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

TABLE OF CONTENTS i
DEDICATION vi
ACKNOWLEDGMENTS vii
LIST OF FIGURES viii
LIST OF TABLES x
Abstract xi
CHAPTER 1 1
1. INTRODUCTION 1
1.1 Background 1
1.2 Research Questions 9
1.3 Research Objectives 11
1.4 Research Outcomes 14
1.5 Significance and Innovations 17
1.5.1 Significance 17
1.5.2 Innovations 19
1.6 Research Methodology 21
1.6.1 Problem Definition 23
1.6.2 Research Planning 23
1.6.3 Method Development 24
1.6.4 Analysis and Validation 25
1.6.5 Evaluation and Revision 29
1.7 Thesis Structure 29
1.8 Publications Related to This Thesis 30
CHAPTER 2 31
2. LITERATURE REVIEW 31
2.1 Property Investment Risk Analysis 31
2.1.1 Dynamic Risk-based Decision-making 32
2.1.2 Types of Risk-based Decision-making Process 37
2.1.3 Decision under Uncertainty 39
2.1.4 Risk Sources and Risk Factors in the Real Estate Industry 40
2.1.5 Existing techniques/models for property investment risk analysis 46
2.1.6 Existing real estate analytic systems 51
2.2 Data-driven Decision-making Approach 53
2.2.1 Knowledge Discovery and Data Mining 54
2.2.2 Intelligent Decision Support System (IDSS) for Risk-based Decision-making 61

2.2.3 A generalized framework for knowledge discovery in business environments 62
2.3 Personalization Technique 64
2.3.1 Adaptive personalization 65
2.4 Multi-criteria Decision-making Techniques 66
2.4.1 Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) Approach 66
2.5 Summary 71

CHAPTER 3 73

3. AN ADAPTIVE PERSONALIZED PROPERTY INVESTMENT RISK ANALYSIS FRAMEWORKS 73

3.1 Introduction 73
3.2 Personalized Multidimensional Process Framework 74
3.2.1 Nine Steps of Personalized Multidimensional Process Framework 76
3.3 Risk-based Decision-making Framework 78
3.3.1 Risk-based Decision-making Concepts 80
3.3.2 Risk-based Decision-making Process 82
3.3.3 Risk Sources and Risk Factors in the Real Estate Industry 84
3.3.4 Risk-based Decision-making Techniques for Real Estate Project Investment 85
3.3.5 Issues and Challenges of Risk-Based Decision-making 91
3.4 Multidimensional Analysis and Data Mining Framework for Property Investment Risk Analysis 93
3.5 Experiment 97
3.5.1 Dataset 98
3.5.2 Application of Risk-based Decision-Making Framework via Select Attribute Function. 104
3.5.3 CfsSubsetEval (CSE) 105
3.5.4 Information Gain Attribute Evaluation 106
3.6 Summary 107

CHAPTER 4 108

4. AN ADAPTIVE PERSONALIZED PROPERTY INVESTMENT RISK ANALYSIS METHOD 108

4.1 Introduction 108
4.2 The Adaptive Personalized Property Investment Risk Analysis (APPIRA) Method 109
4.2.1 Adaptive personalization module 111
4.2.2 Risk analysis module 112
4.2.3 Determinant module 112
4.2.4 Knowledge discovery module using a data-driven approach 113
4.3 Personalized Association Mapping Method 113

4.3.1 Personalization 115
4.3.2 Personalization Model 118
4.3.3 Personalized Property Investment Risk Analysis Model 120
4.3.4 Personalization Session for Risk Analysis 126
4.4 Personalized Multidimensional-Sensitivity Analysis Method 127
4.4.1 Development of Knowledge Using Data Mining Technology 128
4.4.2 Investment Risk Analysis Knowledge Management Development 130
4.4.3 Experiment 132
4.5 APPIRA Process Flow 136
4.5.1 Six Steps of APPIRA 137
4.5.2 TOPSIS method 141
4.5.3 Example 143
4.5.3.1 Decision Tree using J48 Machine Learning Model 144
4.5.3.2 Attribute Selection using Gain Ratio 147
4.6 Summary 149

CHAPTER 5 151

5. AN ADAPTIVE PERSONALIZED PROPERTY INVESTMENT RISK ANALYSIS SYSTEM 151

5.1 Introduction 151
5.2 General Specification 151
5.3 Main Features 153
5.3.1 Personalization technique 153
5.3.2 Data-driven approach 154
5.4 APPIRAS Pseudocode and Algorithm 155
5.5 APPIRAS Main Interfaces 163
5.6 Summary 166

CHAPTER 6 167

6. CASE STUDY 167

6.1 Introduction 167
6.2 Data analysis 167
6.2.1 Heuristic Approach 167
6.2.2 Deterministic Approach 168
6.3 Decision table technique 169
6.4 Data Mining Techniques for Knowledge Discovery 171
6.4.1 Clustering Technique 171
6.4.2 Predictive Technique for Rental Rate Forecasting 174
6.4.3 Predictive Technique for Median House Price Forecasting 176
6.4.4 Association/classification to define patterns of data 186
6.5 Summary 189

CHAPTER 7 190

7. CONCLUSIONS AND FUTURE STUDY 190
7.1 Conclusions 190
7.2 Future Studies 193
References 194
Abbreviations 208

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

Figure 1.1 Relationship between Research Objectives and Research Questions..... 14
Figure 1.2 Relationship between Research Questions and Research Outcomes. 17
Figure 1.3 Four main innovative results in this study. 21
Figure 1.4: Research methodology for this study. 22
Figure 2.1: Risk viewed as a combination of sources and vulnerability (Terje Aven, 2007). 33
Figure 2.2: Static (simultaneous) decision-making process (Rocha et al., 2007). 38
Figure 2.3: Dynamic (sequential) decision-making process (Rocha et al., 2007). 38
Figure 2.4: Knowledge Discovery in Database process (Qi, 2008). 55
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). 57
Figure 2.6: OLAP technical operation diagram (Huang et al., 2006). 58
Figure 2.7: Structure of multi-dimensional analysis technique (Huang et al., 2006). 58
Figure 2.8: Models of multi-dimensional analysis for quantitative data (Huang et al., 2006). 60
Figure 2.9: Models of multi-dimensional analysis for qualitative data (Huang et al., 2006). 61
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). 63
Figure 3.1: The relationships between three different frameworks of an adaptive personalized property investment risk analysis..... 74
Figure 3.2: A personalized multidimensional process (PMP) framework. 75
Figure 3.3: Three main inputs for the PMP framework. 76
Figure 3.4: A personalized multidimensional process (PMP) framework. 77
Figure 3.5: Flow chart for the PMP framework. 78
Figure 3.6: The structure of the proposed risk-based decision-making framework. 80
Figure 3.7: Risk-based decision-making main activity for investment in the real estate industry..... 84
Figure 3.8: The multidimensional analysis framework for property investment risk analysis. 95
Figure 3.9: Screenshot of recognizes attributes and the list of recognized attributes using WEKA application software. 98
Figure 3.10: Relationships between each attribute that contribute to the number of houses owned by the respondents (staffs of Ministry of Housing and Local Government in Malaysia in year 2018). 104
Figure 3.11: Ranking of attributes with respect to CfsSubsetEval method 105
Figure 3.12: Ranking of attributes with respect to GainRatio Attribute evaluation method 106
Figure 4.1: Four main modules in APPIRA method : adaptive personalization, risk analysis, determinants module, and knowledge discovery 110
Figure 4.2: The integration between heuristic and deterministic approaches. 113
Figure 4.3: The use case diagram of the adaptive personalized property investment risk analysis system. 115

Figure 4.4: The personalized property investment risk analysis model for the real estate industry..... 121
Figure 4.5: The transfer of explicit knowledge to the user as tacit knowledge. 126
Figure 4.6: Data flow of the personalization session for property investment risk analysis in the real estate industry..... 127
Figure 4.7: Several questions that investors will face when dealing with investment risk analysis in the real estate industry..... 129
Figure 4.8: Development of knowledge management by using heuristic through deterministic approach for investment risk analysis in the real estate industry..... 130
Figure 4.9.: Relationships between dimensions toward the class (output) namely total numbers of houses owned..... 133
Figure 4.10 : Ranking of attributes with respect to InformationGain Attribute evaluation method..... 134
Figure 4.11: Pattern analysis for the relationship between monthly instalments, household income towards the total number of houses owned 135
Figure 4.12: The process flow of the APPIRA method 136
Figure 4.13: Hierarchical structure of Investor A. 143
Figure 4.14: Result using J48 pruned tree machine learning model. 145
Figure 4.15: Decision tree for number of houses owned using J48 machine learning model. 146
Figure 4.16: Too populated decision tree using REPTree machine learning model..... 147
Figure 4.17: Decision tree after removing the attributes which are not required. 147
Figure 4.18: Ranking of attributes with respect to GainRatio Attribute evaluation method. 148
Figure 4.19: The relationship between both development of knowledge using data mining technology and investment risk analysis knowledge management..... 149
Figure 5.1: APPIRAS use case diagram. 152
Figure 5.2: User Registration Form 163
Figure 5.3: Screenshot of how the user enters their search criteria as described in Step 1 164
Figure 5.4: Decision tree example for dynamic risk analysis for property investment in the real estate industry..... 164
Figure 5.5: Result showing property details sorted by minimum price as described in Step 2..... 165
Figure 6.1: Summary of data selected based on Table 6.2..... 173
Figure 6.2: Clustering technique based on selected Eastlakes property data..... 173
Figure 6.3: Cluster profiles based on Figure 6.2. 174
Figure 6.4: Rental rate prediction until 2017, based on data selected in Table 6.2. 175
Figure 6.5: Median house price in Bellevue Hill, NSW (based on data available from 2004 – 2017). 179
Figure 6.6: The ARIMA filtering box interpretation. 183
Figure 6.7: Screenshot of actual median prices for type of property (Unit and House) in Bellevue Hill, NSW 183
Figure 6.8: Median house price (predicted for the year 2018 - 2028). 185
Figure 6.9: Pattern of Rental at Selected Suburb. 188

LIST OF TABLES

Table 2-1: A summary of risk factors that will affect the dynamic risk analysis for investment in the real estate industry. 44
Table 2-2: Applications of existing techniques/models in property investment risk analysis 47
Table 2-3: Comparison of quantitative and qualitative multi-dimensional analysis. 59
Table 2-4: Comparisons of static vs dynamic multi-dimensional analysis. 60
Table 3-1: Dataset attributes metadata. 99
Table 3-2: Dataset instances descriptive statistic. 101
Table 4-1: Ranking of attributes with respect to InformationGain Attribute evaluation method 134
Table 4-2: Crosstab analysis based on total number of monthly instalment and monthly household income towards total number of houses owned. 135
Table 5-1: Decision matrix of investor A. 156
Table 5-2: List of decision matrices with weight. 158
Table 6-1: A decision table showing a scenario of recommendations for a property investment based on a user's constraints and requirements. 170
Table 6-2: Sample data used for the analysis collected from the Australian Property Monitor domain database. 172
Table 6-3: Parameters of the testing data 176
Table 6-4: Sample of training data for analysis 177
Table 6-5: Algorithm Parameters 178
Table 6-6: Crystal ball report 179
Table 6-7: Forecast range of median house prices for Bellevue Hill, NSW, collected from Australian Property Monitor database 184
Table 6-8: Parameters of the testing data and the prototype. 187
Table 6-9: Summarized the pattern of data for property rental in different suburbs. 188
Table 7-1: Comparison between AHP, ANP and the proposed APPIRA method 192

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.