systematic in the Chinese stock market. They determined that default risk is not a systematic risk factor of the Chinese stock market; however, these two analyses can be used to analyze and identify the systematic risk for investment in an industry.

vi. **Statistical Stepwise Regression Analysis and Neural Network Sensitivity Analysis.** Based on the research study by Wang, Pao and Fu (2000) exploring the relationship between a firm's systematic risk and its long-term investment activities, the results of these techniques show that systematic risk is reduced for investment activities in the fibre industry. For the electronics industry, however, the systematic risk is higher, as firms increase their long-term investment ratio. Companies with a higher portion of long-term investment in assets will show a more significant difference. They use stepwise regression analysis to explore the impact of independent variables on systematic risk and neural network sensitivity analysis to analyze the non-linear relationship between a firm's systematic risk and its long-term investment activities. Their findings indicate that if an industry is somewhat mature and does not have many investment opportunities, the long-term investments in the industry are more diversified.

3.4.1.2 Qualitative RBDM Technique

Qualitative technique is a method for analyzing variables or elements that cannot be measured using numerical data; these variables or elements will instead be given a category such as low, medium or high. Some examples of qualitative RBDM techniques for investment in the real estate industry include the fuzzy comprehensive valuation method and variable precision rough set (VPRS) technique.

i. **Fuzzy Comprehensive Valuation Method.** The fuzzy comprehensive valuation method is used to evaluate the risk degree of a real estate project. Jin (2010) applied this method for comprehensive risk evaluation which would be beneficial and practical for the real estate projects lifecycle. Fuzzy comprehensive valuation is also used to estimate the lifecycle of the project's identified risk factors to confirm the risk level (highest risk, higher risk, general risk, lower risk, and low risk), the first class index evaluation and index weight of all classes. The results of the paper shows that the total risk of each stage of a real estate project reduces gradually with the development of real estate projects, and the risk of the same stage reduces gradually with the development of real estate projects. This would provide foundation data to dynamic deal scheme decisions for risk for real estate projects. This technique has also been used to obtain the value of the risk of real estate investment and has significance in theory and practice for investment risk analysis (Y. Li & Suo, 2009).

ii. **Variable Precision Rough Set (VPRS).** Xie et al. (2010) designed an adaptive algorithm for dynamic risk analysis in a petroleum project investment based on a VPRS technique. They intended to develop a risk ranking technique to measure the degree of risk for individual projects in a portfolio for which experts are invited to identify risk indices and support decision makers in evaluating the risk exposure (RE) of individual projects. Their investigation includes the definition of multiple risks involved in any petroleum project investment using multi-objective programming to obtain the optimal selection of projects with minimum risk exposure. The significance of risk indices is then assigned to each of the corresponding multi-objective functions as a weight.

3.4.1.3 Hybrid RBDM Technique

In order to provide a comprehensive evaluation of risk analysis, the combinations of qualitative and quantitative techniques have been useful for generating better decisions and takes into account all possible uncertainty factors.

i. **Radial Basis Function Neural Network.** Radial Basis Function (RBF) Neural Network is an example of an evaluation model for development risk in the real estate industry. RBFs are embedded in a two-layer feed-forward neural network; the input into an RBF neural network is nonlinear while the output is linear. The RBF neural network consists of one hidden layer of basic functions, or neurons and the chosen RBF is usually a Gaussian. At the input of each neuron, the distance between the neuron's centre and the input vector is calculated. The output of the neuron is then formed by applying the basis function to this distance. The RBF network output is formed by a weighted sum of the neuron outputs and the unity bias shown. Its empirical analysis shows that the evaluation model is characterized as good data approximation, with high stability and normalization. RBF neural networks have the advantage of automatically defining the initial weights and reducing the influence of overlay depending on the experience and knowledge of experts (Ma & Meng, 2009).

ii. **Support Vector Machine (SVM).** The SVM modelling approach has been proposed by Li et. al. (2009) to predict risk for real estate investment. Firstly, the merits of the structural risk minimization principle and the small study sample and non-linear case are used to analyze the risk factors during the investment stage in real estate projects. A model based on SVM in real estate investment risk is then built up. SVM learning training samples are usually based on a project proposal, project feasibility study report, project evaluation reports and other information. According to Liu and Liu

(2010), the main idea of SVM is to transform the input space into a higher dimension space with the nonlinear transformation of inner product function definition, then seeking out the nonlinear relation of the input variables and output variables in the higher dimension. They agreed that SVM can solve such problems as small samples, nonlinear cases, and higher dimensions, and that it provides a global optimal solution since SVM is a convex quadratic programming problem.

iii. **Analytic Hierarchy Process.** AHP is a multi-criteria decision analysis technique that is commonly used for risk analysis. It aims to choose from a number of alternatives based on how well these alternatives rate against a chosen set of qualitative as well as quantitative criteria (Saaty & Vargas, 1994; Schniederjans, Hamaker & Schniederjans, 2005, cited in Angelou & Economides (2009)). AHP has also been employed to determine the weight of every index to deal with the uncertainty of the risk analysis of real estate investment (Y. Li & Suo, 2009).

iv. **Real Option Method.** The application of a real option method seeks to examine the changes in uncertainty that will affect the optimal timing for investment. There are three managerial flexibility criteria of the real option method that influence optimal timing: information gathering, waiting option and abandon option (Rocha et al., 2007).

Kit (2007) examines the effect of uncertainty on investment timing in a canonical real options model. His study shows that the critical value of a project that triggers the exercise of the investment options exhibits a U-shaped pattern against the volatility of the project. It is found that there is a positive relationship between the risk factor and return factor when the volatility of the project increases.

According to Xie et al. (2010), a related factor that makes the timing of a project crucial is the irreversibility of the investments because, for example, the sunk cost cannot be recovered even if market conditions change adversely. One way to avoid regret for irreversible investments under uncertainty is to 'wait and see what happens'.

There are other RBDM techniques, not discussed here, which can be applied in the real estate industry, but this section focuses on the RBDM framework for property investment risk analysis. Each technique has its own limitations and benefits because decision makers must have the knowledge of how to apply the particular technique when making decisions. Decision makers need to choose the best technique to suit their problem-solving situation.

The decision support technology review for the framework indicates that there is much work to be done to develop an intelligent decision support system that can be used for handling risk-based decision-making in business operations, or as a tool for businesses managing their business tasks. Thus, it is vital to explore a new technique for risk analysis using decision support technology such as Intelligent Decision Support System for investment in the real estate industry.

3.3.5 Issues and Challenges of Risk-Based Decision-making

The issues and challenges of RBDM need more consideration and would be a relevant focus for future research. The risk analysis of investment in real estate projects refers to the overall consideration of the risk attributes, the target of risk analysis and the risk bearing capability of risk subjects on the basis of investment risk identification and estimation, which determine the degree of influence of investment risks on the system (X.-L. Tang & Liu, 2009). There are several methodologies or techniques proposed by other researchers to evaluate, analyze, assess or predict the risk. Some of these are the Monte Carlo method, fuzzy set theory (Y. Sun et al., 2008), Markowitz, F-AHP, a real option method (Rocha et al., 2007), and a hidden Markov model. There are a number of issues related to these methods or models. The first is that they have different characteristics, advantages and limitations when applied in different fields (Sun et al., 2008; Lander & Pinches 1998, cited in Rocha et al., (2007)). For example, real option methodology has problems in the practical implementation of risk analysis, such as lack of mathematical skills, restrictive modelling assumptions, increasing complexity and limited power to predict investment in competitive markets (Lander & Pinches, 1998 cited in Rocha et al., (2007)).

Availability of high-quality data is the second issue for RBDM. Zeng, An and Smith (2007) believe that high quality data are a prerequisite for the effective application of sophisticated quantitative techniques. They therefore suggest that it is essential to develop new risk analysis methods to identify major factors, and to assess the associated risks in an acceptable way in various environments in which such mature tools cannot be effectively and efficiently applied.

The third issue is that real estate investment risk evaluation is a complex decision-making problem with multiple factors and multiple targets (S. J. Zhou et al., 2008). The majority of existing real estate investment risk evaluations give priority to single-goal decision-making, use single indices such as the maximum expectation, the largest variance, the minimum standard deviation rate to evaluate the real estate investment. These evaluating methods are easy to understand, but they cannot comprehensively evaluate the quality of an overall program. There are also those who use MAM for real estate investment risk evaluation. The traditional MAM is based on the assumption that the whole of the distribution is subject to the normal distribution, yet the whole distribution of a real estate investment program is uncertain; thus, it is imprecise to use MAM for real estate investment risk analysis. Furthermore, many evaluation programs or models involve many evaluation indices, such that the dimensions are different, and the weights are difficult to determine, and there are therefore difficulties in the practical application.

The fourth issue is related to incomplete risk data availability. In decision-making, the correct methodology is important to ensure that the right decision is made, and that it will be beneficial to investors, users or agents. More formal methodology is thus necessary in decision-making processes (O. K. Hussain et al., 2007; Zeng et al., 2007). Formal methodologies are needed to make sure that any decision can be assessed effectively and efficiently. Many risk analysis techniques currently used in the UK construction industry are comparatively mature, such as Fault Tree Analysis, Event Tree Analysis, Monte Carlo Analysis, Scenario Planning, Sensitivity Analysis, Failure Mode and Effects Analysis, Programme Evaluation and Review Technique (Zeng et al., 2007). In many circumstances, however, the application of these tools may not give satisfactory results due to the incompleteness of risk data.

The fifth issue is the need for an effective and efficient technique. New risk analysis methods to identify major factors and to assess the associated risks in an acceptable way in various environments are needed, as older tools cannot be effectively and efficiently applied.

The sixth issue is the non-scientific method proposed. The methods of risk analysis which have been used by domestic real estate developers so far, such as risk survey, break-even analysis and sensitivity analysis, are based on the discounted cash flow and net present value (NPV). These methods are far from scientific and easily lead to faults, and furthermore, to the severe consequences of failure (Yu & Xuan, 2010a).

This section suggests a risk-based decision-making framework that is applicable for investment in the real estate industry. Risk-based decision-making is an important area of focus in real estate investment, which involves high risk and high cost. Risk with high uncertainties will lead to the occurrence of a higher percentage of probabilities and consequences. The uncertainties of a number of risk factors and risk sources contribute to the level of dynamic risk prediction, which is dependent on what takes place from the initial investment to the later stages of the real estate development.

3.4 Multidimensional Analysis and Data Mining Framework for Property Investment Risk Analysis

The real estate industry is characterized by high risk, high return and long cycle, which needs real estate investors to carefully research to minimize the risk (C. Fang & Marle, 2012; Montgomery & Smith, 2009; A. Piyatrapoomi et al., 2004). Risk occurs at a different stage of the investment process. The risk assessment or risk analysis is part of the process involved when deciding on which project should be given priority for investment with the limited budget and time given. There are many techniques available for risk analysis decision-making. The investors should consider a technique that can provide a comprehensive analysis with a high level of trust and satisfaction.

Risk analysis for property investment decision-making was influenced by many risk factors. The dynamic changes of risk factors such as contract and policy adjustment impress a direct influence on the success of property investment in the real estate industry. Property investment in the real estate industry normally depends on the objective of the investment and is tailored to meet individual preferences and requirements. For example, the main goal for property investment can be divided into three reasons namely for rental, residential or commercial. The investor normally refers to the real estate agents who are experts in the field in giving them suggestions and helping them in making decisions.

Popular technique for indexing the risk factors rank and weight in the real estate industry is known as AHP which takes into account the expert judgment for indexing. The AHP technique, apply the concept of hierarchical data model and have been employed to determine the weight of every index to deal with the uncertainty of the risk analysis of real estate investment (E. C. M. Hui et al., 2010; C. H. Jin, 2010; Ye, 2011). The subjective judgment and objective reasoning were combined to evaluate the real estate risk factors by using AHP (C. Fang & Marle, 2012). The process of achieving the results involved with experts' opinion and different levels of expertise will give different judgment and lead to misinterpretation. Moreover, by referring to expert's or professional's opinions, there is no additional information that might be beneficial to investors. The experts would answer and give suggestions based on the investor's requirements.

In order to solve this issue, this section proposes the application of a multidimensional data model and data mining technique as a deterministic approach to rank and weight the risk factors. The application of the multidimensional model for property investment risk analysis helps the investor to make the decision more reasonable and scientific. The proposed technique helps the investor to analyze data more systematic and provides comprehensive analysis through high dimensionality of data. This section identifies the power and benefit of multidimensional analysis and data mining technique can bring to the practice of determining the rank and weight of the uncertain risk factors. Deterministic approach using massive available data to rank and weight the uncertain risk factors it determines ahead of time which risk factors will be assigned to specific rank and weight.

The implication of studying multidimensional analysis and data mining for property investment risk analysis is to respond to the limitations of expert opinion and AHP that apply the hierarchical data model. By using multidimensional analysis and data mining, the results generated are accurate and based on structured data stored in the database or data warehouse. The main purpose of the proposed technique is to improve the accuracy of risk measurement and to avoid misinterpretation of risk factors rank and weight from expert judgments caused by a different level of expert's experience. The multidimensional analysis and data mining framework proposed includes the following four modules: (1) User module; (2) Information Technology module; (3) End-user access tools module; (4) Application module. The basic framework of multidimensional analysis and data mining framework for property investment risk analysis shown in Figure 3.7.

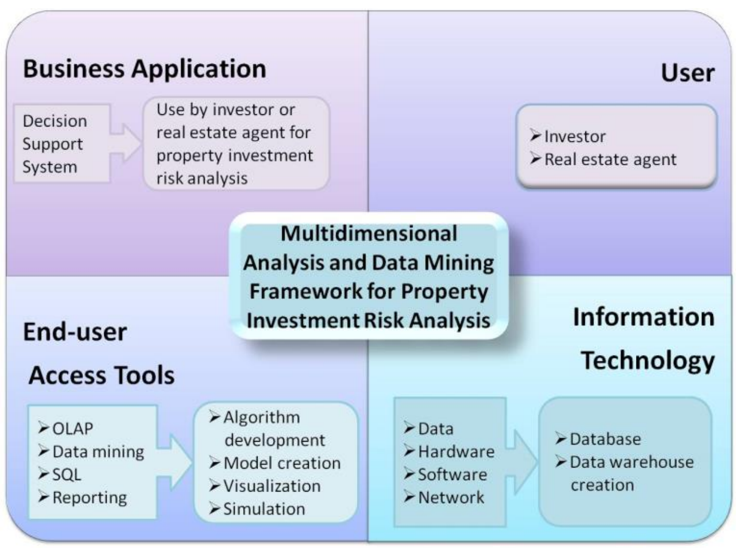


Figure 3.8: The multidimensional analysis framework for property investment risk analysis.

Module 1: User

The user refers to the decision maker particularly the investor and real estate agent who are dealing with the property investment risk analysis. They will need to set their goals and requirements for the investment and will gain knowledge based on results generated by multidimensional analysis and hidden patterns of data based on data mining techniques requested. The explicit knowledge from the decision support tools would be transferred to users as tacit knowledge for them to decide the best investment that aligned with their goals and objectives.

Module 2: Information technology

This module mainly includes the data, hardware, software and network to create a database and data warehouse for the multidimensional analysis and data mining technique. The data as input gathered from different types of form such as report, receipt, and business transaction data either from user or customer, contractor, real estate agent, or investor as a query. The data collected for the analysis must be valid and for multidimensional analysis and data mining technique, a high dimensional data is needed. The hardware, software and network are the technologies used to keep the data, collecting, storing, processing, analyzing the data and providing information as knowledge to the decision maker. The data model apply in this paper is multidimensional data model

Module 3: End-user access tools

The end-user access tools module refers to a set of data warehouse tools and technologies in which determine the success of data warehouse applications. End-user access tools for data warehouse application include OLAP, data mining, structured query language (SQL) and reporting. This module will become the interface for users as in module 1 to integrate with module 4 by using elements in module 2. The user will develop an algorithm and create a model for the property investment risk analysis. The output of the analysis from this module is either displayed as visualization or simulation for the user to understand and analyze.

Module 4: Business Application

The business application refers to how this framework would be helpful and useful in business transactions. The real estate industry applies this technique for property investment risk analysis. The type of information system pertaining to the proposed technique is the DSS which helps the investor to achieve their goals. For the real estate industry, the DSS would be useful for the managerial level to achieve organization targets and gaining competitive advantage. The process of multidimensional analysis and data mining technique for property investment risk analysis includes input collection as a query from the user, data processing using information technology and end-user access tools and results generated will be displayed to users for decision-making.

3.5 Experiment

All investors have different sets of criteria and limitations that will affect their decision. Issues of selecting the best property for investment that fulfill their goals and objectives also need to be considered. Knowledge on significant factors or variables that will affect the decision made need to be analyzed and trend analysis would be helpful for investors to familiar with the current scenario of investment. This section explained the functions of frameworks proposed in this chapter via an experiment using real dataset from the Ministry of Housing and Local Government in Malaysia.

Experiment of the dataset was analyzed using Waikato Environment for Knowledge Analysis (WEKA) application software. WEKA is a data mining system developed by the University of Waikato in New Zealand that implements data mining algorithms. WEKA is a state-of-the-art facility for developing machine learning (ML) techniques and their application to real-world data mining problems. It is a collection of machine learning algorithms for data mining tasks. The algorithms are applied directly to a dataset. WEKA implements algorithms for data pre-processing, classification, regression, clustering, association rules; it also includes a visualization tool. The new machine learning schemes can also be developed with this package. WEKA is open source software issued under the GNU General Public License.

3.5.1 Dataset

The dataset describes the estimated income and home ownership among Ministry of Housing and Local Government staff for the year of 2018. The data has been updated on 14 November 2018 (Data Terbuka, 2020). The dataset retrieved from https://www.data.gov.my/data/ms_MY/dataset/anggaran-pendapatan-dan-pemilikan-rumah-di-kalangan-warga-kpkt. This experiment shows how determinants identification using ranker (machine learning using WEKA) varies among different investor's limitations. Next, simulation data used to simulate on weighing the variable of property features such as house price, security level, surrounding amenities, location, property type and number of houses owned also included to show how the Personalized Multidimensional Process Framework and Risk-based Decision Making Framework that will integrate both heuristic and deterministic approach to map the results to provides solutions for the decision making.

The dataset main goals is to identify factors contributing to the total number of houses owned based on 7 dimensions namely state, age, ethnicity, marital status, number of households living together, monthly household income and total of monthly installment for the house owned. Total number of respondents involved in the dataset is 2392 with 8 attributes as shown in Figure 3.9.

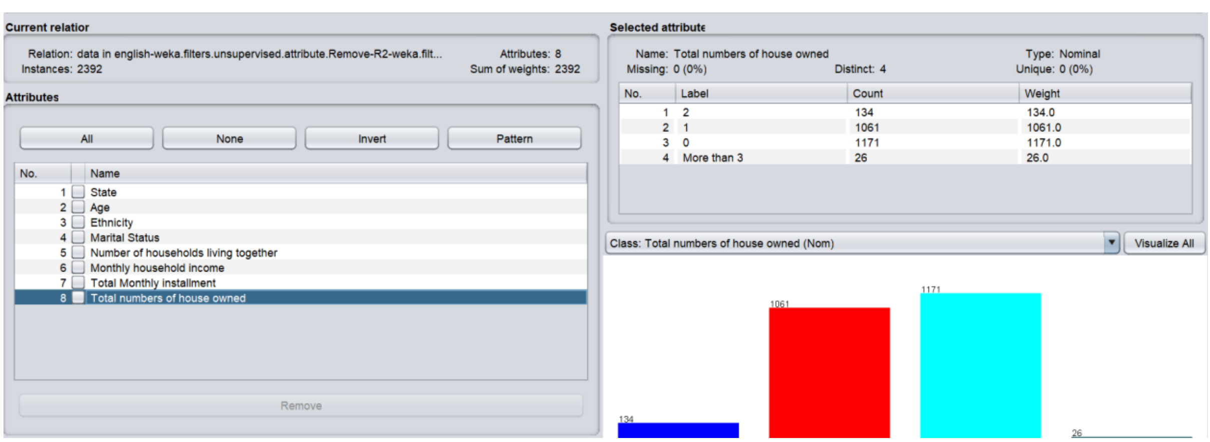


Figure 3.9: Screenshot of recognized attributes and the list of recognized attributes using WEKA application software.

The metadata of the selected dataset and the descriptive analysis of dataset described in Table 3.1 and Table 3.2 respectively.

Table 3-1: Dataset attributes metadata.

No	Attribute	Description	Data Type
1	State	List of State in Malaysia: 1. W.P Putrajaya 2. Selangor 3. Perak 4. Johor 5. W. P Kuala Lumpur 6. Melaka 7. Kedah 8. Negeri Sembilan 9. Pahang 10. Pulau Pinang 11. Sabah 12. Sarawak 13. Perlis 14. Terengganu 15. Kelantan	Nominal 16 groups
2	Age	Age of the respondent: Number discretized to five groups. Minimum age: 19 Maximum age: 60 Category of Age: 1. Below 27.2 2. 27.2-35.4 3. 35.4-43.6	Number to Nominal 5 groups

		4. 43.6-51.8 5. More than 51.8	
3	Ethnicity	List of ethnics related to the respondents. 1. Malay 2. Chinese 3. Indian 4. Bumiputra Sabah 5. Bumiputra Sarawak 6. Others	Nominal 6 groups
4	Marital Status	List of marital status of the respondents. 1. Married 2. Single 3. Widow/Widower	Nominal 3 groups
5	Number of households living together	Number of households living together list. 1. 1 2. 2 3. 3 4. 4 5. 5 6. More than 5	Nominal 6 groups
6	Monthly household income	Total monthly household income. 1. Below RM1,999 2. RM2,000 – RM2,999 3. RM3,000 – RM3,999 4. RM4,000 – RM4,999 5. RM5,000 – RM5,999 6. More than RM6,000	Nominal 6 groups

7	Total Monthly Installment	Total of monthly installment with regards to total number of houses owned by the respondent. 1. 0 2. Below RM599 3. RM600 – RM699 4. RM700 – RM799 5. RM800 – RM899 6. RM900 – RM999 7. RM1000 – RM1,999 8. More than RM2,000	Nominal 8 groups
8	Total number of houses owned	Total number of houses owned is the output (class) for the analysis 1. 0 2. 1 3. 2 4. More than 3	Nominal 4 groups

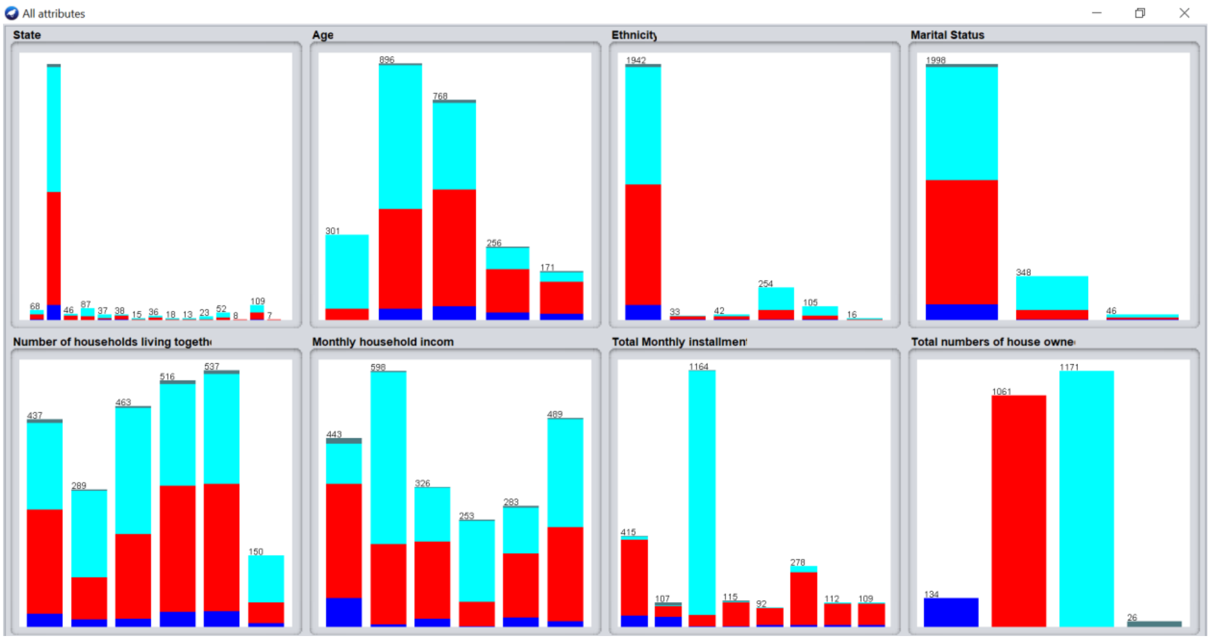
Table 3-2: Dataset instances descriptive statistics.

No	Attribute	Label	Count
1	State	W.P Putrajaya	68
		Selangor	1835
		Perak	46
		Johor	87
		W. P Kuala Lumpur	37
		Melaka	38
		Kedah	15
		Negeri Sembilan	36
		Pahang	18
		Pulau Pinang	13

		Sabah Sarawak Perlis Terengganu Kelantan	23 52 8 109 7
2	Age	Below 27.2 27.2-35.4 35.4-43.6 43.6-51.8 More than 51.8	301 896 768 256 171
3	Ethnicity	Malay Chinese Indian Bumiputra Sabah Bumiputra Sarawak Others	1942 33 42 254 105 16
4	Marital Status	Married Single Widow/Widower	1998 348 46
5	Number of households living together	1 2 3 4 5 More than 5	150 289 463 537 516 437

6	Monthly household income	Below RM1,999 RM2,000 – RM2,999 RM3,000 – RM3,999 RM4,000 – RM4,999 RM5,000 – RM5,999 More than RM6,000	253 598 489 326 283 443
7	Total Monthly Installment	0 Below RM599 RM600 – RM699 RM700 – RM799 RM800 – RM899 RM900 – RM999 RM1000 – RM1,999 More than RM2,000	1164 278 109 112 115 92 415 107
8	Total number of houses owned	0 1 2 More than 3	1171 1061 134 26

As shown in Table 3.2, the majority of the respondents did not own a house (1171) followed by only one house (1061), two houses (134) and more than three houses (26). Now, let us explore the dimensions or factors contributing to the class (total number of houses owned) in the next section.



Color Code	Total number of houses owned
	0
	1
	2
	More than 3

Note: Color code represent the total number of houses owned.

Figure 3.10: Relationships between each attribute that contribute to the number of houses owned by the respondents (staff of Ministry of Housing and Local Government in Malaysia in year 2018).

Figure 3.1 shows the patterns of relationship between each attribute that contribute to the number of houses owned by the respondents. It can be concluded that the patterns of analysis dominated by respondents who did not own any house and only one house.

3.5.2 Application of Risk-based Decision-Making Framework via Select Attribute Function.

Attribute or feature selection plays an important role in the process of determinant identification used to extract the ranking of attributes. The experiment study is implemented in the data mining tool, WEKA. The two approaches are applied to the dataset: i)

CfsSubsetEval (CSE) ii) InformationGainAttribute evaluation (IG). These two approaches are explained below.

3.5.3 CfsSubsetEval (CSE)

This method measures the significance of attributes on the basis of predictive ability of attributes and its degree of redundancy. The subsets which are having less intercorrelation but highly correlated to the target class (total number of houses owned) are preferred. The result shows that the ranking of the first three attributes are taken into account namely age and total monthly installment and ethnicity out of the 7 dimensions listed. Figure 3.11 shows the ranking of attributes with respect to CfsSubsetEval method.

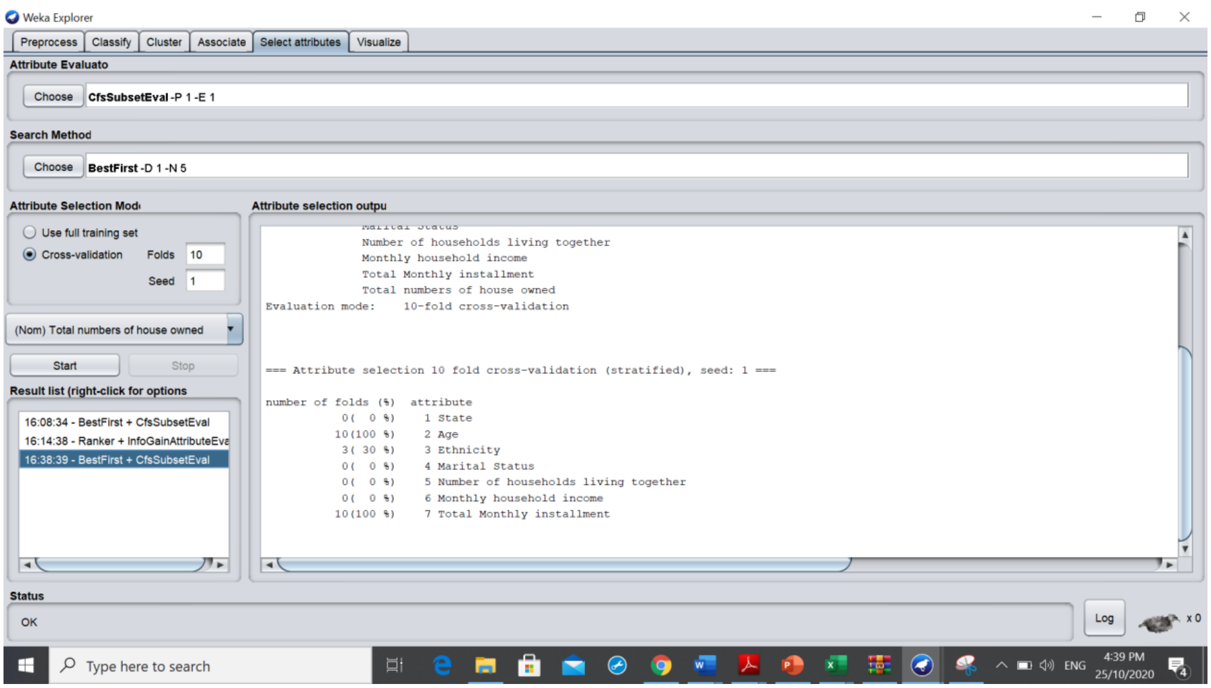


Figure 3.11: Ranking of attributes with respect to CfsSubsetEval method

The aforementioned attribute selection algorithm results in the ranking of attributes. Based on the returned ranking of attributes, the dataset is modified. That is the irrelevant attributes are removed. This modified dataset is involved in the classification task. Then the result of the classifier before and after attribute selection is compared.

3.5.4 Information Gain Attribute Evaluation

This method measures the significance of attributes by the measure of information gain calculated with respect to target class. It can be calculated by the formula

$$\text{InfoGain(Class,Attribute)} = \text{H(Class)} - \text{H(Class} \mid \text{Attribute)}$$

where H represents the Entropy.

By the combination of this attribute evaluator with the Ranking method of Search is applied to the dataset. The ranking of attributes is 1. Total monthly installment, 2. Monthly household income, 3. Age, 4. Marital status, 5. State, 6. Number of households living together and 8. Ethnicity as shown in Figure 3.12.

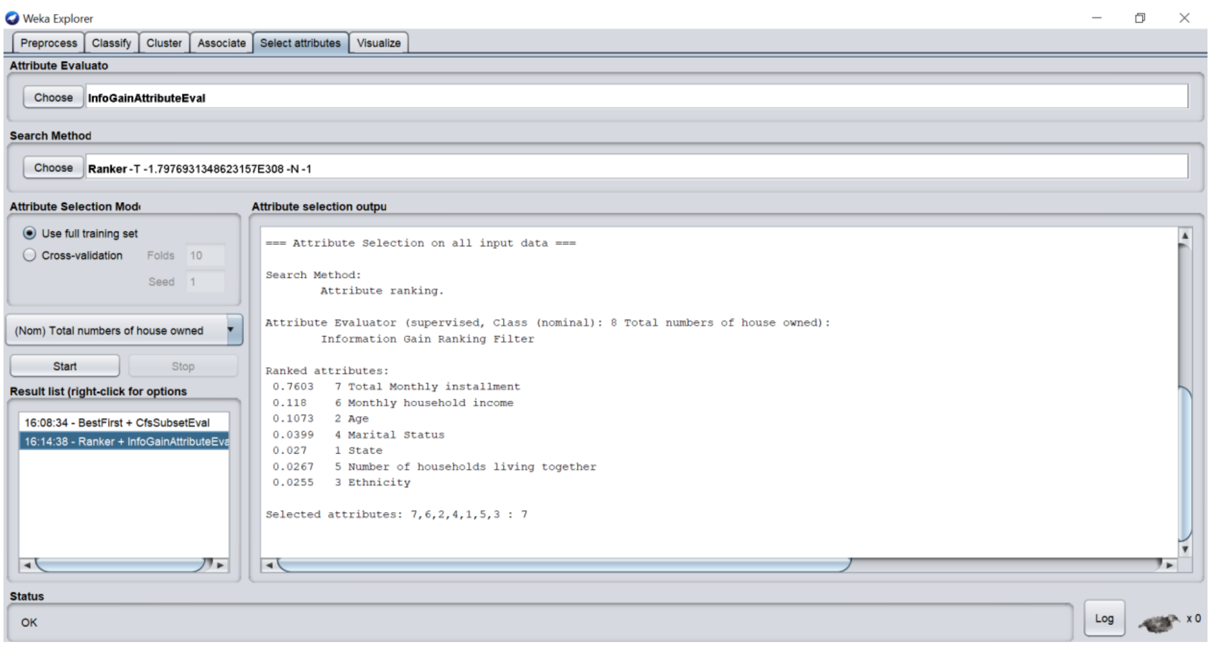


Figure 3.12: Ranking of attributes with respect to GainRatio Attribute evaluation method

The experiment concludes that the determinant identification via heuristic and deterministic approach applied in both risk-based decision-making framework and personalized multidimensional process framework increases the

3.6 Summary

This chapter describes the adaptive personalized property investment risk analysis frameworks to overcome the limitation of existing property investment risk analysis for identifying the determinants and weighting the variable of property features. The adaptive personalized property investment risk analysis framework introduces three new frameworks namely personalized multidimensional process framework, risk-based decision-making framework and multidimensional analysis and data mining framework.

The personalized multidimensional process framework consists of nine major different steps that apply the bottom-up approach to support the decision-making process. The investors need to be ready with their goals and limitations as input for personalization in order to proceed with the data analysis. Risk-based decision-making framework which aims to reduce or mitigate risk in property investment analysis comprises five main elements namely, foundation concepts; process; sources and factors; techniques/approaches; and issues and challenges. Whereas, the multidimensional analysis and data mining framework aims to rank and weight the risk factors which differ among investors as the decision maker which includes the following four modules: (1) User module; (2) Information Technology module; (3) End-user access tools module; (4) Application module. These three frameworks proposed to integrate the adaptive personalization techniques and data-driven approaches such as data mining for risk identification, risk measurement and risk assessment.

CHAPTER 4

AN ADAPTIVE PERSONALIZED PROPERTY INVESTMENT RISK ANALYSIS METHOD

4.1 Introduction

This chapter discusses the proposed adaptive personalized property investment risk analysis (APPIRA) method, which integrates an adaptive personalization approach and a data-driven approach to address the shortcomings of the existing literature that we outline and pointed to in Chapter 2. The adaptive personalized property investment risk analysis frameworks described in Chapter 3 integrates the personalization and data-driven approach to overcome the limitation of existing techniques.

The objective of the personalization approach is to deal with investor's requirements, which is not similar among decision makers. In the personalized multidimensional process framework, personalization technique is essential to achieve optimal solutions and determine the property investment determinant that act as a risk factor for risk identification. Since different investors have different requirements, it is very important to apply a personalization technique to achieve investor's goals and to identify factors to consider before investing. Moreover, data-driven approaches such as data mining techniques are adopted as input for knowledge discovery rather than referring to expert's subjective judgment. The subjective judgment made by the experts in the field contains no formal calculation and only reflects the subject's opinions and past experience.

This chapter presents an adaptive personalized property investment risk analysis method to realize the requirements and introduces two sub methods known as personalized association mapping method and personalized multidimensional sensitivity analysis method. The remainder of this chapter is structured as follows. Section 4.2 briefly describes the adaptive personalized property investment risk analysis method. Two sub methods to be met in order to complete the process referred to personalized association mapping method and personalized multidimensional sensitivity analysis method presented in Section. 4.3 and Section 4.4. The adaptive personalized property investment risk analysis method process flow, how it works, TOPSIS method and example described in Section 4.5. Finally, this chapter is summarized in Section 4.6. The work presented in this chapter has been reported in three of our publications listed in Section 1.8, i.e. publication numbers 2, 3 and 6.

4.2 The Adaptive Personalized Property Investment Risk Analysis (APPIRA) Method

The majority of existing real estate investment risk evaluations give priority to single-goal decision-making. Single indices, such as maximum expectation, largest variance, or minimum standard deviation rate, are used to evaluate real estate investments. Dynamic risk analysis has been studied by many researchers. Each applied a different technique or method to support the decision-making process. For example, studies on risk-based decision-making in real estate investment find that the highest performance index carries the most weight with investors. However, there is a need for studies to justify the use of historical data when weighting and ranking risk in support of decision-making.

Hence, APPIRA ranks and weights risk factors using patterns of data discovered with data mining techniques (deterministic approach) and maps them against investor requirements (heuristic approach) to achieve the investor's goals and objectives. The deterministic approach includes the use of decision support tools, such as data mining techniques. Decision support tools help businesses manage their business tasks efficiently and effectively, especially when managers deal with decision-making processes in their daily routine. Here, we indicate how multidimensional risk factors, and the personalization mechanism are integrated using the APPIRA method. Intended results with this study are to identify the determinants of risk factors for specified goals and limitations of investors. This step refers to the personalization technique to ensure that the options reflected with the investor's requirements. Once the determinants have been identified, the investor will weigh the determinants based on their limitations. Again, personalization involves mapping with data sets available in the system.

This chapter proposed a new way of analyzing risk analysis in property investment using automated tools as existing techniques such as Delphi method, AHP, ANP and real option weight and rank the risk factor based on expertise in the field. In this paper, data-driven approaches were used to explore hidden knowledge to identify the determinants of risk factors based on investor's goals and limitations sets as criteria. Automatic processing is faster and more accurate than manually done. Furthermore, after identifying the determinant, the investor will weigh and rank the determinant personalized based on their requirements. This novelty helps the investors to achieve an optimal solution.

Figure 4.1 shows that APPIRA method consists of four main modules: adaptive personalization, risk analysis, determinants module, and knowledge discovery.

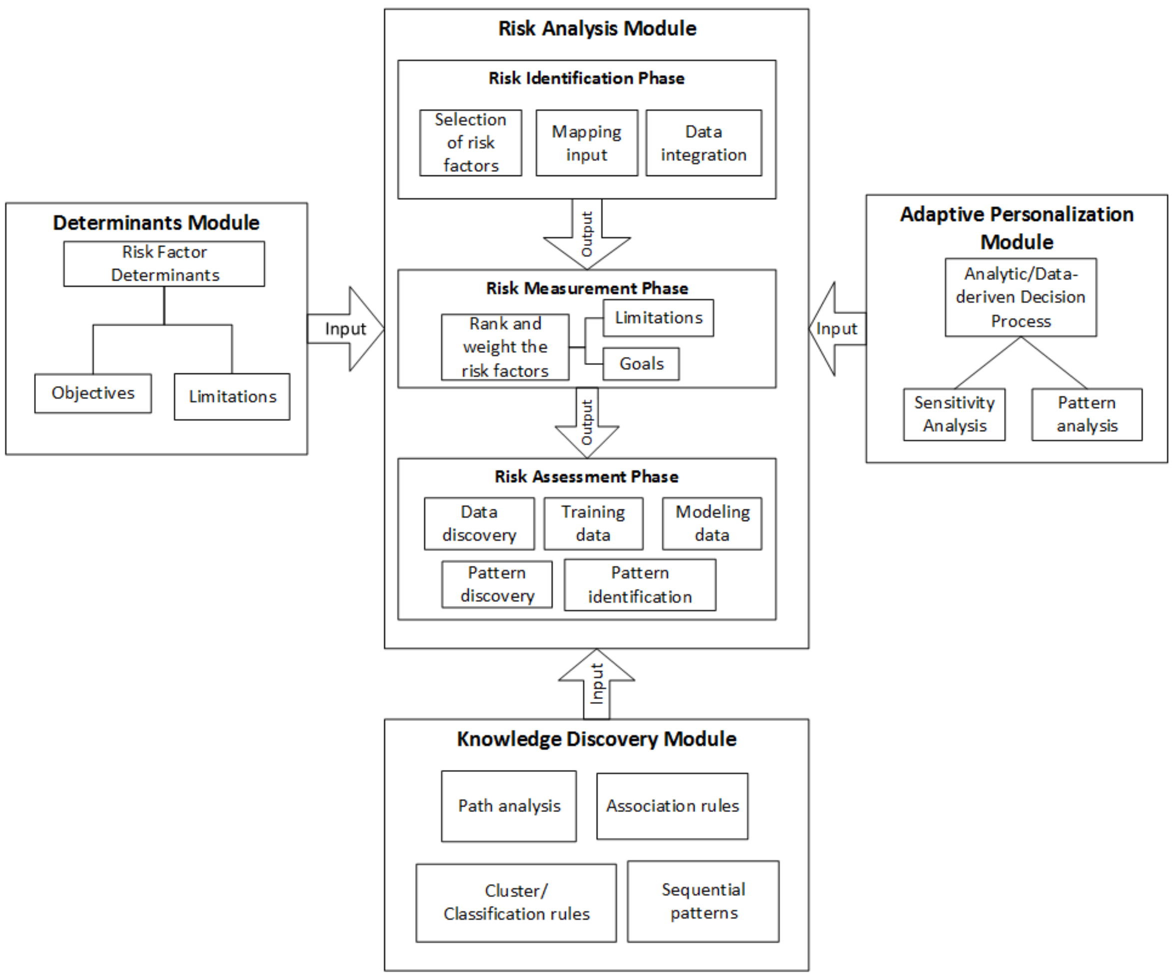


Figure 4.1: Four main modules in APPIRA method : adaptive personalization, risk analysis, determinants module, and knowledge discovery

All investors aim to achieve optimal decision-making in property investment in which decision made relative to current objectives and current limitations/constraints. An adaptive personalization method for property investment risk analysis is needed to optimize and help to achieve investor's goals. Furthermore, it helps to fulfil the requirements of investors based on their limitations. Different investors will have different goals and will set their risk factors differently. Thus, the proposed method helps to fully consider the investors' constraints or limitations such as the amount of capital while achieving their goals. As shown in Figure 4.1, there are four main modules in the proposed adaptive personalization method, integrated to gather input from investors and provide output to investors for decision-making analysis. The input for APPIRA comes from adaptive personalization, risk factor determinants module and knowledge discovery module. Risk analysis module gathered and processed these inputs to produce the output for optimal decision-making, Main characteristics of APPIRA method refers to personalization, automatic processing and user requirements diversification. However, the limitations of the APPIRA method are depending on the datasets available or stored in the system to produce the output and also input from the investors. Timely and valid data are needed to achieve better results and contribute to optimal decision-making. The following sub-sections briefly explain each module and its components.

4.2.1 Adaptive personalization module

The adaptive personalization module consists of a combination of analytic and heuristic approach that produces hidden knowledge based on pattern analysis for investors to do sensitivity analysis. Analytic decision process refers to the input from investors based on their knowledge in the field and heuristic decision process refers to investor's experience. In analytic or heuristic approaches, investors set their goals, objectives, and limitations based on the knowledge they gain when analyzing output from the system. The hidden knowledge gained during the process varies as it is affected by many factors, such as the investors' demographic profiles and objectives. This module provides input for the risk analysis module to optimize and fully use the available sources of the investors. Adaptive personalization module introduces a new way of fulfilling the investor's goal and limitation in which investors set their criteria (risk factor) from a list of property features available and decide weight for each of the criteria selected to be processed using mapping techniques. Moreover, based on the determinants of property investment listed, the investor personalized the weight for each risk factor listed.

4.2.2 Risk analysis module

The risk analysis module is divided into three phases: risk identification, risk measurement, and risk assessment. The risk identification phase includes the selection of risk factors, mapping inputs, and data integration. The selection of risk factors varies among investors, and the system maps the input using the data available in the real estate database. The output of this process gathered from a combination of the data input by the investor and data stored in the system. This output is sent as input to the next step – the risk measurement phase. In the risk measurement phase, the data is ranked and weighed against the risk factors given the investor's limitations and investment goals. The ranking and weighting of risk factors also varies among investors according to their goals and constraints. The output of this step is passed to the next phase for assessment.

The risk assessment phase measures the risk factors for data discovery, pattern discovery, and pattern identification. The data is then trained and, once ready, the data model performs a sensitivity analysis to optimize the results and fulfil the investors' requirements. The data model adapts and changes according to the risk factors identified in the personalization mechanism. The results of this analysis are provided for decision-making.

4.2.3 Determinant module

The determinant module uses a data-driven approach to provide knowledge for investors to identify, rank and weigh the risk factors and property attributes based on investor's requirements. The requirements were useful for producing an optimal decision in risk analysis. Analysis from the determinant module gathered investor's objectives and limitations as input for the risk analysis module. Data mining shows a hidden pattern of data that will become the guidelines for investors to achieve an optimal solution.

4.2.4 Knowledge discovery module using a data-driven approach

The knowledge discovery module adopts the data mining techniques to find a pattern of data and divides it into several techniques. The most popular techniques include sequential patterns, association rules, cluster or classification rules, and path analysis. All results from these techniques are used as inputs for risk analysis to achieve the best results and fulfil the investors' requirements.

4.3 Personalized Association Mapping Method

This section will discuss in detail about the integration of both heuristic approach and deterministic approach applied to the PAM method. Heuristic approach refers to the input from the investors as users of the system developed and deterministic approach refers to the application of data mining as part of knowledge discovery in databases. There is a need for effective property investment risk analysis to help users achieve their best investment. Figure 4.2 depicts the integration between heuristic and deterministic approach applied to PAM method.

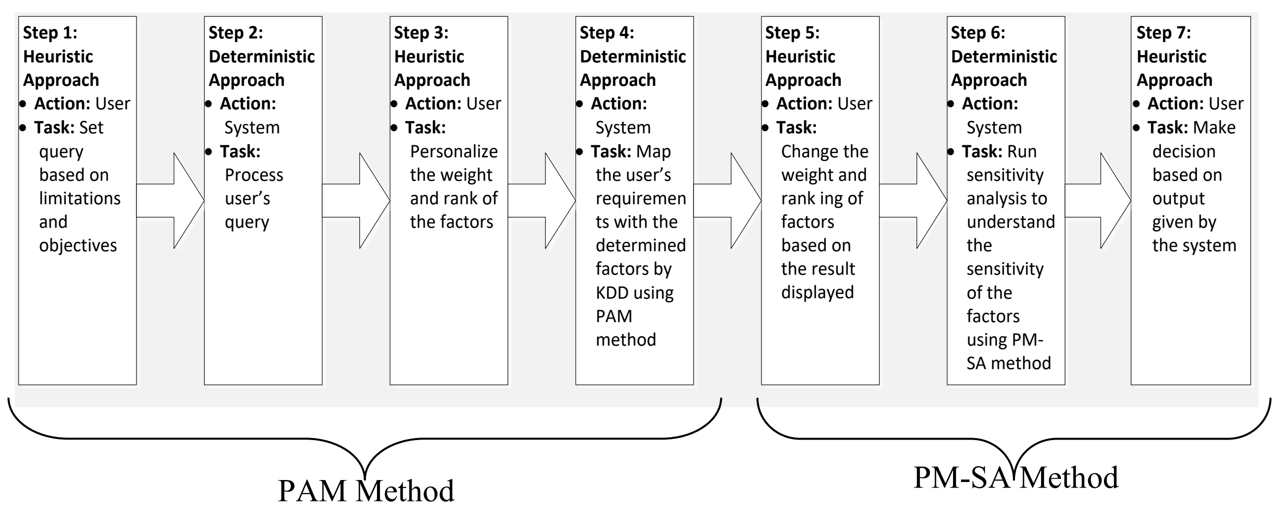


Figure 4.2: The integration between heuristic and deterministic approaches.

The process begins with the heuristic approach where the user will set a query based on their limitations and objectives in finding the best investment which fulfils their requirements.

Next, the system will process the user query by providing a list of potential properties for investment based on user limitations and objectives for the user to analyze. The pattern of output given by the system is based on the user's requirements that provides ranking and weight of risk factors related to the prospect properties for investment. Based on the result displayed from the system, the user will personalize the weight and ranking of the risk factors according to their goals and objectives. Then, the system will map the user requirements with patterns discovered using data mining techniques and provide the output for the user to analyze. After that, the user will make changes to the weight and ranking of the risk factors in order to oversee the variation and sensitivity of risk factors based on their personalization. Subsequently, the system will run the sensitivity analysis to understand the sensitivity of the risk factors towards user's personalization and produce the result. Finally, the result will be made by the user based on the output given by the system.

The function of the PAM method is to map the determined risk factors by KDD with investor's personalization that will affect the decision-making process. The PAM method applied the integration of heuristic approach and deterministic approach to problem solving and decision-making process. This method will help the investors to achieve their goals and objectives by choosing a heuristic approach through the deterministic approach. Deterministic approach is based on historical and pattern data using data mining techniques while the heuristic approach is based on investor's personalization. A system has been developed to validate and run the method proposed.

The heuristic approach will gather information from the decision makers namely users or investors and the deterministic approach will gather information from the developed system. Figure 4.3 shows the use case diagram referring to the investors or decision makers which will determine the criteria and input for the system to process based on the integration of heuristic and deterministic approach.

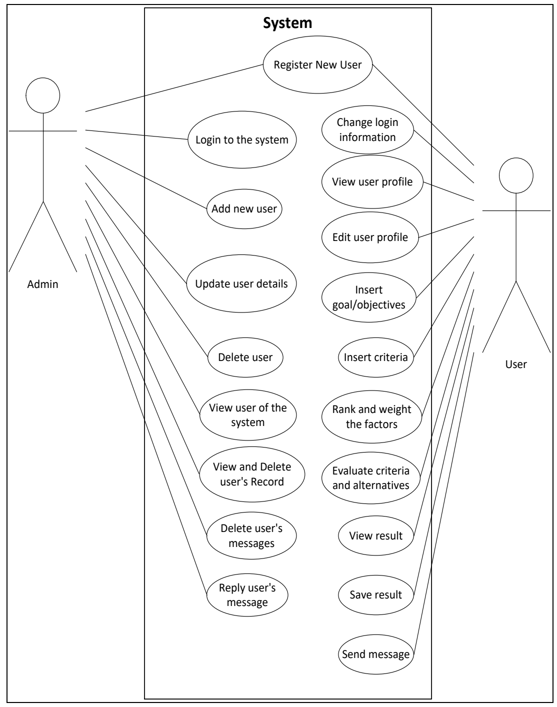


Figure 4.3: The use case diagram of the adaptive personalized property investment risk analysis system.

4.3.1 Personalization

The element of personalization has been researched for many years and many personalization algorithms have been investigated (S. J. Zhou et al., 2008). In this section, the term personalization refers to "the mapping and satisfying of a user's/business's goal in a specific context with a service's/business's goal in its respective context" (D'Urso, 2002). Personalization is motivated by the recognition that a user has needs and meeting them successfully is likely to lead to a satisfying relationship with what they require. Personalization involves a process of gathering user-information during interaction with the user, which is then used to provide appropriate assistance or services, tailor-made to the user's needs (D'Urso, 2002; Yuan et al., 2013).

The technology of *personalized service* has been applied to many different fields. Modern personalized service can provide pertinent service for different users so that their specific demands can be met. In an Internet field the technology of personalized service can improve the quality of a web service and the efficiency of users' access (Peng et al., 2011; Y. Sun et al., 2008). Personalization criteria help to solve several issues because the vast majority of queries to search engines are short and ambiguous, and different users may have completely different information needs and goals when using precisely the same query (S. J. Zhou et al., 2008).

Personalization enables users to fully utilize their constraints, such as budget or capital, and time constraints, for investment in the real estate industry that are characterized as long-term planning. There is a gap in the literature of a need for the justification of risk factor weight and ranking that is based on historical data-driven to decision support using knowledge discovery and investors' personalization for real estate investment risk analysis. The application of data mining techniques to find patterns of data helps to provide accurate and valid information for users to understand, analyse, and use as knowledge to enable them to make better decisions. Personalization helps to achieve a user's goals and requirements by making recommendations automatically, based on data available for analysis. Personalization allows data to be delivered and matched with the user's requirements and interest to fully utilize the user's constraints.

Personalization is a great advantage that enables users to fulfil their requirements with given constraints. For example, in property investment risk analysis, investors can set the rank and weight of the risk factors of the property features that align with their limitations and goals. It is, therefore, a more effective way of meeting the objective of property investment and achieves better results. In addition, no other existing approaches in real estate property investment analysis take into account the investor's personalization and apply the multi-dimensional analysis of the factors that will affect the decision-making process. There are many substantial studies related to the application of AHP and Analytic Network Process for risk analysis in different fields, including analysis of investment in the real estate industry. However, limited study has been undertaken for property investment that does not include the personalization criteria for risk analysis in the real estate industry, based on knowledge discovery from data mining processes. This section proposes a new personalization model for risk analysis in the real estate industry that meets the investor's requirements to achieve their goals and objectives.

Comprehensive risk analysis is needed to help investors to make profits and, at the same time, achieve their goals, since the investment decisions to be made by the investors are complex and risky. The investment decisions are complex because of the uncertain risk factors in risk analysis in the real estate industry. Since uncertainties are involved in the factor determination over a period of time, it is very important to understand how sensitive the factors are to the variation of the investor's personalization. Sensitivity analysis can be used to see the variation of the result if there is variation in the factors personalized by the investors that will affect the real estate property sold in a certain period of time.

Generally, risk analysis in the real estate industry is affected by many factors such as financial and interest rates, schedules, contracts, policies, and location. Property investment in the real estate industry is specifically affected by microanalysis of property features such as price; size; number of bathrooms, bedrooms, number car spaces and/or garages; internal property features, eg. alarm, polished timber floors. In most cases, the property value will increase over time as the development of the surrounding environment of the respective property will contribute to the price of the property.

Investor's goals and limitations influence the measurement of risk factor ranking and weight, and the possible outcomes, or probabilities, will differ from each other. An example is the goal is to invest in either residential or commercial property. If the property investment is intended for commercial use, the investor might plan to rent the property for profit. However, if the investment is for residential property, then the ranking and the weight of the risk factors affected would be different. This is due to the ability to meet the mortgage repayment would be affected by the constraint of investor's capability. The investor's character, for example, whether they are a risk taker or not, also will influence the results of property investment risk analysis. To include this factor in the property investment risk analysis, a personalization criterion is needed.

A great deal of literature and research have been undertaken related to investment in large real estate projects, and various techniques have been deployed to measure risk analysis. The most popular technique to rank and weight risk factors for risk analysis in the real estate industry is an analytical hierarchical process (AHP), developed by Saaty in 1980. However, no risk analysis for property investment to date includes personalization criteria for individual investment. According to the analysis of current researchers, there has not been any attempt to include the personalization element to rank and weight the risk factor for property investment risk analysis. This section aims to redress this lack and discusses in detail the personalization model proposed.

4.3.2 Personalization Model

Property investment in the real estate industry entails high cost and high risk but provides high yield for return on investment. Risk factors in the real estate industry are mostly uncertain and change dynamically with the surrounding developments. There are many existing risk analysis tools or techniques that help investors to find better solutions. Most techniques available refer to expert's opinions in ranking and weighting the risk factors. As a result, they create misinterpretation and varying judgments from the experts. In addition, investment purposes differ between investors for both commercial and residential properties. There is therefore a need for personalization elements to enable investors to interact with the analysis. This chapter presents a personalised risk analysis model that enables investors to analyse the risk of their property investments and make correct decisions. The model has three main components: investor, decision support technologies, and the data. Real world data from the Australian real estate industry is used to validate the proposed model.

Risk analysis is a popular and useful method and tool that enables investors to make decisions on their investments. Property investment involves various risk factors including policy, environment, management, technical issues, schedule, contractual issues, location and finance. The current methods used to solve these issues are Delphi, brainstorming, fault tree analysis and strengths, weaknesses, opportunities and threats analysis (N. Piyatrapoomi et al., 2004). Risk analysis consists of three stages: risk identification, risk estimation and risk assessment. Risk identification is commonly used to minimise the risk of the real estate losses (Montgomery & Smith, 2009). Minimising the risk for property investment in the real estate industry is important as it involves high costs that lead to high risk, if not assessed properly.

Property investment risk analysis in the real estate industry involves decision-making under uncertainty. In the real world, there are many risks and opportunities that can be measured as qualitative or quantitative factors that are subjective to different investors. Some examples of risk factors in the real estate industry include government policies, political risk, social risk, regulatory risk and contract risk. The uncertainty of the risk factor of property investment in the real estate industry will affect the risk analysis results. Existing risk analysis methods do not take into account the personalization criteria for decision-making and most refers to expert's judgment to rank and weigh the risk factors. This is due to investors not having enough information about which property to invest in, given a set of constraints and goals.

Moreover, experts in the field know more about the surrounding environment of the property for investment. Main issue related to the application of experts or professional judgment is that their judgments were not aligned and create misinterpretations. The expert's interest must be aligned with the investors to gain trust and achieve the investor's goals and requirements. Therefore, personalization criterion is needed for property investment risk analysis to achieve an investor's goals and requirements by using decision support technology. The decision support tool or technology helps to provide knowledge and process the input from both investors and data stored in the system.

The application of decision support technology to provide explicit knowledge to investors for property investment risk analysis will be more accurate and trustworthy when compared to an approach that refers to expert's judgment or opinion. However, the knowledge transferred from the system to the investors is fully dependent on data availability and completeness. Again, it depends on the investors' understanding and their level of knowledge to achieve a better result for their investment. Knowledge management such as learner's knowledge, learning material knowledge and learning process knowledge is used to enable personalization (C. Fang & Marle, 2012).

Therefore, it is important to have a personalization model that will provide guidelines for the investor to achieve their goals, mitigate risk and, at the same time, gain benefits from the investment made. The decision made for property investment risk analysis must achieve investor's goals and match with their limitations/constraints. Property investment in the real estate industry normally focuses on individual or single user requirements either for residential or commercial types of investments.

Personalization of risk factor ranking, and measurement will be based on individual requirements. Personalization, as a significant capability to maximize the effectiveness of decision support systems, provides useful information to support individuals' decision-making processes. Individual investors, as the decision makers, interact with the system through personalization. The personalization technique proposed is important to ensure the comprehensive feasibility studies of the risk factors are parallel with investor's limitations/constraints. The significance of the proposed model is the user's ability to interact and be involved in setting their limitations and requirements. Investors rank and weight the risk factor based on their requirements to achieve an optimal decision. This model can be applied to other applications that involve an unstructured decision-making process.

4.3.3 Personalized Property Investment Risk Analysis Model

A personalization model for property investment risk analysis in the real estate industry is needed to fully utilize the investor's limitation and at the same time fulfilling their goals and requirements. Different investors will have different goals and may have completely different information needs. There has been an impressive development of methods for risk analysis in the real estate industry that focus more on real estate projects, rather than individual investment analysis specifically for property investment. The personalization model proposed integrates the knowledge discovery approach with the investor's personalization to enable effective decision-making to deal with the property investment risk analysis and meet investor's requirements. A personalization model for property investment risk analysis in the real estate industry is depicted in Figure 4.4.

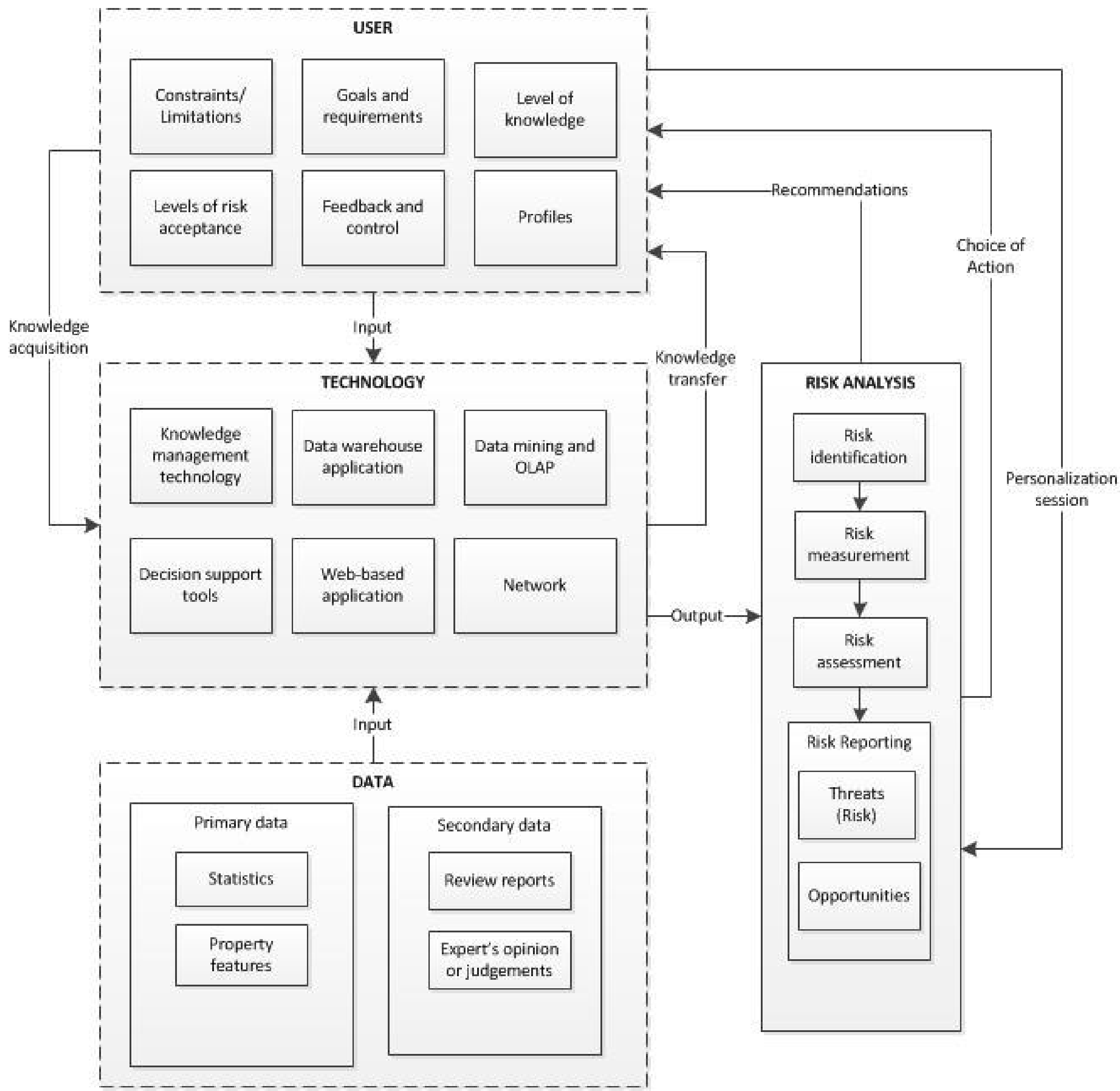


Figure 4.4: The personalized property investment risk analysis model for the real estate industry.

As shown in Figure 4.4, there are three components included in the proposed personalization model:

1) User.
2) Technology, and
3) Data.

Each of these components has its own functions and is interconnected with each of the other components to support decision-making for property investment risk analysis. Based on these three components, the user and data will generate input for the technology to process them and produce the output for the risk analysis.

The user will provide the input for the system that applies the decision support technology to process and match with the available data in the system. By providing input to the system, the user acquires knowledge that will help them to make decisions for the property investment risk analysis. The decision supports technology, or tools, processes the input and data available in the system. The data stored in the system will be processed by this technology as output in the form of information for the user to exploit as explicit knowledge. Based on the explicit knowledge, the user, as investors, will have experience based on a heuristic approach to exploit the tacit knowledge. The knowledge transfer from technology to user involves different types of decision support tools, a web-based application, a network, knowledge management technology, data warehouse application and data mining, and OLAP. The knowledge provided by the system to the user is based on the data, or input, available stored as either primary or secondary data.

The personalization model will provide the pattern of data to determine the factors that contribute to analysis of buying or selling of real estate property, and raise the questions of what, why, and when? For example, the data-driven approach will explain what factors contribute to the short time frame for the property sold. Is it because of the features of the property, location, price, type of property, type of sale, sale result, size of property for a certain period of time, or what real estate agency handles the transactions? This model presents a data-driven system and a process from data to patterns, and from patterns to applicable rules/methods for decision support. The what-if analysis through personalization criteria will determine the effects of any pattern changes made to the risk factor measurement that match the investor's requirements. The output from the technological components will be transferred to the user through the risk analysis process.

The three steps of risk analysis consist of risk identification, risk measurement and risk assessment. The risk identification will be classified based on the data available in the system that matches with the criteria provided by the user. Risk measurement involves ranking and weighting the risk factor for the investor to analyse. Risk assessment will then take place in order to achieve the results. The results, as recommendations, will be displayed as risk reporting, which will comprise the threats (risks) and opportunities for the available selections. Finally, the user (specifically, the investor) will run the personalization session with the rank and the weight of the risk factor until it meets their requirements, and they are satisfied with the results.

4.3.3.1 Components of the Personalized Property Investment Risk Analysis Model

The proposed personalized model for property investment risk analysis in the real estate industry has three main components: user, data, and technology. The user is an independent component and consists of six main sub-components: constraints or limitations, goals and requirements, level of knowledge for the decision made, feedback and control, levels of risk acceptance, and their profiles. The constraints or limitations refer to capital or budget available for the investment and the time period for return on investment. A personalization criterion helps to fully utilize their constraints and achieve better results. The user goals and requirements are all different and personalization is needed to meet their goals.

The level of the investor's knowledge will also affect the result of the decision made, so what-if analysis through personalization will help to provide them with valuable knowledge to make the right choices based on data available in the system. The explicit knowledge transfer from the system to the user as tacit knowledge will help the investor to provide feedback and maintain control of the what-if analysis to enable them to understand and assess the decision-making process. Moreover, the level of risk acceptance will also impact the input given for the personalization process as different investors will have different levels of risk acceptance. Different investors will have different profiles because they have different priorities and different risk tolerances that will reflect the risk analysis. Different investors will also define their goals and explicit knowledge transfer from the system in different ways. Each of these components is related and will impact the level of risk measurement. For example, an investor profile is a reflection of an investor's goals and objectives.

The technology component is dependent on the user and data for processing and consists of six main sub-components: knowledge management technology, data mining and OLAP technologies, data warehouse application, decision support tools, web-based application, and network. All of these technologies are integrated to process the input from users and match their inquiries with available data stored in the system. The technology refers to the tools used to process the input from the user and stored primary or secondary data for the analysis.

The knowledge management technology applied in the personalization model is a vital sub-component because it will support new strategies, processes, methods and techniques to better disseminate and apply the best knowledge at any time and in anyplace. The web application to support decision-making will be used to disseminate the knowledge. Data warehouse application, specifically the data mining technology, is applied to mine the data and provide hidden knowledge for investors to analyse. For example, data mining operations include link analysis (association), predictive modelling (regression), database segmentation (clustering) and deviation detection (visualization and statistics) (Y.-M. Fang et al., 2009).

All these techniques will provide an output or explicit knowledge for investors to personalize the risk factor analysis. OLAP, which applies the multi-dimensional data model, will enable investors to analyse data with more than two dimensions. The network is important to ensure that the data travel is input for the technology to be processed and transferred to the user. The personalization session integrates all six main components to provide the best result by using the investor's limitation to meet their requirements. The technology collects, gathers and prepares data for analysis to build predictive models and make recommendations for investors to analyse.

The data can be categorized into two types, namely primary data and secondary data. Primary data includes statistics and property features, while secondary data includes review reports and expert's opinions. It is important to ensure that the data is valid and of a high dimension for accurate results and analysis. Features of data that provide accurate analysis should be up to date, standardized, integrated and include historical information for better analysis. As with the user, the data component is independent and gathered from different types of sources. Accuracy of prediction is dependent on the data available and stored in the system. The data itself must be correct, valid and integrated.

4.3.3.2 Main activities of the Personalized Property Investment Risk Analysis Model

Five main activities involved with the personalization session are: input, process, output, feedback and control. Two main sources of input for the personalization model include input from the user and input from data collected and stored in the system. The input gathered from the user as knowledge acquisition for the system consists of the six sub-components of the user, discussed earlier. Input from data storage will be used to match the user's criteria. The data collected and stored in the system uses a web-based application. The input gathered from users and data stored in the system provides the technology to process and produce the output.

The technology is then applied to process the input and match with the data available in the system to produce the output to be given to the user as risk reporting. Processing of data and input from users is dependent on the technology applied to solve the problem. For example, data mining technology uses the data to build a data mining model and produce a hidden pattern of data as an output for users, while the OLAP technology will produce aggregate information. The output of the system as explicit knowledge will be displayed in the form of recommendations and choice of actions to the user as tacit knowledge, as shown in Figure 4.5.

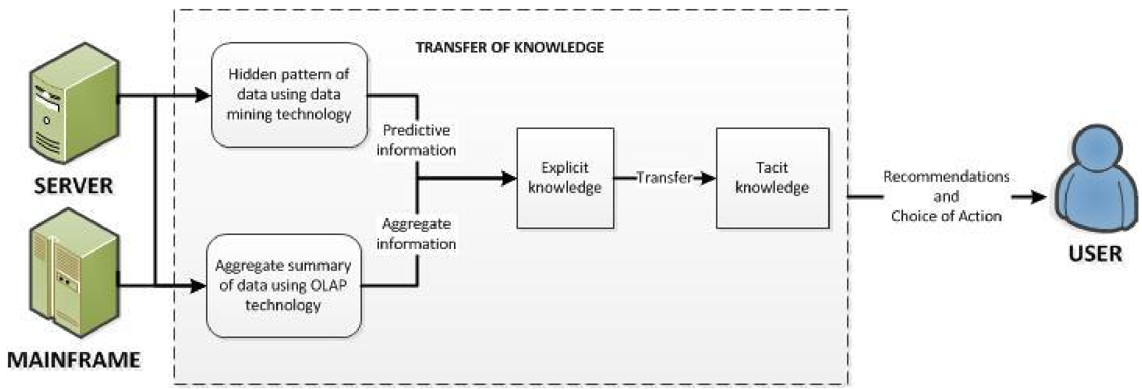


Figure 4.5: The transfer of explicit knowledge to the user as tacit knowledge.

The user will then provide feedback through a personalization session for the choice of action presented by the system as the output of risk reporting. The what-if analysis will be applied to personalize the rank and weight of the risk factors that meet with the investor's requirements. The user will have control of the degree level of the risk factor in order to achieve user goals, with available budget as a limitation.

4.3.4 Personalization Session for Risk Analysis

Personalization helps to meet the user's goals and fully utilize their constraints. Their profiles and preferences will be matched with data stored in the system. The personalization session deals with the justification of risk factor weight and ranking, which is based on historical data-driven by decision support using the knowledge discovery approach. The investor's personalised recommendations and choice of action are provided, fulfilling their requirements. The personalization session starts with the user identifying their goals and limitations for processing by the technology, as shown in Figure 4.6.

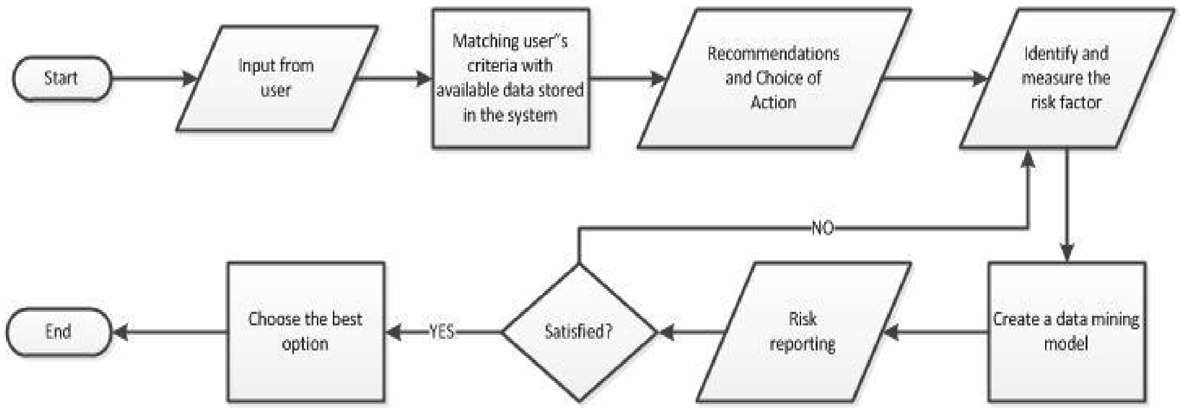


Figure 4.6: Data flow of the personalization session for property investment risk analysis in the real estate industry.

The system will then process the input by matching the user's criteria with available data stored in the system. The system produces the output as recommendations and choice of action to be analysed by the user, who will identify the risk factor for the risk analysis. The user will provide a degree level of the risk factor by rank and weight of each risk factor, which is based on the percentage; different users will definitely provide different measurements. Based on the input given by the user, the system will process the data by creating a data mining model and producing the risk report, which consists of the threats or risk, and opportunity. The predictive modelling of the data mining category – for example, the decision table technique – is applied to discover the pattern of data to support the decision-making process. After that, the user will analyse the results displayed and, if satisfied with the analysis, will finally make a decision.

4.4 Personalized Multidimensional-Sensitivity Analysis Method

Sensitivity analysis is defined as the study to understand uncertainty in the output of a model due to different sources of uncertainty in the input model (Apostolakis, 1995; Borgonovo 2007; Saltelli 2002). In general, sensitivity analysis is a very simple idea, which incorporates two simple tasks i.e. changes the input parameters and observes the impacts to the model. In sensitivity analysis, three important things need to be clearly understood prior to analyse the system sensitivity, i.e. 1) what to be varies in the determinants as risk factors; 2) what to be observed in the list of parameters given; and 3) what experimental design to be performed for sensitivity analysis. It is used to evaluate system quantitative parameters after modifying the input and can determine how sensitive system parameters to the change of the weight given (Ferdous et al 2007) The sensitivity analysis caters the diversifying investment portfolio of investors in order to achieve their goals and optimal solution that meets with investor's requirements. The result of the sensitivity analysis can be used to support decision-making, to ease communication among decision makers (Pannell 1997). This is the most useful and most widely analyzed technique used by decision makers to achieve optimal solutions (Huang & Chang 2007).

4.4.1 Development of Knowledge Using Data Mining Technology

Knowledge management for investment risk analysis in the real estate industry is an important field that needs to be focused on as it involves high cost and high risk. An investor needs to know the vital process involved as it is dealing with different kinds of information either structured, semi-structured and unstructured. The information gathered from the risk analysis must be reliable and can be trusted. The application of the deterministic approach for disseminating the knowledge by discovering the hidden pattern of data using the data mining technique is more reliable as it refers to valid data stored in the data warehouse. This section will explain in detail the development of knowledge management for investment risk analysis.

It is important to think carefully about how to gain knowledge for investment risk analysis. The decision support tools and technologies could help to minimize the risk impacts since investments in the real estate industry involves high capital. There are several questions that need to be asked at the initial step. Figure 4.7 depicts the several questions that the investors would face when dealing with investment risk analysis in the real estate industry.

Figure 4.7: Several questions that investors will face when dealing with investment risk analysis in the real estate industry

As shown in Figure 4.7, the 'Wh' questions namely what, where, when, why, how and who are the common types of issues that the decision maker would deal with when trying to figure out an idea of how to start with the analysis. Example of questions that would be asked when dealing with the investment risk analysis as follows:

- What method to use?
- What features of property to choose for the investment such as how many bedrooms, bathroom and car park?
- How to run the investment risk analysis?
- Why do you need to have knowledge in the field?
- Who to refer?
- When you need to run the investment risk analysis?
- Where to get the knowledge about the investment risk analysis?

In order to answer these questions, the investor must have in-depth knowledge to decide on investment in the real estate industry incurred with high cost and high risk. Based on the several questions highlights above, it is very important to have the correct method, tools and technologies to handle this situation. A reliable and accurate decision support tool and technology is needed in determining the rank and weight of risk factors for investment in the real estate industry. The application of the deterministic approach as proposed in this paper helps the investor to gain knowledge and answer the questions highlighted.

4.4.2 Investment Risk Analysis Knowledge Management Development

The development of knowledge management for investment risk analysis proposed in this chapter is focused on knowledge embedded in individuals, specifically the investor. The transfer of knowledge here refers to the explicit knowledge generated by the system transferred to the investor as tacit knowledge. The tacit knowledge transferred as an experience for the investor to understand and evaluate the risk factors in the decision-making process. The development of knowledge management by using heuristic through deterministic approach for investment risk analysis in the real estate industry consists of three different parts as illustrated in Figure 4.8.

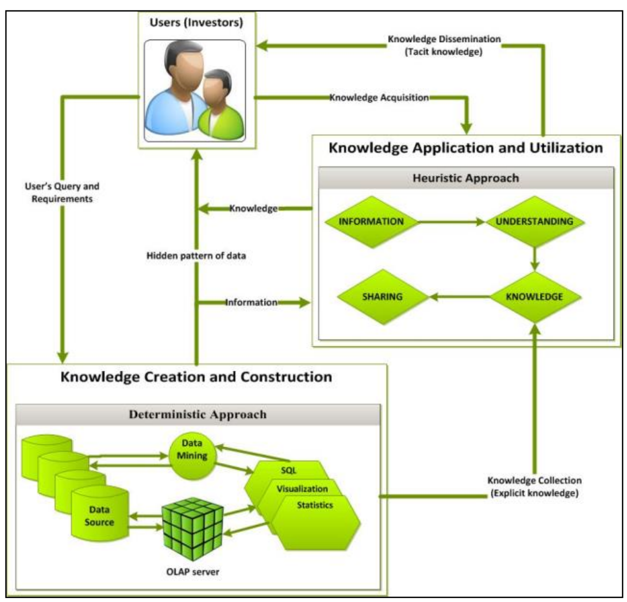


Figure 4.8: Development of knowledge management by using heuristic through deterministic approach for investment risk analysis in the real estate industry

The first element is the investor as the decision maker that will have problems and need in-depth knowledge to solve their problems. Second, the deterministic approach that processes the data based on the investor's requirements and provides knowledge by showing hidden patterns of data using data mining techniques. Third, the application of the heuristic approach in which the investors will gain knowledge through a deterministic approach to solve their problems that meets their requirements.

The process begins with the investor will set their query and requirements to the decision support technology. The user specifically the investor needs to define their goals, limitations and requirements as a query to the system that would be processed using a deterministic approach. Next, the result will be generated by the system that matches with the investor's requirement. The deterministic approach will process the input gathered from the investor to prepare the knowledge related to investor's requirements and limitations. By using a deterministic approach, the transfer from explicit knowledge to tacit knowledge refers to the experience that the investor will get to analyze the risk analysis for investment. Data warehouse end user applications such as data mining, online analytical processing (OLAP), structured query language (SQL), visualization and statistics will create and construct the knowledge as explicit knowledge. For instance, the data mining techniques such as association and prediction techniques will try to find the hidden pattern of data and make forecasts of house price as knowledge collection. Based on these results, the investor will receive, understand and analyze this information heuristically as tacit knowledge. The heuristic approach refers to the knowledge acquisition, application and utilization in making decisions. The investor will rank and weigh the risk factor analysis based on hidden knowledge generated by the system and map it with their requirements for the investment risk analysis.

The evolution of knowledge management applications for stage 5 (future age) important activities is to support business intelligence (Germanakos et al., 2007). The paper moved towards this milestone in which the development of knowledge management proposed comprehension with the personalization technique. The application of the deterministic approach solved the problem of fulfilling the investor demand better than referring to the experts in the field. The main contributions of this chapter are (1) it proposes a new technique to produce explicit knowledge through deterministic approach through personalization model; (2) it proposes the acquisition of tacit knowledge through heuristic approach; (3) it proposes a novel knowledge management development for investment risk analysis in the real estate industry by using heuristic through deterministic approach; (4) risk measurement of ranking and weighting the risk factors personalized by investors that meets with their requirements. The application of decision support technology for deterministic approach helps to speed up the transfer of knowledge hence faster decision can be made. Moreover, the personalization technique applied helps the investor to achieve their goals, within their limitations and fulfil their requirements.

4.4.3 Experiment

The objective of this experiment was to develop markedly improved risk prediction models for property investment. There are many factors that will affect the decision on property investment which can be categorized into controllable and uncontrollable factors and each factor contributes a different level of risk which varies among different investors. Not only that, different investors also will have different goals and limitations. A sensitivity analysis takes into account several factors that might impact property's investment decision and computes how a change in each factor would affect the results. This is valuable since it considers factors that are uncontrollable such as the property price in which permitting you to sort out alternatives to maximize profits.

According to Westwood Net Lease Advisors (2020), a sensitivity analysis will help investors to answer questions like:

- What is the lowest rent can I charge and still turn a profit?
- What would my net operating income be if I bought the property for $X and charged $Y for rent?
- At what vacancy rate will I start losing money?

- I want to buy the property for not more than \$X; what would the cap rate and net operating income be in such a case?
- What is my debt service likely to be if I paid \$X for the property with \$Y interest?
- Will I have enough to pay my debt service if my vacancy rate is \$X?
- At what point will the property start losing money?

Thus, this experiment explores the sensitivity analysis by deciding what variables or dimensions that will affect most of the total number of houses owned using select attributes function using similar dataset discussed in Chapter 3. However, for this section, respondents who did not own a house were removed from the dataset to increase accuracy of results. As shown in Figure 4.9, the relationships between dimensions toward the class (output) namely total numbers of houses owned dominated by houses owned equal to 1 which represented in blue color.

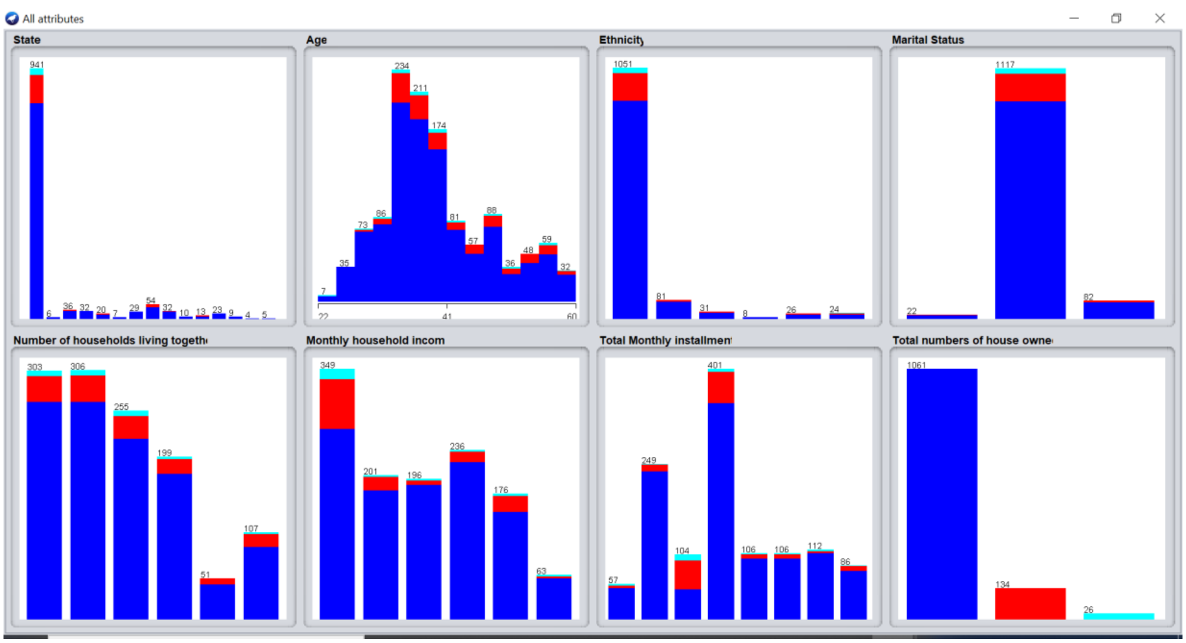


Figure 4.9: Relationships between dimensions toward the class (output) namely total numbers of houses owned

For feature selection, InformationGainAttributeEval method used to identify determinants or factors that affect the output namely the total number of houses owned. The result of the information gain attribute evaluation depicted in Figure 4.10 and Table 4.1 respectively.

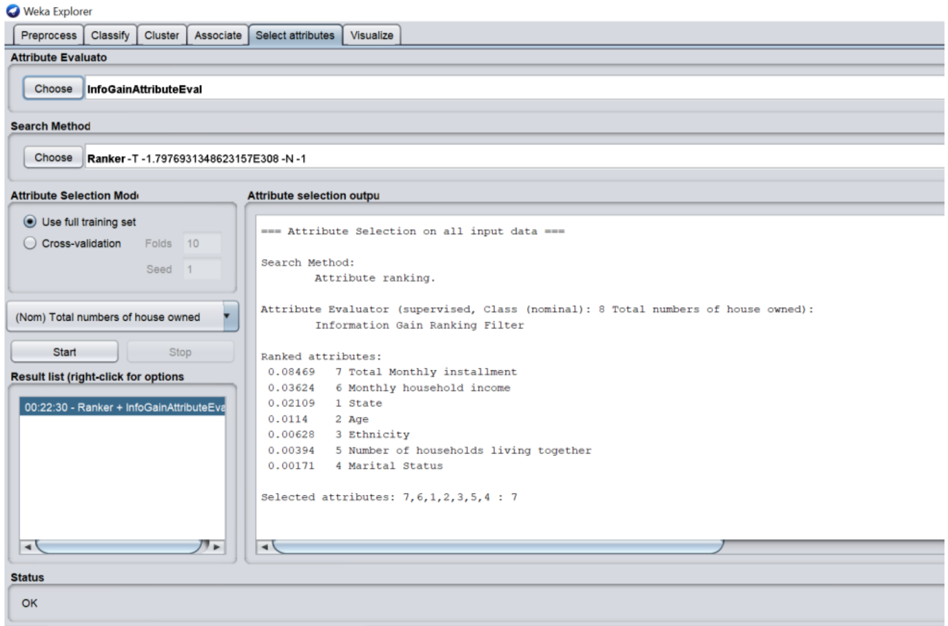


Figure 4.10 : Ranking of attributes with respect to InformationGain Attribute evaluation method

Table 4-1: Ranking of attributes with respect to InformationGain Attribute evaluation method

No	Attribute	Ranker Value
1	Total Monthly instalments	0.08469
2	Monthly household income	0.03624
3	State	0.02109
4	Age	0.0114
5	Ethnicity	0.00628
6	Number of households living together	0.00394
7	Marital Status	0.00171

Plot analysis based on the dataset shown in Table 4.2 tabulated in Figure 4.11 to visualize the trend analysis of total number of houses owned. From the Figure, it shows that the most highlighted area is referring to the monthly instalment with the value of RM1,000 – RM1,999 and total household income value more than RM6000 with the total number of respondents equal to 147.

Table 4-2: Crosstab analysis based on total number of monthly instalment and monthly household income towards total number of houses owned.

Count of Total numbers of house owned	Total Number of Monthly Instalments								
Monthly Household Income	0	Below RM599	More than RM2,000	RM1,000 - RM1,999	RM600 - RM699	RM700 - RM799	RM800 - RM899	RM900 - RM999	Total
Below RM1,999	3	19	1	11	14	5	5	5	63
More than RM6,000	11	30	72	147	12	27	25	25	349
RM2,000 - RM2,999	22	63		32	28	20	19	12	196
RM3,000 - RM3,999	10	60	1	65	29	25	33	13	236
RM4,000 - RM4,999	9	44	10	74	15	14	17	18	201
RM5,000 - RM5,999	2	33	20	72	8	15	13	13	176
Grand Total	**57**	**249**	**104**	**401**	**106**	**106**	**112**	**86**	**1221**

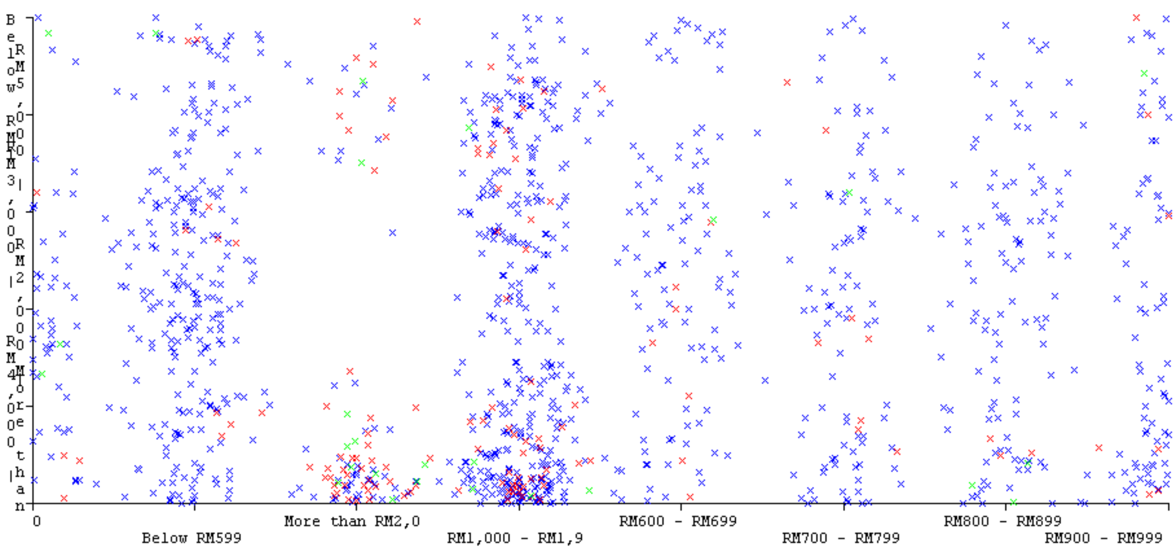


Figure 4.11: Pattern analysis for the relationship between monthly instalments, household income towards the total number of houses owned

As conclusions, sensitivity analysis has become an integral part of property investment risk analysis to identify sensitive factors that will affect the investor's decision. The sensitivity analysis is carried out to determine how changes in cost and revenue items affect the cash flow estimate in property development and investment. Studying the overall expected outcome to diversify the changes in input variables will affect the identification of key variables or determinants of property investment risk analysis.

4.5 APPIRA Process Flow

The adaptive personalization approach gathers inputs from the investors to be manipulated in the system and the data mining approach produces patterns of data also known as knowledge discovery from the database. Figure 4.12 depicts how the APPIRA method integrates these two approaches.

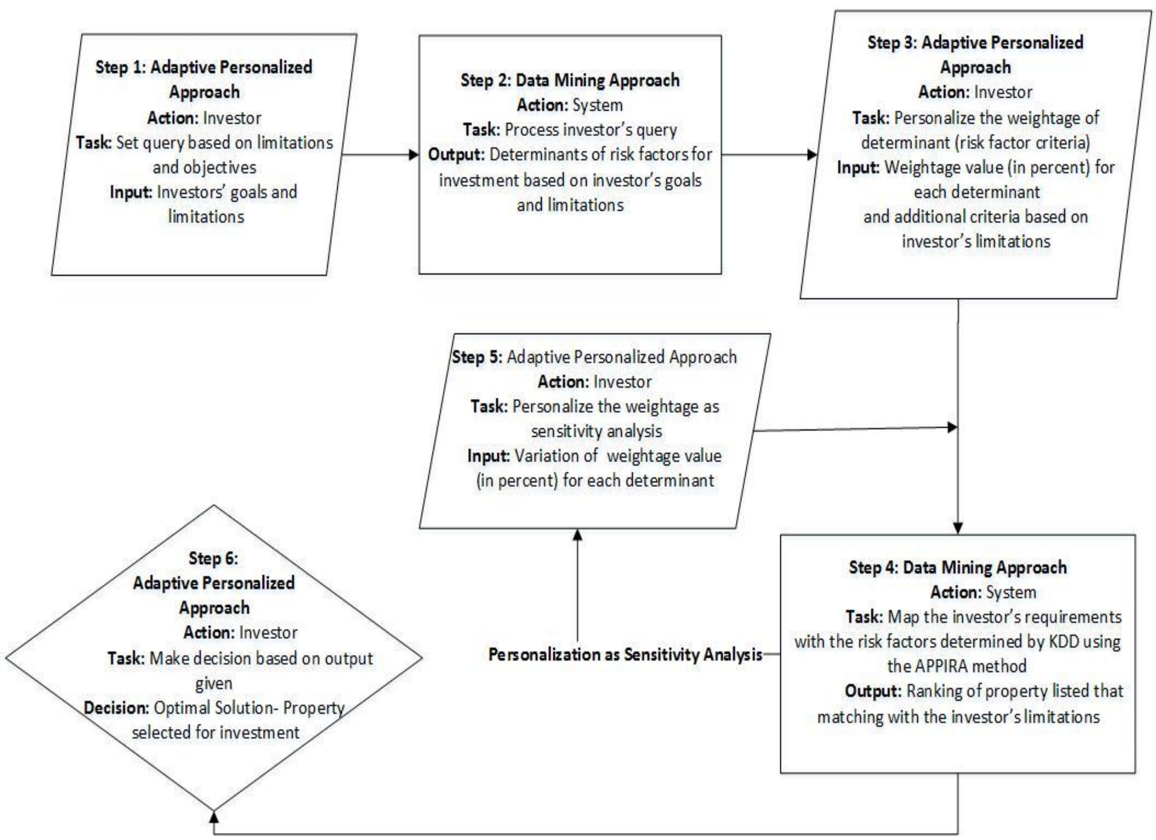


Figure 4.12: The process flow of the APPIRA method

Figure 4.12 shows steps of the APPIRA method where the investor sets their goals and objectives. Once the investors have set their risk criteria and its weight, the system will map the criteria with data available in the database. However, mapping the input gathered from the investors' depending on the availability of the quality of the data stored in the database. Some properties might not have values for certain attributes that will limit the results displayed. Thus, the APPIRA method is dependent on the quality and validity of data available in the system to achieve optimal decision-making. Moreover, data mining crawls for information related to the property mapping with the investor's criteria and spatial analysis needed to view hidden patterns or knowledge.

The APPIRA method maps the determined risk factors based on output gathered from data mining operations in order to find a pattern of data which is also knowledge data discovery using criteria personalized by investors. It helps investors achieve their goals and objectives by gathering information as hidden knowledge through a heuristic approach and analytic approach, based on patterns of data from the system identified through a deterministic approach as shown in Step 4. Investors gain hidden knowledge embedded in action when personalizing the risk level through a heuristic/analytic approach. The pattern of data from the system based on the data mining approach produces explicit knowledge. The principle of the APPIRA method relies on a weight for the personal risk level set by investors. Adaptive personalization is the key to these proposed methods and based on the investors' tacit knowledge gained from the combination of both the proposed APPIRA method and the proposed personalized sensitivity analysis methods.

4.5.1 Six Steps of APPIRA

The APPIRA method adopts the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) method which determines the best solution from a set of alternatives with certain attributes as shown below. TOPSIS is widely used in various multi-attribute decision-making problems such as supply chain logistics, marketing management, environmental management or chemical engineering (Yadav, Karmakar, Kalbar, & Dikshit, 2019). The algorithm of APPIRA method for dynamic risk prediction comprises six steps explained in details steps describe as follows:

Step 1: Obtain investor's goals and limitations (adaptive personalization approach)

The investor sets their goals and sends corresponding information as queries to the APPIRA system for processing. Different investors will have different sets of criteria as risk factor and different value of parameters given for each criterion selected. This personalization technique helps the decision maker to achieve their goals given limitations which vary among the investors.

Step 2: Generate results (output – property investment determinant) based on the investor's query (criteria-risk factor) using a data-driven approach

The APPIRA system generates initial results that meet the investor's search criteria. These results are generated from an analysis that uses association rules to show patterns of data for a given set of attributes as shown in the case study chapter. The data mining techniques help to identify the determinants of risk factors as automated tools for investors to personalize based on their goals and limitations. This association technique identifies unforeseen factors that may affect the investor's decision-making process. For example, the results might classify properties according to their related attributes, such as a suburb profile which includes transport to work; occupation; education level; and country of birth. These association rules are integrated with the APPIRA method to help investors evaluate investment properties by finding relationships between the property's characteristics and market value. The data mining techniques help to identify the determinants of risk factors as automated tools for investors to personalize based on their goals and limitations. Comparing the evaluations of properties with many features and specifications would be time-consuming if done manually and may result in an unclear recommendation. As compared to analysis which is done manually such as referring to experts' judgment for analysis takes longer and is more expensive as you have to pay for the expert services. The more information you requested the more cost involved. Moreover, the outputs from the expertise were depending on their level of knowledge in the field and availability to consult. Thus, this method helps to save time, quicker analysis and save money for the decision-making process. Furthermore, the results or output given is more structured and subject oriented.

Examples of situations for analysis include: a) association rules – find all frequently rented houses with two bedrooms; b) classification – find all houses with a rental rate below $400 per week located in Botany Bay; and c) clustering – identify houses with similar additional features. Hidden patterns of data stored in the database will be revealed based on the parameter (property features) given by the investor. The identification of determinant in the selected area will be revealed through a hidden pattern or knowledge based on a dataset available in the system. Thus, it is important to ensure that the system keeps up-to-date and valid data.

Step 3: Weight and rank risk factors (determinant) based on the output from the system (adaptive personalization approach)

The investor weights and ranks risk factors based on their investment goals. The number of criteria provided by the investor determines the number of fields that need to be analyzed by the APPIRA system. The investor then weights and ranks the chosen criteria based on the results (determinant variable) displayed by the APPIRA system using the deterministic approach as discussed in Step 2. For example, the investor might rank the price of the property (financial risk) as the highest risk factor, followed by location (location risk), property features (design risk), and suburb profile (social risk / socioeconomic environment risk).

Step 4: Mapping weight and ranking factors with investor requirements and available data (integration between adaptive personalization approach and data-driven approach) using TOPSIS model

APPIRA system then makes a comparison between investors' weightage (heuristic approach) and results produced from the data stored in the system (deterministic approach) and maps the investor's personalization with the data extracted to the data warehouse. At this stage, extracted data processed using decision-tree-based association rules to find relevant patterns of data. Comparisons of input between investors' weightage (heuristic approach) and results produced from the data stored in the system (deterministic approach) created both before and after personalization by the investor. Results are displayed based on the query set by the investor's (personalization) mapping using the deterministic approach as in Step 3.

Step 5: Make changes to the weight and rank of the risk factors (adaptive personalization approach)

The investor can vary the weight and rank of each risk factor using the sensitivity analysis as required for adaptive personalization. This is an iterative process, similar to that of Step 3. Each adjustment the investor makes helps demonstrate the impact of specific risk factors and this process continues until the investor is satisfied that the results meet their objectives.

Given that risk factors are uncertain, and the investor's requirements may vary over time, a sensitivity analysis is used to dynamically determine the sensitivity of risk factors, to provide the most up-to-date property data at all times. Through an iterative process, similar to that of Step 4, a combination of the results from Steps 3 and 4 were compared with the investor's current objectives to achieve an optimal set of recommendations. The investor will repeat Step 3 and 4 where they will change the weight of risk factors in both Step 3 and 4 until they are satisfied with the results produced by the system. Sensitivity analysis is important to determine how different values of an independent variable affect a particular dependent variable under a given set of assumptions. Cong-cong et.al., (2019) applied the concept of preference relation, which is constructed by comparing alternatives using pre-established scales, is one of the most commonly used representation formats of decision makers' preferences to simulate the sensitivity analysis. The reciprocal preference relation (RPR) is a powerful tool to represent decision makers' preferences in decision-making problems. In some complex decision-making situations, it is evident that decision makers may be unable to describe 'the exact degree of preferences' between pairs of alternatives and they would prefer the use of interval values or more than one discrete value to express their preferences.

Step 6: Make decisions based on the analysis of results (adaptive personalization approach)

Lastly, the investor chooses an option from the results that best fulfils their requirements and mitigates their risk. The proposed APPIRA method supports both qualitative and quantitative measurements of the risk factors for risk prediction.

4.5.2 TOPSIS method

The TOPSIS method, a multi-criteria decision-making method, was proposed by Hwang and Yoon in 1981 as cited in Cong-cong et al., (2019). It provides the best alternatives which are as close as possible to the best solution. TOPSIS have been adopted and applied in different kinds of industries to solve simple to complex problems. The procedure of the TOPSIS model can be described as shown in the following steps.

Step 1: Create an evaluation matrix consisting of m alternatives and n criteria, with the intersection of each alternative and criteria given as $\tilde{x}_{ij}$, we therefore have a matrix $\left(\tilde{x}_{ij}\right)_{m*n}$

As stated above, a multi-criteria decision-making which can be concisely expressed in matrix format as

$$\widetilde{D} = \begin{bmatrix} \tilde{x}_{11} & \tilde{x}_{12} & \dots & \tilde{x}_{1n} \\ \tilde{x}_{21} & \tilde{x}_{22} & \dots & \tilde{x}_{2n} \\ \vdots & \vdots & \dots & \vdots \\ \tilde{x}_{m1} & \tilde{x}_{m2} & \dots & \tilde{x}_{mn} \end{bmatrix}$$

$$\widetilde{W} = [\widetilde{w}_1, \widetilde{w}_2, \dots, \widetilde{w}_n]$$

where $\tilde{x}_{ij}$, $\forall i, j$ and $\widetilde{w}_j$, j= 1,2,…,n are value from data mining processing using Monte Carlo simulations to identify the determinant ($\widetilde{D}$) and the weight value ($\widetilde{W}$).

Step 2: Define and normalize the decision matrix $\widetilde{D} = \left(\tilde{x}_{ij}\right)$

The importance weight of each criterion as determinant obtained using automated tools of data-driven techniques as shown in the case study chapter of this thesis. Assume that the investor A has K variables, then the importance of the criteria and the rating of alternatives (A) with respect to each criterion can be calculated as

$$\tilde{x}_{ij} = \frac{X_{ij}}{\sqrt{\sum_{j=1}^{n} X_{ij}^2}} \tag{1}$$

Different investors will have a different set of decision matrices based on their goals and limitations. Different K variables will produce a different number of alternatives (A) depending on data available for the current situation.

Step 3: Aggregate the weights to the decision matrix by making $v_{ij} = \widetilde{w}_n \tilde{x}_{ij}$.

Step 4: Define the positive ideal solution (PIS), v_j^+, and the negative ideal solution (NIS) v_j^-, for each criterion. Usually, $v_j^+ = max\{v_{ij}, \dots, v_{mj}\}$ and $v_j^- = min\{v_{ij}, \dots, v_{mj}\}$ for benefit criteria, and $v_j^+ = min\{v_{ij}, \dots, v_{mj}\}$ and $v_j^- = max\{v_{ij}, \dots, v_{mj}\}$ for cost criteria.

Step 5: Calculate the separation measures for each alternative.

$$S_J^+ = \sqrt{\sum_{j=1}^{n} \left(v_j^+ - v_{ij}\right)^2},\ i = 1,2,\dots,m \tag{2}$$

$$S_J^- = \sqrt{\sum_{j=1}^{n} \left(v_j^- - v_{ij}\right)^2},\ i = 1,2,\dots,m \tag{3}$$

Step 6: Calculate the closeness coefficients to the ideal solution for each alternative.

$$CC_i = \frac{S_i^-}{S_i^- + S_i^+} \tag{4}$$

Step 7: Rank the alternatives according to CC_i. The bigger CC_i is, the better alternative A_i will be.

4.5.3 Example

Suppose that the Investor A plans to invest in the property investment with the goal to rent the property with its capital limitations. The system will extract this input and identify the determinant of property for rental such as number of bedrooms, type of property, location, rental rate and number of bathrooms as the determinants for rental pattern. Different goals and limitations will produce a different number of criteria as determinant that map with investor's requirements as follows:

Investor n, $I_n = \{C_1, C_2, C_3, C_4, C_5, \dots C_n\}$

The hierarchical structure of Investor A goals and limitations shown in the following Figure 4.13.

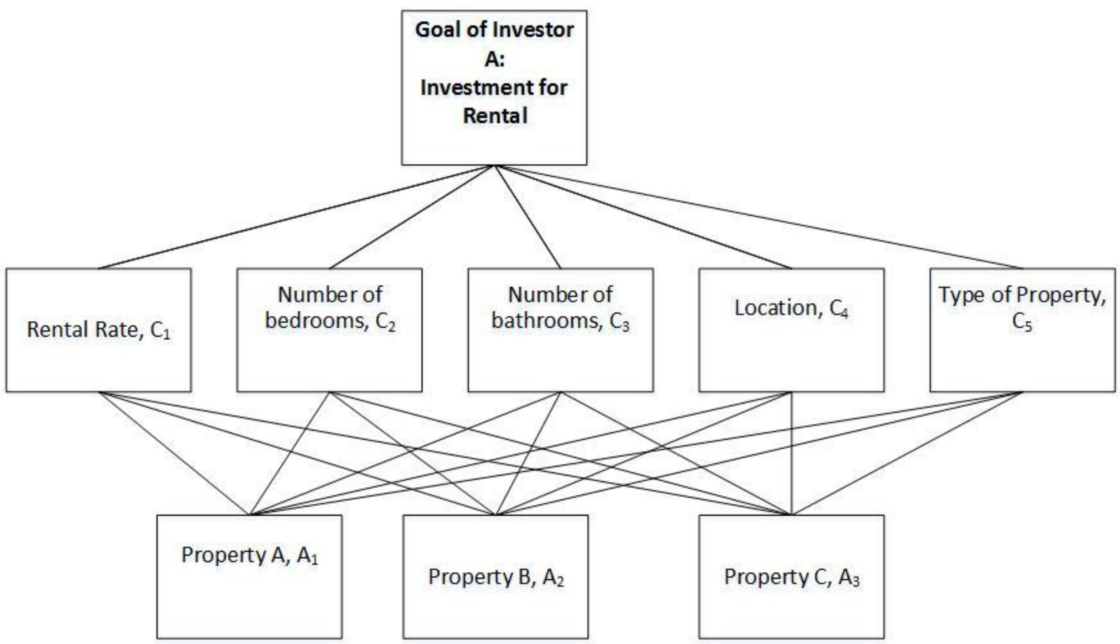


Figure 4.13: Hierarchical structure of Investor A.

Then, the investor will provide weight for each of these determinants (criteria) and mapping with their limitations such as the range price of the property for the system to mine. After processing the query, m alternatives A_1, A_2 and A_m remain for further evaluation. The investor personalized the weight value of the determinant to run sensitivity analysis until an optimal solution was achieved.

Thus, the decision matrix for Investor n, I_n as follows:

$$I_n = \begin{array}{c|c} & \begin{matrix} C_1 & C_2 & \dots & C_n \end{matrix} \\ \hline \begin{matrix} A_1 \\ A_2 \\ \vdots \\ A_m \end{matrix} & \begin{bmatrix} \tilde{x}_{11} & \tilde{x}_{12} & \dots & \tilde{x}_{1n} \\ \tilde{x}_{21} & \tilde{x}_{22} & \dots & \tilde{x}_{2n} \\ \vdots & \vdots & \dots & \vdots \\ \tilde{x}_{m1} & \tilde{x}_{m2} & \dots & \tilde{x}_{mn} \end{bmatrix} \end{array}$$

$$\widetilde{W}_n = [\widetilde{w}_1, \widetilde{w}_2, \dots, \widetilde{w}_n]$$

Where $\widetilde{W}_n$ refers to personalized weightage value given by the investors based on their requirements.

$$\begin{matrix} A_1 \\ A_2 \\ \vdots \\ A_m \end{matrix} \begin{bmatrix} \tilde{x}_{11} & \tilde{x}_{12} & \dots & Max^{A^+}, \tilde{x}_{1n} \\ \tilde{x}_{21} & \tilde{x}_{22} & \tilde{x}_{23} & \dots \tilde{x}_{2n} \\ \vdots & \vdots & \vdots & \dots \vdots \\ Min^{A^-} \tilde{x}_{m1} & \tilde{x}_{m2} & \dots & \tilde{x}_{mn} \end{bmatrix}$$

The APPIRA method helps investors make faster and better decisions in practice.

4.5.3.1 Decision Tree using J48 Machine Learning Model

As an extension of an experiment in Chapter 3, this section applied the same dataset to classify the determinants that reflect the output or class namely the total amount of houses owned based on state, age, marital status, ethnicity, number of households living together, monthly household income and total monthly instalment.

Decision Tree is the classification technique that consists of three components: root node, branch (edge or link), and leaf node. Root represents the test condition for different attributes, the branch represents all possible outcomes that can be there in the test, and leaf nodes contain the label of the class to which it belongs. The root node is at the starting of the tree which is also called the top of the tree (Software testing help, 2020).

WEKA offers many classification algorithms for decision trees. J48 is one of the popular classification algorithms which outputs a decision tree. Using the Classify tab the user can visualize the decision tree. Figure 4.14 shows the results of classification algorithms for decision trees using the J48 machine learning model. The result shows there are 8 number of leaves with 9 trees as listed below:

1. Total Monthly installment = RM1,000 - RM1,999: 1 (415.0/69.0)
2. Total Monthly installment = More than RM2,000: 1 (107.0/59.0)
3. Total Monthly installment = 0: 0 (1164.0/57.0)
4. Total Monthly installment = RM800 - RM899: 1 (115.0/9.0)
5. Total Monthly installment = RM900 - RM999: 1 (92.0/14.0)
6. Total Monthly installment = Below RM599: 1 (278.0/40.0)
7. Total Monthly installment = RM700 - RM799: 1 (112.0/15.0)
8. Total Monthly installment = RM600 - RM699: 1 (109.0/11.0)

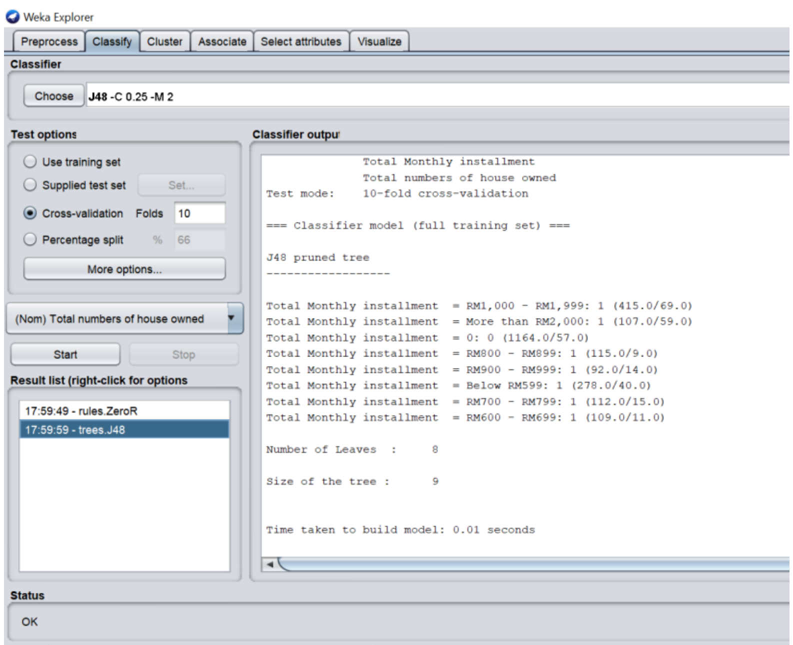


Figure 4.14: Result using J48 pruned tree machine learning model.

Analytics Vidhya (2020) interpreting the decision tree nodes as below:

- The values on the lines joining nodes represent the splitting criteria based on the values in the parent node feature
- In the leaf node:
  - The value before the parenthesis denotes the classification value

- The first value in the first parenthesis is the total number of instances from the training set in that leaf. The second value is the number of instances incorrectly classified in that leaf

The first value in the second parenthesis is the total number of instances from the pruning set in that leaf. The second value is the number of instances incorrectly classified in that leaf

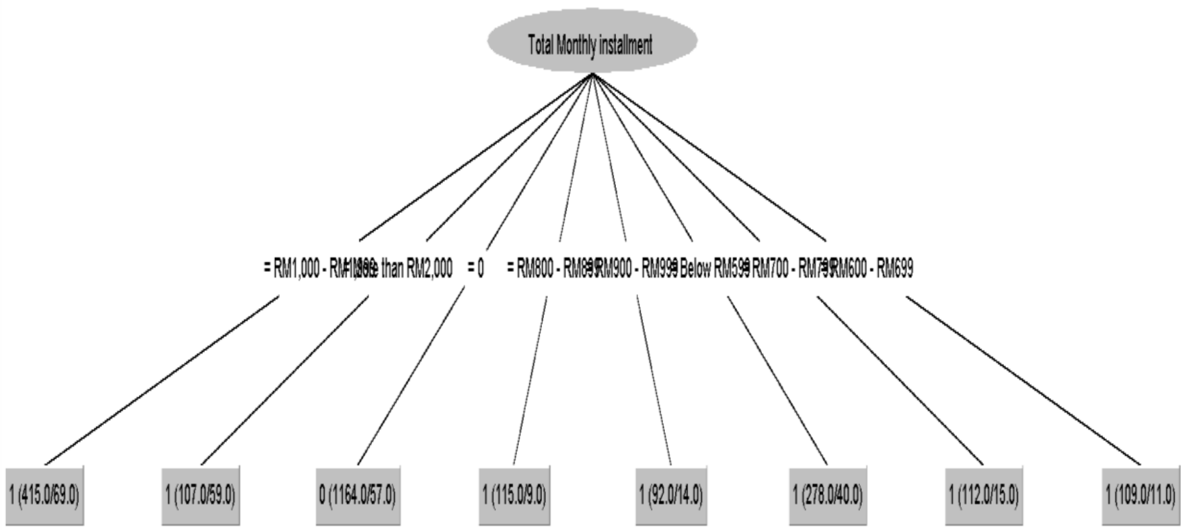


Figure 4.15: Decision tree for number of houses owned using J48 machine learning model.

As can be seen in Figure 4.15, the decision tree for the number of houses owned is affected by the total amount of monthly instalment. For all categories of total monthly instalment, the number of houses owned by the respondent equal to 1 except for RM0 instalment (1164). The leaf dominated by monthly instalment equal to RM1,000 – RM1,999 with the total number of instances from the training set in that leaf equal to 415 followed by below RM599 (278), , RM800 – RM899 (115), RM700 – RM799 (112), RM600 – RM699 (109), more than RM2,000 (107) and RM900 – RM999 (92) respectively. If the decision tree is too populated again as shown in Figure 4.16, tree pruning can be applied from the Preprocess tab by removing the attributes which are not required and start the classification process again as shown in Figure 4.16 analyzed using REPTree machine learning model. After removing not required attributes, only 4 attributes namely 1. Marital Status, 2. Monthly household income,

3. Total Monthly installment and 4. Total number of houses owned with test mode:10-fold cross-validation. The result shown in Figure 4.17.

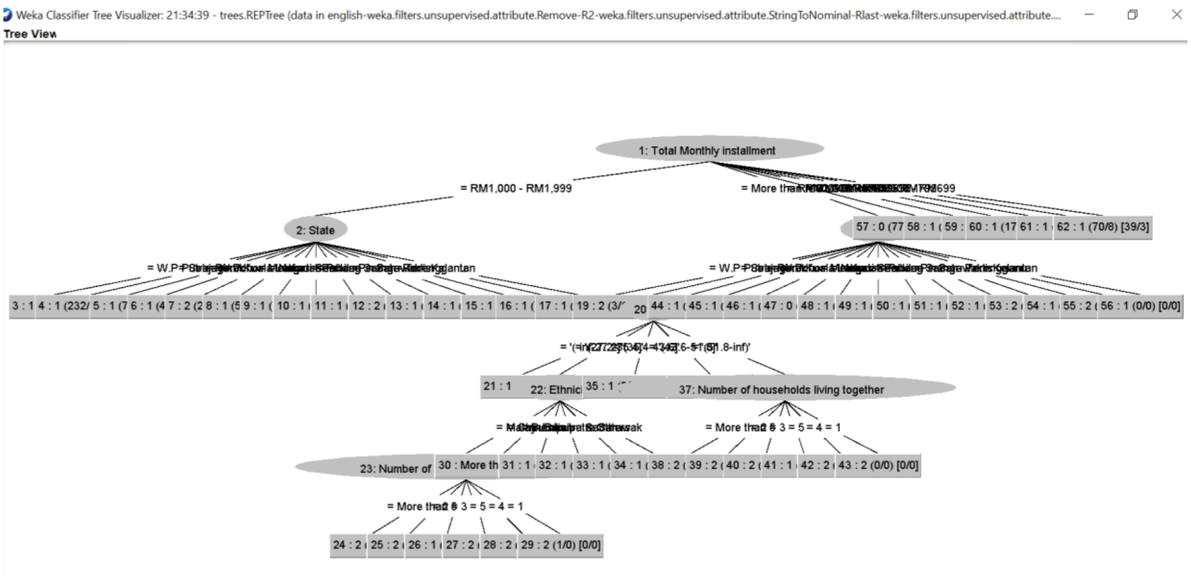


Figure 4.16: Too populated decision tree using REPTree machine learning model.

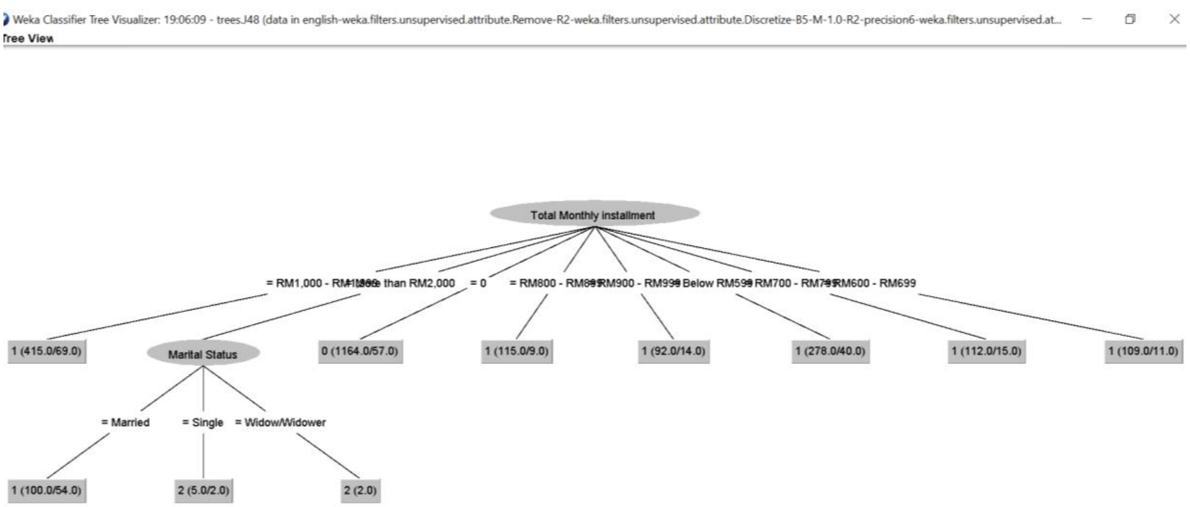


Figure 4.17: Decision tree after removing the attributes which are not required.

4.5.3.2 Attribute Selection using Gain Ratio

Application of data mining for feature subset selection is a great contribution for business analytics. In the real situation for property investment, there are many uncertainties factors that will affect the decision making made by the investors. Due to the high number of dimensions that will affect the decision, it creates difficulties in testing and training the dataset. Thus, an approach namely Gain Ratio and Correlation based feature selection have been used to illustrate the significance of feature subset selection for classifying the determinants that affect the investment alternatives as shown in Figure 4.18. The C4.5 tree uses gain ratio to determine the splits and to select the most important features.

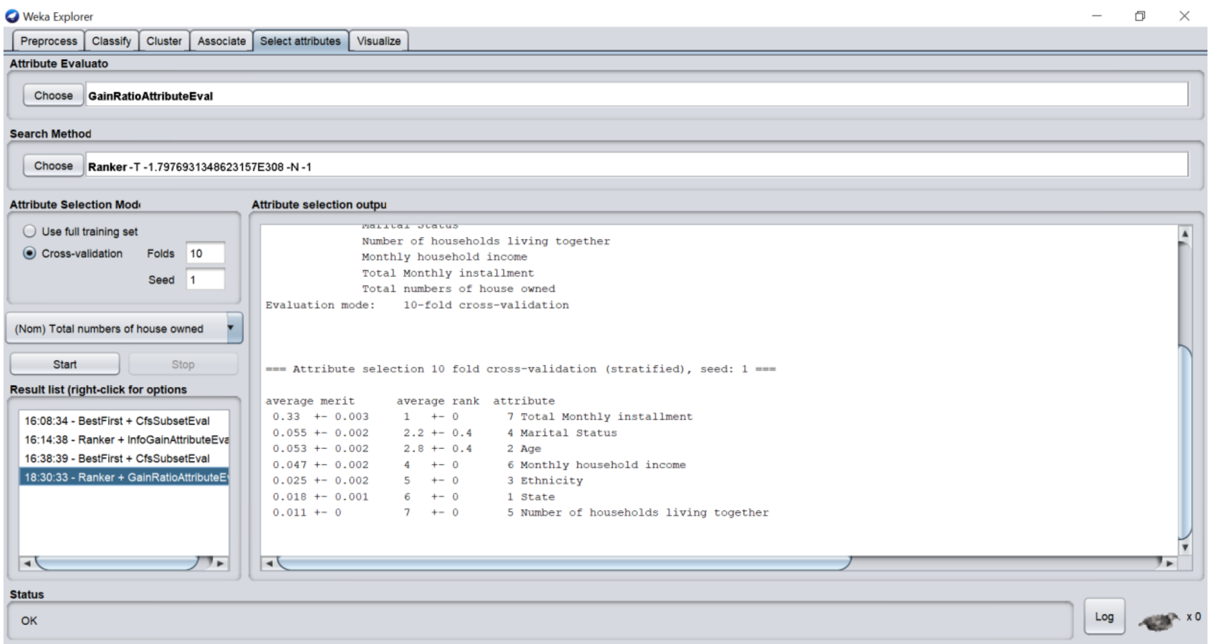


Figure 4.18: Ranking of attributes with respect to GainRatio Attribute evaluation method.

This method measures the significance of attributes with respect to target class on the basis of gain ratio. It can be calculated by the following formula,

$$\text{GainR(Class, Attribute)} = (H(\text{Class}) - H(\text{Class} \mid \text{Attribute})) / H(\text{Attribute})$$

Where H represents the Entropy. This attribute evaluator with the Ranker Searching is applied to the dataset explained in Section 3.5.1. The ranking of first five attributes are taken into accounts which are 1. Total monthly instalment, 2. Marital status, 3. Age, 4. Monthly household income and 5. Ethnicity.

The classification and features selection using machine learning model of data mining experiment conducted to highlight the application of PAM and PM-SA methods proposed.

4.6 Summary

In this chapter, we describe an adaptive personalized property investment risk analysis method to identify the determinant of risk factor, weigh the variable of property investment criteria and rank the alternatives available based on the investor's goals and limitations. We define two essential methods namely PAM and PM-SA that need to be embedded in the APPIRA method in order to achieve an optimal solution. The relationship between both development of knowledge using data mining technology and investment risk analysis knowledge management proposed depicted in Figure 4.19.

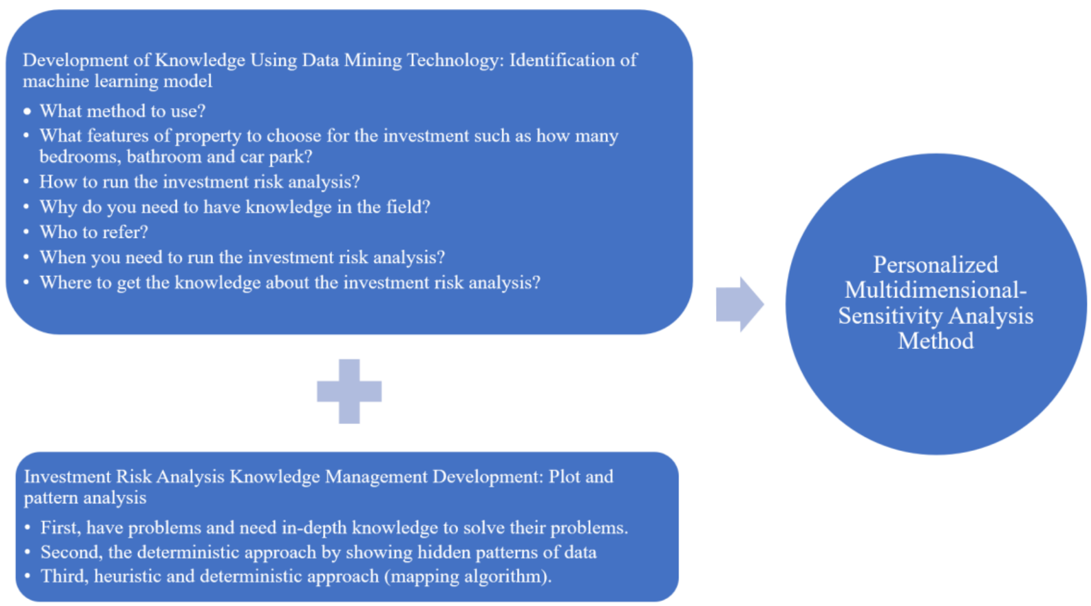


Figure 4.19: The relationship between both development of knowledge using data mining technology and investment risk analysis knowledge management

This chapter has detailed a personalization model for property investment risk analysis in the real estate industry. The main objective of proposing this model is to improve the measurement of risk factors that align with investor's goals and limitations. The proposed personalization model provides a guideline for investors to achieve their goals because the ability to accept different levels of risk varies significantly from one investor to another. The investor must prepare and specify valid, accurate information of their limitations and requirements before proceeding with the property investment risk analysis. Different goals and limitations have different values for risk factor rank and weight for property investment risk analysis in the real estate industry. It is important to have a computer system in place for personalized property investment risk analysis that achieves investor's goals, taking into account their limitations. The proposed personalization model for property investment risk analysis introduces a new perspective to investors to measure the risk analysis factor. This model will also be used as a decision support tool in property or real estate investment risk analysis. Further research on personalization algorithms is needed to evaluate more clearly, systematically and mathematically how effectively the personalization model may be applied to property investment risk analysis.

CHAPTER 5

AN ADAPTIVE PERSONALIZED PROPERTY INVESTMENT RISK ANALYSIS SYSTEM

5.1 Introduction

In this chapter, a newly developed property investment risk analysis prototype software called adaptive personalized property investment risk analysis system (APPIRAS), which introduced the adaptive personalization technique and data-driven approach to achieve investor's goals and at the same time fulfil the investor's requirements. All the necessary primary features for property investment risk analysis have been implemented in friendly graphical user interfaces. Several experiments are performed to verify the effectiveness of the developed prototype software system explained in Chapter 6.

The rest of the chapter is organized as follows. The general specifications of the APPIRAS are briefly presented in Section 5.2. Section 5.3 describes its main features and an algorithm to subjectively assess the adaptive personalized property investment risk analysis is given in Section 5.4. The main interface of APPIRAS presented in Section 5.5. Finally, the chapter is summarized in Section 5.6.

5.2 General Specification

The current version of an APPIRAS has implemented the primary features of the APPIRA method such as personalization technique, determinant identification, weight the determinant factors, and mapping with the datasets available in the system. It also has the capability for expansion and can be easily improved for web mining capabilities to add more parameters to be listed as determinants of the risk analysis. The use case diagram of APPIRAS is shown in Figure 5.1.

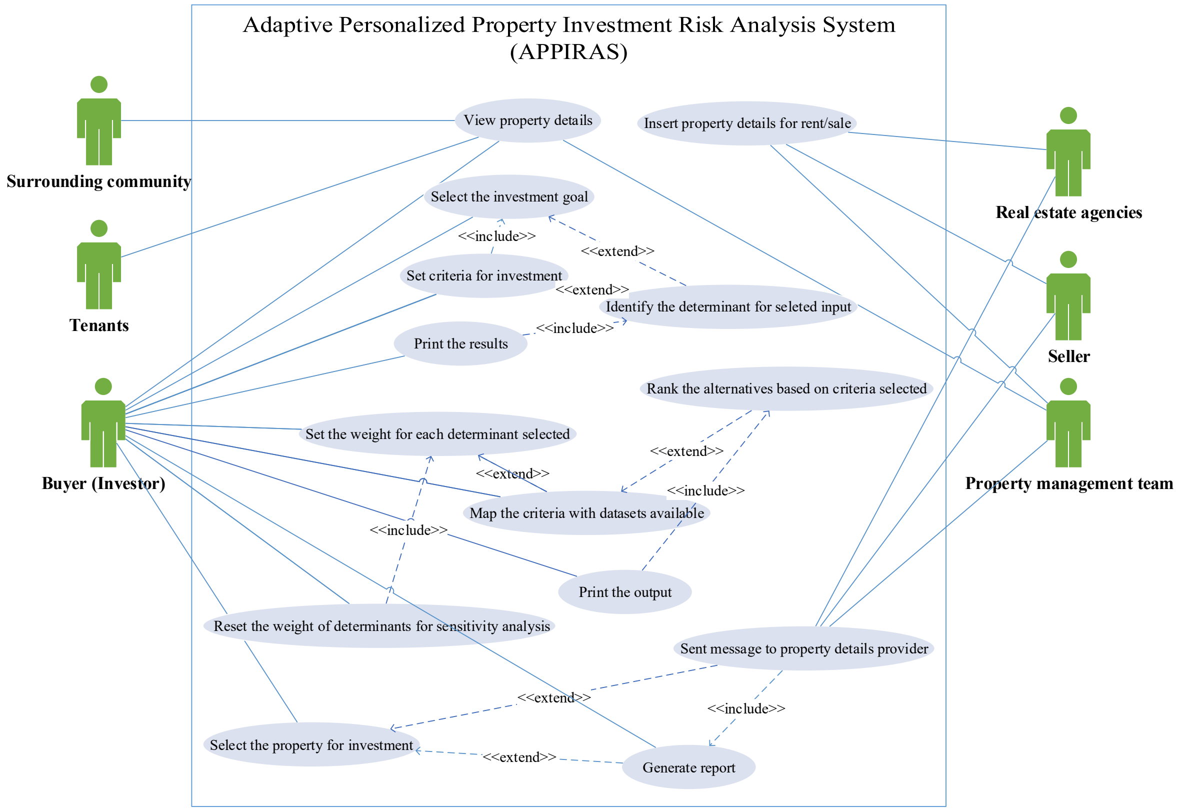


Figure 5.1: APPIRAS use case diagram.

APPIRAS provides a number of use cases to enable users of the system to conduct a variety of tasks and provide report for property investment risk analysis into three categories namely:

(1) INPUT

a. Insert property details for sale/rent
b. Select the investment goal
c. Set criteria for investment
d. Set the weight for each determinant selected
e. Reset the weight of determinants for sensitivity analysis
f. Select the property for investment

(2) ANALYSIS

a. Identify the determinant for selected input
b. Map the criteria with datasets available
c. Rank the alternatives based on criteria selected

d. Sent message to property details provider

(3) REPORT

a. View property details
b. Print the output/results for the user to view
c. Generate reports/print the results

5.3 Main Features

APPIRAS is the realization of the APPIRA frameworks and APPIRA methods explained in Chapter 3 and Chapter 4 respectively. This section briefly describes the eleven main features of APPIRAS which can be grouped into two main categories: personalization technique and data-driven approach.

5.3.1 Personalization technique

(1) To enable the investor to select the investment goal

Investment goals among investors are different and each of the goals will result in different criteria to be analysed as determinant for property investment risk analysis. Thus, input for investment goals is mandatory for the system to provide accurate results and attain the investor's requirements.

(2) To enable investor to set criteria for investment

Once the investment goals have been set, the investor also needs to set other criteria to be analysed for the property investment such as number of bedrooms, number of parking, number of bathrooms and location of the property.

(3) To enable investor to set the weight for each determinant selected

Once the determinants have been identified by the system, the investor needs to measure or weigh the determinant selected for mapping purposes. This is an imperative task to ensure the options selected align with investor's goals and limitations.

(4) To enable investor to reset the weight for sensitivity analysis

Sensitivity analysis will be conducted to see the variation of results by resetting the weight of determinant for selected property investment risk analysis goal. Different weight values given will result in different alternatives of listed property to be ranked and selected in order to achieve optimal decision-making.

(5) To enable the investor to select the property for investment

The investor will select the best property for investment based on ranking measured using the TOPSIS approach.

(6) To enable investor to contact the property management selected for investment

The system will provide information about property owners to be contacted for further action needed.

5.3.2 Data-driven approach

(1) To gather input of property details for rent/sale from real estate stakeholders

APPIRAS will act as a platform or marketplaces for all property investment stakeholders to get information about property details. Property owners or property management teams can post their property details for investment in this system which will be one of their tools for digital marketing.

(2) To provide information about property details for users to view

The information available in the system depending on the number of users specifically refers to the property owner, property management team, real estate agencies and developers that provide input to the system to be available for viewing. Registered users view the data based on their user roles either as seller, property management team, tenant, investor, real estate agencies and surrounding community. Each of these users have different modules as users depending on their roles when they register into the system.

(3) To identify the determinants of property investment risk analysis based on investor's criteria/goal

Identification of determinants of property investment risk analysis depending on investment goals set by the investor. Different goals of investment will result in different determinants. Three main goals of investment normally set by the investor namely: 1) for rental; 2) for own stay; and 3) for future investment (resell). Property types that act as determinant for these specific goals can be divided into five categories namely: 1) Residential; 2) Commercial; 3) Retail; 4) Industrial; and 5) real estate investment trusts (REITs). Besides, a data-driven approach using data mining techniques will provide patterns of data that can help in identifying the determinant for different investment goals.

(4) To map the weight of determinants with datasets available in the system

Different investors will have different goals and limitations. Thus, based on determinants provided using a data-driven approach, hidden patterns of data provide guidelines for investors to weigh the determinants align with their limitations such as capital and other property details. Mapping the weight of determinants with datasets available will provide different results among investors as different investors have different limitations.

(5) Generate report

The data-driven approach using data mining and TOPSIS model for ranking will help the investors to select the best option of property investment to achieve an optimal solution. The APPIRAS will generate a report for the investors to view and run in-depth analysis.

5.4 APPIRAS Pseudocode and Algorithm

Investors set criteria for their decision-making based on property parameters or variables as shown in Table 5.1 and Table 5.2. The investor will need to identify the criteria, value for each criterion and the weight for each criterion selected. Total weight must be equal to 100 percent.

Note***: The algorithm adopts the TOPSIS model for ranking the alternatives.

Pseudocode and Algorithm 1: INPUT

Pseudocode 1:

GET investor property criteria (parameter/variable)

GET the value for each property criteria selected (parameter/variable)

GET investor weight (percentage) for each criteria selection

COMPUTE the total weight must be equal to 100 percent

Table 5-1: Decision matrix of investor A.

Investor's criteria based on their limitation and objectives	**Value and weight for each criterion**	**Attributes weights**
m_1	n_1	w_1
m_2	n_2	w_2
m_3	n_3	w_3
m_4	n_4	w_4
m_5	n_5	w_5
m_i	n_i	w_i
	Maximum total weightage:	**100 %**

Algorithm1:

Define

List of criteria (property parameter such as number of bedrooms), ListCriteria: $m_1, m_2, m_3, m_4, m_5, m_i$

Integer value for each criterion selected such as number of bedroom equal to 2, ValueCriteria: $n_1, n_2, n_3, n_4, n_5, n_i$

Percentage value for weight for each criterion selected such as priority to have these features is 35%, WeightValue (in percent):

$\boldsymbol{w_1, w_2, w_3, w_4, w_5, w_i}$

Input

$\boldsymbol{m_1, m_2, m_3, m_4, m_5, m_i}$[ ], $\boldsymbol{n_1, n_2, n_3, n_4, n_5, n_i}$[ ] ,

$\boldsymbol{w_1, w_2, w_3, w_4, w_5, w_i}$[ ];

Process

Get $\boldsymbol{m_1, m_2, m_3, m_4, m_5, m_i}$[ ], $\boldsymbol{n_1, n_2, n_3, n_4, n_5, n_i}$[ ] ,

$\boldsymbol{w_1, w_2, w_3, w_4, w_5, w_i}$[ ];

Calculate $\boldsymbol{w_1 + w_2 + + w_4 + w_5 + w_i = 1}$ [];

DetermineCriteria(ListCriteria, ValueCriteria[ ])

{

ListCriteria = Criteria [ ].length; //total number of criteria selected by the investor

ValueCriteria = ListCriteria =ValueCriteria[ ].length; //value of criteria selected by the investor

}

DetermineWeightValue(ListCriteria, ValueCriteria, WeightageValue[ ])

{

m: criteria: $\boldsymbol{m_1, m_2, m_3, m_4, m_5, m}$ ...,$\boldsymbol{m_r}$

n: integer: $\boldsymbol{n_1, n_2, n_3, n_4, n_5, n_{\dots}, n_r}$

weightage: $\boldsymbol{w_1, w_2, w_3, w_4, w_5, w_{\dots}, w_r}$

$i = 1,2,3,4,5 \dots r$

$r = total\ number\ of\ variable\ i$

$i = 1\ (index\ of\ summation)$

n value of m for weightage ($\boldsymbol{w}$) %

For WeightageValue: [$\mathbf{w}$]

Add the value:

For ($\boldsymbol{w_1 + w_2 + w_3 + w_4 + w_5 + w_{\dots} = 100}$ %)

r: number of n(integer)

$$\boldsymbol{w_1 + w_2 + w_3 + w_4 + w_5 + w_r}$$

$$\sum_{i=1}^{r} w_i = \sum_{i=1}^{r} (\boldsymbol{mxn})\boldsymbol{i} = \sum_{i-1}^{r} \boldsymbol{m_i n_i} = 100$$

$$\boldsymbol{m_1 n_1} + \boldsymbol{m_2 n_2} + \boldsymbol{m_3 n_3} + \boldsymbol{m_4 n_4} + \boldsymbol{m_r n_r}$$

}

Output

{

Print ListCriteria [$\boldsymbol{m}$]; //list of criteria selected by the investor

Print ValueCriteria [$\boldsymbol{n}$]**;** //list of value for each criteria selected by the investor

Print WeightageValue [**w**]**;** //list of weightage value selected by the investor

}

Table 5-2: List of decision matrices with weight.

List of property that match the investor's criteria	Investor's criteria based on their limitation and objectives ($\boldsymbol{m_i n_i}$)						Average Weight for each alternatives of, (if all parameters have equal weight) $\boldsymbol{a_i}$
	$\boldsymbol{m_1 n_1} = \boldsymbol{b_1}$	$\boldsymbol{m_2 n_2} = \boldsymbol{b_2}$	$\boldsymbol{m_3 n_3} = \boldsymbol{b_3}$	$\boldsymbol{m_4 n_4} = \boldsymbol{b_4}$	$\boldsymbol{m_5 n_5} = \boldsymbol{b_5}$	$\boldsymbol{m_{\dots} n_{\dots}} = \boldsymbol{b_i}$	
$\boldsymbol{a_1}$	$\boldsymbol{y_{11}}$	$\boldsymbol{y_{12}}$	$\boldsymbol{y_{13}}$	$\boldsymbol{y_{14}}$	$\boldsymbol{y_{15}}$	$\boldsymbol{y_{1\dots}}$	$\frac{\boldsymbol{y_{11} + y_{12} + y_{13} + y_{14} + y_{15} + y_i \cdots}}{\boldsymbol{n_i}} = \overline{\boldsymbol{x_1}}$
$\boldsymbol{a_2}$	$\boldsymbol{y_{21}}$	$\boldsymbol{y_{22}}$	$\boldsymbol{y_{23}}$	$\boldsymbol{y_{24}}$	$\boldsymbol{y_{25}}$	$\boldsymbol{y_{2\dots}}$	$\frac{\boldsymbol{y_{21} + y_{22} + y_{23} + y_{24} + y_{25} + y_i \cdots}}{\boldsymbol{n_i}} = \overline{\boldsymbol{x_2}}$
$\boldsymbol{a_3}$	$\boldsymbol{y_{31}}$	$\boldsymbol{y_{32}}$	$\boldsymbol{y_{33}}$	$\boldsymbol{y_{34}}$	$\boldsymbol{y_{35}}$	$\boldsymbol{y_{3\dots}}$	$\frac{\boldsymbol{y_{31} + y_{32} + y_{33} + y_{44} + y_{35} + y_i \cdots}}{\boldsymbol{n_i}} = \overline{\boldsymbol{x_3}}$
$\boldsymbol{a_4}$	$\boldsymbol{y_{41}}$	$\boldsymbol{y_{42}}$	$\boldsymbol{y_{43}}$	$\boldsymbol{y_{44}}$	$\boldsymbol{y_{45}}$	$\boldsymbol{y_{4\dots}}$	$\frac{\boldsymbol{y_{41} + y_{42} + y_{43} + y_{44} + y_{45} + y_i \cdots}}{\boldsymbol{n_i}} = \overline{\boldsymbol{x_4}}$
$\boldsymbol{a_5}$	$\boldsymbol{y_{51}}$	$\boldsymbol{y_{52}}$	$\boldsymbol{y_{53}}$	$\boldsymbol{y_{54}}$	$\boldsymbol{y_{55}}$	$\boldsymbol{y_{5\dots}}$	$\frac{\boldsymbol{y_{51} + y_{52} + y_{53} + y_{54} + y_{55} + y_i \cdots}}{\boldsymbol{n_i}} = \overline{\boldsymbol{x_5}}$
$\boldsymbol{a_{\dots}}$	$\boldsymbol{y_{\dots 1}}$	$\boldsymbol{y_{\dots 2}}$	$\boldsymbol{y_{\dots 3}}$	$\boldsymbol{y_{\dots 4}}$	$\boldsymbol{y_{\dots 5}}$	$\boldsymbol{y_{\dots\dots}}$	$\frac{\boldsymbol{y_{\dots 1} + y_{\dots 2} + y_{\dots 3} + y_{\dots 4} + y_{\dots 5} + y_i \cdots}}{\boldsymbol{n_i}} = \overline{\boldsymbol{x_i}}$

Pseudocode and Algorithm 2: OUTPUT

Pseudocode 2:

READ investor property criteria (parameter/variable), $\boldsymbol{m}$

READ the value for each property criteria selected (parameter/variable), $\boldsymbol{n}$

READ investor weightage (percentage) for each criteria selection, $\boldsymbol{w}$

MAPPING ListCriteria, ValueCriteria, with dataset available in the database, $\boldsymbol{a}$

COMPUTE the weighted value based on ValueCriteria from Investor and value available in the dataset, $\boldsymbol{b}$

COMPUTE the total weightage must be equal to 100 percent): $\boldsymbol{y}$

MAP the input with sets of data available in the database

RANK the alternatives matching the criteria

PRINT list of alternatives with its determinants

PRINT report for selection

Algorithm 2:

Define

List of criteria (property parameter), ListCriteria: $\boldsymbol{m_1, m_2, m_3, m_4, m_5, m}$ ...

Integer value for each criterion selected, ValueCriteria: $\boldsymbol{n_1, n_2, n_3, n_4, n_5, n}_{\ldots}$

Percentage value for weightage for each criterion selected, WeigtageValue (in percent): $\boldsymbol{w_1, w_2, w_3, w_4, w_5, w}_{\ldots}$

List of property (property parameter), ListofProperty: $\boldsymbol{a_1, a_2, a_3, a_4, a_5, a}$ ...

Integer value for each criterion (result from datasets), Investor'sPropertyCriteria: $\boldsymbol{b_1, b_2, b_3, b_4, b_5, b}_{\ldots}$

Percentage value for weightage for each criterion (matching from investor input with value in datasets available), PropertyWeightedValue (in percent): $\boldsymbol{y_{11}, y_{12}, y_{13}, y_{14}, y_{15}, y_{ij}}$

Average weight for each property listed as alternatives, AverageWeight (in percent): $\boldsymbol{\overline{x_i}}$

Input

Let us denote by $D = \{1,2,\dots,K\}$ a set of decision makers. The multi criteria problem can be expressed in $k -$ matrix format in the following way:

Let $X = (x_{ij})$ be a decision matrix and $W = [w_{1,} w_{2,} \dots, w_n]$ a weight vector, where $x_{ij} \in \Re$.

$\boldsymbol{m_i n_i = b_i}$: criteria,

$\boldsymbol{a_i}$: alternatives,

$\boldsymbol{y_i}$: weight = $\boldsymbol{y_{a_i b_i}}$= a decision-making matrix $D = (d_{ab})_{m*n}$

$\overline{\boldsymbol{x_i}}$: Average weight for ranking and decision

$\boldsymbol{m_1, m_2, m_3, m_4, m_5, m}$ ... [ ], $\boldsymbol{n_1, n_2, n_3, n_4, n_5, n}_{\dots}$[ ] ,

$\boldsymbol{w_1, w_2, w_3, w_4, w_5, w}_{\dots}$[ ];

$\boldsymbol{a_1, a_2, a_3, a_4, a_5, a}$ ... [ ], $\boldsymbol{b_1, b_2, b_3, b_4, b_5, b}_{\dots}$[ ] , $\boldsymbol{y_1, y_2, y_3, y_4, y_5, y}_{\dots}$[ ];

Process

Get $\boldsymbol{m_1, m_2, m_3, m_4, m_5, m}$ ... [ ], $\boldsymbol{n_1, n_2, n_3, n_4, n_5, n}_{\dots}$[ ] ,

$\boldsymbol{w_1, w_2, w_3, w_4, w_5, w}_{\dots}$[ ];

Get $\boldsymbol{a_1, a_2, a_3, a_4, a_5, a}$ ... [ ], $\boldsymbol{b_1, b_2, b_3, b_4, b_5, b}_{\dots}$[ ] ,

$\boldsymbol{y_1, y_2, y_3, y_4, y_5, y}_{\dots}$[ ];

Determine ListofProperty(ListofProperty, Investor'sPropertyCriteria[ ])

{

ListofProperty = Criteria [ ].length; //total number of properties matching the criteria selected by the investor

Investor'sPropertyCriteria = ListCriteria =ValueCriteria[ ].length; //value of the property matching the criteria selected by the investor

}

DeterminePropertyWeightValue(ListCriteria, ValueCriteria, WeightageValue[ ])

{

$\boldsymbol{m}$: investor criteria: $\boldsymbol{m_1}, \boldsymbol{m_2}, \boldsymbol{m_3}, \boldsymbol{m_4}, \boldsymbol{m_5}, \boldsymbol{m} \ldots, \boldsymbol{m_r}$

$\boldsymbol{n}$: investor value (integer): $\boldsymbol{n_1}, \boldsymbol{n_2}, \boldsymbol{n_3}, \boldsymbol{n_4}, \boldsymbol{n_5}, \boldsymbol{n}_{\ldots}, \boldsymbol{n_r}$

$\boldsymbol{w}$: investor weightage: $\boldsymbol{w_1}, \boldsymbol{w_2}, \boldsymbol{w_3}, \boldsymbol{w_4}, \boldsymbol{w_5}, \boldsymbol{w}_{\ldots}, \boldsymbol{w_r}$

$\boldsymbol{a}$: list of property that match the investor;s criteria:

$\boldsymbol{a_1}, \boldsymbol{a_2}, \boldsymbol{a_3}, \boldsymbol{a_4}, \boldsymbol{a_5}, \boldsymbol{a} \ldots, \boldsymbol{a_r}$

$\boldsymbol{b}$: investor criteria to be map with the dataset: $\boldsymbol{b_1}, \boldsymbol{b_2}, \boldsymbol{b_3}, \boldsymbol{b_4}, \boldsymbol{b_5}, \boldsymbol{b}_{\ldots}\boldsymbol{b_r}$

$\boldsymbol{y}$: $w$eight after mapping with the investor's criteria: $\boldsymbol{y_1}, \boldsymbol{y_2}, \boldsymbol{y_3}, \boldsymbol{y_4}, \boldsymbol{y_5}, \boldsymbol{y}_{\ldots}$

$\boldsymbol{t}$: total criteria set by the investor

$$i = 1,2,3,4,5 \ldots r$$

$$r = total\ number\ of\ variable\ i$$

$$i = 1\ (index\ of\ summation)$$

n value of m for weightage ($\boldsymbol{w}$) %

For WeightageValue: [$\mathbf{w}$]

Add the value:

For ($\boldsymbol{w_1} + \boldsymbol{w_2} + \boldsymbol{w_3} + \boldsymbol{w_4} + \boldsymbol{w_5} + \boldsymbol{w}_{\ldots} = \mathbf{100}$ %)

r: number of n(integer)

$$\boldsymbol{w_1} + \boldsymbol{w_2} + \boldsymbol{w_3} + \boldsymbol{w_4} + \boldsymbol{w_5} + \boldsymbol{w_r}$$

$$\sum_{i=1}^{r} w_i = \sum_{i=1}^{r} (\boldsymbol{mxn})\boldsymbol{i} = \sum_{i-1}^{r} \boldsymbol{m_i n_i} = \mathbf{100}$$

$$\boldsymbol{m_1 n_1} + \boldsymbol{m_2 n_2} + \boldsymbol{m_3 n_3} + \boldsymbol{m_4 n_4} + \boldsymbol{m_r n_r}$$

}

DetermineAverageWeight(Investor'sPropertyCriteria, PropertyWeightageValue, [ ])

$$\frac{\boldsymbol{y_{11}} + \boldsymbol{y_{12}} + \boldsymbol{y_{13}} + \boldsymbol{y_{14}} + \boldsymbol{y_{15}} + \boldsymbol{y_{ij}}}{\boldsymbol{t_i}} = \overline{\boldsymbol{x_i}}$$

Step 1- standardize the decision matrix.

Each column of the decision matrix is divided by the root of the sum of squares (RSS).

$$RSS = \sqrt{{x_{11}}^2 + {x_{12}}^2 + {x_{13}}^2 + {x_{1i}}^2}$$

Standardization of decision matrix (SD):

$$SD = (x_{ij}) \div RSS$$

Step 2 – weighted standardization decision matrix

$$WSD = \bar{x_\iota} \times SD$$

Step 3 – determination of ideal solution and negative ideal solution

A set of maximum values for each criterion is an ideal solution (Positive Ideal Solution, PIS).

A set of minimum values for each criterion is a negative ideal solution (Negative Ideal Solution, NIS).

Step 4 – determination of separation from ideal solution. S$_i$*

$${S_i}^* = (WSD - IS)^2$$

Step 5 - determination of separation from negative ideal solution. S$_i$*

$${S_i}^* = (WSD - NIS)^2$$

$$S_{i'} = \sqrt{{S_{11}}^* + {S_{21}}^* + {S_{31}}^* + {S_{ij}}^*}$$

$${S_i}^*$$

$$S_{i'}$$

$${S_i}^* + S_{i'}$$

$$S_{i'}/({S_i}^* + S_{i'})$$

Output

{

Print ListofProperty [$\boldsymbol{a}$]; //list of property based matching the criteria selected by the investor

Print ValueofPropertyCriteria [$\boldsymbol{b}$]**;** //list of value of the property matching the criteria selected by the investor based on dataset available in the database

Print PropertyWeightageValue [**y**]**;** //list of weightage value based on dataset available in the database that map the weightage value set by the investor

Print AverageWeight [$\overline{\boldsymbol{x_i}}$]; //list of average weight value based on total criteria, $\boldsymbol{t}$ and PropertyWeightValue, $\mathbf{y}$.

}

5.5 APPIRAS Main Interfaces

The adaptive personalized property investment risk analysis system implements the PAM method proposed to help investors make faster and better decisions. The APPIRA method helps investors make faster and better decisions in practice. The screenshots on how the system works are shown in Figure 5.2 up to Figure 5.4 as follows:

Figure 5.2: User Registration Form

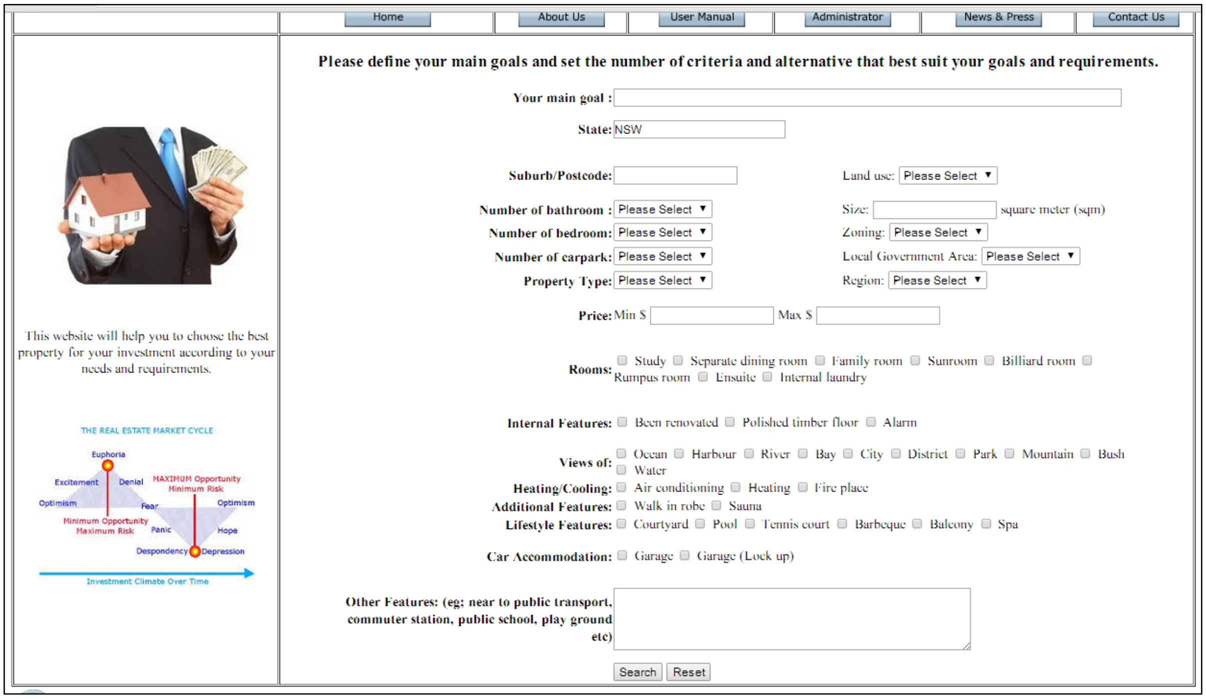


Figure 5.3: Screenshot of how the user enters their search criteria as described in Step 1

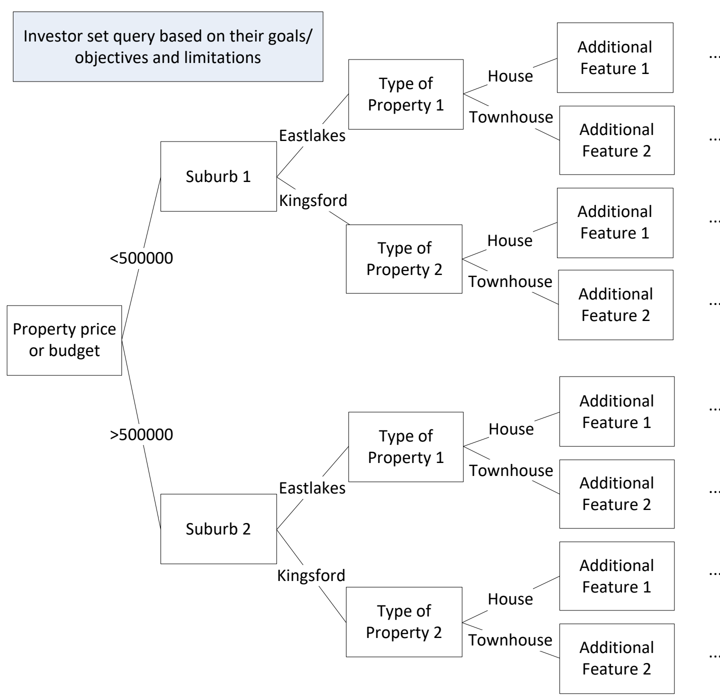


Figure 5.4: Decision tree example for dynamic risk analysis for property investment in the real estate industry.

Figure 5.4 shows the decision tree for one scenario where the investor has ranked property price first, followed by the type of property and other additional features during the process of selecting a suitable property. Once the system has completed mapping and matching the criteria, APPIRAS displays the results, as outlined in Step 2 of the APPIRA method.

These are the results of available property for investment based on your cri

Comparable Properties on the Market

Property ID	Address Number	Street Name	State	Suburb/Postcode	Number of Bathroom	Number of Bedroom	Number of Carpark	Zoning	Minimum Price Range (AUD)	Maximum Price Range (AUD)	Size (sqm)	Type	A			
16	68	Tooronga Tce, Beverly Hills	NSW	2209	1	3	1	Residential	689000	703000	541	House		Separate dining room		Sunroom
20	83	Bay St, Glebe	NSW	2037	0	0	0	Residential	710000	737000	116	House				
30	7	Griffith St, Ashfield	NSW	2131	1	4	1	Residential	744000	781000	590	House				
14	12	Cobden St, Enfield	NSW	2136	1	3	1	Residential	831000	853000	544	House		Separate dining room		

Figure 5.5: Result showing property details sorted by minimum price as described in Step 2.

Figure 5.5 illustrates the displayed results after mapping and matching the criteria set by the investor. Investors gain hidden knowledge from these results which is based on knowledge discovered using data mining which shows the pattern of data through deterministic approach as shown in Step 2, then personalize the risk factors to meet their needs. Steps 3 to 6 are repeated until the investor is satisfied with the results. The proposed system helps reduce information overload by personalizing property searches using advanced criteria.

As compared to other methods available such as Delphi method, AHP, ANP and real options, the proposed method, APPIRA provides personalization on the risk factor criteria, weight of the criteria selected and hidden knowledge or patterns of data that will help the investors to achieve their goals which have different limitations and objectives. Optimal decision-making can be made depending on data availability and timeliness. Thus, it is important to have updated and valid data to ensure optimal decision-making can be made and providing strategic decision-making. By using digital technology which refers to the deterministic approach that adopts the data-driven technique, faster decisions can be made, and automatic and valid results can be applied.

Both heuristic and deterministic approaches have been integrated in both the PAM and PM-SA method proposed in the previous chapter. These two methods are related to each other in providing input for APPIRAS for decision making. APPIRAS gathers input from these two methods, maps with the dataset available and analyzes the data using machine learning available or set in the system.

PAM method provides input from the investor regarding weight level for attributes selected as determinants which varies among different investors. Whereas PM-SA method runs sensitivity analysis which allows the changing of weight level for determinants selected to evaluate the effect of that change. Input for both methods selected were depending on knowledge gathered based on pattern or trend analysis conducted using different machine learning model mapping with investors goals and limitations. The output of trend analysis using visualization and plotting tools provide an accessible way to see and understand trends, outliers and patterns in data. Good data visualization leads to better data-driven decisions and helps to tell stories by curating data into a form which is easier to understand, highlighting the trends and outliers. The machine learning model via visualization approach dashboard adopted in the APPIRAS system for effective user interaction in achieving optimal decisions.

5.6 Summary

This chapter describes an adaptive personalized property investment risk analysis software system to identify the determinants, weight and rank the property listed that meet the investor's requirements. The newly developed system, APPIRAS, introduces the concept of adaptive personalization technique, which is expressed in PAM and PM-SA method of APPIRA to achieve optimal solution and align with investor's goal of investment. The data-driven approach using data mining techniques are used to map the investor's criteria or determinant selected with the datasets available in the system.

CHAPTER 6

CASE STUDY

6.1 Introduction

This chapter presents the data sets and experiment design for the deterministic approach to property investment risk prediction. Section 6.2 explains in detail the data analysis using heuristic and deterministic approaches that are employed in both PAM and PM-SA methods proposed. This chapter discusses results based on experiments with sample data collected from APM domain property data. The remainder of this chapter is structured as follows. The decision table technique to demonstrate how the personalization criterion affects experiments results presented in Section 6.3. The data sets description is given in Section 6.4. Section 6.5 elaborates the results of experiments of three data mining techniques chosen namely: clustering technique, link or association analysis and two experiments of predictive technique for forecasting to discover the hidden pattern of data. Finally, this chapter is summarized in Section 6.6. The work presented in this chapter has been reported in four of our publications listed in Section 1.8, i.e. publication number 3, 4, 5 and 6.

6.2 Data analysis

In order to validate the effectiveness of the proposed PAM method, four sets of data collected from Australian Property Monitor domain database with 138 attributes for the data analysis. Data collected from Australian Property Monitor domain databases stored in a database and analysed for identifying the matching criteria of property available with investor's requirements and hidden knowledge of selected properties using MS SQL Server 2008 integrated with Microsoft SQL Data Mining Add-ins, Crystal ball software, TIBCO Spotfire and WEKA tools. Sample data were used to analyze the data and show how the system works based on heuristic approach integrated with deterministic approach.

6.2.1 Heuristic Approach

The investor will personalize the rank and weight of risk factors based on results displayed using the multidimensional data model and data mining analysis (deterministic approach) of data stored in the database. The rank and weight of the criteria of factors will be customized or personalized based on the decision maker's objectives and limitations. Some examples of the limitations might include available budget (financial risk factor), time series (scheduled risk), and location. The decision-making process will consider all the related criteria stored in the database with a high level of data for the property investment risk analysis. Instead of referring to experts in the field to rank and weigh the risk factors as applied to the AHP method which is the most popular in the real estate industry, the personalization technique applied in the PAM method will be helpful for decision makers to produce accurate and reliable decisions that meet their requirements. In terms of time consuming to produce the results, the PAM method is more accurate, faster and simpler in generating results for the user to analyze.

6.2.2 Deterministic Approach

In order to explain how the deterministic approach functions, two sets of data have been analyzed using Microsoft SQL 2008 Data Mining Add-ins were used to train the data. Microsoft SQL 2008 Data Mining Add-ins use a single algorithm, Auto-Regressive Tree With Cross Prediction (ARTXP) and second algorithm AutoRegressive Integrated Moving Average (ARIMA) to analyze data for predictions. The ARTXP algorithm was optimized for short-term predictions and the ARIMA is optimized for long-term prediction. By default, the Microsoft Time Series algorithm uses a mix of the algorithms when it analyses patterns and making predictions. The algorithm trains two separated models on the same data; one model uses the ATRXP algorithm and one model uses the ARIMA algorithm. The algorithm then blends the results of the two models to yield the best prediction over a variable number of time slices. Because ATRXP is best for short-term predictions, it is weighted more heavily at the beginning of a series of predictions. However, if predicting further into the future, ARIMA is weighted more heavily. To give an example of how the deterministic approach functions, two data mining techniques have been chosen for the analysis namely the predictive technique for forecasting and association technique to define patterns of data as discussed in the next subsection.

6.3 Decision table technique

This section describes the application of the decision table technique to demonstrate how the personalization criterion affects both of the results. An example of a decision table, as depicted in Table 6.1, is created to examine the combination of inputs, which produce different results by using this technique. The user's limitation will be the conditions for the decision table and each condition has a different number of values. For example, the cost of investment can be defined as three values, known as low (L), medium (M) or high (H); the level of risk acceptance also can be defined as three values, known as low (L), medium (M) or high (H) and the objective of the investment can be defined as two values, known as residential (R) or commercial (C). The combination of these conditions will generate a number of rules with respective actions. For each combination, the action can be single or multiple. The investor profiles or constraints will be linked with their goals, objectives and strategy for investment. For example, as shown in Table 6.1, the conditions, or user's limitations, include the cost of the investment, level of risk acceptance and objective of investment. The combination of these constraints will produce different recommendations for the user to choose.

The decision table is one of the best techniques to model complicated logic as personalization because property investment risk analysis in the real estate industry has many limitations. Each action in the decision table corresponds to associate conditions, as shown in Table 6.1. The what-if analysis is described through the decision table to identify the risk level of the risk analysis factor that will affect the analysis results. The recommendations and choice of actions given are a guideline for the user to understand and discover the alternatives that meet their goals and requirements. Based on this scenario using the decision table technique, the investors achieve better results if they employ a personalization session for risk measurement to achieve their goals.

Table 6-1: A decision table showing a scenario of recommendations for a property investment based on a user's constraints and requirements.

Conditions (User's constraints/limitation/ profiles)	Rules																	
Cost of investment	L	M	H	L	M	H	L	M	H	L	M	H	L	M	H	L	M	H	…
Level of risk acceptance	L	L	L	M	M	M	H	H	H	L	L	L	M	M	M	H	H	H	…
Objective of investment	R	R	R	R	R	R	R	R	R	C	C	C	C	C	C	C	C	C	…
. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
Actions																			
Highly recommended					X		X						X						…
Moderately recommended	X			X				X		X	X	X		X		X	X	X	…
Low recommended		X	X			X			X						X				…
. . .	. . .	. . .	. . .	. . .	Recommendations and Choice of	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .

Legend:

Cost of investment: L-Low, M-Medium, H-High

Level of risk acceptance: L-Low, M-Medium, H-High

Objective of investment: R-Residential, C-Commercial

6.4 Data Mining Techniques for Knowledge Discovery

Three data mining techniques, namely the clustering technique, predictive technique for forecasting analysis and link or association rule mining and were chosen to discover the hidden pattern of data. The sample data used for the experiments were gathered from Australian Property Monitor domain database and stored in the develop prototype system. About 138 attributes were chosen and stored in the prototype system developed for the analysis. For data mining techniques, the Microsoft SQL Server 2008 integrated with Microsoft SQL Data Mining Add-ins were used to analyze the data and provide the output. For multidimensional analysis, it provides historical value of house price or rental rate for property in different suburbs based on sample data extracted. This section explains in detail how the application of data mining techniques helps to discover hidden patterns of data. Three data mining techniques have been chosen for the analysis, namely the clustering technique to group or cluster similar records, link analysis to define patterns of data and predictive technique for forecasting analysis to predict future events. To explain how the deterministic approach functions, four sets of data were analyzed to produce a data model. Three data mining techniques, relevant to real estate analysis, were chosen; however, other techniques would also be applicable depending on the user's requirements.

6.4.1 Clustering Technique

A clustering technique organizes data by abstracting underlying structures, either as a grouping of individuals, or as a hierarchy of groups. The representation can then be investigated to see if the data group is in accordance with preconceived ideas, or to suggest new experiments. Cluster analysis groups data objects into clusters so that objects belonging to the same cluster are similar, while those belonging to different ones are dissimilar [33]. The resulting reports derived from an experimental data which consist of 619 rows of data selected for the analysis with four attributes as shown in Table 6.2. Data was collected from the Australian Property Monitor domain database and analysed using MS SQL Server 2008, integrated with Microsoft SQL Data Mining Add-ins.

Table 6-2: Sample data used for the analysis collected from the Australian Property Monitor domain database.

Property Type	Land Size (Sqm)	Year	Rental
Commercial	250	2005	750
Commercial	5312	2006	132800
Commercial	80	2006	31200
House	342	2004	350
House	334	2004	430
House	567	2005	650
Industrial	2801	2012	60
Industrial	48	2012	24800
Other Residential	929	2012	420
Semi	171	2006	360
Terrace	134	2012	485
Terrace	108	2012	500
Townhouse	149	2012	700
Unit	464	2003	225
Unit	930	2003	230
Unit	976	2003	190
Unit	1882	2003	220
Villa	811	2012	380
. . .	. . .	. . .	. . .

The sample of property data located in the Eastlakes suburb of Sydney in New South Wales was selected for the experiments. Four attributes have been chosen for the experiments, namely the type of property, land size, year and rental rates.

Variables	Values	Probability
Property Type	Unit	80 %
Rental	4,524 - 19,775	25 %
Land Size _Sqm_	3,697 - 12,084	25 %
Rental	19,776 - 72,364	25 %
Land Size _Sqm_	12,085 - 41,004	25 %
Year	2012	21 %
Year	2006	20 %
Year	2004	15 %
Land Size _Sqm_	18 - 3,696	12 %
Year	2011	11 %
Property Type	House	10 %
Rental	60 - 4,523	8 %
Year	2009	8 %
Year	2007	8 %
Year	2005	7 %
Property Type	Commercial	6 %
Year	2008	5 %
Year	2010	5 %
Year	2003	2 %
Property Type	Terrace	1 %
Property Type	Industrial	1 %
Property Type	Villa	1 %
Property Type	Semi	1 %

Figure 6.1: Summary of data selected based on Table 6.2.

As shown in Figure 6.1, the most common type of property available for rental at the Eastlakes suburb is dominated by units, followed by house, commercial, terrace, industrial, villa and semi-detached. This information will give investors an idea of what type of property they should focus on if they intend to invest in commercial or business real estate.

Based on sample data collected for the analysis, nine clusters have been created and the associations between clusters are linked using lines as shown in Figure 6.2.

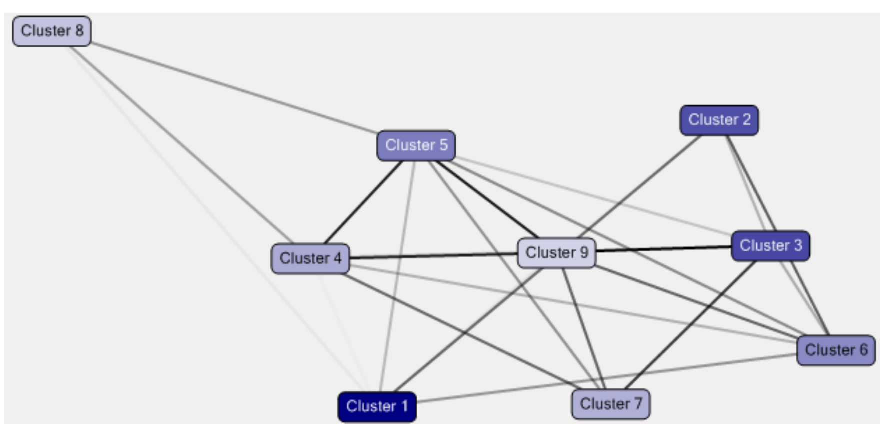


Figure 6.2: Clustering technique based on selected Eastlakes property data.

The associations between clusters provide knowledge about the most influential features of property for rental that the investors need to choose when planning a rental property investment.

Variables	States	Population (All)	Cluster 1	Cluster 3	Cluster 2	Cluster 5	Cluster 6	Cluster 4	Cluster 7	Cluster 8	Cluster 9
Size		173	39	28	27	20	18	13	12	9	7
Land Size _Sqm_	Mean	3,697.00	907.3	897.87	745.02	401.64	1,288.20	370.34	6,316.00	1,563.78	63,156.00
Land Size _Sqm_	Deviation	12,435.90	353.43	262.69	299.54	250.15	975.49	216.33	6,615.72	2,176.20	
Property Type	Unit	139	99 %	97 %	100 %	45 %	98 %	24 %	87 %	0 %	100 %
Property Type	House	18	1 %	0 %	0 %	44 %	2 %	61 %	0 %	0 %	0 %
Property Type	Commercial	11	0 %	0 %	0 %	3 %	0 %	6 %	4 %	100 %	0 %
Property Type	Terrace	2	0 %	1 %	0 %	2 %	0 %	8 %	0 %	0 %	0 %
Property Type	Industrial	1	0 %	0 %	0 %	0 %	0 %	0 %	9 %	0 %	0 %
Property Type	Villa	1	0 %	2 %	0 %	0 %	0 %	1 %	0 %	0 %	0 %
Property Type	Semi	1	0 %	0 %	0 %	6 %	0 %	0 %	0 %	0 %	0 %
Property Type	...	...	...	...	...	...	...	...	...	...	...
Rental	Mean	4,523.00	230.32	387.59	216.5	305	246.12	390.85	483.98	80,738.36	580
Rental	Deviation	22,613.70	25.77	50.76	28.38	112.83	64.86	200.48	197.38	63,461.45	50
Year	2012	36	0 %	55 %	0 %	11 %	0 %	24 %	82 %	11 %	57 %
Year	2006	34	48 %	0 %	17 %	24 %	0 %	6 %	0 %	67 %	0 %
Year	2004	26	1 %	0 %	64 %	11 %	15 %	5 %	0 %	0 %	0 %
Year	2011	19	0 %	30 %	0 %	27 %	4 %	18 %	0 %	0 %	29 %
Year	2009	13	5 %	4 %	1 %	6 %	46 %	14 %	0 %	0 %	0 %
Year	2007	13	23 %	0 %	1 %	6 %	4 %	12 %	0 %	11 %	0 %
Year	2005	12	22 %	0 %	1 %	0 %	5 %	11 %	8 %	0 %	0 %
Year	2008	8	1 %	1 %	12 %	2 %	14 %	1 %	0 %	11 %	0 %
Year	...	...	...	...	...	...	...	...	...	...	...

Figure 6.3: Cluster profiles based on Figure 6.2.

Figure 6.3 illustrates the cluster's profile, in which cluster 1 dominates the population, followed by cluster 3, 2, 5, 6, 4, 7, 8 and 9, respectively. The rental rate for cluster 1 is 230 dollars per week, with a standard deviation 25.77. Based on this result, the investor can calculate the mortgage instalment if they plan to invest in property for rental and need to organise a loan for the capital. The investor should utilize this information as knowledge and guidelines in order to buy the best property for investment, if their objective is to invest in property for rent.

6.4.2 Predictive Technique for Rental Rate Forecasting

The predictive technique is used for forecasting based on time series and historical data available to generate further analysis. The forecast data generated by the system will provide more valuable information and knowledge to the investor. Out of 619 rows of data, 239 rows are characterized as a 'unit' type of property that has been selected to forecast the rental rate. Figure 6.4 depicts the five-year forecast of rental rate predicted up to 2017 for investors to analyse, based on selected data shown in Table 6.2.

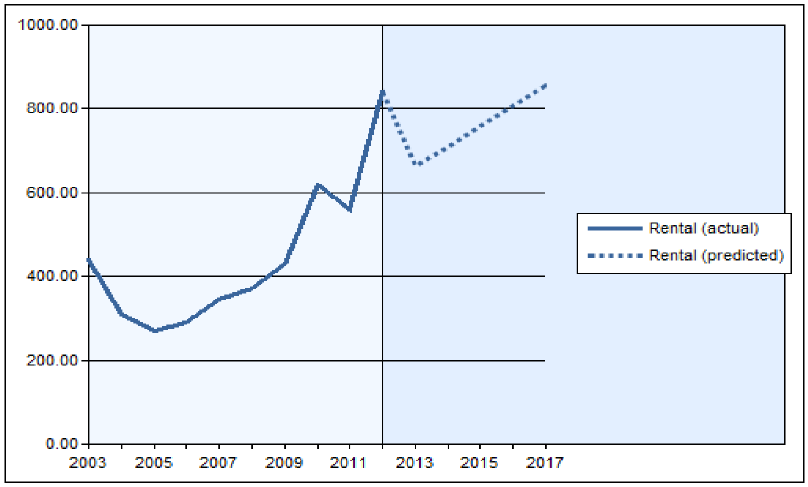


Figure 6.4: Rental rate prediction until 2017, based on data selected in Table 6.2.

As shown in Figure 6.4, the historical information appears to the left of the vertical line (straight line), which represents the data that the algorithm uses to create the model, while predicted information appears to the right of the vertical line (dotted line) and represents the forecast that the model makes. The forecasting value will help investors make better decisions based on their limitations and goals or objectives.

Based on the experiment and the results shown, it is important to consider the time series for investment and forecasting, depending on the investor's requirements. By using an expert survey method, there is no customization on the time frame that fits with decision makers' requirements. Based on the historical data and valid data available in the database, investor's confidence is increased by using the output generated by the system.

Origin of the knowledge is coming from the data stored in the system and created using data mining techniques to discover and find the hidden pattern of data that would be useful for the investors as the heuristic approach. The application of the heuristic through deterministic approach to disseminate knowledge and its implications provide reliable and accurate information. Investors as the decision maker needs to have knowledge in the field to speed up the decision-making process. Heuristic commonly utilized in the process of decision-making to help users obtain relevant ideas, experience and gain knowledge in managing problems they dealt with. In addition, the hidden pattern of data discovered will enable the investor to make efficient decisions with the assistance of a heuristic approach to personalize the criteria based on their requirements. The technology of data mining could be used for daily practice of analyzing property for investment, and those in the field of analyzing uncertain factors for decision-making process. The proposed model can be easily understood by the investor thus providing a practical assessment tool for decision-making about investment risk analysis.

6.4.3 Predictive Technique for Median House Price Forecasting

The predictive technique for forecasting analyzes facts to make predictions about future events. The predictive model was used to identify investment risk for the selected property. Time is the independent variable and is based on an annual analysis of median house prices for that year using selected sample data. The median house price is the dependent variable to be measured and is based on sample data extracted from the APM domain property data, as shown in Table 6.3 and Table 6.4 respectively.

Table 6-3: Parameters of the testing data

Parameter	Value
Year	2004–2017
Median House Price	Collected from Australian Property Monitor (APM) domain property data
Suburb/Postcode	Bellevue Hill, New South Wales

Table 6-4: Sample of training data for analysis

Year	Median House Price (AUD)
2004	2,500,000
2005	2,229,000
2006	2,825,000
2007	2,725,000
2008	3,350,000
2009	3,105,000
2010	2,500,000
2011	3,346,000
2012	2,958,000
2013	3,328,000
2014	3,470,000
2015	4,150,000
2016	4,400,000
2017	5,301,000

All extracted data was then exported to Microsoft Excel, and data mining add-ins and oracle crystal ball were used for data analysis. This study included data for 500 properties, based on the suburb or postcode of the property. The median house price and the year were used for the first experiment. To validate the effectiveness and the applicability of the proposed APPIRA method, sample data was collected from the Australian Property Monitor database (www.apm.com.au). The data set included 500 properties in Bellevue Hill, NSW, Australia, each with 138 attributes. MS SQL Server 2008, with integrated Microsoft SQL Data Mining add-ins, was used to analyze the data and extract relevant patterns of data. The parameters of the testing data and the sample data for forecasting analysis are shown in Tables 6.3 and 6.4, respectively.

Table 6-5: Algorithm Parameters

Parameter	Default	Range
AUTO_DETECT PERIODICITY	0.6	[0,0,1,0]
COMPLEXITY_PENALTY	0.1	(...,1,0)
FORECAST_METHOD	MIXED	ARIMA,ARTXP,MIXED
HISTORIC _MODEL_COUNT	1	[0.100]
HISTORIC_MODEL_GAP	10	[1,...)
INSTABILITY_SENSITIVITY	1.0	[0,0,1,0]
MAXIMUM_SERIES_VALUE	+1E308 (single type of data)	[column maximum,...]
MINIMUM_SERIES_VALUE	-1E308 (single type of data)	[...,column minimum]
MINIMUM_SUPPORT	10	[1,...]
MISSING_VALUE_SUBSTITUTION	None	None,Previous,Mean
PERIODICITY_HINT	(Zalewska-Turzyńska)	{...list of integers...}
PREDICTION_SMOOTHING	0.5	[0,0,1,0]

Parameter description:

Table 6.5 specifies the algorithm parameters using a numerical value between 0 and 1 used to detect periodicity. Setting this value closer to 1 favors discovery of many near-periodic patterns and automatic generation of periodicity hints. Dealing with a large number of periodicity hints will likely lead to significantly longer model training times. If the value is closer to 0, periodicity is detected only for strongly periodic data. The data from Australian Property Monitor domain database is chosen and collected as the testing data sources; the parameters and the algorithm parameters of the testing data and the prototype can be seen in Table 6.3 and Table 6.5 respectively. Table 6.4 shows the input for the training or testing data.

Figure 6.5 shows median house prices, based on the sample data of median house price of Bellevue Hill, New South Wales shown in Table 6.4.

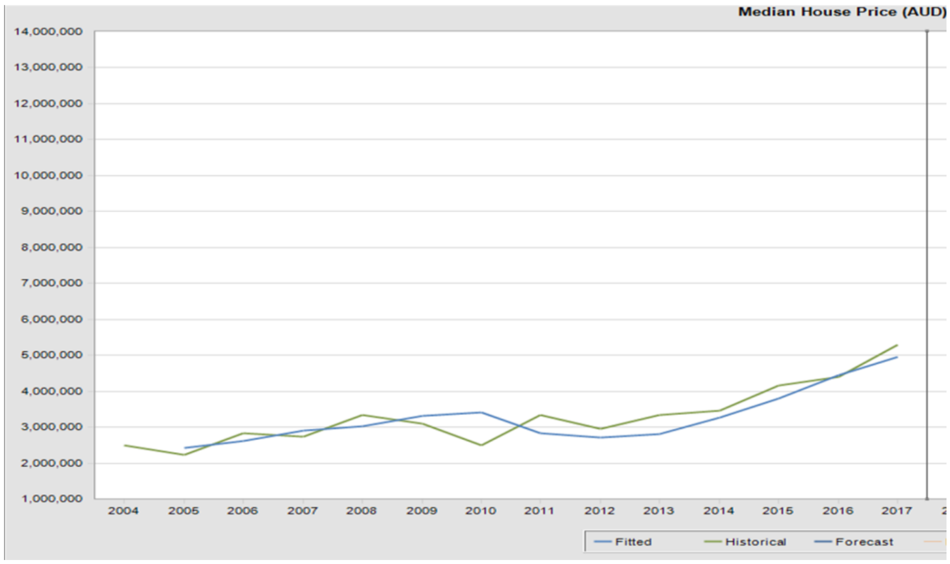


Figure 6.5: Median house price in Bellevue Hill, NSW (based on data available from 2004 – 2017).

Table 6.6 respectively shows the details of the Crystal Ball report that uses Monte Carlo simulation to forecast the result.

Table 6-6: Crystal ball report

Run preferences:		
Periods to forecast	12	
Fill-in missing values	On	
Adjust outliers	Off	
	Non-seasonal methods	
		ARIMA methods
	Forecasting technique	Standard forecasting
	Error measure	RMSE
	Summary:	
	Best method	ARIMA(1,1,2)
Method used	Error measure (RMSE)	391,894

Forecast results:			
Date	**Lower: 2.5%**	**Forecast**	**Upper: 97.5%**
2018	4,795,551	5,563,648	6,331,745
2019	5,120,909	6,056,074	6,991,239
2020	5,070,385	6,430,778	7,791,171
2021	4,810,670	6,715,902	8,621,134
2022	4,446,789	6,932,862	9,418,936
2023	4,030,214	7,097,955	10,165,695
2024	3,588,673	7,223,579	10,858,485
2025	3,138,298	7,319,170	11,500,043
2026	2,688,805	7,391,909	12,095,013
2027	2,246,062	7,447,259	12,648,455
2028	1,813,526	7,489,376	13,165,226
2029	1,393,106	7,521,424	13,649,742

Historical data:		
Statistic	**Historical data**	
Data Values	14	
Minimum	2,229,000	
Mean	3,299,071	
Maximum	5,301,000	
Standard Deviation	837,077	
Ljung-Box	9.96	(Detrended)
Seasonality	Non-seasonal	(AutoDetect)
Screened Values	0	

ARIMA statistics:		
ARIMA		**Statistic**
Transformation Lambda	1.00	
BIC		**26.35**
AIC		26.22
AICc		26.42
* Used for model selection		

ARIMA model coefficients:	
Variable	**Coefficient**
AR(1)	0.7609
MA(1)	1.07
MA(2)	-0.8243

Forecast accuracy:			
Method	**Rank**	**RMSE**	**Standard Error**
ARIMA(1,1,2)	**Best**	**391,894**	
Double Exponential Smoothing	2nd		465,695
Damped Trend Non-Seasonal	3rd		466,551

Method				
ARIMA(1,1,2)			**Theil's U**	**Durbin-Watson**
Double Exponential Smoothing	0.8892	2.05	0.7369	1.99
Damped Trend Non-Seasonal	0.9016	2.06		

Method parameters:				
Method			**Parameter**	**Value**
ARIMA(1,1,2)			---	---
Double Exponential Smoothing	Alpha	0.5586	Beta	0.5653
Damped Trend Non-Seasonal	Alpha	0.4675	Beta	0.9990
			Phi	0.9221

Crystal Ball Report:

Note: Crystal Ball uses Monte Carlo simulation to display results in a forecast chart that shows the entire range of possible outcomes and the likelihood of achieving each of them. In addition, it keeps track of the results of each scenario.

The forecast results were based on 12 periods of data starting from 2004 until 2017. Non-seasonal ARIMA method, an advanced modelling technique for time-series analysis were used for forecasting while to measure the error RMSE (root mean squared error) is an absolute error measure that squares the deviations to keep the positive and negative deviations from cancelling out one another were used. Based on the result shown in Table 6.6, the summary of crystal ball report ARIMA(1,1,2) without constant (damped-trend linear exponential smoothing) depicted the AR (auto regressive = p) value is 1 which predicts the change in prices as an average change, plus some fraction of the previous change, plus a random error; non-seasonal difference (d) value is 1 which represent periods of differencing equal to one year and MA (moving average = q) value is 2 depicts damped-trend linear exponential smoothing, $\hat{Y}t = Yt\text{-}1 + \phi 1\ (Yt\text{-}1 - Yt\text{-}2\ ) - \theta 1et\text{-}1 - \theta 1et\text{-}1$. The interpretation of the value for the ARIMA method is shown in Figure 6.6.

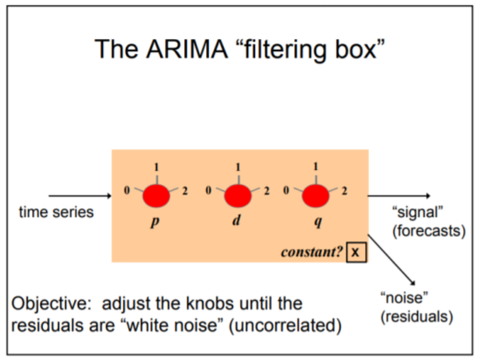


Figure 6.6: The ARIMA filtering box interpretation.

Figure 6.7 shows the actual median house price from 2015 -2018 for Bellevue Hill, NSW, retrieved from the APM property database.

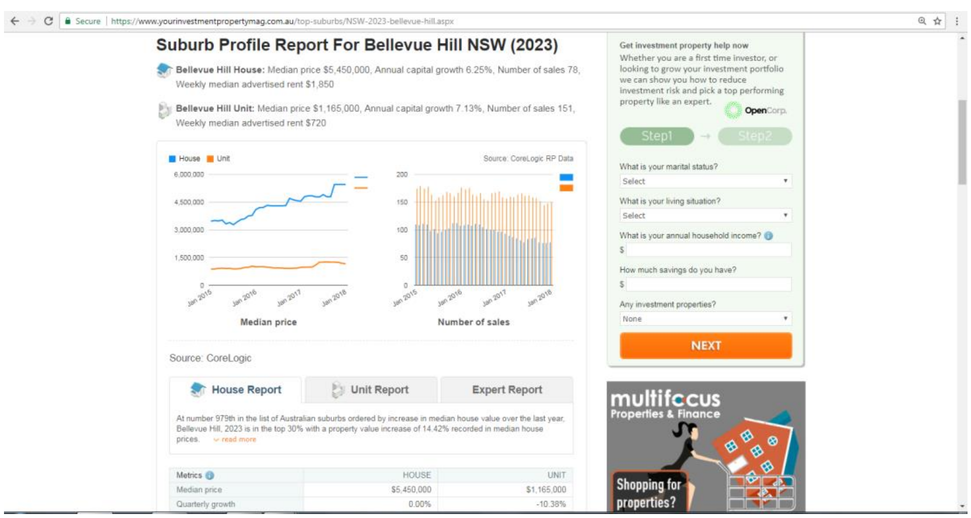


Figure 6.7: Screenshot of actual median prices for type of property (Unit and House) in Bellevue Hill, NSW

Table 6-7: Forecast range of median house prices for Bellevue Hill, NSW, collected from Australian Property Monitor database

Year	Median House Price (AUD)	Changes (AUD)	Percentage (%)
2004	2,500,000	-	8.70
2005	2,229,000	-271,000	-10.84
2006	2,825,000	596,000	26.74
2007	2,725,000	-100,000	-3.54
2008	3,350,000	625,000	22.94
2009	3,105,000	-245,000	-7.31
2010	2,500,000	-605,000	-19.48
2011	3,346,000	846,000	33.84
2012	2,958,000	-388,000	-11.60
2013	3,328,000	370,000	12.51
2014	3,470,000	142,000	4.27
2015	4,150,000	680,000	19.60
2016	4,400,000	250,000	6.02
2017	5,301,000	901,000	20.48

The comparison between the forecast value and the actual value is shown in Table 6.7 visualized in Figure 6.8. The iterative process of re-weighting and re-ranking the personalization criteria based on displayed results allows investors to perform 'what-if' style analyses using multiple searches. This also helps to convert hidden knowledge into explicit knowledge which, in turn, helps investors make better decisions. The multidimensional data model (data mining approach) is combined with personalization (heuristic approach) to generate better results to meet the objectives of decision-makers.

Figure 6.8 predicts the median house price up to the year 2028 derived from the training data shown in Table 6.4.

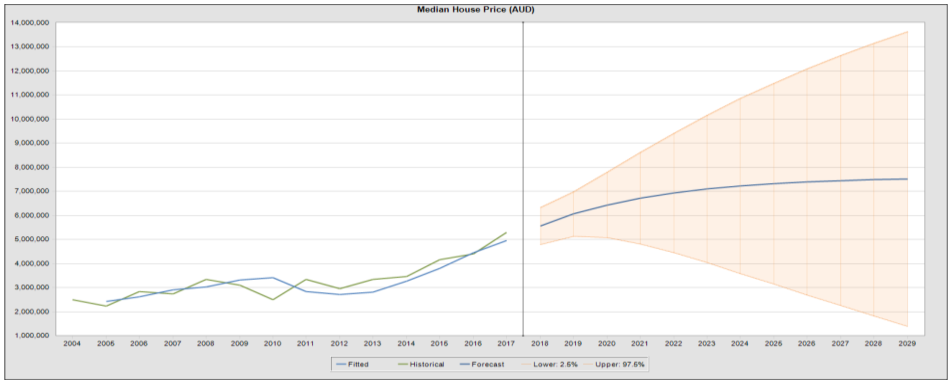


Figure 6.8: Median house price (predicted for the year 2018 - 2028).

Based on the experiment and the results shown, it is important to consider the time period of the investment and make forecasts according to the decision maker's personal preferences. Unlike the APPIRA method, expert survey methods do not allow the time frame to be customized to suit the investor's requirements. The multidimensional data model and data mining technique used by the APPIRA method also generates intuitive, visual output to make analysis easier. Moreover, APPIRA interprets investors' requirements better than the expert survey method. Based on the historical data and valid data available in the database, it increases decision-makers' confidence in the output generated by the system.

Figure 6.8 depicts the predicted median house price up to 2012 based on training data shown in Table 6.4. The historical information appears to the left of the vertical line (straight line) which represents the data that the algorithm uses to create the model while predicted information appears to the right of the vertical line (dotted line) which represents the forecast that the model makes.

By using an expert survey method, there is no customization on the time frame that fits with decision makers' requirements. As a result, the PAM method solves this problem by providing a personalization technique. The multidimensional data model and data mining technique applied to the PAM method also generate intuitive and visually type of output for easy analysis. Moreover, the PAM method creates a better understanding of decision makers' requirements and propels better decisions compared with expert survey methods. Based on the historical data and valid data available in the database, it increases the decision maker's confidence towards the output generated by the system. Based on the experiment and results shown, the time dimension in multidimensional analysis is compulsory for investment risk analysis especially for predictive techniques for forecasting.

6.4.4 Association/classification to define patterns of data

Association rule mining finds interesting associations and correlation relationships among a large set of data items. Association rule shows attribute value conditions that occur frequently together in a given dataset (He & Song, 2009). An association technique is a tool used to infer or mine patterns from data. Revealing hidden patterns in data brings to light valuable knowledge and information for investors that may influence the decision-making process, or even change their overarching objectives. For this reason, APPIRA provides the opportunity to adjust the weight and rank of risk factors based on the data patterns from APPIRAS. To illustrate the use of association techniques to define the patterns of data, a set of 339 properties were extracted to produce a model. The parameters of the testing data set include the suburb name; the number of bathrooms, bedrooms, and parking spaces; the property type and size; the rental rate; and additional features, as listed in Table 6.8.

Association technique is used to define the patterns of data and generate more analysis. The hidden patterns of data generated by the system will give more valuable information and knowledge to the decision makers. They might not be aware about certain criteria surrounding the property that match with their query or limitations which will influence the overall decision-making process. Thus, by analyzing the hidden knowledge, it will manipulate the decision maker's requirements and objectives. They might change their objectives after looking at the hidden knowledge and personalize the rank and weight of the criteria for property investment analysis. Besides, predicting the patterns of features of the property in different locations will help the decision makers to achieve better results. To give an example of an association technique to define the patterns of data, a set of 339 rows of data is extracted and transformed to train and test the data as shown in Table 6.8.

Table 6-8: Parameters of the testing data and the prototype.

Parameter	Value
Total number of case/rows	339
Suburb	36 suburbs
No. Of Bathroom	Minimum = 1, Maximum = 7
No. Of Bedroom	Minimum = 1, Maximum =6
No. Of Carpark	Minimum = 1, Maximum = 8
Property Type	3 types (Townhouse, House, Unit)
Property Size	Minimum = 55 sqm, Maximum = 13779 sqm
Rental Rate	Minimum = 100, Maximum = 6500
Additional Features	33 features (study room, Separate dining room, family room, billiard room, rumpus room, ensuite, internal laundry, walk in robe, sauna, air conditioning, heating, fire place, courtyard, pool, tennis court, barbeque, balcony, spa, been renovated, polished timber floor, alarm, garage, garage (lock up), ocean, harbour, river, bay, city, district, park, mountain, bush, water)

This experiment included data on 399 properties, with specific property features, in 36 different suburbs. Some of the data in the APM website contained missing values for property features, which can produce misleading results, so the data needed to be cleaned and standardized before selecting the sample data set. The parameters for link analysis to define patterns of data used in the second experiment are shown in Table 6.8.

Based on the testing data summarized in Table 6.9, the pattern of rental in three different suburbs is associated with different attributes as shown in Figure 6.9.

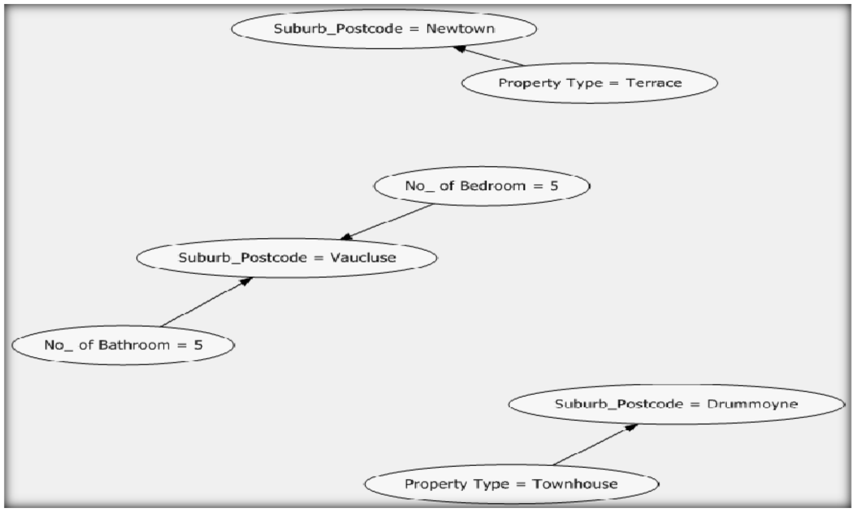


Figure 6.9: Pattern of Rental at Selected Suburb.

Based on Figure 6.9, different suburbs have a different pattern of property for rent features as summarized in Table 6.9.

Table 6-9: Summarized the pattern of data for property rental in different suburbs.

Suburb	**Popular features of property rental**
Drummoyne	Property type = Townhouse
Newtown	Number of bedrooms = 5 and Number of Bathroom = 5
Vaucluse	Property type = Terrace

Based on these patterns, investors can make a better decision on what are the suitable properties to buy if they want to make investments for property rental in different suburbs. For example, if the investor would like to buy a property for rental in Drummoyne, they have to find a townhouse as a type of property.

Using these patterns, an investor can make better decisions about which properties to buy as rental investments in different suburbs. For example, if an investor wants to buy a rental property in Drummoyne, a townhouse provides the highest probability for rental.

This personalization mechanism during decision-making helps the investor undertake a more detailed analysis, according to their needs, as shown in Experiment 1. A variation of the forecast value against the actual value provides a strong indicator that more detailed analysis is required before making a decision. Without personalization, the information is static and no further analysis can be undertaken.

6.5 Summary

Property investment risk analysis is a vital process that needs more attention by individuals as investors and real estate agents. The application of multidimensional analysis and data mining techniques to measure the risk factors of property investment through a decision support system helps the investors to avoid and reduce risk. The investors can discover more knowledge about the features of available property for investment by applying this technique. Reflecting on the limitations and disadvantages of expert opinion to measure the risk factors for investment, the application of multidimensional analysis and data mining technique through deterministic approach helps the investors to achieve optimal solutions that meet their requirements. The historical data kept using a multidimensional data model helps to provide forecasting analysis and aggregate data in providing foundations for the investment decision maker. The proposed technique is helpful in measuring the risk factor of the property investment in the real estate industry. Even though the cost is higher to implement the multidimensional analysis and data mining technique, the results generated would be more accurate and useful for investment risk analysis. A prototype decision support system that applies the concept of multidimensional analysis, data mining technique for property investment risk analysis and web mining will be developed for future works to cater complex demand by users of the system.

CHAPTER 7

CONCLUSIONS AND FUTURE STUDY

7.1 Conclusions

This study presents a new method for improving problem solving, decision-making, and risk analysis for real estate investment. Comprehensive risk analysis is a crucial part of real estate investment, and all related risk factors, both qualitative and quantitative, should be determined in the early stages of risk analysis before the risk factors are weighted and ranked. Mapping the weight and rank of the risk factors between an investor's requirements and patterns of data using data mining operations is an innovation of this study. It integrates both heuristic and deterministic approaches for weighting and ranking the risk factors in order to help investors achieve their goals and objectives. Experiments show that the proposed APPIRA method is very useful for identifying, weighting and ranking the risk factors associated with property investment.

Judging the rank and weight of risk factors in an actual risk evaluation is difficult and the most popular technique used in the real estate industry is AHP and ANP. However, these two techniques require expert judgments that might not align with investor's goals and limitations. Thus, this paper presented a new method for improving problem solving and decision-making process for dynamic risk analysis investment which incurred a very high cost in the real estate industry. The key idea of the proposed method is to map the weight and ranking of risk factors between investor's requirements (heuristic approach) with the output of pattern discovery using the multidimensional data model and data mining techniques (deterministic approach). A comprehensive risk analysis should be prepared to ensure the success of an investment. All related risk factors either qualitative or quantitative should be discerned at an early stage of the risk analysis before estimating the ranking and weighting of the risk factors. Mapping the weight and ranking of risk factors between investor's and patterns of data using data mining operations is the innovation of this study. It integrates both heuristic and deterministic approaches for weighting and ranking the risk factors and at the same time helps investors to achieve their goals and objectives. The case study experiments show that the PAM method proposed can be used for ranking and weighting the risk factor analysis for property investment in the real estate industry. An APPIRA method proposed in this study solves the lack of risk prediction, lack of personalization, and information overload problems in real estate investment decision-making. This research makes several contributions as follows. Firstly, we propose a novel mechanism to overcome the drawback of depending on expert judgments which is done manually when making decisions by instead integrating data-driven techniques to identify the determinant of risk factors and personalized the weight of various risk factors against the user's requirements using mapping techniques automatically. Secondly, a personalization mechanism that allows investors to interact with the property risk analysis system helps to reduce information overload, thus improving the quality of investor decisions. An adaptive personalization mechanism, based on data-driven approaches and TOPSIS model, are also integrated to set the risk level of a property investment. Thirdly, practically this study responds to the demand for investors to be able to make better decisions and gain hidden knowledge through a deterministic approach, as big data analytics becomes more important to businesses such as real estate.

APPIRA method proposed is based on a generalized framework for knowledge discovery in business environments by Fong and Hui (2010) The proposed framework uses an application that integrates data mining processes with online analytical processing (OLAP) to present an easy-to-use solution that can be used by people who do not necessarily have much IT knowledge and still be able to perform the tasks of knowledge discovery effectively. It extends existing generalized frameworks to meet investors' goals (Fong & Hui, 2010).

It is recommended that the decision maker must be able to identify their specific goals and limitations as these requirements are a vital input for the risk analysis measurement. Moreover, the features of property data must be made available and up-to-date to achieve optimal results and strategic decision-making.

The APPIRA method allows users to personalize the weight and rank of risk factors based on a set of search results. These results act as a guideline, or reference, and assist the investor to personalize the rank and weight of each risk factor, based on discovered data patterns, so the results more closely match their goals and limitations. The data patterns are derived from the database through knowledge discovery, using a multidimensional data model, which is more accurate because all property-related factors are considered. Furthermore, the decision-making process is made faster and easier as long as high-level data are available. Thus, the APPIRA method provides simple and concise property analyses to match investors' requirements.

Table 7.1 compares the features and innovations of the APPIRA method with published AHP and ANP techniques.

Table 7-1: Comparison between AHP, ANP and the proposed APPIRA method

Features	**Method**		
	AHP	**ANP**	**APPIRA**
Risk identification (property features set as risk factors determinants)	Expert survey (human)	Expert survey (human)	Data-driven + adaptive personalization (digital)
Risk estimation (weight the determinants of risk factors)	Expert survey (human)	Expert survey (human)	Data-driven + adaptive personalization (digital)
Risk Assessment Model (Mapping techniques using deterministic approach integrated with personalization techniques)	Hierarchical data model	Network data model	Multidimensional data model + data-driven + adaptive personalization (digital)

The APPIRA method, AHP and ANP are all designed to solve complex multi-criteria decision problems. The comparison and advances of the APPIRA method over AHP/ANP are:

- Risk identification: Delphi technique, both AHP and ANP use expert surveys, the APPIRA method uses data-driven and personalization.

- Risk estimation: Delphi technique, both AHP and ANP use expert surveys, the APPIRA method uses data-driven and personalization.
- Risk analysis: Delphi technique uses multiple rounds of questionnaires sent to a panel of experts, AHP uses a hierarchical data network, while ANP uses a network data model. The APPIRA method uses an adaptive personalized data model reflected using heuristic/analytic decision-making processes.

7.2 Future Studies

Future work will focus on the development of a new hybrid method that can, independent of human judgment, understand and assess the impact of uncertainty in risk factors. Additionally, a deterministic approach, using an intelligent decision support system that can be used for handling risk-based decision-making in business operations, is also an essential area to be explored. For this reason, it is essential to explore new techniques for risk analysis using decision support technology such as IDSS for investment in the real estate industry and also social network contexts. Lastly, an effective tool for automating risk analysis and expanding these analyses to larger markets is needed.

An important extension of this approach, a subject of future work, is to develop the new hybrid method to understand and assess the sensitivity of the uncertainty of risk factors which is independent from human judgment for risk estimation. Moreover, the deterministic approach using an intelligent decision support system that can be used for handling risk-based decision-making in business operations is also an essential area to be explored. Therefore, a strong tool in analyzing the real estate market can be used in the automation of evaluations and the extension of the analysis on larger markets is needed.

References

Almeida, R. J., Baştürk, N., & Golan, R. (2017). *Intraday value-at-risk estimation for directional change events and investment strategies.* Paper presented at the 2017 IEEE Symposium Series on Computational Intelligence (SSCI), Honolulu, USA.

Analytics Vidhya. (2020). Build a Decision Tree in Minutes using WEKA (No Coding Required!). Retrieved 5 October 2020 at https://www.analyticsvidhya.com/blog/2020/03/decision-tree-weka-no-coding/#:~:text=Visualizing%20your%20Decision%20Tree%20in,Choose%20the%20%E2%80%9CVisualise%20tree%E2%80%9D%20option

Angelou, G. N., & Economides, A. A. (2009). A multi-criteria game theory and real-options model for irreversible ICT investment decisions. *Telecommunications Policy, 33*(10-11), 686-705. doi:https://doi.org/10.1016/j.telpol.2009.07.005

Arentze, T. A. (2013). Adaptive Personalized Travel Information Systems: A Bayesian Method to Learn Users' Personal Preferences in Multimodal Transport Networks. *IEEE Transactions on Intelligent Transportation Systems, 14*(4), 1957-1966. doi:10.1109/TITS.2013.2270358

Aven, T. (2007). A unified framework for risk and vulnerability analysis covering both safety and security. *Reliability Engineering & System Safety, 92*(6), 745 - 754. doi:https://doi.org/10.1016/j.ress.2006.03.008

Aven, T., Vinnem, J. E., & Wiencke, H. S. (2007). A decision framework for risk management, with application to the offshore oil and gas industry. *Reliability Engineering & System Safety, 92*(4), 433 - 448. doi:https://doi.org/10.1016/j.ress.2005.12.009

Beracha, E., Downs, D. H., & MacKinnon, G. (2017). The 4% rule: Does real estate make a difference? *Journal of Property Research, 34*(3), 181-210. doi:10.1080/09599916.2017.1293134

Bergmann, P., Kamaras, E., Gleibner, W., & Guenther, E. (2020). Enhanced Cash Flow Valuation in Real Estate Management by Integrating Innovative Materials and Risk Assessment. *Sustainability,* 12, 2201-2226. doi:10.3390/su12062201

Bolger, F., & Wright, G. (1994). Assessing the quality of expert judgment: Issues and analysis. *Decision Support Systems, 11*(1), 1-24. doi:https://doi.org/10.1016/0167-9236(94)90061-2

Busemeyer, J., & Pleskac, T. (2009). Theoretical tools for understanding and aiding dynamic decision making. *Journal of Mathematical Psychology, 53*, 126-138. doi:10.1016/j.jmp.2008.12.007

Caron, E. A. M., & Daniels, H. (2008). Explanation of exceptional values in multi-dimensional business databases. *European Journal of Operational Research, 188*, 884-897. doi:10.1016/j.ejor.2007.04.039

Cassaigne, N., Lorimier, L., Gupta, J., A. Forgionne, G., & Mora T, M. (2006). A Challenging Future for i-DMSS. 401-422. doi:10.1007/1-84628-231-4_21

Chaoyang, J. (2013). The application of web data mining in personalized learning systems. *Journal of Chemical and Pharmaceutical Research, 5*(12), 865-869.

Chauveau, T., & Gatfaoui, H. (2002). Systematic risk and idiosyncratic risk: a useful distinction for valuing European options. *Journal of Multinational Financial Management, 12*(4), 305 - 321. doi:https://doi.org/10.1016/S1042-444X(02)00013-0

Chen, C., Lo, K., Tsang, D., & Zhang, J. (2020). Understanding accounting discretion in China: An analysis of fair value reporting for investment property. *Journal Account Public Policy*, 39, 106766.

Chen, L. (2010). *Research on the risk identification and evaluation of real estate development.* Paper presented at the 2nd International Conference on Information Science and Engineering (ICISE).

Chen, Z., & Khumpaisal, S. (2009). An analytic network process for risks assessment in commercial real estate development. *Journal of Property Investment & Finance, 27*(3), 238-258. doi:10.1108/14635780910951957

Chong, W., Guo, Y., & Wang, D. (2008). Study on Capital Risk Assessment Model of Real Estate Enterprises Based on Support Vector Machines and Fuzzy Integral. *Control and Decision Conference*, 2317 - 2320. doi:10.1109/CCDC.2008.4597737

Cong-Cong, L., Dong, Y., Xu, Y., Chiclana, F., Herrera-Viedma, E., & Herrera, F. (2019). An overview on managing additive consistency of reciprocal preference relations for consistency-driven decision making and fusion: Taxonomy and future directions. *Information Fusion, 52*, 143 - 156. doi:https://doi.org/10.1016/j.inffus.2018.12.004

Crosby, N., & Henneberry, J. (2016). Financialisation, the valuation of investment property and the urban built environment in the UK. *Urban Studies, 53*(7), 1424-1441. doi:10.1177/0042098015583229

D'Alpaos, C., & Canesi, R. (2014). Risks assessment in real estate investments in times of global crisis. *WSEAS Transactions on Business and Economics, 11*, 369-379.

D'Alpaos, C., & Canesi, R. (2015). Risks Assessment In Real Estate Investments: An AHP Approach.

D'Urso, V. T. (2002). Home Buyer Search Duration and the Internet(MIT Sloan School Working Paper No. 4271-02). Retrieved from https://ssrn.com/abstract=360460 or http://dx.doi.org/10.2139/ssrn.360460

Data Terbuka. (2020). Anggaran Pendapatan Dan Pemilikan Rumah Di Kalangan Warga KPKT Retrieved 5 October 2020 at https://www.data.gov.my/data/ms_MY/dataset/anggaran-pendapatan-dan-pemilikan-rumah-di-kalangan-warga-kpkt

Delen, D., & B. Pratt, D. (2006). An integrated and intelligent DSS for manufacturing systems. *Expert Systems with Applications, 30*, 325-336. doi:10.1016/j.eswa.2005.07.017

Demong, N. A. R., & Lu, J. (2011). *Personalized Multidimensional Process Framework For Dynamic Risk Analysis In The Real Estate Industry.* Paper presented at the 6th International Conference on Construction in the 21st Century, Malaysia.

Deng, C., & Ma, Y. (2008). Study of Chinese Real Estate Stock Daily Return Under Different Market Condition. doi:10.1109/WiCom.2008.2274

Ding, S., Wang, Z., Wu, D., & Olson, D. L. (2017). Utilizing customer satisfaction in ranking prediction for personalized cloud service selection. *Decision Support Systems, 93*, 1-10. doi:https://doi.org/10.1016/j.dss.2016.09.001

Dong, Y., & Li, J. (2010). Personalized distance education system based on Web mining. doi:10.1109/ICEIT.2010.5607495

Durica, M., Guttenova, D., Pinda, L., & Svabova, L. (2018). Sustainable Value of Investment in Real Estate: Real Options Approach. *Sustainability, 10*(12), 4665.

Ebert, S., Wei, W., & Zhou, X. Y. (2020). Weighted discounting - On group diversity, time-inconsistency, and consequences for investment, *Journal of Economic Theory,* 189, 105089. https://doi.org/10.1016/j.jet.2020.105089

Fang, C., & Marle, F. (2012). A simulation-based risk network model for decision support in project risk management. *Decision Support Systems, 52*(3), 635-644. doi:https://doi.org/10.1016/j.dss.2011.10.021

Fang, Y.-M., Lin, L.-Y., Huang, C.-H., & Chou, T.-Y. (2009). An integrated information system for real estate agency-based on service-oriented architecture. *Expert Systems with Applications, 36*(8), 11039-11044. doi:https://doi.org/10.1016/j.eswa.2009.02.082

Fong, A. C. M., & Hui, S. C. (2010). *A generalized framework for knowledge discovery in business environments.* Paper presented at the 2010 Second International Conference on Communication Systems, Networks and Applications, Hong Kong.

Gao, C., & Wang, J. (2010). *Direct mining of discriminative patterns for classifying uncertain data*. Paper presented at the Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, Washington, DC, USA.

Gao, J., & Wang, Z. (2009). Research on Real Estate Supply Chain Risk Identification and Precaution Using Scenario Analysis Method. doi:10.1109/ICIEEM.2009.5344449

Germanakos, P., Tsianos, N., Lekkas, Z., Mourlas, C., Belk, M., & Samaras, G. (2007). *A Semantic Approach of an Adaptive and Personalized Web-Based Learning Content-The Case of AdaptiveWeb.* Paper presented at the Second International Workshop on Semantic Media Adaptation and Personalization (SMAP 2007).

Glorio, O., Mazón, J.-N., Garrigós, I., & Trujillo, J. (2012). A personalization process for spatial data warehouse development. *Decision Support Systems, 52*(4), 884-898. doi:https://doi.org/10.1016/j.dss.2011.11.010

Gumparthi, S., & Venkatachalam, M. (2010). Risk classification based on discriminant analysis for smes. *International Journal of Trade, Economics and Finance, 1*, 242-246. doi:10.7763/IJTEF.2010.V1.44

Halbert, L., Henneberry, J., & Mouzakis, F. (2014). The Financialization of Business Property and What It Means for Cities and Regions. *Regional Studies: The Journal of the Regional Studies Association, 48*. doi:10.1080/00343404.2014.895317

Horeis, T., & Sick, B. (2007). Collaborative Knowledge Discovery & Data Mining: From Knowledge to Experience. *Proceedings of the 2007 IEEE Symposium on Computational Intelligence and Data Mining, CIDM 2007*, 421 - 428. doi:10.1109/CIDM.2007.368905

Huang, C.-C., Tseng, B., Li, M.-Z., & R. Gung, R. (2006). Models of multi-dimensional analysis for qualitative data and its application. *European Journal of Operational Research, 174*, 983-1008. doi:10.1016/j.ejor.2005.07.021

Hui, E. C. M., Yu, C. K. W., & Ip, W.-C. (2010). Jump point detection for real estate investment success. *Physica A: Statistical Mechanics and its Applications, 389*(5), 1055-1064. doi:https://doi.org/10.1016/j.physa.2009.11.022

Hui, S., Zhi Qing, F., & Ye, S. (2009). Study of impact of real estate development and management risk on economic benefit. doi:10.1109/ICIEEM.2009.5344419

Hussain, & Asghar, S. (2013). Web mining: approaches, applications and business intelligence. *International Journal of Academic Research, 5*(2), 2011-2217. doi:10.7813/2075-4124.2013/5-2/A.32

Hussain, O. K., Chang, E., Hussain, F. K., & Dillon, T. S. (2007). *Quantifying Failure for Risk Based Decision Making in Digital Business Ecosystem Interactions*. Paper presented at the Second International Conference on Internet and Web Applications and Services (ICIW'07), Morne.

Hutchison, N., Adair, A., & Leheny, I. (2005). Communicating Investment Risk to Clients: Property Risk Scoring. *Journal of Property Research, 22*, 137-161. doi:10.1080/09599910500453764

J. Williams, D., & M. Noyes, J. (2007). How does our perception of risk influence decision-making? Implications for the design of risk information. *Theoretical Issues in Ergonomics Science, 8*. doi:10.1080/14639220500484419

Jin, C.-h. (2010). RETRACTED ARTICLE: The lifecycle comprehensive risk evaluation for real estate projects-application. *ICAMS 2010 - Proceedings of 2010 IEEE*

International Conference on Advanced Management Science, 2. doi:10.1109/ICAMS.2010.5552958

Jin, C. H. (2010). *Lifecycle Comprehensive Risk Evaluation for Real Estate Projects – Principle*. Paper presented at the International Conference on Management and Service Science (MASS), Wuhan, China.

Juhong, G., & Zihan, W. (2009). *Research on real estate supply chain risk identification and precaution using Scenario analysis method.* Paper presented at the 2009 16th International Conference on Industrial Engineering and Engineering Management, Beijing.

Kao, S.-C., Tseng, Y.-F., & Lee, T.-Z. (2010). The design of personalized knowledge integration platform using digitized information resources. doi:10.1109/ICCIE.2010.5668376

Karacapilidis, N., Gupta, J., A. Forgionne, G., & Mora T, M. (2006). An Overview of Future Challenges of Decision Support Technologies. 385-399. doi:10.1007/1-84628-231-4_20

Kechadi, T., & Savvas, I. (2010). Cooperative Knowledge Discovery & Data Mining CKDD. 96-97. doi:10.1109/WETICE.2010.21

Khumpaisal, S., & Chen, Z. (2010). Risk assessment in real estate development: an application of analytic network process. *Journal of Architecture/Planning Research and Studies, 7*(1), 103-116.

Khumpaisal, S., Ross, A., & Abdulai, R. J. (2010). An examination of Thai practitioners' perceptions of risk assessment techniques in real estate development projects. *Journal of Retail and Leisure Property, 9*(2). doi:10.1057/rlp.2010.3

Kit, P. W. (2007). The effect of uncertainty on investment timing in a real options model. *Journal of Economic Dynamics and Control, 31*, 2152-2167. doi:10.1016/j.jedc.2006.07.002

Koller, G. (2005). *Risk Assessment and Decision Making in Business and Industry: A Practical Guide* (Second ed.): Chapman and Hall/CRC

Krishna, V., Jose, J., & Suri, N. N. R. R. (2014). Design and development of a web-enabled data mining system employing JEE technologies. *Indian Academy of Sciences, 39*(6), 1259-1270.

Kwon, K., & Kim, C. (2012). How to design personalization in a context of customer retention: Who personalizes what and to what extent? *Electronic Commerce Research and Applications, 11*(2), 101-116. doi:https://doi.org/10.1016/j.elerap.2011.05.002

Lafleur, J. M. (2011). *Probabilistic AHP and TOPSIS for multi-attribute decision-making under uncertainty.* Paper presented at the 2011 Aerospace Conference, Big Sky, MT.

Lee, J.-S., & Jang, S. (2007). The systematic-risk determinants of the US airline industry. *Tourism Management, 28*(2), 434 - 442. doi:https://doi.org/10.1016/j.tourman.2006.03.012

Li, J., Wang, J.-f., Wu, C.-Y., Yan-tao, Y., Ji, Z.-t., & Wang, H.-b. (2007). Establishment of a Risk Assessment Framework for Analysis of the Spread of Highly Pathogenic Avian Influenza. *Agricultural Sciences in China, 6*(7), 877 - 881. doi:https://doi.org/10.1016/S1671-2927(07)60125-4

Li, W., Zhao, Y., Meng, W., & Xu, S. (2009). Study on the Risk Prediction of Real Estate Investment Whole Process Based on Support Vector Machine. 167-170. doi:10.1109/WKDD.2009.40

Li, Y., & Suo, J. (2009). Model on Risk Evaluation of Real Estate Investment. 138-140. doi:10.1109/FSKD.2009.60

Liang, T.-P., Lai, H.-J., & Ku, Y.-C. (2007). Personalized Content Recommendation and User Satisfaction: Theoretical Synthesis and Empirical Findings. *J. of Management Information Systems, 23*, 45-70. doi:10.2753/MIS0742-1222230303

Lieser, K., & Groh, A. P. (2014). The Determinants of International Commercial Real Estate Investment (April 29, 2014). Vol. 48, No. 4, 2014. Available at SSRN: . *Journal of Real Estate Finance and Economics,, 48*(4), 611-659.

Lin, C., Meng, L., & Pan, H. (2001). Applications and Research on Gis for the Real Estate.

Lin, H., & Chen, X.-p. (2008). Is default risk a systematic risk of Chinese stock markets? *2008 International Conference on Management Science and Engineering 15th Annual Conference Proceedings, ICMSE*. doi:10.1109/ICMSE.2008.4669058

Lin, R.-S., Wang, Q.-Y., Hu, J.-H., Gao, L., & Lu, L. (1996). *An intelligent decision support system applied to the investment of real estate.* Paper presented at the IEEE International Conference on Industrial Technology (ICIT'96), Shanghai, China, China.

Lin, Y.-H., Lee, P.-C., & Ting, H.-I. (2008). Dynamic multi-attribute decision making model with grey number evaluations. *Expert Systems with Applications, 35*, 1638-1644. doi:10.1016/j.eswa.2007.08.064

Liow, K. H., & Kwame.Addae-Dapaah. (2010). Idiosyncratic risk, market risk and correlation dynamics in the US real estate investment trusts. *Journal of Housing Economics, 19*(3), 205 - 218. doi:https://doi.org/10.1016/j.jhe.2010.06.001

Liu, L., Zhao, E., & Liu, Y. (2007). *Research into the Risk Analysis and Decision-Making of Real Estate Projects.* Paper presented at the International Conference on Wireless Communications, Networking and Mobile Computing.

Liu, T., & Liu, Y. (2010). A Risk Early Warning Model in Real Estate Market Based on Support Vector Machine. doi:10.1109/ICICCI.2010.65

Lobur, M., Stekh, Y., Kernytskyy, A., & Sardieh, F. M. E. (2008). *Some trends in Knowledge Discovery and Data Mining.* Paper presented at the 2008 International Conference on Perspective Technologies and Methods in MEMS Design.

Lu, J., Zhang, G., Ruan, D., & Wu, F. (2007). *Multi-Objective Group Decision Making: Methods, Software and Applications with Fuzzy Set Techniques*. UK: Imperial College Press.

Lu, Xiaomeng, Zhang, Y., Zhang, X. & Wang, L. (2020). Can investment advisors promote rational investment? Evidence from micro-data in China. *Economic Modelling*, 86, 251-263.

Ma, Z.-q., & Meng, Q.-b. (2009). The Research on Risk Evaluation of Real Estate Development Project Based on RBF Neural Network. doi:10.1109/ICIC.2009.180

Meng, Z.-q., Xiao-fen.Yu, Jiang, M., & Gao, H. (2007). Risk Measure and Control Strategy of Investment Portfolio of Real Estate based on Dynamic CVaR Model. *Systems Engineering - Theory & Practice, 27*(9), 69 - 76. doi:https://doi.org/10.1016/S1874-8651(08)60058-7

Mintah, K. (2016). *'Real options and application to Australian property development: A conceptual analysis.* Paper presented at the Proceedings of the 22nd Annual Pacific-Rim Real Estate Society Conference (PRRES 2016), Sunshine Coast, Australia.

Montgomery, A. L., & Smith, M. D. (2009). Prospects for Personalization on the Internet. *Journal of Interactive Marketing, 23*(2), 130-137.

Mora, M., Forgionne, G., Gupta, J., Garrido, L., Cervantes-Pérez, F., & Gelman, O. (2006). A Strategic Descriptive Review of the Intelligent Decision-making Support Systems Research: the 1980–2004 Period. doi:10.1007/1-84628-231-4_23

Morente-Molinera, J. A., Kou, G., Pang, C., Cabrerizo, F. J., & Herrera-Viedma, E. (2019). An automatic procedure to create fuzzy ontologies from users' opinions using sentiment analysis procedures and multi-granular fuzzy linguistic modelling methods. *Information Sciences, 476*, 222-238. doi:doi.org/10.1016/j.ins.2018.10.022

Mustajoki, J., & P. Hämäläinen, R. (2007). Smart-Swaps - A decision support system for multicriteria decision analysis with the even swaps method. *Decision Support Systems, 44*, 313-325. doi:10.1016/j.dss.2007.04.004

Niu, L., Lu, J., & Zhang, G. (2009). *Cognition-Driven Decision Support for Business Intelligence*.

Olsson, R. (2007). In search of opportunity management: Is the risk management process enough? *International Journal of Project Management, 25*(8), 745 - 752. doi:https://doi.org/10.1016/j.ijproman.2007.03.005

Park, S., D'Angelo, C., & Gunashekar, S. (2018). *Citizen science: Generating ideas and exploring concensus*. Retrieved from

Peng, Y., Zhang, Y., Tang, Y., & Li, S. (2011). An incident information management framework based on data integration, data mining, and multi-criteria decision making. *Decision Support Systems, 51*(2), 316-327. doi:https://doi.org/10.1016/j.dss.2010.11.025

Pires, A. S. C., Ferrerira, F. A. F., Jalali, M. S., & Chang, H. C. (2018). Barriers to real estate investments for residential rental purposes: Mapping out the problem. International Journal of Strategic Property Management, 22 (3), 168-178. https://doi.org/10.3846/ijspm.2018.1541.

Pivo, G. (2008). Responsible property investment criteria developed using the Delphi Method. *Building Research and Information, 36*, 20-36. doi:10.1080/09613210701574795.

Piyatrapoomi, A., Kumar, A., & Setunge, S. (2004). Framework for Investment Decision-Making under Risk and Uncertainty for Infrastructure Asset Management. *Research in Transport in Economics, 8*(10), 193-209.

Piyatrapoomi, N., Kumar, A., & Setunge, S. (2004). Framework for Investment Decision-Making under Risk and Uncertainty for Infrastructure Asset Management. *Research in Transportation Economics, 8*, 199 - 214. doi:https://doi.org/10.1016/S0739-8859(04)08010-2

Pomerol, J.-C. (2001). Scenario development and practical decision making under uncertainty. *Decision Support Systems, 31*(2), 197-204. doi:https://doi.org/10.1016/S0167-9236(00)00131-7

Qi, L. (2008). Advancing Knowledge Discovery and Data Mining. *Proceedings - 1st International Workshop on Knowledge Discovery and Data Mining, WKDD*, 3 - 5. doi:10.1109/WKDD.2008.153

Ren, H., & Yang, X. (2009). *Risk Measurement of Real Estate Portfolio Investment Based on CVaR Model*.

Rinner, C., & Martin, R. (2004). Personalized Multi-Criteria Decision Strategies in Location-Based Decision Support. *Geography Publications and Research, 10*. doi:10.1080/10824000409480666

Rocha, K., Salles, L., Garcia, F. A. A., Sardinha, J. A., & Teixeira, J. P. (2007). Real estate and real options — A case study. *Emerging Markets Review, 8*(1), 67-79. doi:https://doi.org/10.1016/j.ememar.2006.09.008

Roszkowska, E. (2011). Multi-Criteria Decision Making Models By Applying The Topsis Method To Crisp And Interval Data. *Multiple Criteria Decision Making, 6*, 200-230.

Saleem, K., & Vaihekoski, M. (2008). Pricing of global and local sources of risk in Russian stock market. *Emerging Markets Review, 9*(1), 40 - 56. doi:https://doi.org/10.1016/j.ememar.2007.08.002

Sdino, L., Rosasco, P., & Magoni, S. (2018). Real Estate Risk Analysis: The Case of Caserma Garibaldi in Milan. *International Journal of Financial Studies, 6*. doi:10.3390/ijfs6010007

Shi, T., & Zhu, J. (2009). Demonstration of Book Issue Chain Decision Support and Multi-dimensional Data Model in E-commerce Era. 68 - 71. doi:10.1109/WCSE.2009.20

Shinzato, T. (2018). Maximizing and minimizing investment concentration with constraints of budget and investment risk. *Physica A: Statistical Mechanics and its Applications, 490*, 986-993. doi:https://doi.org/10.1016/j.physa.2017.08.088

Shiwang, Y., Jianping, W., & Na, G. (2009). *Application of project portfolio management in the real estate corporations.* Paper presented at the 16th International Conference on Industrial Engineering and Engineering Management, Beijing.

Software testing help. (2020). WEKA Dataset, Classifier And J48 Algorithm For Decision Tree. Retrieved 5 October 2020 at https://www.softwaretestinghelp.com/weka-datasets/

Soujanya, M., & Kumar, S. (2010). Personalized IVR system in Contact Center. *1*. doi:10.1109/ICEIE.2010.5559673

Stair, R., & Reynolds, G. (2008). *Fundamentals of Information Systems* (5th ed.): Cengage Learning.

Storesletten, K., Telmer, C. I., & Yaron, A. (2007). Asset pricing with idiosyncratic risk and overlapping generations. *Review of Economic Dynamics, 10*(4), 519 - 548. doi:https://doi.org/10.1016/j.red.2007.02.004

Strong, R., Paasi, J., Zhou, R., & Rantala, T. (2009). Systematic Risk Management for the Innovative Enterprise. 1 - 9. doi:10.1109/HICSS.2009.414

Sumon Shahriar, M., & Anam, S. (2008). Quality Data for Data Mining and Data Mining for Quality Data: A Constraint Based Approach in XML. *Future Generation Communication and Networking Symposia, International Conference on, 2*, 46-49. doi:10.1109/FGCNS.2008.74

Sun, H., Fan, Z.-q., & Shi, Y. (2009). Study of impact of real estate development and management risk on economic benefit. doi:10.1109/ICIEEM.2009.5344419

Sun, Y., Huang, R., Chen, D., & Li, H. (2008). Fuzzy set-based risk evaluation model for real estate projects. *Tsinghua Science and Technology, 13*(1), 158-164. doi:10.1016/S1007-0214(08)70143-3

Syalim, A., Hori, Y., & Sakurai, K. (2009). Comparison of Risk Analysis Methods: Mehari, Magerit, NIST800-30 and Microsoft's Security Management Guide. 726 - 731. doi:10.1109/ARES.2009.75

Tang, G., & Cheong Shum, W. (2003). The relationships between unsystematic risk, skewness and stock returns during up and down markets. *International Business Review, 12*, 523-541. doi:10.1016/S0969-5931(03)00074-X

Tang, X.-L., & Liu, G.-D. (2009). A tentative study on risk estimation and evaluation of investment in real estate projects. 561-566. doi:10.1109/ICMLC.2009.5212531

Thilini, M., & Wickramaarachchi, N. (2019). Risk assessment in commercial real estate development: An application of analytic network process. *Journal of Property Investment & Finance, 37*(5), 427-444.

Vahdatmanesh, M., & Firouzi, A. (2017). Real Options Valuation for Residential Real Estate Development Projects.

Vesanen, J. (2007). What is personalization? A conceptual framework. *European Journal of Marketing, 41*(5/6), 409-418. doi:doi:10.1108/03090560710737534

Wang, G., Hu, J., Zhang, Q., Xianquan, L., & Zhou, J. (2008). Granular Computing Based Data Mining in the Views of Rough Set and Fuzzy Set. *Novel Developments in Granular Computing: Applications for Advanced Human Reasoning and Soft Computation*, 67 - 67. doi:10.1109/GRC.2008.4664791

Wang, K., Pao, H.-T., & Fu, H. C. (2000). Using statistical and neural network methods to explore the relationship between systematic risk and firm's long term investment activities. *2*, 851 - 858 vol.852. doi:10.1109/NNSP.2000.890165

Westwood Net Lease Advisors. (2020). What is a Sensitivity Analysis? Retrieved 10 October 2020 at https://westwoodnetlease.com/what-is-a-sensitivity-analysis/

Wilhelmsson, M., & Zhao, J. (2018). Risk Assessment of Housing Market Segments: The Lender's Perspective. *Journal of Risk and Financial Management,* 11(69), 1-22. doi.10.3390/jrfm11040069.

Xiaoyan, W., & Ping, D. (2011). *A study of personalized learning system based on XML and WEB mining.* Paper presented at the Proceedings of 2011 Cross Strait Quad-Regional Radio Science and Wireless Technology Conference, Harbin.

Xie, G., Yue, W., Wang, S., & Lai, K. K. (2010). Dynamic risk management in petroleum project investment based on a variable precision rough set model. *Technological Forecasting and Social Change, 77*, 891-901. doi:10.1016/j.techfore.2010.01.013

Xiong, X., Zhang, X.-T., Zhang, W., & Li, C.-Y. (2005). Wavelet-based Beta estimation of China stock market. 3501 - 3505 Vol. 3506. doi:10.1109/ICMLC.2005.1527548

Xiong, Y.-n., & Geng, L.-x. (2010). Personalized Intelligent Hotel Recommendation System for Online Reservation--A Perspective of Product and User Characteristics. 1 - 5. doi:10.1109/ICMSS.2010.5576790

Yadav, V., Karmakar, S., Kalbar, P. P., & Dikshit, A. K. (2019). PyTOPS: A Python based tool for TOPSIS. *Software X, 9*, 217-222. doi:https://doi.org/10.1016/j.softx.2019.02.004

Yan, L., & Lei, H. (2008). Financial constraints and systematic risk: Theory and evidence from China. *2008 International Conference on Management Science and Engineering 15th Annual Conference Proceedings, ICMSE*. doi:10.1109/ICMSE.2008.4669050

Yang, X. (2008). Real Estate Financial Risk Prevention and Countermeasure. doi:10.1109/WiCom.2008.2286

Ye, X. (2011). Risk Analysis in the Process of Real Estate Enterprise Project Investment. 731 - 736. doi:10.1109/CSO.2011.227

Yu, J., & Xuan, H. (2010a). *'Study of a Practical Method for Real Estate Investment Risk Decision Making.* Paper presented at the International Conference on Management and Service Science (MASS), Wuhan, China.

Yu, J., & Xuan, H. (2010b). Study of a Practical Method for Real Estate Investment Risk Decision Making. doi:10.1109/ICMSS.2010.5577537

Yuan, X., Lee, J.-H., Kim, S.-J., & Kim, Y.-H. (2013). Toward a user-oriented recommendation system for real estate websites. *Information Systems, 38*(2), 231-243. doi:https://doi.org/10.1016/j.is.2012.08.004

Zalewska-Turzyńska, M. M. (2016). The Communication Patterns of Leaders And Employees. Evidence From Medium And Large Enterprises. *Academy of Contemporary Research Journal*, 1-6.

Zeng, J., An, M., & Smith, N. J. (2007). Application of a fuzzy based decision making methodology to construction project risk assessment. *International Journal of Project Management, 25*(6), 589-600. doi:https://doi.org/10.1016/j.ijproman.2007.02.006

Zhai, B., Chen, H., & Chen, A. (2018). Study on investment risk asessement model of real estate project based on Monte Carlo Method. 2018 International Conference on Civil and Hydraulic Engineering (IComCHE 2018). doi:10.1088/1755-1315/189/4/042016

Zhang, M., Guo, X., & Chen, G. (2016). Prediction uncertainty in collaborative filtering: Enhancing personalized online product ranking. *Decision Support Systems, 83*, 10-21. doi:https://doi.org/10.1016/j.dss.2015.12.004

Zhang, P., & Zeng, Z. (2006). *A Framework for Personalized Service Website based on TAM.* Paper presented at the 2006 International Conference on Service Systems and Service Management.

Zhang, X., & Ji, W. (2010). Study on the Risk Assessment of Real Estate Project Based on BP Neural Network. doi:10.1109/ICIFE.2010.5609416

Zhang, X., & Li, H. (2009). *Extension Evaluation on Risks in Earlier Stage of Real Estate Project.* Paper presented at the International Conference on Industrial Engineering and Engineering Management.

Zheng, J., Yao, J., & Niu, J. (2008). Web user de-identification in personalization. 1081-1082. doi:10.1145/1367497.1367666

Zhou, J., Wu, Z., & Xin, X. (2006). *Recognizing the Pattern of Systematic Risk Based on Financial Ratios and Rough Set-Neural Network System.* Paper presented at the 2006 International Conference on Machine Learning and Cybernetics, Dalian, China.

Zhou, S., & Li, S. (2010). Projection Pursuit Model Based on PSO in the Real Estate Risk Evaluation. *Intelligent Computing and Cognitive Informatics, International Conference on, 0*, 235-238. doi:10.1109/ICICCI.2010.48

Zhou, S., Wang, F., & Li, Y. (2010). *Risk assessment of real estate investment.* Paper presented at the 2010 2nd International Asia Conference on Informatics in Control, Automation and Robotics (CAR 2010).

Zhou, S. J., Li, Y. C., & Zang, Z. D. (2008). *Self-adaptive Ant Algorithm for Solving Real Estate Portfolio Optimization.* Paper presented at the 2008 International Symposium on Computational Intelligence and Design, Wuhan.

Abbreviations

AHP	**A**nalytical **H**ierarchy **P**rocess
AI	**A**rtificial **I**ntelligent
ANP	**A**nalytic **N**etwork **P**rocess
APM	**A**ustralian **P**roperty **M**onitor
APPIRA	**A**daptive **P**ersonalized **P**roperty **I**nvestment **R**isk **A**nalysis
APPIRAS	**A**daptive **P**ersonalized **P**roperty **I**nvestment **R**isk **A**nalysis **S**ystem
APT	**A**rbitrage **P**ricing **T**heory
ARIMA	**A**uto**R**egressive **I**ntegrated **M**oving **A**verage
ARTXP	**A**uto-**R**egressive **T**ree With **C**ross **P**rediction
CAPM	**C**apital **A**sset **P**ricing **M**odel
CVaR	**C**ondition **V**alue-**a**t-**R**isk
DCF	**D**iscounted **C**ash **F**low
DSS	**D**ecision **S**upport **S**ystems
EIS	**E**xecutive **I**nformation **S**ystems
ELECTRE	**El**imination and **C**hoice **E**xpressing **Re**ality
F-AHP	**F**uzzy-**A**nalytical **H**ierarchy **P**rocess
GDP	**G**ross **D**omestic **P**roduct
HPAI	**H**ighly **P**athogenic **A**vian **I**nfluenza
ICT	**I**nformation and **C**ommunication **T**echnology
i-DMSS	**i**ntelligent **Decision-making** **S**upport **S**ystem
IDSS	**I**ntelligent **D**ecision **S**upport **S**ystem
KDD	**K**nowledge **D**iscovery and **D**ata Mining
LLN	**L**aw of **L**arge **N**umbers
MAM	**M**ulti-element **A**nalysis **M**odel
MCDM	**M**ulti-**C**riteria **Decision-making**
MIS	**M**anagement **I**nformation **S**ystems
MODWT	**M**aximal **O**verlap **D**iscreet **W**avelet **T**ransform
NAR	**N**ational **A**ssociation of **R**ealtors
OLAP	**O**n**l**ine **A**nalytical **P**rocessing
OLTP	**O**n**l**ine **T**ransaction **P**rocessing

PAM	**P**ersonalized **A**ssociation **M**apping
PMP	**P**ersonalized **M**ultidimensional **P**rocess
PM-SA	**P**ersonalized **M**ultidimensional **S**ensitivity **A**nalysis
PSO	**P**article **S**warm **O**ptimization
RBDM	**R**isk-**B**ased **Decision-making**
RBF	**R**adial **B**asis **F**unction
RE	**R**isk **E**xposure
RPI	**R**esponsible **P**roperty **I**nvesting
SAW	**S**imple **A**dditive **W**eighting
SEET	**S**ocial, **E**conomic, **E**nvironmental, and **T**echnological
SMART	**S**imple **M**ulti **A**ttribute **R**ating **T**echnique
SVM	**S**upport **V**ector **M**achine
TOPSIS	**T**echnique for **O**rder **P**reference by **S**imilarity to the **I**deal **S**olution
UML	**U**nified **M**odeling **L**anguage
UTS	**U**niversity of **T**echnology **S**ydney 3
VPRS	**V**ariable **P**recision **R**ough **S**et
WBS	**W**ork **B**reakdown **S**tructure
WEKA	**W**aikato **E**nvironment for **K**nowledge **A**nalysis