A multi-objective optimization of online real estate property search

Chit, Lin Su

2019

Chit, L. S. (2019). A multi-objective optimization of online real estate property search. Master's thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/136577>

<https://doi.org/10.32657/10356/136577>

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

Downloaded on 26 Apr 2023 22:43:25 SGT

NANYANG TECHNOLOGICAL UNIVERSITY

A MULTI-OBJECTIVE OPTIMIZATION OF ONLINE REAL ESTATE PROPERTY SEARCH

CHIT LIN SU

School of Computer Science and Engineering

2019

NANYANG TECHNOLOGICAL UNIVERSITY

A MULTI-OBJECTIVE OPTIMIZATION OF ONLINE REAL ESTATE PROPERTY SEARCH

A thesis submitted to the
Nanyang Technological University
in partial fulfilment of the requirement
for the degree of Master of Engineering

CHIT LIN SU

Supervisor: PROF. ONG YEW SOON

School of Computer Science and Engineering

2019

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research, is free of plagiarised materials, and has not been submitted for a higher degree to any other University or Institution.

15 November 2019

.....

Date

.....

CHIT LIN SU

Supervisor Declaration Statement

I have reviewed the content and presentation style of this thesis and declare it is free of plagiarism and of sufficient grammatical clarity to be examined. To the best of my knowledge, the research and writing are those of the candidate except as acknowledged in the Author Attribution Statement. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

16 November 2019

.....
Date

.....
Prof. ONG YEW SOON

Authorship Attribution Statement

(A) This thesis **does not** contain any materials from papers published in peer-reviewed journals or from papers accepted at conferences in which I am listed as an author.

15 November 2019

.....
Date

.....
CHIT LIN SU

ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincere gratitude to Prof. Ong Yew Soon of the School of Computer Science and Engineering at Nanyang Technological University, my supervisor, for his valuable guidance in the right direction of the dissertation work, and utmost kindness and encouragement given throughout my studies of Master of Engineering, and this dissertation work. His supervision and support gave me the motivation to complete this dissertation work. I am honoured to have the excellent opportunity to work under his supervision.

Secondly, I would like to show my great appreciation to Dr. Abhishek Gupta of the Singapore Institute of Manufacturing Technology, for his invaluable guidance and feedback provided throughout my research works. His research works in the areas of optimization and machine learning gave me a great experience study of multi-objective optimization and its application in the real estate industry.

Thirdly, I would like to thank Assistant Prof. Feng Liang of Chongqing University, for his kind guidance in learning the fundamental concepts of multi-objective optimization. His research works in computational intelligence, and artificial intelligence provided me a crucial starting point on the problem formulation of the real-world case scenarios.

Fourthly, I would like to thank Dr. Alan Tan Wei Min, for his guidance in explaining the fundamental concepts of artificial neural networks and helpful suggestion on various libraries in the development of price estimation model.

Furthermore, I would like to show my appreciation to Dr. Iti Chaturvedi of Data Science & Artificial Intelligence Research Centre, for her wonderful guidance and feedback in writing the dissertation work, and her time and efforts spent in the proofreading of the overall dissertation report.

I would also like to thank Ms. Lee Kee Fong, Shirley of Graduate Research Office, for her utmost kindness in organizing and helping me in the administration of the confirmation of candidature and the final thesis submission process. Moreover, I would like to thank Ms. Ho-Ang Lye Choon, Grace of Graduate Research Office, for her kind advice and feedback to my inquiry regarding the confirmation of candidature.

I would like to show my appreciation to the internal/external examiners for the time and efforts spent in the evaluation of this dissertation work.

Last but not least, I would like to show my special gratefulness to my family for their support and encouragement throughout my graduate studies.

TABLE OF CONTENTS

Statement of Originality	i
Supervisor Declaration Statement	ii
Authorship Attribution Statement	iii
Acknowledgements	iv
Table of Contents.....	v
Table of Figures	ix
Table of Tables	xiv
Table of Equations.....	xv
Abstract	xvi
1. Introduction.....	1
1.1. Background: PropTech.....	1
1.2. Problem: Challenges in the Property Listing and Search Service	4
1.3. Inspiration & Motivation	7
1.3.1. Can we find our perfect dream home?.....	7
1.3.2. Can we make a successful business contract?	10
1.4. Contributions: Three Types of Data Analytics.....	11
1.5. Outline.....	12
2. PropTech Market Analysis	14
2.1. Investment Trend on Property Technology	14
2.2. PropTech Sectors.....	18
2.3. Residential Real Estate Market.....	20
2.4. Current Property Listing and Search Services	23
2.5. Related Academic Research Works	26
3. Literature Review	29
3.1. Multi-Objective Optimization	29

3.1.1.	Multi-Objective Optimization Problem.....	30
3.1.2.	Pareto Optimality and Dominance.....	31
3.1.3.	Pareto Optimal Set and Pareto Front	31
3.1.4.	Optimization Search Techniques/Algorithms.....	32
3.2.	Evolutionary Computation.....	34
3.2.1.	Evolutionary Algorithm.....	34
3.2.2.	Fundamental Design of Evolutionary Algorithm	35
3.2.3.	Performance Measure of Evolutionary Algorithm	37
3.3.	Multi-Objective Optimization Evolutionary Algorithm	38
3.3.1.	Different Approaches to MOEA	38
3.3.2.	Performance Measures of MOEA.....	39
3.4.	Non-Dominated Sorting Genetic Algorithm (NSGA)	40
3.5.	Related Academic Research Works	42
3.6.	Artificial Neural Networks	43
3.6.1.	Fundamental Design of Artificial Neural Networks.....	43
3.6.2.	Architectures of Neural Networks	45
3.6.3.	Training of Artificial Neural Networks	47
3.7.	Related Academic Research Works	48
4.	Data Exploration.....	49
4.1.	Data Collection	49
4.1.1.	Singapore's Public Housing Estates	49
4.1.2.	Rental Statistics of Singapore HDB Flats	51
4.1.3.	Spatial Dataset of Map of Singapore	52
4.2.	Descriptive Analytics	53
4.2.1.	Univariate Statistical Data Analysis	53
4.2.2.	Bivariate Statistical Data Analysis	59
4.2.3.	Multivariate Statistical Data Analysis	63

4.3.	Summary	68
5.	System Design	69
5.1.	Multi-Objective Optimization Problem	69
5.1.1.	Problem Formulation	69
5.1.2.	Exhaustive Search (Baseline).....	74
5.1.3.	Multi-Objective Optimization Evolutionary Algorithm Search	76
5.2.	Price Estimation Model.....	79
5.2.1.	Design of Artificial Neural Networks	79
5.3.	Web-based Property Listing and Search Platform	82
5.3.1.	System Architecture Design.....	82
5.3.2.	Software Architecture Design	83
5.3.3.	Database Design.....	84
5.3.4.	User Interface Design	86
6.	System Implementation.....	87
6.1.	Web-based Property Listing and Search Platform	87
6.1.1.	Web Application Framework	87
6.1.2.	Database Management System	87
6.1.3.	Integrated Development Environment.....	88
6.1.4.	Google Maps APIs.....	88
6.1.5.	MOEA Framework	89
6.1.6.	User Interface	90
6.2.	Price Estimation Model.....	94
6.2.1.	Integrated Development Environment.....	94
6.2.2.	Keras	94
6.2.3.	Training and Validation of Neural Networks.....	94
7.	System Testing	96
7.1.	Experimental Results.....	96

7.1.1.	Experimental Setup.....	96
7.1.2.	Initial Performance Assessment.....	97
7.1.3.	Improvement in Performance Assessment.....	99
7.2.	Web-based Property Listing and Search Demonstration	101
7.2.1.	Local Environment Setup.....	101
7.2.2.	Test Cases.....	101
7.3.	Summary	119
8.	Conclusion	120
9.	Future Works	121
	References	122
	Appendix	126
	Author's Publications	126

TABLE OF FIGURES

Figure 1: Early-Stage Real Estate Tech Market Map provided by CB Insights [2]	2
Figure 2: Residential Real Estate Tech Market Map provided by CB Insights [3]	3
Figure 3: PropTech Financial Funding Trend (in \$ million) between 2008 and 2012 provided by CB Insights [4]	14
Figure 4: PropTech Financial Funding Trend (in \$ billion) between 2013 and 2018 provided by CB Insights [4]	15
Figure 5: PropTech Financial Investment (in US\$ million) between Asia Pacific Regions and Global excluding Asia Pacific provided by JLL [5]	16
Figure 6: PropTech Financial Investment (in US\$ million) on Start-ups in Asia Pacific Regions by PropTech Sectors [5]	16
Figure 7: PropTech Market Sectors – Verticals	18
Figure 8: Association of PropTech Verticals and Horizontals	19
Figure 9: Technology Landscape of Commercial Real Estate Market in the year 2018 provided by Thomvest Ventures [6]	20
Figure 10: Technology Landscape of Residential Real Estate Market in the year 2018 provided by Thomvest Ventures [7]	20
Figure 11: Contributions of PropTech Start-ups in Residential Real Estate Market [1]	21
Figure 12: Financial Status of PropTech Start-ups in Asia Pacific regions [5]	21
Figure 13: Publications related to Real Estate Industry listed in Google Scholar	26
Figure 14: Publications related to Finance (left) and Construction (right) Industries listed in Google Scholar	26
Figure 15: Distribution of Research Index Terms in Real Estate Related Research Publications	27
Figure 16: Association between Pareto Optimal Set in Decision Space and Pareto Front in Objective Space [19]	31
Figure 17: Various Types of Optimization Search Techniques [18]	32
Figure 18: Four Paradigms of Evolutionary Algorithm (EA)	34
Figure 19: Selection Procedure of NSGA-II Algorithm [29]	41
Figure 20: General Architecture of Artificial Neural Networks with two Hidden Layers	44
Figure 21: General Computation of a Single Neuron from the Neural Networks [30]	44
Figure 22: Activation Functions commonly used in Artificial Neural Networks [31]	45
Figure 23: Overall Architecture Designs of Artificial Neural Networks [32]	46

Figure 24: Step by Step Process of Web Scraping Procedure for HDB Flat Rental Dataset Collection	49
Figure 25: Schedule of Web Scraping Process for Data Collection	50
Figure 26: Singapore's HDB Flat Rental Dataset with 24 Features	50
Figure 27: Step by Step Process of Data Collection Procedure for HDB Flat Rental Statistics	51
Figure 28: Singapore's HDB Rental Statistics Dataset with 6 Features	51
Figure 29: Step by Step Process of Data Collection Procedure for Spatial Dataset of Singapore.....	52
Figure 30: Spatial Dataset of Map of Singapore.....	52
Figure 31: Summary of Data Distribution of Rental Price Feature	53
Figure 32: Data Distribution of Rental Price Feature	54
Figure 33: Most Frequent Groups of Living Facilities provided in Property Rental	55
Figure 34: Most offered and Least offered Living Facilities in Property Rental	55
Figure 35: Location of HDB Rental Flats in Singapore	56
Figure 36: Boundary of Singapore	56
Figure 37: Data Categorization according to HDB Flat Type	57
Figure 38: HDB Rental Offers in Singapore based on different District Areas	58
Figure 39: Data Distribution of HDB Rental Offers in different District Areas	58
Figure 40: Data Distribution of Rental Price by HDB Flat Type	59
Figure 41: Data Distribution of Rental Price on Singapore Geographic Map	60
Figure 42: Results of K-Means Clustering on Rental Price	60
Figure 43: Data Clustering of Rental Price and Visualization on Singapore Geographic Map	61
Figure 44: Data Distribution of Rental Price in Each District Area	62
Figure 45: Data Distribution of Rental Price in Each District based on HDB Flat Type	63
Figure 46: Statistical Trend of HDB Rental Price by Flat Type from the Past Decade in Quarterly Manner.....	64
Figure 47: Statistical Trend of Average Median Rental Price by Flat Type from the Past Decade	65
Figure 48: 10-Year Timeline of Rental Price Trend in Town Areas by 2-room and 3-room HDB Flat Types	66
Figure 49: 10-Year Timeline of Rental Price Trend in Town Areas by 4-room and 5-room HDB Flat Types	67

Figure 50: 10-Year Timeline of Rental Price Trend in Town Areas by executive HDB Flat Type	67
Figure 51: Overall Algorithm Workflow of Multi-Objective Optimization Evolutionary Algorithm Search.....	76
Figure 52: Overall Algorithm Workflow of Fast Non-dominated Sorting Genetic Algorithm (NSGA-II)	77
Figure 53: 2-Layer Neural Networks Design of Price Estimation Model.....	80
Figure 54: ReLU Activation Function.....	80
Figure 55: System Architecture Design of Web-based Property Listing and Search Platform	82
Figure 56: Software Architecture Design of Web-based Property Listing and Search Platform	84
Figure 57: Database Design of Web-based Property Listing and Search Platform.....	85
Figure 58: User Interface Design of Web-based Property Listing and Search Platform	86
Figure 59: User Interface of Interactive Map Page	90
Figure 60: Control Panel of Interactive Map Page	91
Figure 61: Map Viewer of Interactive Map Page	91
Figure 62: Ranking of Best-Known Optimal Solutions according to Various Preference Priorities	92
Figure 63: Travel Scheduler of Interactive Map Page.....	92
Figure 64: Visualization of Recommended Routes among Property Listings and Locations specified by the user	93
Figure 65: Best Known Property Listings of Interactive Map Page.....	93
Figure 66: Property Listings displayed in Best Known Property Listings section with relevant information.....	93
Figure 67: Training Results of Neural Networks in Different Epoch Settings	95
Figure 68: 20 Handpicked Geographic Coordinate Points on Map for Performance Analysis	97
Figure 69: Search for Price and Living Facilities Operation Button in Control Panel	102
Figure 70: Result of Bi-Objective Based Test Case with Price Ranking on Map.....	102
Figure 71: Result of Bi-Objective Based Test Case with Living Facilities Ranking on Map	103
Figure 72: Good HDB Flat recommended for Bi-Objective Based Test Case	103

Figure 73: Search for Price, Living Facilities and Distance Operation Button in Control Panel	104
Figure 74: Result of Multi-Objective Based Test Case with Price Ranking on Map.....	105
Figure 75: Result of Multi-Objective Based Test Case with Living Facilities Ranking on Map	105
Figure 76: Result of Multi-Objective Based Test Case with Distance Ranking on Map.....	105
Figure 77: Good Options for the Customers who prioritize the Location	106
Figure 78: A Good Option for the Customers who are Price conscious	106
Figure 79: Search for Price, Living Facilities and Duration Operation Button in Control Panel	107
Figure 80: Result of Multi-Objective Based Test Case with Duration Ranking on Map	107
Figure 81: Good Options for the Customers who prioritize the Location nearby Workplace	108
Figure 82: A Good Option for the Customers who prefers Lower Price	109
Figure 83: Setting of Price Range and Location Distance Range, and Search for Price, Living Facilities and Distance Operation Button in Control Panel	110
Figure 84: Setting of Facilities in Control Panel and Location Points in Travel Scheduler ..	110
Figure 85: Result of Property Listings based on the Case Study	111
Figure 86: Result of Multi-Objective Based Test Case with User's Preference in Price Ranking on Map.....	111
Figure 87: Result of Multi-Objective Based Test Case with User's Preference in Living Facilities Ranking on Map	112
Figure 88: Result of Multi-Objective Based Test Case with User's Preference in Distance Ranking on Map.....	112
Figure 89: Good HDB Flat recommended for Multi-Objective Based Test Case with User's Preference.....	113
Figure 90: Result of Property Listings in the table of Best-Known Property Listings ranked by Price	113
Figure 91: Search of Driving Directions from Property Listings to the specified Locations in Travel Scheduler	114
Figure 92: Visualization of Driving Directions from Property Listings to the specified Locations on the Map	114
Figure 93: Driving Direction from a selected Property Listing to the specified Locations ...	115

Figure 94: Result of Property Listings in the table of Best-Known Property Listings ranked by Distance.....	116
Figure 95: Setting of Price Range and Time Duration Range, and Search for Price, Living Facilities and Duration Operation Button in Control Panel	117
Figure 96: Result of Property Listings based on the Case Study with Duration Criteria	117
Figure 97: Result of Multi-Objective Based Test Case with User's Preference in Duration Ranking on Map.....	117
Figure 98: Result of Property Listings in the table of Best-Known Property Listings ranked by Price	118

TABLE OF TABLES

Table 1: A Brief Review of PropTech in Singapore’s Residential Real Estate Industry	24
Table 2: List of Research Papers published in terms of Real Estate and Multi-Objective Optimization extracted from IEEE Xplore	28
Table 3: Commonly Used Encoding Schemes for Chromosome Representation	35
Table 4: Major Domain Areas in which research works of MOEA applications are mostly focused on [18]	42
Table 5: Domain Areas in which research works of ANNs are mostly focused on [24]	48
Table 6: List of Living Facilities provided in Real Estate Property and their Weights of Frequency Distribution	72
Table 7: Input Features of Price Estimation Model	79
Table 8: Local Environment Setting for Performance Assessment	96
Table 9: Parameters Setting for Multi-Objective Evolutionary Algorithm	96
Table 10: Performance Assessment of Multi-Objective Optimization using Confusing Metrix	98
Table 11: Improved Performance Assessment of Multi-Objective Optimization using Confusing Metrix	100
Table 12: Web Browser Setting for Performance Assessment of Web-based Property Listing and Search Platform	101

TABLE OF EQUATIONS

Equation (1):	Decision Variables.....	30
Equation (2):	Constraints in Mathematical Inequality.....	30
Equation (3):	Constraints in Mathematical Equality.....	30
Equation (4):	Objective Functions.....	30
Equation (5):	Multi-Objective Optimization Problem.....	30
Equation (6):	Interquartile Range (IQR).....	53
Equation (7):	Lower Fence.....	53
Equation (8):	Upper Fence.....	53
Equation (9):	Decision Variables for Problem Model.....	69
Equation (10):	Constraints for Problem Model.....	70
Equation (11):	Minimum and Maximum Constraints Values for Problem Model.....	70
Equation (12):	Objective Function for Problem Model.....	71
Equation (13):	Alternative Objective Function for Problem Model.....	71
Equation (14):	Minimization Objective Function of Rental Price.....	71
Equation (15):	Alternative Minimization Objective Function of Rental Price.....	71
Equation (16):	Definition of α and β in Rental Price Objective Function.....	71
Equation (17):	Maximization Objective Function of Living Facilities.....	72
Equation (18):	Minimization Objective Function of Travel Distance.....	73
Equation (19):	Definition of distance function d	73
Equation (20):	Minimization Objective Function of Travel Duration.....	73
Equation (21):	Definition of duration function D	73
Equation (22):	Constraints on the Objective Functions.....	79
Equation (23):	Input Vector of Artificial Neural Networks Model.....	79
Equation (24):	ReLU Activation Function.....	80
Equation (25):	Minimization of Mean Squared Error.....	81

ABSTRACT

The search for the property listings is a time-consuming task. Traditionally, a person who wants to buy or rent a house will search through the tremendous amount of property listings advertised in the local newspapers or brochures. After the preferred property listings have been selected, it is necessary to connect with the property agent for the house viewing and make a price negotiation with the house owner. Once the price negotiation is successful, the contract signing and further legal works for the ownership are processed. The real estate industry had been nurturing such a conventional business model for more than a few decades. Gradually, the technological advancements allow the entrepreneurs to adopt the innovative technologies in the development of the property listing and search services to provide intelligent solutions more efficiently and effectively. Property search on the online web-based platforms is common because it significantly reduces the level of time consumption on the search and increases the search efficiency. Consequently, various kinds of search methods are developed in the online web-based platforms. However, it is discovered that current search methods require the contribution of the customers' preferences in the search process. It can lead to a situation where some *good* property listings, which customers might favor, can be filtered out due to the constraint of the preference criteria.

Therefore, in this dissertation, a new kind of property search system is proposed as a decision support system, which can be differentiated from existing property search methods. With an adoption of multi-objective optimization techniques, an online web-based property listing and search system is designed to consider multiple criteria in the search with the minimum preference input from the customers and recommend the property listings, which are the ideal possible options for the customers to make an intelligent decision in the property selection. Moreover, in order to achieve the goal of a convenient transition from the selection of a dream home to a successful business contract between the customer and house owner, a price negotiation model is cooperated in the decision support system to perform the appropriate price estimation of the real estate property. The whole dissertation work is mainly organized into three types of data analytics: descriptive analytics, predictive analytics, and prescriptive analytics to go through the lifecycle of design and development of an online web-based property listing and search system. According to the performance assessment, it is discovered that the property listing and search system can perform a good recommendation of the property listings considering three multiple criteria in the search performance: 1) *minimizing* the price expense, 2) *maximizing* the facilities offered in the real estate property, and 3) *minimizing* the distance/duration it takes to go to the specified locations.

❧ CHAPTER 1 ❧

1. INTRODUCTION

In this chapter, the student introduced the background of PropTech, and the challenges observed in the property listing and search services currently offered by the web-based search platforms, which led to the inspiration and motivation of this dissertation work. The student described how this dissertation work was categorized into three types of data analytics and constructed the outline for the readers to achieve the convenient readability.

1.1. Background: PropTech

The search for the property listing is a time-consuming task. Traditionally, a person who wants to buy or rent a house searches through the tremendous amount of property listings advertised in the local newspapers or brochures. It is particularly difficult for a non-local who intends to buy or rent a house in a foreign country. For instance, a person who decides to migrate to a foreign country for a new job opportunity may require the local assistance for the property search. Moreover, after the preferred property listings have been selected, it is necessary to connect with the property agent for the house viewing and make a price negotiation with the house owner. Once the price negotiation has been successful, the contract signing and further legal works for the ownership are processed. Real estate industry had been nurturing such a conventional business model for more than a few decades when a similar sector such as finance, started to adopt the technology-based innovations in its operation processes. Gradually, the entrepreneurs had found the lack of technology adoptions as an opportunity and started applying the advanced technologies in various operations of the real estate business models to provide the solutions more efficiently and effectively. The general term used to describe the application of innovative technology in the real estate property industry, is coined as Property Technology or also known as *PropTech*.

Nowadays, PropTech has gradually become mature with the creative technology innovations, venture investments, and entrepreneurial business operations in three major sectors, namely [1]:

1. Smart Real Estate (*focus on the operation and management of real estate assets*),
2. The Shared Economy (*focus on the use of real estate assets*), and
3. Real Estate FinTech (*focus on the ownership of real estate assets*).

Current PropTech companies are adopting various innovative technologies in different PropTech areas for the improvement in the business operations under these three major sectors. For instance, *Property Management* is under Smart Real Estate sector, *Corporate and Shared Housing* falls under The Shared Economy sector, and *Listing/Search Services* is under Real Estate FinTech sector for PropTech segmentations. Figure 1 shows the early-stage real estate tech market map provided by CB Insights [2] in which PropTech companies were categorized into various PropTech areas between the commercial and residential real estate markets. It is discovered that PropTech companies find the residential real estate market more compelling than the commercial real estate market in some PropTech areas. Furthermore, in the residential real estate market, there are some areas in which numerous PropTech companies are interested in, such as *Listing/Search Services*, *Mortgage and Lending*, and *Leasing and Renting*. It proves that PropTech has become the growing area for the entrepreneurs, technology-based service providers, and research scientists to tackle various challenges encountered during the revolutionary change of the conventional real estate business operations into digitization.

Figure 1: Early-Stage Real Estate Tech Market Map provided by CB Insights [2]

According to the market map, numerous PropTech companies are found to be interested in the residential real estate market especially in the listing and search services area due to the availability of public data sources, enormous demand size of customer base and an ample supply of residential real estate assets. The proof can be found in Figure 2, created by CB Insights [3], where out of 96 PropTech companies, 27 companies (28%) focus on the listing and search services area. With the advances in Internet Technology (IT), PropTech companies make a great use of web-based technologies and platforms to facilitate the property listing and search services. Hence, the traditional search in a local newspaper or brochure has evolved into an online web-based search. Property agents and house owners are now advertising their property listings online, and the customers are searching them by using different types of search methods, which makes the property search process more convenient and save time. The competitive advantages of a PropTech company which focuses on the property listing and search services area are 1: the vast number of property listings posted by the property agents and house owners (data availability), and 2: the convenient and efficient search techniques provided to the customers (technology differentiation).

Figure 2: Residential Real Estate Tech Market Map provided by CB Insights [3]

1.2. Problem: Challenges in the Property Listing and Search Service

Unlike the conventional property search in the local newspaper or brochure, property search on the online web-based platforms significantly reduces the level of time consumption on the search and increases the search efficiency. These two factors attract a large customer base and property agents. Consequently, various kinds of property listings and search methods are offered to the customers to achieve these two goals. Every web-based property search platform is found to adopt at least one search method, ranging from providing the simple search methods (e.g., criteria-based search) to more advanced search methods (e.g., personalized recommendation system). Since the real estate property assets naturally involve the geographical information, web mapping technologies (e.g., Google Maps) are greatly utilized in the current search platforms to visualize the location of property listings and integrate with more sophisticated search methods (e.g., location-based search). Additionally, a tremendous amount of public data sources applicable to the real estate industry are freely provided by the government, individual organization, and community, which can be utilized to improve the property search capabilities.

However, even with the use of web-based property search platforms, it still takes a considerable amount of time for a customer to make a selection on the property listings and to proceed with the purchase or rental process due to a large number of property listings posted online. A brief study of three different types of search methods mentioned earlier was conducted to understand the general concept and application in the web-based property search platforms. First of all, a criteria-based search, which is a simple search method, provides the property listings according to the customer's preferences. However, it can provide the unbalanced search results in which either a vast amount of property listings (due to simplified preference inputs) or a tiny number of property listings (due to greatly customized preference inputs) can be returned from the search method. The former case results in more time consumption on the search through the extensive list, and the latter one leads to a situation in which some *better* options might be filtered out.

Secondly, a personalized recommendation system or recommender system is an advanced search method that uses various types of machine learning algorithms during the search. It generally provides 1: the property listings which are similar to the listings that a customer has already viewed (content-based recommendation), and 2: the property listings which are viewed

by other customers who have the similar preferences with the customer (collaborative filtering). Recommender systems are often discovered to be integrated in the current web-based property search platforms. However, unlike the product recommendation which are usually found in the online shopping/retail platforms, the recommender system in the real estate property has some limitations. One of them is that a customer who has already purchased or rented a house is unlikely to come back to the web-based property search platform in a few months or even in years, unless he or she wants to move to a new house again (the lack of long-term customer relationship). It is different from a typical online shopping platform where there are frequent site visits by the customer from which the customer's online activities can be efficiently utilized for both content-based recommendation and collaborative filtering techniques to build a strong customer relationship. In order to compensate for the lack of data for long-term customer relationship, current recommender system makes use of the customers' recent browsing activities and directly recommends the similar property listings, which may not match with the customer's preferences. Another challenge found in the recommender system is that once the purchase or rental transaction of a particular house has been made, it can no longer be considered in the recommendation process for the next customers (dynamic change of data). In this case, the collaborative filtering technique may not work efficiently. Likewise, a newly added real estate property can be at a disadvantage due to the lack of customers' browsing activities (cold-start problem).

A location-based or map-based property search method is currently found to be the most adopted search method in the web-based property search platforms. It is due to the advancement in the spatial data analysis, efficient visualization of the geographical information, and higher perception capability of a human mind in the visual data than textual data. Most web-based property search platforms adapt to the innovative map-based search techniques with various geographical information to deliver much more useful knowledge and deep insights to the customers during the property selection process. For instance, data availability of the local amenities (such as school, clinic or train station) allows a customer to know nearby amenities to a particular property or easily find nearby property listings to a particular amenity on the geographical map. However, in most web-based property search platforms, the location-based search method is treated as another kind of criteria-based data filtering method, i.e., it requires the customer's preferences on the distance/duration range. For example, the search for property listings that are within 500 meters around a nearby train station will filter the property listings that are out of the specified range or the property listings with more affordable price.

Based on the brief analysis of property listing and search services currently adopted in the web-based property search platforms, it was discovered that the property listing and search techniques greatly rely on the contribution of customers' preference inputs in the initial stage of the search process. In a typical web-based property search platform, the customer is required to provide the initial preference criteria for the search. In this case, some *good* property listings, which might be favored by the customer, can be filtered out due to the constraint of preference criteria. Besides, when the customer chooses a particular property for detailed review, the personalized recommendation is made based on the selected property. It results in recommending the property listings, which are out of the preference criteria initially set by the customer. Moreover, with the manual adjustment of distance/duration range in the location-based or map-based property search, the property listings with a better option, i.e., more affordable choice, can be filtered out. It leads to the never-ending search cycle, which is time-consuming when different types of search methods are not cooperative among each other and focus only on a specific criterion. Therefore, in the search of the property listings, it is essential to consider how a reasonable amount of *good* property listings should be provided to the customer while bearing in mind an understanding of what the customers might *want* to achieve from the search performance.

1.3. Inspiration & Motivation

After the study on different types of search methods usually adopted in the current property listing and search services for the web-based property search platforms, two key questions lingered to be answered.

1.3.1. Can we find our perfect dream home?

Since the customer is a final decision maker in the selection of property listings, the property search platform should be considered to be a **decision support system**, which provides an efficient and effective property search service with the minimum contribution requests from the customer (i.e., minimum input of the initial preference criteria). It should deliver the *good* property listings which might be *ideal* for the customer in several aspects (i.e., achieve several preference criteria which are naturally set by a typical customer) and provide the customer with the ability to make his/her own decision with freedom. For a property search system to achieve this goal, the real-world case scenarios were studied.

One of the case scenarios was that for a customer who wants to purchase or rent the real estate property, the first preference criterion he/she wants to consider is the sale/rental price of the property. Property listings with the lower price are more favorable. The search can be done with a simple criteria-based search method if the price is the only preference criterion that the customer considers. However, in the real world, there are various preference criteria set by each customer, and they conflict with each other. Considering two preference criteria that are relevant to the property rental: price and living facilities (i.e., furniture, air condition, internet accessibility, etc.), these two criteria conflict with each other because the more the living facilities are included in the rental, the higher the rental price is. Therefore, it will be impossible to find the house with the low price and many living facilities provided. However, it can be achieved if there is a trade-off. A property search system should consider these two criteria during the search and provide the houses, which are relevant and reasonable in both criteria for the customer to make the final decision on the most suitable house for him/her. For example, the customer's decision could be a house that costs the rental price of \$2,000 with three living facilities offered or a house with the rental price of \$2,200 but providing five living facilities.

Another case scenario for the multiple conflicting preference criteria was that when a customer wants to search the real estate property, which is convenient for his/her workplace. Convenient transportation is found to be one of the favorite preference criteria during the

property search. In continuation of the previous case scenario, among three preference criteria (i.e., price, living facilities, and transportation), price and transportation are two conflicting criteria since the property in the central business district is undoubtedly high priced although it is very convenient for the workplace. Including the conflict with living facilities, it seems impossible to find the property which is the best in all preference criteria. Therefore, the property search system should have the capability to perform the criteria fine-tuning to recommend the property listings which are the ideal possible options for the customer to make his/her criterion adjustments. Possible options for this case scenario could be a house that costs the rental price of \$4,000 in the central business district with a shorter transportation time of 10 minutes to the workplace and a house with the low rental price of \$2,500 in the outskirts of the central area that takes 45 minutes of transportation time to the workplace. Based on these possible options, the customer who prefers to prioritize the location and transportation time can choose the former option, and the price-conscious customer can choose the latter option.

From the study on real-world case scenarios, it was inspired to develop a property search system that can be differentiated from existing property listing and search methods in terms of *efficient knowledge-support search performance* considering *multiple criteria* in the *decision-making* process. With the automatic adjustments of multiple preference criteria, the proposed property search system will be able to solve the problems found in the criteria-based search, which provides the unbalanced search results due to the level of preference customization defined by the customers. Instead of filtering out the property listings which violate the preference criteria, the proposed search system will consider all preference criteria and make the appropriate criteria tuning based on the objectives of the search, i.e., the search towards lower price, more living facilities and nearer to the specified locations.

Moreover, the proposed property search system will be able to overcome the challenges encountered in the personalized recommendation system, which are the lack of long-term customer relationship, dynamic change of data and cold-start problem. This is due to the capability that the proposed search system can find the *best-known available* property listings which satisfy all multiple criteria and provide the customers without the need of long-term customer relationship with the web-based property search platform. Additionally, the proposed search system is not prone to the dynamic change of data and cold-start problem because it considers the property listings currently available in the property search platform. Similar to the case with the criteria-based search, the proposed search system will be able to handle the limitations of the location-based or map-based property search due to its capability of

appropriate adjustment in the distance/duration range, instead of filtering out the property listings which are out of the specified range.

To achieve this goal of the proposed property search system, one of the advanced techniques were adopted in this dissertation work: *Multi-Objective Optimization*, the performance of searching one or more optimal solutions that can satisfy all defined constraints and correspond to achieve the minimization or maximization of the specified objectives or goals.

With the adoption of Multi-Objective Optimization techniques in the property listing and search service, this dissertation work inspired to achieve as follows:

1. to design a new kind of property search system as a decision support system which differs from existing property search methods and overcome the challenges encountered in existing search methods
2. to recommend the property listings which are the best-known optimal solutions with the minimum preference criteria inputs from the customers
3. to develop a web-based property search platform which can perform the optimization process efficiently and effectively

1.3.2. Can we make a successful business contract?

Although a customer has already found his/her dream home for the purchase or rental, *successful* decision-making is only achieved after he/she has made a business contract with the house owner. The first step in achieving a successful business contract is to make an appropriate price negotiation between the customer and house owner. Following the previous real-world case scenario of the property rental search, if the rental price of a house in the central business district is quoted as \$4,000 by the house owner, the customer might want to make a price negotiation based on the living facilities provided in this rental or the current market price. It may be possible that the rental price quotation is overvalued in the current market. Therefore, a property search system should have the capability to support the price negotiation in which it can make an appropriate price approximation for both customer and house owner.

From the perspective of the customer, he/she can gain knowledge and make an intelligent decision during the property search. A successful negotiation with the house owner, such as the price bargain or any request for additional living facilities, can be achieved effectively. For instance, a price bargain of \$3,600 for the non-air con rooms or a request of additional portable air cooler can be made. From the perspective of the house owner, he/she can avoid setting an overvalued or undervalued price quotation and make a better price estimation to attract the attention of the customers. Based on the knowledgeable information, the house owner can further improve the house into a fully renovated and a well-furnished home with the appropriate living facilities.

Therefore, in this dissertation work, a segment of price negotiation, which can make the appropriate *price estimation* based on a price quotation set by the house owner, will be incorporated into the decision support system to achieve the goal of a convenient transition from the selection of a dream home to a successful business contract between the customer and house owner. For this purpose, one of the machine learning techniques will be applied: *Artificial Neural Networks*, a computational model biologically inspired by the human brain, which can perform various tasks of pattern recognition, classification, and clustering.

With the application of Artificial Neural Networks model in the price negotiation, this dissertation work encouraged to achieve as follows:

1. to design the price estimation system which can approximate the price of the real estate property based on the features of the house and the current real estate market
2. to provide both customers and house owners with the intelligent suggestions for the price negotiation

1.4. Contributions: Three Types of Data Analytics

This dissertation work was categorized into three different types of data analytics, namely: descriptive analytics, predictive analytics, and prescriptive analytics. In descriptive analytics, historical data sources were visualized and studied to discover the insights into the past. Real world residential real estate property data set was collected using the web-crawling techniques for the experimental purposes and analyzed to understand the data. Various types of data visualization techniques were applied for the knowledge discovery.

Based on the findings from the descriptive analytics, in predictive analytics, different machine learning and data mining techniques were applied to understand the potential future outcomes. Moreover, data pre-processing processes (i.e., data cleansing, data transformation) were performed to solve any flaw in the data (i.e., missing values, anomaly outliers). Predictive analytics prepared the data into an appropriate structure to have incorporated in the prescriptive analytics. Predictive analytics were briefly described and incorporated in the system implementation.

As for the main contribution, in the prescriptive analytics, a decision support system was designed, and a web-based property search platform was implemented to advise the best-known optimal solutions for each customer based on the different real-world problems. Moreover, a price estimation model was developed as a support for the price negotiation between the customer and house owner. In this section, a multi-objective optimization technique was adopted as the core model for the property search, and the artificial neural networks model was applied for the price estimation model, which was incorporated in the web-based property search platform.

With the statistical analysis of the academic research works on the real estate industry reviewed in **Chapter 2: 2.5 Related Academic Research Works**, this dissertation work can be considered as one of the earlier works which proposes and introduces a novel *Decision Support System* adopting a *Multi-Objective Optimization* technique in an *online* real estate property search system. With this proposal of the optimization techniques to be applied in the real estate industry, this dissertation work encourages the research community to contribute more advanced optimization techniques and innovative technologies (in theoretical perspective) to the future research works of PropTech to achieve the better efficient and effective performance in the operations of various PropTech areas and sectors (in practical real-world problems).

1.5. Outline

The outline of the dissertation started with **Chapter 1: *Introduction***, which introduced the background of PropTech, the challenges in the current property listing and search services offered by the web-based property search platforms, which led to the inspiration for this dissertation work and explained how the dissertation was categorized into three types of data analytics.

Chapter 2: *PropTech Market Analysis* mentioned the investment trend on PropTech market during the last decade, different types of major PropTech sectors, and the market analysis of the current PropTech companies, which offer the property listing and search services in the residential real estate market. Some notable search methods provided by the predominant PropTech companies were briefly studied. Furthermore, academic research works related to the property listing and search services were reviewed.

Chapter 3: *Literature Review* provided the introduction of multi-objective optimization, evolutionary computation, and evolutionary multi-objective optimization algorithms that were adopted in the decision support system. Moreover, the introduction of artificial neural networks, applied for the price estimation, was described. Relevant academic research works related to the real estate property industry, and similar industries were reviewed.

Chapter 4: *Data Exploration* analyzed various types of data sources that were used in this dissertation work to explore and discover the insightful knowledge. Two types of data analytics: descriptive analytics and predictive analytics were presented in this chapter.

Chapter 5: *System Design* included the descriptions of problem models, algorithm designs of multi-objective optimization based search system, and price estimation model. System architectures designed for the development of a web-based property search platform were explained. Prescriptive analytics was also presented in this chapter.

Chapter 6: *System Implementation* explained the development of a web-based property search platform in the client-side and server-side system environments. The implementation of the price estimation model was provided. Technologies used in the development such as web development framework, database management system, programming language, open source libraries, and web services were mentioned in this chapter.

Chapter 7: *System Testing* provided detailed assessments of multi-objective optimization search method and price estimation model. A workflow analysis of a web-based property search platform was shown in this chapter.

Chapter 8: *Conclusion* concluded the dissertation works and described any potential future research works, which could be extended from this dissertation.

CHAPTER 2

2. PROPTech MARKET ANALYSIS

In this chapter, the non-exhaustive research on the investment trend of PropTech market during the last decade, major PropTech sectors, and market analysis of current PropTech companies, which offer the property listing and search services in the residential real estate market were conducted. Some notable real estate property search methods currently provided by the predominant PropTech companies were studied. Furthermore, a brief review of the academic research works that are relevant to the property listing and search services were performed.

2.1. Investment Trend on Property Technology

According to Figure 1 and Figure 2 from *1.1 Background: PropTech* section which are produced by CB Insights, it is discovered that PropTech (Property Technology), also known as Real Estate Technology, has been rapidly emerging in the recent years among various operational areas in both commercial and residential real estate markets. It is due to the investment from the venture capital investors, blossoming technopreneurship, and advances in technology.

Figure 3: PropTech Financial Funding Trend (in \$ million) between 2008 and 2012 provided by CB Insights [4]

Figure 4: PropTech Financial Funding Trend (in \$ billion) between 2013 and 2018 provided by CB Insights [4]

Figure 3 and Figure 4, provided by CB Insights [4], show the history of investment funding between the year 2008 and 2012, and a rising trend of investments and business deals occurred between the year 2013 and third quarter of 2018 respectively. It is found that there were lesser investments in PropTech before the year 2012 (approximately less than \$90 million). From the year 2013 onward, a rapidly increasing amount of venture capital investments (from \$519 billion dollars in the year 2013 to \$3,945 billion dollars in the third quarter of the year 2018) and the number of business deals (from 128 deals in the year 2013 to 335 deals in the third quarter of the year 2018) occurred in the real estate industry. It proves that PropTech area is a recently emerging and fruitful area for the investors to focus on the strong returns.

Moreover, according to the report from JLL Investment Management Company [5], which provides the commercial real estate services, the financial funding of PropTech in Asia Pacific regions contributed a large proportion of the global investment in PropTech from the year 2014 onward in terms of the number of business deals as shown in Figure 5. It proves that Asia Pacific regions have the potential for PropTech start-ups to explore the technology engagement in the real estate industry, especially in China and India, which are found to possess the most dynamic markets in PropTech.

Figure 5: PropTech Financial Investment (in US\$ million) between Asia Pacific Regions and Global excluding Asia Pacific provided by JLL [5]

Figure 6: PropTech Financial Investment (in US\$ million) on Start-ups in Asia Pacific Regions by PropTech Sectors [5]

Figure 6, extracted from a report published by JLL Investment Management Company [5], reflects overall financial funding of PropTech start-ups in Asia Pacific regions from the year 2012 onward among the different types of PropTech sectors. It is discovered that China (including Hong Kong) has surpassed other countries in terms of the financial investments (US\$ 3,040 million dollars) and India has possessed the largest number of business deals (approximately 75 deals). Furthermore, most of the investment deals are made under the *Brokerage and Leasing* PropTech sector in all Asia Pacific regions.

Based on the study on the financial investment trend on PropTech, it is discovered that PropTech has become an emerging area to attract the attention from venture capital investors in the recent years, especially in Asia Pacific regions where the financial investments contribute the most to the global real estate PropTech investments. It leads to the blossoming technopreneurship in various PropTech areas.

2.2. PropTech Sectors

Generally, there are three main sectors in Property Technology, namely: Smart Real Estate, The Shared Economy, and Real Estate FinTech [1]. In *Smart Real Estate* sector, the technology-based platforms provide information about the real estate assets to facilitate the operation and management of the real estate assets efficiently. PropTech under this sector incorporates the real estate assets with the built-in sensor technology and the support of technology platforms, smart cities, on-site sustainable energy supply, etc.

In *The Shared Economy* sector, the technology-based platforms provide information for the prospective customers and sellers and effect the fee-based transactions to facilitate the efficient use of the real estate assets. PropTech under this sector entails the short-term housing rental, co-living, shared workspace, co-working, etc.

In *Real Estate FinTech* sector, the technology-based platforms provide information for the prospective buyers and sellers and affect the transactions of the ownership or leases with a capital value to facilitate the trading of real estate asset ownership. PropTech under this sector supports the real estate capital markets, residential sales and lease, debt and mortgage, commercial real estate lease, portfolio management, etc. These three sectors of PropTech are known as PropTech verticals, and every technology-based platform developed for either commercial or residential real estate market is categorized into one of them.

Figure 7: PropTech Market Sectors – Verticals

Figure 7 visualizes three main sectors of PropTech market and their associations with FinTech (Financial Technology) and ConTech (Construction Technology). It can be found that Real Estate FinTech is the association with PropTech and FinTech, and Smart Real Estate facilitates the association with PropTech and ConTech. Moreover, Figure 8 depicts how each PropTech sector or the vertical relates to three PropTech horizontals: information, transactions, and management/control. It can be described that the technology-based platforms under Real Estate FinTech sector and The Shared Economy sector focus on providing the services or solutions which are relevant to information and/or transactions of the real estate assets. Similarly, the technology-based platforms under Smart Real Estate sector provides the services or solutions related to information and/or management/control of the real estate assets.

Figure 8: Association of PropTech Verticals and Horizontals

Based on the analysis on the classification of PropTech sectors and areas, this dissertation work was categorized into *Listing/Search Services* PropTech area that falls under the *Real Estate FinTech* vertical sector and *Information* horizontal sector due to its focus on the design and development of a decision support search system which finds the real estate assets for the long-term sale/rental.

2.3. Residential Real Estate Market

Between the commercial and residential real estate markets, the residential real estate market was focused on due to the estimation, provided by Savills Research, that the size of the global residential real estate market is approximate to be around five or six times the size of global commercial real estate market. Nevertheless, the technology landscapes of both commercial and residential real estate markets are blossoming and fruitful in recent years according to the market maps provided by Thomvest Ventures, a venture capital firm which specializes in different stages of technological and financial investment, as shown in Figure 9 [6] and Figure 10 [7]. They depict the technology landscapes in the year 2018, where various business operational areas of both commercial and residential real estate markets are tackled by PropTech firms.

Figure 9: Technology Landscape of Commercial Real Estate Market in the year 2018 provided by Thomvest Ventures [6]

Figure 10: Technology Landscape of Residential Real Estate Market in the year 2018 provided by Thomvest Ventures [7]

Furthermore, Figure 11 provides the relative contributions of PropTech start-ups in the residential real estate market [1] in which 28% of current PropTech start-ups around the world focus on *Listing and Search Services*, followed by 11% on *Mortgage Tech*, 10% on *Marketplace* and 9% on *Investment/Crowdfunding*.

Figure 11: Contributions of PropTech Start-ups in Residential Real Estate Market [1]

Figure 12: Financial Status of PropTech Start-ups in Asia Pacific regions [5]

In the Asia Pacific regions, it is evident that most PropTech start-ups that had raised the funding focus on the listing and search service in the residential real estate market as proved in Figure 12, which is produced by JLL Investment Management Company [5]. It presents the financial status of PropTech start-ups according to different PropTech vertical subsectors. Based on the figure, in terms of the number of business deals, it can be found that most PropTech start-ups with various funding stages (from the earliest stage of funding to the mature stage before going public) focus on the listing and search services (25 deals on *list & search*, 15 deals on *brokerless list & search* and *tech-enabled brokerage*).

According to the statistical analysis, the listing and search services area is considered to be the current major focus area by both venture capital investors and entrepreneurs due to the enormous size of customer demands (house owners, buyers, property agents, etc.) and the variety of residential real estate assets around the world (apartment, condominium, detached house, etc.). Moreover, the availability of public data sources related to the residential real estate market enhances the level of technology adoption in the real estate industry. Some notable veteran PropTech companies focusing on the listing and search services in the residential real estate market are Zillow, 2006 [8] from the United States and Zoopla, 2007 [9] from the United Kingdom.

2.4. Current Property Listing and Search Services

According to the Disrupt Property [10] which discovers and tracks the global PropTech start-ups, among 296 PropTech start-ups listed to date from the residential real estate market, 71 PropTech start-ups (24%) are focusing on the listing and property search services. Among the countries in the Southeast Asia region, Singapore is considered to be the leader of PropTech because of its supportive start-up ecosystem. In Singapore's residential real estate market, 12 out of 25 PropTech start-ups listed (48%) are interested in the listing and search services. It proves that the listing and search services area is mature enough, yet can be considered to be an emerging area for the advanced technology adoption to grow to attract the current PropTech start-ups to tackle the challenges previously mentioned in the *1.2 Problem: Challenges in the Property Listing and Search Service* section.

Due to the tremendous amount of PropTech firms that are focusing on the property listing and search services, a non-exhaustive review were performed to discover a variety of search methods currently adopted in the web-based property listing and search services. Based on the review, there are a few notable PropTech web-based search platforms that can provide the innovative property listing and search services to the customers to explore a significant number of property listings. Table 1 reviews the current PropTech companies in Singapore's residential real estate industry whose listing and search services are prominent to be examined for this dissertation work. Review analysis was based on three main search methods commonly found in the majority of PropTech web-based property search platforms: criteria-based search, a personalized recommendation system, and location-based or map-based search.

PropTech Companies	Criteria-based Personalized Recommendation	Map-based	Distinguished Features	Shortcomings
PropertyGuru [11]	😊	😊	<ul style="list-style-type: none"> – sophisticated criteria filter 	<ul style="list-style-type: none"> – map-based search is for criteria filtering – more interactive map-based search is provided after each property has been selected
99.co [12]	😊	😊	<ul style="list-style-type: none"> – sophisticated criteria filter – innovative map-based search 	<ul style="list-style-type: none"> – different map-based search methods are provided for property listings page and individual property page
EdgeProp [13]	😊	😊	<ul style="list-style-type: none"> – sophisticated criteria filter 	<ul style="list-style-type: none"> – independent property listing results are provided for each map-based search filter – more interactive map-based search is provided after each property has been selected
keylocation.sg [14]	😊	😊	<ul style="list-style-type: none"> – sophisticated criteria filter – innovative map-based search – decision-support system 	<ul style="list-style-type: none"> – only act as a decision-support system for condominiums, and it redirects to the actual property web portals for further listing search

Table 1: A Brief Review of PropTech in Singapore's Residential Real Estate Industry

From a non-exhaustive review on the several PropTech firms and the comprehensive analysis of aforementioned web-based search platforms, it is discovered that most web-based search platforms integrate three major search methods into their search systems. Moreover, to provide the customers with more useful and knowledgeable results, the sophisticated criteria filters, and innovative location-based search methods are applied. Some web-based platforms exploit the property listings from various web-based property search portals and perform the data analytics to act as a decision-support system for the customers. They cross-refer the customers to various web-based property search portals for further search, which is similar to the metasearch engine model.

Although most web-based property search platforms provide the major search methods, the challenges mentioned in the *1.2 Problem: Challenges in the Property Listing and Search Service* remain to be tackled for the purpose of providing the customers with the intelligent assistance in making the best-informed decision to find their dream home and competent guidance in making a successful business contract which can accomplish the *1.3 Inspiration & Motivation* of this dissertation.

2.5. Related Academic Research Works

Statistical analysis of the academic research works on the real estate industry were reviewed to understand various related works published in the academic research journals. According to Google Scholar [15], publications on the Real Estate industry and their h5-index citation metrics are listed as displayed in Figure 13. It shows that the research community for the real estate industry is smaller than other related industries (such as finance, construction) as found in Figure 14 in terms of both numbers of publications and citation impact (h5-index).

	Publication	<u>h5-index</u>	<u>h5-median</u>
1.	The Journal of Real Estate Finance and Economics	<u>25</u>	31
2.	Real Estate Economics	<u>20</u>	29
3.	Journal of Real Estate Research	<u>16</u>	18
4.	Journal of Real Estate Literature	<u>10</u>	16
5.	Real Estate Management and Valuation	<u>10</u>	11
6.	International Symposium on Advancement of Construction Management and Real Estate	<u>6</u>	8
7.	Corporate Real Estate Journal	<u>4</u>	7

Figure 13: Publications related to Real Estate Industry listed in Google Scholar

Publication	<u>h5-index</u>	<u>h5-median</u>
1. Journal of Financial Economics	115	192
2. The Journal of Finance	109	189
3. The Review of Financial Studies	92	163
4. Journal of Banking & Finance	78	111
5. Journal of Financial and Quantitative Analysis	52	87
6. Journal of International Money and Finance	52	76
7. Review of Finance	49	74
8. Journal of Corporate Finance	44	89
9. Journal of Financial Stability	42	64
10. International Review of Financial Analysis	38	58

Publication	<u>h5-index</u>	<u>h5-median</u>
1. Construction and Building Materials	81	98
2. Automation in Construction	63	92
3. Journal of Construction Engineering and Management	39	53
4. International Conference on Tools and Algorithms for the Construction and Analysis of Systems (TACAS)	36	55
5. Journal of Composites for Construction	35	46
6. Construction Management and Economics	28	40
7. Journal of Performance of Constructed Facilities	22	27
8. Engineering, Construction and Architectural Management	19	29
9. Construction Innovation	17	23
10. Construction Economics and Building	14	20

Figure 14: Publications related to Finance (left) and Construction (right) Industries listed in Google Scholar

According to IEEE Xplore Digital Library [16], which publishes for research related to computer science, electrical engineering and electronics, and allied fields, it is found that there are approximately 1,700 research papers published in IEEE conferences, journals, magazines, and books. Figure 15 shows the distribution of research publications related to the real estate industry in which 16% of the research papers are related to *real estate data processing*, 12% on the *property market* and 8% on *pricing*. Moreover, information technology related research papers for the real estate industry have been published such as *regression analysis* (4%), *fuzzy set theory* (3%), *neural nets* (3%), *data mining* (2%), *artificial intelligence* (2%) and *optimization* (2%). From this review, it is discovered that a small number of research papers that adopted advanced techniques in the real estate industry are published among the research

communities. With the use of terms *Real Estate* and *Multi-Objective Optimization*, it is found that there are only five research papers published in IEEE Xplore, as shown in Table 2.

Figure 15: Distribution of Research Index Terms in Real Estate Related Research Publications

Research Paper	Focused Areas
1 Research on Optimization of Debt Financing Source Structure of Real Estate Smes Based on Multi-Objective Programming Model – Taking Representative Enterprises In Hunan Province As An Example Huikai Zheng 2018 3rd International Conference on Smart City and Systems Engineering (ICSCSE)	Debt Financing Finance
2 Adaptive Genetic Algorithm for Multi-objective Sustainable Land Use Planning Qian Xiang and Biao Liu 2015 11th International Conference on Natural Computation (ICNC)	Sustainable Land Use Planning Construction
3 Multi-Criteria Decision Support Systems: A Glorious History and a Promising Future Aouni Belaid and Jamil Razmak 2013 5th International Conference on Modeling, Simulation and Applied Optimization (ICMSAO)	Building and Construction
4 The Analysis on Residential Real Estate Development Multi-objective Linear Programming and Decision Xinan Li 2011 2nd International Conference on Artificial Intelligence, Management Science and Electronic Commerce (AIMSEC)	Risk Estimation of Real Estate Development

5	The Research Focused on Multiple Criteria Decision-making in the Second-hand House transaction	Evaluation of Second-hand House
	Bingnan Liu and Hui Liu	
	<i>2011 International Conference on Uncertainty Reasoning and Knowledge Engineering</i>	

Table 2: List of Research Papers published in terms of Real Estate and Multi-Objective Optimization extracted from IEEE Xplore

After the review of five research papers, it is discovered that there are very few research works that adopted multi-objective optimization techniques in the real estate industry. Hence, this dissertation work had the opportunity to introduce the property search based on a novel multi-objective optimization technique, which can be applied in the property listing and search services.

3. LITERATURE REVIEW

In this chapter, literature reviews of multi-objective optimization, evolutionary computation, and evolutionary multi-objective optimization algorithms that are adopted in the decision support system, were prepared. Moreover, the concept of artificial neural networks that are applied for the price estimation, were studied. Relevant academic research works that are related to the real estate property industry and similar industries, were reviewed.

3.1. Multi-Objective Optimization

Optimization is defined as a task of searching for one or more solutions that satisfy all stated constraints and at the same time, corresponds to minimizing or maximizing the specified objectives or goals [17] of a problem. Generally, a single-objective optimization problem consists of one objective or one goal to be achieved, which should be either minimized or maximized and results in a single solution that is optimal or the best. For instance, the search for a house with a minimum price will result in a house with the lowest price. However, real-world problems are complicated, with more than one objective or goal to be accomplished. In this case, the multi-objective optimization problem is formed in which the simultaneous optimization tasks are performed considering several objectives that might be conflicting with each other to achieve all of them. The result of a multi-objective optimization problem is not usually a single solution, but a set of solutions that are the best-known and incomparable with each other in considering all specified objectives. They are known as Pareto optimal solutions or non-dominated solutions. For instance, the search for a house with minimum price, maximum living facilities provided, and minimum travel distance/time to the preferred locations can result in more than one house with different trade-offs among these three objectives for the customer to make the final decision by himself/herself.

3.1.1. Multi-Objective Optimization Problem

A multi-objective optimization problem can be defined with 1) *decision variables*, 2) *constraints*, and 3) *objective functions*. **Decision variables** are the numerical values which are selected for the optimization problem, and a solution with n decision variables is represented by:

$$x = [x_1, x_2, \dots, x_n]^T \quad (1)$$

Constraints are the conditions that must be satisfied with the decision variables or objective functions during the optimization process to evaluate the feasible solutions, and they are represented in either mathematical inequality or equality, respectively as:

$$g_i(x) \leq 0 \text{ where } i = 1, 2, \dots, m \quad (2)$$

$$h_j(x) = 0 \text{ where } j = 1, 2, \dots, p \text{ and } p < n \quad (3)$$

Objective functions are the quantifiable evaluation functions of the decision variables to represent the quality of a solution (i.e., how good the solution is for the problem), and they are to be either minimized or maximized to achieve the optimal solutions. k objective functions in the optimization problem can be defined as:

$$f(x) = [f_1(x), f_2(x), \dots, f_k(x)]^T \quad (4)$$

Therefore, a multi-objective optimization problem can be constructed as either minimization or maximization of all specified objective functions. If an objective function is required to be maximized (assuming for the minimization problem), it corresponds to the minimization of its negative value. A multi-objective optimization problem can be represented as:

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t. } & x \in X \end{aligned} \quad (5)$$

3.1.2. Pareto Optimality and Dominance

Multi-objective optimization produces a set of non-dominated solutions (i.e., solutions with trade-offs) in which there is no feasible solution that can increase one objective function value without decreasing another objective function value [18]. It is called Pareto Optimality. By definition: a solution $x' \in X$ is called a Pareto optimal solution if there is no solution $x \in X$ such that $f_i(x) \leq f_i(x')$ for all $i = 1, 2, \dots, k$ and $f_j(x) < f_j(x')$ for at least one objective function index j . Dominance comparison is performed between two solutions to search for a better Pareto optimal solution. Solution $x^* \in X$ is weakly Pareto optimal (i.e., weakly dominates) if there is no solution $x \in X$ such that $f_i(x) < f_i(x^*)$ for all $i = 1, 2, \dots, k$, and strongly Pareto optimal (i.e., strongly dominates) if there is no solution $x \in X$ and $x \neq x^*$ such that $f_i(x) \leq f_i(x^*)$ for all $i = 1, 2, \dots, k$.

3.1.3. Pareto Optimal Set and Pareto Front

In an optimization problem, there are two search spaces: 1) the decision space where the solutions for the problem are defined with the decision variables and 2) the objective space where their corresponding objective function values are evaluated. During the multi-objective optimization, a set of solutions is found which are not dominated by any other solutions within the set based on their corresponding objective function values. This set is known as a non-dominated solution set or Pareto optimal set in the decision space. The boundary formed by the objective function values of the Pareto optimal set is known as the Pareto front in the objective space [19] as represented in Figure 16.

Figure 16: Association between Pareto Optimal Set in Decision Space and Pareto Front in Objective Space [19]

3.1.4. Optimization Search Techniques/Algorithms

There are various types of optimization search techniques or algorithms adopted in solving optimization problems, as shown in Figure 17 [18]. They are categorized into three main types, namely: enumerative, deterministic, and stochastic. Enumerative search algorithms are the simple search strategies that consider all feasible solutions within a finite search space. They can perform a complete search activity; however, it may be inefficient and computationally intensive if the search space becomes too large. Deterministic search algorithms are mostly considered to be the graph-based or tree-based search algorithms by incorporating the problem domain knowledge to reduce the search space for a faster search performance than the enumerative search algorithms. They are applied to solve various types of optimization problems.

Figure 17: Various Types of Optimization Search Techniques [18]

However, multi-objective optimization problems usually have the characteristics of a high-dimensional search space, discontinuous nature of Pareto Front, multimodal, or NP-Complete problem (non-deterministic polynomial time). Deterministic search algorithms would not be suitable to solve them efficiently and effectively due to their requirement of the problem domain knowledge for the search space restriction. Therefore, the stochastic search algorithms are designed and applied to solve the multi-objective optimization problems. Stochastic search

algorithms design the evaluation function to assign the fitness values to the possible solutions in the search space and construct a mapping mechanism to perform the encoding/decoding between the problem domain and algorithmic domain. Although there is no guarantee that the optimal solutions will be found, the best known optimal or good solutions can be achieved in most optimization problems.

Among various types of available stochastic search techniques, Evolutionary Computation (EC) search technique was applied to search the optimal solutions in this dissertation work.

3.2. Evolutionary Computation

Evolutionary Computation (EC) is an abstraction of algorithms for solving the global optimization problems. Concepts of Evolutionary Computation are based on the natural evolutionary biological process and Darwin's Theory of Evolution: *Survival of the Fittest*. In Evolutionary Computation, a population of individuals (i.e., candidate solutions) is generated and evaluated according to their fitness measure (i.e., minimization or maximization of the objective function values). Based on the fitness evaluation, better individuals are selected to perform the biological reproduction process (i.e., crossover or mutation). Competition between the newly generated individuals and existing individuals is performed for the selection of the next generation of the population. The whole cycle of processes is repeated until the best individuals are found or a predefined time limit has reached [20].

3.2.1. Evolutionary Algorithm

The idea of Darwin's Theory of Evolution had been applied to the problem-solving during the 1940s. Since then, different variants of algorithms had been invented based on the concepts of Evolutionary Computation and were termed under the area of Evolutionary Algorithm (EA). *Evolutionary Programming* was introduced by Fogel, Owens, and Walsh in 1966 while Rechenberg and Schwefel developed *Evolution Strategies* in 1971. *Genetic Algorithm* was introduced by Holland in 1975, followed by *Genetic Programming*, which was developed by Cramer and Koza in 1985 [21] [22]. Figure 18 represents four major classes of the Evolutionary Algorithm.

Figure 18: Four Paradigms of Evolutionary Algorithm (EA)

Algorithm 1 describes the general concepts of Evolutionary Algorithm [23].

Algorithm 1: Evolutionary Algorithm	
1:	initialize population with random candidate solutions
2:	evaluate each candidate solution
3:	repeat until the termination condition is satisfied do
4:	select parents
5:	crossover pairs of parents
6:	mutate generated offspring
7:	evaluate new candidate solutions
8:	select better candidate solutions for the next generation
9:	end repeat

3.2.2. Fundamental Design of Evolutionary Algorithm

The general design of an Evolutionary Algorithm consists of 1) *chromosome*, 2) *population*, 3) *fitness function*, and 4) *genetic operators*. Different variants of the Evolutionary Algorithm (i.e., above four paradigms) follow this fundamental design idea.

First of all, a ***chromosome*** is the representation of an individual or a candidate solution to a problem. It is composed of several *genes*, which define the functional units of the inheritance; in other words, the features of an individual [24]. Various types of encoding schemes are available for the chromosome representation, such as *binary coding* which encodes the features of an individual into a binary string, *real-valued coding* which uses the real values, *hybrid coding* which combines various data structures. Table 3 shows the commonly used encoding schemes for the chromosome representation in which three encoding schemes and their respective chromosome representations are provided.

Encoding Scheme	Chromosome Representation
Binary Coding	11001001
Real-Valued Coding	499.5, 0.8945, 9.993
Hybrid Coding	{(0110), (499.5), (A)}

Table 3: Commonly Used Encoding Schemes for Chromosome Representation

Once the chromosome has been defined, a set of chromosomes is generated to construct the search space for the problem. It is called a **population**, a set of individuals, and is generated randomly in the initial stage. The size of a population is defined as the number of individuals in the population, which is an essential factor in the performance of the evolutionary algorithm. The minimal size of the population will lead to the limitation of population diversity for the search, and the enormous size will make the search slow due to the computational time.

In order to evaluate the quality of individual in the population, the evaluation function or **fitness function** is required to assess and assign the fitness value to each individual. Individuals with better fitness value have a higher chance of survival to the next generation of the population. Fitness functions are defined according to the problem to be solved by the evaluation algorithm.

After the assessment of individuals from a population, a new population is produced with the use of **genetic operators** for the next generation. Basic genetic operators commonly applied in the evolutionary algorithm are *selection*, *crossover*, and *mutation* operators. **Selection** operators are used to choosing the individuals based on their fitness values to perform the crossover or mutation process. Various selection operators are applied depending on the problem to be solved. A few of selection methods are *proportional selection* which selects the individuals according to the probability distribution of the fitness value, *tournament selection* which chooses a group of k individuals randomly for the tournament and selects the individuals with the best fitness values, and *rank-based selection* which ranks the individuals in the order of fitness values and determines the selection probability.

After the individuals have been selected from the population, two parents from the individuals are randomly selected for the crossover process. **Crossover** operators perform the genetic blending of the information from two parent chromosomes to produce a new offspring chromosome, which might have a higher fitness value to survive to the next generation. Different types of crossover schemes are available such as *single-point crossover* where one point is randomly set on the chromosomes, and the segments of genes are swapped between two parents to generate the offspring, *n-point crossover* where n points are randomly selected for the swapping.

The **Mutation** process is performed to ensure the diversity of the individuals for the entire problem space. Various mutation operators are used in the search such as *inversion* for the binary coding which flips the bit value and *uniform mutation* for the real-valued coding where the value of the gene is converted into a random number that is uniformly generated within the specified lower and upper bounds, and so on.

3.2.3. Performance Measure of Evolutionary Algorithm

The performance of an evolutionary algorithm is evaluated by measuring the rate of convergence, which is the average number of generations required to achieve the optimal solution with a high fitness value. A simple way to measure the rate of convergence is to observe the average fitness value in relation to the best fitness value. The value of the difference between them seems to be small for a population that has converged to an optimal solution than for a population whose individuals are scattered in the entire solution space [24]. Due to the stochastic nature, the performance of the evolutionary algorithm is evaluated with several experiments conducted to observe the results.

3.3. Multi-Objective Optimization Evolutionary Algorithm

Evolutionary Algorithms (EAs) are applied to the multi-objective optimization problems due to their nature of simultaneously dealing with the candidate solutions in the population. It leads EAs to find a set of non-dominated optimal solutions in a single run. Moreover, they can solve the optimization problems with the discontinuous Pareto fronts due to the ability to search for different regions of the solution space simultaneously [25]. Therefore, considering the nature of the biological genetic operations, the evolutionary approaches are commonly applied in the multi-objective optimization tasks to approximate the optimal solutions for the problem.

3.3.1. Different Approaches to MOEA

Various techniques of Multi-Objective Optimization Evolutionary Algorithms (MOEAs) were designed and proposed. The earliest MOEA technique was proposed in 1985, which is called *Vector Evaluated Genetic Algorithm* (VEGA). It is the first MOEA approach in which the subpopulations are generated through the proportional selection concerning each objective function and evaluated according to the biological process of the Genetic Algorithm. Some well-known MOEAs were briefly reviewed [26] [27] for a better understanding of various evolutionary approaches that are adopted in the multi-objective optimization tasks.

Multi-Objective Genetic Algorithm (MOGA) adopted the Pareto ranking approach, which ranks the individual solutions according to the non-dominance level and used a niche-formation method for the diversity preservation of population. *Weight-Based Genetic Algorithm* (WBGA) computed the weighted objective function values for the fitness assignment and adopted the niching method to maintain the diversity in the weight vectors. *Niched-Pareto Genetic Algorithm* (NPGA) applied the tournament selection approach based on the Pareto dominance and proposed the equivalence class sharing, which computes the niche count to determine the winner in the tournament selection.

Non-dominated Sorting Genetic Algorithm (NSGA) adopted the ranking of the individual solutions based on the level of non-dominance, and the fitness values are shared by niching to ensure better distribution of the individuals in the population. In order to achieve the computational efficiency, an improved version of NSGA was proposed, also known as *Fast Non-dominated Sorting Genetic Algorithm* (NSGA-II). In addition to the original design of NSGA, NSGA-II performed niching by using the crowding distance approach to keep the diversity of the population and adopted the elitism to achieve better convergence.

Strength Pareto Evolutionary Algorithm (SPEA) applied the concepts of the ranking based on the non-domination strength values with the use of an external archive and clustering techniques to maintain the external archive. A revised version of SPEA was invented, which is called SPEA2, to enhance the fitness assignment strategy by considering the domination strength values of the individuals back and forth. Moreover, it adopted the nearest neighbor density estimation technique to improve the diversity of the population.

3.3.2. Performance Measures of MOEA

Comparisons among various MOEAs were performed in the literature and research works in terms of the efficiency (computational performance to search the optimal solutions) and the effectiveness (accuracy and convergence of the optimal solutions) [18]. Various performance measures are applied in the research works depending on the nature of the multi-objective optimization problems (either benchmark problems or the real-world scenarios) and the types of MOEAs to be compared to evaluate MOEA techniques. *Generational Distance* analyses the distance between the actual non-dominated Pareto optimal solutions and Pareto optimal solutions found by MOEA. Measurement of the *Hypervolume* is another way to evaluate the area coverage of Pareto optimal solutions in the objective space. *Maximum Pareto Front Error* measures the largest minimum distance between the set of Pareto optimal solutions found by MOEA and the set of true non-dominated Pareto optimal solutions.

In this dissertation work, one of the most widely used MOEAs, ***Fast Non-Dominated Sorting Genetic Algorithm*** (NSGA-II) was adopted in the multi-objective optimization-based search on the real estate property listings.

3.4. Non-Dominated Sorting Genetic Algorithm (NSGA)

Non-Dominated Sorting Genetic Algorithm (NSGA) was proposed by N. Srinivas and K. Deb in 1994. In this algorithm, the population is ranked according to the level of non-dominance before the selection is made. All non-dominated solutions are ranked into the same category with the fitness values, which are computed based on the proportion of the population size to achieve the same opportunity for survival. These classified solutions are shared with their fitness values to maintain the diversity of the population. Once non-dominated solutions within the same level have been ranked, the next level of non-domination is determined for the rest of the solutions and ranking proceeds until all of the solutions in a population are classified into the different levels of non-dominance. Solutions at the first level of non-dominance have been assigned with the highest fitness value and possess the highest chance of selection. It improves the search of the Pareto front regions and achieves the convergence of population towards those regions. Although the fitness sharing mechanism assists in the distribution of the population over Pareto front regions, it becomes a computational bottleneck in the non-dominance ranking.

Therefore, K. Deb, S. Agrawal, A. Pratap, and T. Meyarivan proposed an improved version of NSGA, known as NSGA-II, which is based on the design of NSGA. The general algorithm of NSGA-II is as shown in Algorithm 2 [28]. In NSGA-II, the offspring population is initially generated from the parent population at every generation. Both parent and offspring populations are combined into one, and the individuals are ranked according to the level of non-dominance, i.e., non-dominated individuals are classified into the same level. Afterward, NSGA-II will create a new population by selecting the individuals from the first level of non-domination, followed by the individuals from the second level and so on. Due to the limitation of the population size, not all domination levels can be added to the new population. When the individuals from the last allowed level are considered, NSGA-II will determine and choose only those individuals from the level, which contribute the most to the diversity of the population. In this case, the crowding distance values are used for sorting the individuals from the last level that cannot be fully added to the population [29]. The schematic procedure of the NSGA-II algorithm is displayed in Figure 19, in which a step by step selection process of the individuals is described.

Algorithm 2: Fast Non-Dominated Sorting Genetic Algorithm (NSGA-II)

```
1:      initialize population
2:      generate a random population of size N
3:      evaluate the objective function values of candidate solutions
4:      generate child population of size N
5:      binary tournament selection
6:      crossover and mutation
7:      combine parent and child populations of size 2N
8:      sort candidate solutions based on the level of non-dominance
9:      until a new population of size N is filled repeat
10:         add all individuals from a higher level of non-dominance
11:         if all individuals from the same level cannot be added do
12:             determine crowding distance within the same level
13:             add individuals which are in a lesser crowded region
14:         end if
15:     end repeat
16:     Create a population for the next generation
17:     binary tournament selection
18:     crossover and mutation
```

Figure 19: Selection Procedure of NSGA-II Algorithm [29]

3.5. Related Academic Research Works

Multi-Objective Optimization Evolutionary Algorithms (MOEAs) are greatly applied in various areas of real-world applications and developments. In academic research, new techniques of MOEA are proposed for better performance in the optimization problems. Moreover, the applications of MOEAs in real-world optimization problems are also presented in academic research. Table 4 provides four major areas of interest in which the academic research works of MOEA applications are focused on [18].

Engineering	Scientific	Industrial	Miscellaneous
Environmental, Naval and Hydraulic Engineering	Geography	Design and Manufacture	Finance
Electrical and Electronics Engineering	Chemistry	Scheduling	Classification and Prediction
Telecommunications and Network Optimization	Physics	Management	
Robotics and Control Engineering	Medicine	Grouping and Packing	
Structural and Mechanical Engineering	Ecology		
Civil and Construction Engineering	Computer Science and Computer Engineering		
Transport Engineering			
Aeronautical Engineering			

Table 4: Major Domain Areas in which research works of MOEA applications are mostly focused on [18]

Among four main domain areas, *Engineering* is found to be the most popular domain area within MOEAs literature due to its nature of having the good mathematical models that can directly be associated with the optimization search. Moreover, under the *Scientific* domain area, *Computer Science and Computer Engineering* subdomain is the most popular area of interest where most research works on MOEAs are proposed and applied in the real-world optimization problems. MOEA applications are found in machine learning, image processing, natural language processing, and so on under this subdomain. The *Finance* domain area is discovered to adopt MOEAs applications in various financial operations, such as investment portfolio optimization, time series analysis, stock ranking, and bank loan management [18]. From this review, it can be found that MOEAs are applied in various kinds of real-world optimization problems in different domain areas. For this dissertation work, MOEAs were applied in the *Real Estate* domain area in which very few research works were proposed and developed according to the analysis in 2.5 Related Academic Research Works from 2 PropTech Market Analysis.

3.6. Artificial Neural Networks

The concept of Artificial Neural Networks (ANN) is inspired by the mechanism of the human brain in which the cognitive processes are naturally performed. The human brain consists of approximately 100 billions of neuron cells that connect to form the networks for the ability to perform various operations in daily life. A biological neuron cell receives the information from other neurons, accomplishes the particular actions, and produces the result to the next neurons via the electrochemical pathways. The architecture of artificial neural networks is similar to the networks of biological neuron cells in the human brain to process the information efficiently and effectively.

The first model of artificial neural networks was designed by McCulloch and Pitts in 1943 [24], and Hebb proposed the learning scheme of the neural pathways for the reinforcement of the neural networks in 1949. In 1957, Rosenblatt invented the model of simple perceptron for the classification problems. Widrow and Hoff developed the first neural networks model, which was successfully applied to the real-world problem in 1959, known as ADALINE (Adaptive Linear Elements) and MADALINE (Multiple ADALINE). Since then, the design and development of artificial neural networks models are improved in both academic research works and the application in real-world problems.

3.6.1. Fundamental Design of Artificial Neural Networks

The architecture of artificial neural networks is designed with three main layers: 1) *input layer*, 2) *output layer*, and 3) *hidden layer*. The ***input layer*** consists of a set of neurons, represented by the features of data to be processed. The ***output layer*** consists of a set of neurons that produces the result of the problem to be solved. Between these two layers, there is one or more ***hidden layer*** in which processing of the artificial neurons has occurred through the connections among the layers. Figure 20 shows the example of the general architecture of neural networks in which there is an input layer with three neurons, two hidden layers with five neurons in each layer and an output layer with two neurons for solving the problems.

Figure 20: General Architecture of Artificial Neural Networks with two Hidden Layers

In the computation of the neural networks, the strength of the connection between any two neurons is considered and processed, which is represented by the **weights** of the neural networks. Generally, the result produced by each neuron is the summation of values produced by the neurons from the previous layer multiplied by their respective weights connected to the current neuron. In some neural networks models, a threshold value called the **bias** is included in the summation process. Once the summation is completed, an **activation function** is applied in each neuron to achieve the nonlinearity of the neural networks. The activation function is selected based on the nature of the problems to be solved. Figure 21 describes how each neuron in the neural networks computes the incoming data from the preceding layer and produces the output result to the succeeding layer [30].

Figure 21: General Computation of a Single Neuron from the Neural Networks [30]

Depending on the nature of the problem to be solved (i.e., classification problem, regression problem) and the selection of the network topology (i.e., number of neurons, layers, connections), various kinds of activation functions are applied to the neurons in the hidden layers and output layers. Figure 22 lists the activation functions, which are commonly used in the design of the neural networks [31]. *Sigmoid* activation function computes the input data and generates the probability value between 0 and 1, which is commonly applied for the binary classification problems. *Tanh* activation function is similar to the sigmoid activation function.

However, it generates an output value between -1 and 1. *ReLU* or Rectified Linear Units activation function is usually used in the hidden layers in which it directly passes the output value if it is greater than 0 and passes the output value of 0 if otherwise. *Leaky ReLU* activation function is the improved version of ReLU, which uses the non-horizontal component for the output value that is less than 0.

Activation function	Equation	Graph
Sigmoid	$S(x) = \frac{1}{1 + e^{-x}}$
Tanh	$\tanh x = \frac{e^x - e^{-x}}{e^x + e^{-x}}$
ReLU	$RELU(x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } x \geq 0 \end{cases}$
Leaky ReLU	$f(x) = \begin{cases} x & \text{if } x > 0 \\ 0.01x & \text{otherwise} \end{cases}$

Figure 22: Activation Functions commonly used in Artificial Neural Networks [31]

3.6.2. Architectures of Neural Networks

Altogether with the aforementioned components, the neural network architectures are designed and constructed based on the nature of the problem to be solved. Two basic types of artificial neural networks are 1) *Feedforward Neural Network* and 2) *Recurrent Neural Network*. ***Feedforward Neural Network*** is a simple form of network in which the information is fed from the input layer, through the hidden layers, and to the output layer. Neurons from one layer are fully connected to the neurons from the succeeding layer. ***Recurrent Neural Network*** is similar to the feedforward neural network; however, it adopts the connection between the passes, which gives the feedback information from the output layer back to the input layer. Figure 23, created by Fjodor van Veen [32], gives an overall neural networks architectures designed and constructed in the research community of the neural networks for solving various types of problems.

Figure 23: Overall Architecture Designs of Artificial Neural Networks [32]

3.6.3. Training of Artificial Neural Networks

Artificial Neural Networks are required to be trained to perform the adjustments of the weight and bias values of the whole network. It is also known as the learning process of the neural networks. Various learning methods are designed and applied in the training of the neural networks to modify the weight and bias values. Generally, there are two significant types of learning: 1) *supervised learning*, and 2) *unsupervised learning*.

Supervised learning is defined as learning with supervision. In order to train the neural networks with a supervised learning method, a training data set, which includes a set of input and output pairs, is required. In supervised learning, the inputs are applied to the neural networks, and the comparison between the network's current outputs and the actual outputs is made to observe the errors. The learning method minimizes these errors through the adjustments of the weight and bias values of the whole network until an acceptable result is achieved. Commonly applied supervised learning method for the training of the neural networks is the *gradient descent* learning algorithm with the *backpropagation*. Moreover, in order to observe the errors between the predicted output from the neural networks and the actual output, different loss functions are available based on the nature of problems to be solved, such as *mean squared error* loss function for the regression problem and *cross-entropy* loss function for the classification problem. In the training of the neural networks, the goal is to minimize the value of loss function of the networks by adjusting the weight and bias values.

Unsupervised learning is known as learning without supervision. In unsupervised learning, the actual outputs are not available in the training data set. Therefore, the unsupervised learning method analyses the features and identifies the patterns and trends in the training data set and is applied to the clustering problems and feature extraction. Commonly applied unsupervised learning method for the training of the neural networks is *Hebbian learning rule*. Between supervised and unsupervised learning methods, the supervised learning methods are commonly applied to the training of the neural networks.

3.7. Related Academic Research Works

Similar to Multi-Objective Optimization Evolutionary Algorithms (MOEAs), the area of Artificial Neural Networks (ANNs) is greatly interested in academic research and is tremendously applied to various fields of real-world applications and developments. Table 5 lists some domain areas in which the research works of artificial neural networks are commonly focused on and are applied to the real-world applications [24].

Domain	Area of Interest
Aerospace Engineering	Aircraft Control System
	Fault Detection System
Automotive	Automobile Automatic Guidance System
Finance	Credit Application Evaluation
	Credit Card Fraud Detection
Defense	Weapon Steering
	Facial Recognition
Design and Manufacture	Machine Diagnosis
	Quality Inspection
Medicine	EEG and ECG Signal Analysis
	Image Processing
Speech	Speech Recognition
	Text-to-Speech Synthesis
Telecommunication	Speech Processing
	Real-time Translation of Spoken Language

Table 5: Domain Areas in which research works of ANNs are mostly focused on [24]

Considering the academic research works on the price estimation with the use of artificial neural networks, it can be observed that there are a significant number of research articles (approximately around 2,000 research articles), published on IEEE Xplore Digital Library [16] in various domain areas. Moreover, the research works of artificial neural networks are more commonly applied to the real estate industry, compared to those of multi-objective optimization techniques. According to Figure 15 which provides the distribution of research index terms in the real estate related research publications, among approximately 1,700 research papers published on IEEE Xplore Digital Library, 8% are the research papers focused on the *pricing* of the real estate property and 3% are the papers focused on the *neural nets*.

In this dissertation work, artificial neural networks model was designed and developed to estimate the price of the real estate property based on the features of the house and the current real estate market. The estimated price will help both customers and the house owners in the price negotiation process.

4. DATA EXPLORATION

4.1. Data Collection

4.1.1. Singapore's Public Housing Estates

In this dissertation, data sources from the public housing estates of Singapore were used for experimental purposes. Housing & Development Board (HDB) is the authority of building and providing public housing estates to over 80% of the population in Singapore [33]. There are more than one million HDB flats completed in 23 towns of Singapore to date. For experimental purpose, the real-world rental data set of HDB flats was collected from one of the web-based property listing and search platforms: *99.co* [12]. Web scraping techniques were used to extract the relevant information and stored in the database management system, as shown in Figure 24 which provides the step by step procedures to collect the data from the web-based property listing and search platform.

Figure 24: Step by Step Process of Web Scraping Procedure for HDB Flat Rental Dataset Collection

XML sitemap files contain a list of URLs (Uniform Resource Locator) of the web pages from the real-world web-based property listing and search platform. Web pages in HTML were collected via the URLs, and the page parsing was performed to extract the essential features relevant to the HDB rental information such as the specification of the HDB flat, monthly rental fee, living facilities provided, images of the HDB flat. With the geocoding technology, the full address of each HDB flat and its respective geographic coordinates were achieved. However, due to the restriction of the daily quota for the use of geocoding technology, a list of URLs was

divided into the sub-groups with less than 2000 HTML web pages in order to follow the limitation of the daily quota of 2000 request calls to geocoding API during the page parsing process. A total of six days were spent on the web scraping process to collect the data from 8706 URLs, as observed in Figure 25. After the data collection with the minor errors in the web scraping process, out of a total of 8706 HDB flats, 8463 data records with 24 valuable features were successfully extracted. Figure 26 provides a list of features collected. All extracted information was stored in the local database for further analysis.

Figure 25: Schedule of Web Scraping Process for Data Collection

Field Name	Table	Remote Field Name	Field Name	Table	Remote Field Name
# Property ID	property	propertyID	# Sqft	property	sqft
Abc Url	property	url	# Sqm	property	sqm
Abc Address	property	address	# Psf	property	psf
Abc Full Address	property	fullAddress	# Price	property	price
📍 Latitude	property	latitude	# Is HDB	property	isHDB
📍 Longitude	property	longitude	# For Rent	property	forRent
Abc Geolocation	property	geolocation	Abc Lease	property	lease
# District	property	district	# Pets Allowed	property	petsAllowed
# Num Of Rooms	property	numOfRooms	Abc Availability	property	availability
# Num Of Beds	property	numOfBeds	# Fully Furnished	property	fullyFurnished
# Num Of Baths	property	numOfBaths	Abc Facilities	property	facilities
			📅 Created Date	property	createdDate
			📅 Modified Date	property	modifiedDate

Figure 26: Singapore's HDB Flat Rental Dataset with 24 Features

4.1.2. Rental Statistics of Singapore HDB Flats

To assist with a tenancy agreement, Singapore HDB publishes the rental statistics quarterly since the year 2007, in which the median rental price of various HDB flat types in different town areas [34] are recorded. This rental statistics data set was extracted from Singapore Housing & Development Board web portal in a yearly manner in order to discover the valuable knowledge and insights about the median rental price of 6 HDB flat types (i.e., 1-room, 2-room, 3-room, 4-room, 5-room and executive) from the year 2007 to 2018 in 26 different town areas. After the data collection procedure, as shown in Figure 27, a total of 6864 data records were successfully extracted along with the list of features provided in Figure 28 and stored in the local database for further analysis.

Figure 27: Step by Step Process of Data Collection Procedure for HDB Flat Rental Statistics

Field Name	Table	Remote Field Name
# Statistics ID	hdb_rental_statistics	statisticsID
# Year	hdb_rental_statistics	year
# Quarter	hdb_rental_statistics	quarter
Abc Town	hdb_rental_statistics	town
Abc Flat Type	hdb_rental_statistics	flatType
# Price	hdb_rental_statistics	price

Figure 28: Singapore's HDB Rental Statistics Dataset with 6 Features

4.1.3. Spatial Dataset of Map of Singapore

GADM, a database of global administrative areas, offers the maps and spatial data sets for all countries and their sub-divisions [35] to be used in various GIS (Geographic Information System) applications. Spatial data set for the map of Singapore was downloaded from the GADM web portal in the form of a shapefile format, which is a standard geospatial vector data format for the GIS software. Data parsing was performed to extract the latitude and longitude points of the border of Singapore, as shown in Figure 29 and Figure 30.

Figure 29: Step by Step Process of Data Collection Procedure for Spatial Dataset of Singapore

Field Name	Table	Remote Field Name
Longitude	singapore.csv	Longitude
Latitude	singapore.csv	Latitude
Gid 0	singapore.csv	GID_0
Boundary	singapore.csv	Boundary
Area	singapore.csv	AREA
Perimeter	singapore.csv	PERIMETER
Xcoord	singapore.csv	XCOORD
Ycoord	singapore.csv	YCOORD
Sequence	singapore.csv	Sequence

Figure 30: Spatial Dataset of Map of Singapore

4.2. Descriptive Analytics

Various types of data visualization techniques were applied for the descriptive analytics in order to achieve a better understanding of the statistical data and to discover the knowledgeable insights, which can contribute to defining the problem and designing the solution framework.

4.2.1. Univariate Statistical Data Analysis

Univariate data analysis was performed on *rental price*, *living facilities*, *location of HDB flat*, *HDB flat type*, and *district area* in order to understand the data distribution based on each feature.

i. Rental Price

Figure 31: Summary of Data Distribution of Rental Price Feature

Figure 31 presents the data distribution of the rental price feature in the box plot, in which a summary of rental price is provided: minimum: S\$500, first quartile (Q1): S\$1,900, median (Q2): S\$2,200, third quartile (Q3): S\$2,450, maximum: S\$8,480 and average: S\$2,210.6 with the standard deviation of S\$429.6. From the box plot, the outliers can be easily identified with the calculations of the lower fence: S\$ 1,075 and the upper fence: S\$ 3,275. According to the values of the lower fence and upper fence, it can be found that there are outliers in the data set, and further analysis is required. Moreover, it can be observed that the values of the median and average are very similar in this data distribution.

$$\text{Interquartile Range (IQR)} = Q3 - Q1 \quad (6)$$

$$\text{Lower Fence} = Q1 - 1.5 * IQR \quad (7)$$

$$\text{Upper Fence} = Q3 + 1.5 * IQR \quad (8)$$

The histogram in Figure 32 provides the distribution of numerical continuous rental prices in 20 bins (intervals). From the histogram, it can be found that the most significant data point concentration (3559 data records) is within the rental price between S\$2,100 and S\$2,520, and the deficient number of data points (9 data records) are found with the rental price higher than S\$5,000.

Figure 32: Data Distribution of Rental Price Feature

ii. Living Facilities

A bubble chart was prepared to visualize the group of living facilities provided in the property rental as shown in Figure 33 in which the considerable amount (37.78%) of the property rental (3,197 data records) do not provide any information about the living facilities, and 684 data records belong to a group of air con, bed, fridge, stove, tv, and washer.

Living Facilities Group

Facilities (color) and count of PropertyID (size).

Figure 33: Most Frequent Groups of Living Facilities provided in Property Rental

Further analysis was done on the different categories of the living facilities in which air con, fridge, washer, stove, bed, and tv are the most commonly offered living facilities in the property rental according to the horizontal bar chart in Figure 34, with air con being the largest provided living facility (4,875 data records). Moreover, it can be found that bathtub, walk-in closet, audio system, and wireless internet seem to be the least offered living facilities in the property rental, with wireless internet being the lowest provided living facility (3 data records).

Count as an attribute for each Living Facility. Color shows details about Living Facility.

Figure 34: Most offered and Least offered Living Facilities in Property Rental

iii. Location of HDB Flat

Figure 35 visualizes the location of HDB rental flats placed on the Singapore geographic map according to their latitude and longitude points. Data points can be found uniformly distributed around Singapore town areas. Moreover, Figure 36 displays the latitude and longitude points on the boundary of Singapore to ensure that the locations of all HDB flats are correctly placed within the boundary of Singapore.

Figure 35: Location of HDB Rental Flats in Singapore

Figure 36: Boundary of Singapore

iv. HDB Flat Type

Figure 37: Data Categorization according to HDB Flat Type

Data categorization according to 5 HDB flat types (1-room, 2-room, 3-room, 4-room, and 5-room) can be visualized in a pie chart shown in Figure 37, in which 3-room HDB flat, and 4-room HDB flat are the most commonly found in Singapore followed by 5-room HDB flat. There is no 1-room HDB flat available in the current data set. Furthermore, there are data records with no information of the flat type (1,451 data records with the value -1 in the number of rooms) which need to be analyzed further.

v. District Area

Figure 38 and Figure 39 visualize the data distribution of HDB rental flats in Singapore based on different district areas. As shown in the data visualizations, it is found that District 19 has the most significant number of HDB rental offers (1,513 data records) while there is only one data record available in District 17.

HDB Rental Flats in District

Figure 38: HDB Rental Offers in Singapore based on different District Areas

Figure 39: Data Distribution of HDB Rental Offers in different District Areas

4.2.2. Bivariate Statistical Data Analysis

Bivariate data analysis was performed in order to understand the relationship between the two features. In this section, three pairs of features: *rental price and flat type*, *rental price and geolocation*, and *rental price and district area* were analyzed.

i. Rental Price and Flat Type

Data distribution of the rental prices in different flat types was explored as visualized in Figure 40. According to the data visualization, it is found that the average rental price for all flat types is around S\$2,000, and the more the number of rooms, the higher the average rental price is. Moreover, there are potentially overpriced or underpriced rental offers among all flat types, which are required to be explored further. As mentioned in *iv HDB Flat Type* from the 4.2.1 Univariate Statistical Data Analysis, there are data records with a missing value of the flat type (i.e., -1 in the number of rooms). It can be misled into the fact that there is a flat type called -1 and will lead to a severe error in the optimization process due to the correlation between the rental price and flat type.

Figure 40: Data Distribution of Rental Price by HDB Flat Type

ii. Rental Price and Geolocation

Analysis of the data distribution of the rental price on the actual Singapore geographic map was performed to discover the insightful distribution patterns of the rental price. As shown in Figure 41, it can be found that there is no recognizable pattern available between the rental price and geolocation. Therefore, further data analysis is required to search for more interesting patterns. For this purpose, data clustering was performed on the rental price using the k-means clustering algorithm as provided in Figure 42. The result of clustering was visualized on Singapore geographic map, as shown in Figure 43 in which 5 clusters are generated with various rental price ranges.

Figure 41: Data Distribution of Rental Price on Singapore Geographic Map

Inputs for Clustering

Variables: Price
Level of Detail: Latitude, Longitude, Price
Scaling: Normalized

Summary Diagnostics

Number of Clusters: 5
Number of Points: 6999
Between-group Sum of Squares: 17.909
Within-group Sum of Squares: 3.5075
Total Sum of Squares: 21.417

Clusters	Number of Items	Centers Price
Cluster 1	3417	1885.1
Cluster 2	2656	2358.7
Cluster 3	809	2874.9
Cluster 4	108	3545.6
Cluster 5	9	6564.4
Not Clustered	0	

Figure 42: Results of K-Means Clustering on Rental Price

According to the analysis of 5 data clusters, it can be observed that cluster 1 and cluster 2 are the groups with the most data points, which means that most of the data records are categorized into cluster 1 with the rental price range from S\$500 to S\$2,120 (3,417 data records) and cluster 2 with the rental price range from S\$2,150 to S\$2,600 (2,656 data records). Moreover, their spatial data points are generally distributed on Singapore geographic map. However, the spatial data points of cluster 3 with the rental price range from S\$2,648 to S\$3,200 mostly occupy in the southern region of Singapore (lower region of Singapore map) and the spatial data points of cluster 4 with the rental price range from S\$3,250 to S\$4,500 are only available around the central town areas of Singapore. It proves that HDB flats with the higher priced rental fee are around the central town areas and the southern region of Singapore. Cluster 5 consists of only 9 data points, and there is no spatially related pattern found on Singapore geographic map which can interpret them as either a special case or an outlier which proves the analysis of the *i Rental Price* from 4.2.1 Univariate Statistical Data Analysis.

Cluster 1: price range from S\$500 to S\$2,120

Cluster 2: price range from S\$2,150 to S\$2,600

Cluster 3: price range from S\$2,648 to S\$3,200

Cluster 4: price range from S\$3,250 to S\$4,500

Cluster 5: price range from S\$5,100 to S\$8,480

Figure 43: Data Clustering of Rental Price and Visualization on Singapore Geographic Map

iii. Rental Price and District Area

In order to explore the data distribution of the rental price in each district area, a box and whisker plot was used to visualize the statistical population, as shown in Figure 44. The average price of each district area was analyzed. Based on the analysis of the box and whisker plot, it is discovered that there are a few data points, which are far away from the box and whisker plot, which describes that there are some HDB flats with possible overpriced or underpriced rental offers in 14 district areas, which need to be further analyzed for any anomaly outlier case. Moreover, some district areas have very few data points, i.e., District 17, due to the lack of data available during the data collection period. Overall, there is no distinct relationship found between the rental price and district area.

Rental Price in District

Figure 44: Data Distribution of Rental Price in Each District Area

4.2.3. Multivariate Statistical Data Analysis

Multivariate data analysis was performed in order to understand the relationship among more than two features. In this section, multivariate data analysis on *rental price*, *district area*, *town area*, *HDB flat type*, and historical *timeline* was conducted.

i. Rental Price, District Area, and HDB Flat Type

Figure 45: Data Distribution of Rental Price in Each District based on HDB Flat Type

In order to analyze the possible overpriced or underpriced rental offers, data distribution of the rental price in each district based on the HDB flat type was explored and visualized in Figure 45. Based on the data visualization, it is found that the data collection from the web-based property listing and search platform includes the anomaly outlier cases (i.e., overpriced rental fee) and the missing values in the HDB flat type (i.e., the value of -1 in the number of rooms). For example, as for the former case, a rental price of 4-room HDB flat in district 3 is quoted as S\$8,480 which is extremely overpriced compared to the other 4-room HDB flats in the same district area which is between S\$2,000 and S\$4,000. Therefore, it is essential to handle the outliers with care for better performance of the optimization tasks. As for the latter case, it is found that a large number of data records do not seem to have the value of HDB flat type. It may lead to the problem of misleading a new flat type during the later stages of the optimization tasks. Therefore, it is critical to perform data cleansing in the initial stage of data analytics.

ii. Rental Statistics (Rental Price, HDB Flat Type, and Timeline)

Rental statistics were explored to understand the historical and current rental price of the public housing estates in Singapore. Historical data of the HDB rental price trend from the past decade were visualized according to 6 HDB flat types namely: 1-room, 2-room, 3-room, 4-room, 5-room, and executive as shown in Figure 46.

Figure 46: Statistical Trend of HDB Rental Price by Flat Type from the Past Decade in Quarterly Manner

Based on the data visualization of quarterly median rental price by the flat types in 26 town areas, it can be observed that there is not enough statistical rental price information for 1-room HDB flat from the past decade and the limited statistical information for 2-room HDB flat in a few town areas. Moreover, there are some data points at the price value of S\$0K which shows that there is no statistical information provided for some town areas in a particular quarter which might affect the performance of data analysis. Therefore, it is essential to handle the missing values or invalid data points with care to achieve an accurate rental price trend line and valuable insights. Currently, there is no distinct pattern found in the rental price trend line, according to the above data visualization.

iii. Rental Statistics after Data Cleansing (Rental Price, HDB Flat Type, and Timeline)

During the data cleansing process of the rental statistics data set, all missing values and invalid data points were excluded in calculating the average value of the median rental price. Based on the data visualization depicted in Figure 47, it is found that the historical trend of the average median rental price is consistent among all HDB flat types which prove that the rental prices of all HDB flat types either increase or decrease altogether in each particular quarter regardless of the HDB flat type. Hence, it can represent the overall rental price trend line of the public housing estates in Singapore. Moreover, it can be found that the average median rental price is consistent with the HDB flat type among the whole rental price trend line, i.e., the more the number of rooms, the higher the rental price is.

Quarterly Average Median Rental Price in Town Area

Average of Price for each Year & Quarter (Combined) broken down by Flat Type. Color shows details about Flat Type. The data is filtered on Price, which includes values greater than or equal to 1. The view is filtered on Flat Type, which keeps executive, room2, room3, room4 and room5.

Figure 47: Statistical Trend of Average Median Rental Price by Flat Type from the Past Decade

iv. Rental Statistics (*Rental Price, HDB Flat Type, Town Area, and Timeline*)

Within various district/town areas, the rental price trend was visualized according to the HDB flat type as shown in three figures below: Figure 48 for 2-room and 3-room flats, Figure 49 for 4-room and 5-room flats, and Figure 50 for executive flats. In the case of a 2-room HDB flat type, there are very few data records (38 data points) available in Bukit Merah and Queenstown town areas. As for the case of a 3-room HDB flat type, it is found that Central town areas have the highest rental price records and Woodlands town area has the lowest rental price records. 4-room and 5-room HDB flat types are found to possess the most concentration of data points in the past decade, and according to the general pattern, Bukit Merah town area has the highest rental price, and Bukit Panjang town area has the lowest rental price in both flat types. Although there seems to be no distinct price trend for executive HDB flat type, Tampines, Jurong East, and Bedok town areas have a high rental price trend, and Bukit Panjang and Sembawang town areas have a low rental price trend occasionally.

Figure 48: 10-Year Timeline of Rental Price Trend in Town Areas by 2-room and 3-room HDB Flat Types

Figure 49: 10-Year Timeline of Rental Price Trend in Town Areas by 4-room and 5-room HDB Flat Types

Figure 50: 10-Year Timeline of Rental Price Trend in Town Areas by executive HDB Flat Type

4.3. Summary

In this chapter, the descriptive analytics was conducted with the use of various types of data visualization techniques in order to achieve a better understanding of the statistical data and to discover the knowledgeable insights which can contribute to defining the problem and designing the solution framework. Firstly, univariate data analysis was performed on *rental price*, *living facilities*, *location of HDB flat*, *HDB flat type*, and *district area* with different data visualization techniques: box plot, histogram, bubble chart, horizontal bar chart, pie chart and geographic map. It can be discovered that the most significant data point concentration is within the rental price between S\$2,100 and S\$2,520, with the average of S\$2,210.6. The facilities group of air con, bed, fridge, stove, tv, and washer is the most commonly offered living facilities in the property rentals. Moreover, 3-room HDB flat and 4-room HDB flat are the most commonly found in Singapore. Based on the locations of HDB rental flats, it can be found that District 19 has the most significant number of HDB rental offers.

Secondly, bivariate data analysis was performed on three pairs of features to observe the relationship between the two features: *rental price and flat type*, *rental price and geolocation*, and *rental price and district area*. Complex data visualization techniques were used. Based on the data visualization, the average rental price for all flat types is around S\$2,000, and the number of rooms is highly correlated with the average rental price. Furthermore, with the data clustering method, it can be observed that HDB flats with the higher priced rental fee are around the central town areas and the southern region of Singapore. And, there was no distinct relationship found between the rental price and district area.

Finally, multivariate data analysis was conducted to understand the relationship among more than two features: *rental price*, *district area*, *town area*, *HDB flat type*, and *historical timeline*. Data cleansing was performed, and it is found that the historical trend of the average median rental price is consistent among all HDB flat types and the rental prices of all HDB flat types either increase or decrease altogether in each particular quarter regardless of the HDB flat type. Moreover, it can be found that the average median rental price is consistent with the HDB flat type among the whole rental price trend line.

5. SYSTEM DESIGN

5.1. Multi-Objective Optimization Problem

In this chapter, a real-world property listing and search method was designed as a multi-objective optimization problem, and a step by step procedure of the problem definition and search designs were described.

5.1.1. Problem Formulation

Problem definition was designed to formulate the multi-objective optimization problem for the property listing and search services. Three essential parts of the optimization problem were defined in this section, namely: *decision variables*, *constraints*, and *objective functions*.

i. Decision Variables

Decision variables for the multi-objective optimization problem were defined as the district area (*district*), the number of rooms (*room*) and the living facilities (*living*) included in each real estate property. Mathematically, a solution for a multi-objective optimization problem is a vector of 3 decision variables x in the solution space X and is generally defined as:

$$x = [x_{district}, x_{room}, x_{living}^L]^T \quad (9)$$

where each decision variable is represented as:

$x_{district}$ is the district area and encoded as an integer value

x_{room} is the number of rooms and encoded as an integer value

x_{living}^L is a set of living facilities included in the property and encoded as a bit string with the length L

ii. Constraints

Constraints were set to make sure that all solutions considered for the multi-objective optimization problem are feasible and acceptable. Constraints can be dynamically defined according to the preference criteria set by the customer. There were two types of constraints defined in the multi-objective optimization problem: 1) *Constraints on Decision Variables* and 2) *Constraints on Objective Functions*. By default, **constraints on decision variables** for a multi-objective optimization problem are mathematically defined as:

$$\begin{aligned}
 x &\in X \\
 \text{s. t. } &x_{district}^{min} \leq x_{district} \leq x_{district}^{max} \\
 &x_{room}^{min} \leq x_{room} \leq x_{room}^{max} \\
 &x_{living}^{min} \leq x_{living} \leq x_{living}^{max} \\
 &x_{living}^{i \in L} = \begin{cases} 1 & \text{if facility is included} \\ 0 & \text{otherwise} \end{cases}
 \end{aligned} \tag{10}$$

where the minimum and maximum values are defined as:

$$\begin{aligned}
 x_{district}^{min} &= 1 \\
 x_{district}^{max} &= 28 \\
 x_{room}^{min} &= 1 \\
 x_{room}^{max} &= 5 \\
 x_{living}^{min} &= x_1 x_2 \dots x_L, x_i = 0 \\
 x_{living}^{max} &= x_1 x_2 \dots x_L, x_i = 1
 \end{aligned} \tag{11}$$

iii. Objective Functions

In this multi-objective optimization problem model, three objectives were formulated which are to *minimize* the price expense f_{price} , *maximize* the number of living facilities provided in each property $f_{facilities}$ and *minimize* the estimated distance to the specified location $f_{distance}$. Therefore, mathematically, a multi-objective optimization problem is defined as:

$$\text{minimize } f(x) = [f_{price}(x), -f_{facilities}(x), f_{distance}(x)]^T \quad (12)$$

Alternatively, the time taken to travel to the specified location while taking traffic condition of the roads into account can be considered as the objective function $f_{duration}$. In this case, the multi-objective optimization problem can mathematically be defined as:

$$\text{minimize } f(x) = [f_{price}(x), -f_{facilities}(x), f_{duration}(x)]^T \quad (13)$$

iv. Minimization of Price Expense

In this multi-objective optimization problem model, the price expense was to be minimized. The mathematical form of minimizing the price expense is defined as:

$$\min f_{price}(x) = x_{price} \quad (14)$$

Price expense can be calculated based on the current monthly price value quoted by the house owner, the period of the lease contract (i.e., 6 months, 12 months or 24 months) and the predicted price based on the district/town area. The mathematical form of minimizing the price expense is defined as:

$$\min f_{price}(x) = \alpha x_{price} + \beta \quad (15)$$

where α and β are defined as:

$$\alpha = \text{period of the lease contract in months} \quad (16)$$

$$\beta = \text{predicted monthly price} \times \text{period of the lease contract in months}$$

v. Maximization of Living Facilities

In this multi-objective optimization problem model, a total number of living facilities offered in each real estate property was to be maximized. The mathematical form of maximizing the living facilities is defined as:

$$\max f_{facilities}(x) = \sum_{i=1}^{23} w_i x_{living}^i \quad (17)$$

$$\text{where } x_{living}^i \begin{cases} 1 & \text{if facility is included} \\ 0 & \text{otherwise} \end{cases} \text{ and } \sum_{i=1}^{23} w_i = 100$$

A list of living facilities commonly offered in each real estate property is provided in a non-exhaustive manner with their respective weights of frequency distribution computed based on Figure 34, as shown in Table 6.

Index i in x_{living}^i	Living Facilities	Weights
1	Aircon	14.72
2	Audio System	0.12
3	Bathtub	0.43
4	Bed	9.59
5	Closet	3.23
6	Corner Unit	1.53
7	Dining Room Furniture	4.06
8	Dryer	0.49
9	Fridge	11.05
10	Low Floor	0.56
11	Oven	0.90
12	Sofa	4.25
13	Stove	10.30
14	TV	7.14
15	Walk-in Closet	0.31
16	Washer	10.33
17	Wireless Internet	0.01
18	Bomb Shelter	0.85
19	High Floor	2.20
20	Renovated	2.25
21	Utility Room	0.65
22	Pets Allowed	0.55
23	Fully Furnished	14.51

Table 6: List of Living Facilities provided in Real Estate Property and their Weights of Frequency Distribution

vi. *Minimization of Distance or Duration*

In this multi-objective optimization problem model, the distance or duration between the location of each real estate property and the specified geographical location points was to be minimized. The mathematical form of minimizing the distance traveled is defined as:

$$\min f_{distance}(x) = d(x, p) \quad (18)$$

where the distance function d from point x to point p can be formulated as:

$$d(x, p) = ST_Distance_Sphere(x.geolocation, p.geolocation) \quad (19)$$

in which $ST_Distance_Sphere$ is a spatial function provided by MySQL that calculates the estimated spherical distance in meter (m) between two points on the earth's surface [36]. As for the case of minimizing the duration, it is defined as:

$$\min f_{duration}(x) = D(x, p) \quad (20)$$

where the duration function D from point x to point p can be formulated as:

$$D(x, p) = DistanceMatrixAPI(origin_x, destination_p, departure, traffic, mode) \quad (21)$$

in which $DistanceMatrixAPI$ is Distance Matrix API provided by Google that calculates the estimated real-life travel time between two location points and provides the travel time of the recommended route in second (s) considering the traffic conditions of the roads [37].

Constraints on the objective functions can be dynamically defined according to the preference criteria set by the customer. By default, there was no constraint set on the objective functions. Custom ***constraints on objective functions***, which can be dynamically set by the customer, on the multi-objective optimization problem are mathematically defined as:

$$f_{price}^{min} \leq f_{price}(x) \leq f_{price}^{max} \quad (22)$$

$$f_{distance}(x) \leq f_{distance}^{max}$$

$$f_{duration}(x) \leq f_{duration}^{max}$$

5.1.2. Exhaustive Search (Baseline)

Exhaustive search was initially designed as a baseline search algorithm to observe all possible non-dominated solutions and evaluated the appropriate performance measure of the multi-objective optimization evolutionary algorithm search. In the exhaustive search, all feasible solutions are linearly analyzed according to the three objectives (i.e., *minimize price expense*, *maximize living facilities*, and *minimize distance/duration*).

The analogy of the exhaustive search is similar to the search of the maximum/minimum number in a list of numbers, in which the current best-known maximum/minimum number observed is kept during the search. In a multi-objective optimization problem, a list of best-known optimal solutions which are non-dominated with each other was kept. The algorithm of an exhaustive search which finds a set of non-dominated solutions among all feasible solutions is described in Algorithm 3 in which the linear search is performed among all feasible solutions in the set X and the set of non-dominated optimal solutions X^* is returned.

Algorithm 3: Exhaustive Search (Baseline)

```
input:     $X$ : the set of feasible solutions
output:   $X^*$ : the set of non-dominated optimal solutions
1:         initialize  $X^* \leftarrow \emptyset$ 
2:         for each solution  $x_i$  in  $X$  do
3:             if  $x_i$  is first solution and  $X^*$  is empty set
4:                  $X^* \leftarrow \{x_i\}$ 
5:             else
6:                 for each non-dominated solution  $x_j^*$  in  $X^*$  do
7:                     CompareDominance( $x_i, x_j^*$ )
8:                     if  $x_i$  strongly/weakly dominates  $x_j^*$ 
9:                         remove  $x_j^*$  from  $X^*$ 
10:                    end if
11:                end for
12:                if  $x_i$  and remaining solutions in  $X^*$  are non-dominated
13:                     $X^* \leftarrow \{x_i\}$ 
14:                end if
15:            end if
16:        end for
```

Algorithm 4 describes the comparison of dominance between two feasible solutions which was applied in an exhaustive search, in which two feasible solutions x_1 and x_2 are compared in terms of three objective functions. A strong dominance is defined as a situation in which one solution dominates another solution in all of the objective functions. A weak dominance occurs when one solution dominates another solution in at most $k - 1$ objective functions and two solutions tie in the remaining objective function(s). Non-dominance happens when neither one of the feasible solutions does not dominate another in any objective function.

Algorithm 4: Dominance Comparison

input: x_1 : feasible solution 1
 x_2 : feasible solution 2

output: x_i : a solution which dominates strongly/weakly ($i = 1$ or 2)
 k : an indicator for non-dominance

```

1:  case 1: strong dominance
2:      if  $x_1 > x_2$  in all three objectives
3:          return  $x_1$ 
4:      else return  $x_2$ 
5:      end if
6:  case 2: weak dominance
7:      if  $x_1 > x_2$  in two objectives and  $x_1 = x_2$  in one objective
8:          return  $x_1$ 
9:      else return  $x_2$ 
10:     end if
11:     if  $x_1 > x_2$  in one objective and  $x_1 = x_2$  in two objectives
12:         return  $x_1$ 
13:     else return  $x_2$ 
14:     end if
15:  case 3: non-dominance
16:      if  $x_1 = x_2$  in all three objectives
17:          return non-dominance indicator  $k$ 
18:      end if

```

5.1.3. Multi-Objective Optimization Evolutionary Algorithm Search

Exhaustive baseline search provided a list of global optimal solutions after a linear search of all feasible solutions in the solution space. It was computationally intensive and took time and space incrementally if the search space becomes larger. Therefore, the search method using an Evolutionary Algorithm (EA) was designed to solve the multi-objective optimization problem which can achieve an efficient time and space search performance. Among various multi-objective optimization evolutionary algorithms, an improved version of the Non-dominated Sorting Genetic Algorithm, also known as NSGA-II, was adopted in this dissertation work.

Figure 51 shows a graphical representation of an overall algorithm workflow designed for solving a multi-objective optimization problem with the evolutionary algorithm. Concepts of NSGA-II algorithm were adapted from the MOEA Framework [38]: non-dominated ranking, offspring generation, and the candidate selection for the next generation of the population to achieve the best known optimal solutions. A graphical representation of the overall algorithm workflow of NSGA-II algorithm is described in Figure 52 [39] in which the general step by step procedures of the NSGA-II algorithm is displayed.

Figure 51: Overall Algorithm Workflow of Multi-Objective Optimization Evolutionary Algorithm Search

Figure 52: Overall Algorithm Workflow of Fast Non-dominated Sorting Genetic Algorithm (NSGA-II)

The algorithm of the candidate evaluation component was designed to perform a mapping from a candidate solution from the solution search space, generated by the NSGA-II algorithm, to an actual solution from the real-world property data set. Evaluation of three objective functions was defined in Algorithm 5 in which an actual solution is selected as the nearest point to the candidate solution based on three decision variables. Once the actual solution has been found, three objective functions are computed from the actual solution and applied in the evaluation of a candidate solution for its survival to the next generation.

Algorithm 5: Candidate Evaluation

input: x_{can} : candidate solution

output: $f(x_{can})$: objective function values of a candidate solution

- 1: initialize x_{actual}
 - 2: find the nearest point to x_{can}
 - 3: $x_{actual} = FindNearestPoint(x_{can})$
 - 4: evaluate x_{can}
 - 5: $f_{price}(x_{can}) = CalculatePriceObjFun(x_{actual})$
 - 6: $f_{facilities}(x_{can}) = CalculateFacilitiesObjFun(x_{actual})$
 - 7: $f_{distance}(x_{can}) = CalculateDistanceObjFun(x_{actual})$
 - 8: save ID of x_{actual} for an audit trail of x_{can}
-

The search of the nearest point to the candidate solution based on three decision variables are provided in Algorithm 6 in which the nearest point is searched in a sequential manner: 1) the search of the district area, 2) the search of the HDB flat type, and 3) the search of the living facilities offered. α , β and θ are the constant weightages which attributes to the distance in terms of the decision variables. The condition in which there is more than one nearest point is handled by performing the dominance comparison of living facilities.

Algorithm 6: Find Nearest Point to the Candidate Solution

input: x_{can} : candidate solution
 X_{actual} : the set of actual solutions

output: x_{actual} : actual solution which is the nearest to x_{can}

```

1:   for each solution  $x_i$  in  $X_{actual}$  do
2:      $x_{nearest}$ : current nearest point
3:     find the nearest point
4:       if  $x_{can}^{district} == x_i^{district}$ 
5:         if  $x_{can}^{room} == x_i^{room}$ 
6:            $x_i^l = FindDistance(x_{can}^{living}, x_i^{living}) \times \alpha$ 
7:            $x_i^{dist} = x_i^l$ 
8:         else
9:            $x_i^r = FindDistance(x_{can}^{room}, x_i^{room}) \times \beta$ 
10:           $x_i^{dist} = x_i^l + x_i^r$ 
11:        else
12:           $x_i^d = FindDistance(x_{can}^{district}, x_i^{district}) \times \theta$ 
13:           $x_i^{dist} = x_i^d + x_i^l + x_i^r$ 
14:        end if
15:        if  $x_i^{dist} < x_{nearest}^{dist}$ 
16:           $x_{nearest} = x_i$ 
17:        else if  $x_i^{dist} == x_{nearest}^{dist}$ 
18:          if  $FacilitiesObjFun(x_i) > FacilitiesObjFun(x_{nearest})$ 
19:             $x_{nearest} = x_i$ 
20:          end if
21:        end if
22:      end for

```

5.2. Price Estimation Model

In this section, a price estimation model was designed as a regression problem, and a step by step procedures of the design of artificial neural networks were described.

5.2.1. Design of Artificial Neural Networks

Artificial neural networks was designed to formulate the regression problem for the price estimation model. Five essential components of the neural networks were defined in this section, namely: *input/output variables, the architecture of neural networks, activation function, learning method, and loss function.*

i. Input/Output Variables

Input data set was defined as X with n features. In the price estimation model, there are total of 34 input features available. Each input vector x with n features is defined as:

$$x = [x_1, x_2, \dots, x_n] \quad (23)$$

Table 7 describes 34 input features of the real estate property dataset to be fed in the artificial neural networks model.

Index i in x_i	Features	Index i in x_i	Features
1	Latitude	18	Dining Room Furniture
2	Longitude	19	Dryer
3	District	20	Fridge
4	Number of Rooms	21	Low Floor
5	Number of Beds	22	Oven
6	Number of Baths	23	Sofa
7	Sqft (square foot)	24	Stove
8	Psf (price per square foot)	25	TV
9	HDB	26	Walk-in Closet
10	For Rent	27	Washer
11	Lease Type	28	Wireless Internet
12	Aircon	29	Bomb Shelter
13	Audio System	30	High Floor
14	Bathtub	31	Renovated
15	Bed	32	Utility Room
16	Closet	33	Pets Allowed
17	Corner Unit	34	Fully Furnished

Table 7: Input Features of Price Estimation Model

Output data was defined as Y with one value, which is the rental price of the real estate property.

ii. Architecture of Neural Networks

Two layers of feedforward neural networks model were designed for the price estimation model, which includes 1) the input layer with 34 neurons for the input features, 2) one hidden layer with 18 neurons and 3) the output layer with one neuron. Figure 53 displays the architecture design of the neural networks model to be trained for the price estimation.

Figure 53: 2-Layer Neural Networks Design of Price Estimation Model

iii. Activation Function

ReLU or Rectified Linear Units activation function, as shown in Figure 54, was used in the hidden layer of the neural networks. The ReLU activation function is defined as:

$$y = \max(0, x) \quad (24)$$

Figure 54: ReLU Activation Function

iv. Learning Method and Loss Function

The supervised learning method, *Gradient Descent* learning algorithm, was applied to the feedforward neural network. The difference between the output produced by the neural networks model and the actual output of the training data is defined as the error which is required to be minimized during the training of the neural networks. One of the loss functions, called *Mean Squared Error*, was used for the optimization process, and it is defined as:

$$\min MSE = \frac{1}{2} \sum_i^m (y_i - \hat{y}_i)^2 \quad (25)$$

v. Learning in Neural Networks

A step by step procedures of the learning process in the artificial neural networks is described in Algorithm 7. The trained neural network model was saved for the offline price estimation of the real estate property.

Algorithm 7: Learning Process of Artificial Neural Networks

input: (X, Y) : a set of input-output pairs
output: *model*: trained neural networks model

- 1: **load** dataset: X and Y
- 2: **do** data normalization
- 3: **prepare** three datasets
- 4: x_{train}, y_{train} : training dataset (70%)
- 5: x_{val}, y_{val} : validation dataset (15%)
- 6: x_{test}, y_{test} : test dataset (15%)
- 7: **construct** a neural network model
- 8: *input layer*: 34 neurons
- 9: *hidden layer*: 18 neurons with *ReLU* activation function
- 10: *output layer*: 1 neuron
- 11: optimizer: *stochastic gradient descent*
- 12: loss function: *mean squared error*
- 13: **train** and **validate** neural network model
- 14: epoch: 100
- 15: $\text{train}(\text{model}, x_{train}, y_{train})$
- 16: $\text{validate}(\text{model}, x_{val}, y_{val})$
- 17: **test** neural network model
- 18: $\text{predict}(\text{model}, x_{test}, y_{test})$
- 19: **save** *model*

5.3. Web-based Property Listing and Search Platform

5.3.1. System Architecture Design

Figure 55 presents an overall system architecture design of the web-based property listing and search platform to be developed for this dissertation work. Development of the web-based property listing and search platform was divided into two major components: the server-side component and the client-side component. In the server-side component, multi-objective optimization tasks are performed dynamically according to the incoming requests from the client-side component. In the client-side component, interactions with the users (i.e., the decision maker) are processed through a web-based search platform. Computationally intensive operations, such as price prediction in different district areas, were designed to perform in an offline pre-processing manner. Moreover, the price estimation model was designed as the offline model, which will be run offline periodically to update the estimated price of the real estate property dataset stored in the database management system.

Figure 55: System Architecture Design of Web-based Property Listing and Search Platform

5.3.2. Software Architecture Design

An overall software architecture design of the web-based property listing and search platform was constructed as displayed in Figure 56 in which the Model-View-Controller (MVC) architectural pattern was adopted for the development of the web-based search platform. The main components responsible for the performance of the multi-objective optimization tasks were presented. In the *Controllers* component package, *MainController* is responsible for managing the overall workflow of the web-based property listing and search platform, *MOOController* is designed for handling the multi-objective optimization tasks, *ObjFunController* is defined to assist in the computation of three objective functions and *DBController* is responsible for the data requests between the web-based search platform and the database management system.

In the *Models* component package, *PropertyProblem* represents a candidate solution for the multi-objective optimization problem, and it stores the information about the decision variables, constraints, and evaluation function that processes the objective function values. *PropertyObjFun* stores the objective function values of each candidate solution, and *Property* represents the real estate property. The *Controllers* package will interact with the *Models* package to store and process the data during the runtime or online.

In the *Views* component package, *Map* will assist in the display of the interactive map-based page view for the users to perform the property listing and search activities. It is responsible for the interactive visualization of the search results. *Listing* will provide the users with a list-based page view. The *Views* package will interact with the *Controllers* package to perform appropriate property listing and search activities depending on the requests from the users. Based on the incoming request from the user, the *Controllers* package will interact with the *Database* for any necessary data to be processed for the multi-objective optimization tasks.

Figure 56: Software Architecture Design of Web-based Property Listing and Search Platform

5.3.3. Database Design

Figure 57 graphically presents the database design prepared for the data management and the storage of data sources that are utilized for the web-based property listing and search platform. In the database design, various data tables were created, such as `property`, `image`, `hdb_rental_statistics`, `rental_prediction`, `district`, `distance_matrix`, `pareto_popularity`, and so on to assist in a computationally efficient multi-objective optimization performance.

Data sources related to the real estate property listings and their corresponding images collected during the data collection stage are stored in the `property` and `image` data tables. Historical statistics related to the rental price is stored in the `hdb_rental_statistics` data table while the pre-processed price prediction data is stored in the `rental_prediction` data table. The `distance_matrix` data table is used for the data storage of the real-world distance and duration between two geographical locations which are collected during the multi-objective optimization tasks. The `pareto_popularity` data table is designed to keep track of the best-known optimal solutions provided by the multi-objective optimization evolutionary algorithm for further data analytics. Additionally, the data tables such as `district` and `train_station` are used to store the publicly available data sets which are applicable in the improvement of the multi-objective optimization tasks.

Figure 57: Database Design of Web-based Property Listing and Search Platform

5.3.4. User Interface Design

Figure 58 displays the user interface design of an online web-based property listing and search platform in which the main pages of the search platform were designed and drafted to observe the workflows of the interactions among pages and components to be included in each page. **Home Page (Dashboard)** page displays the visualization of the data analytics previously described in the 4.2 *Descriptive Analytics* to provide the valuable knowledge about the real estate property listings offered in the web-based property listing and search platform.

Figure 58: User Interface Design of Web-based Property Listing and Search Platform

Interactive Map page allows the user to perform a search on the property listings which satisfy all preference criteria (i.e., objective functions) efficiently. In order to achieve user convenience, various tools are provided within a single page. **Property Listings** page allows the user to perform the criteria-based search effortlessly using a user-friendly control panel. **Price Estimation** page will provide an approximate price value of each real estate property for the price negotiation between the house owner and customers. **Property Bank** page manages the current property listings offered in the search platform, which is only accessible by the administrator.

6. SYSTEM IMPLEMENTATION

6.1. Web-based Property Listing and Search Platform

6.1.1. Web Application Framework

In order to achieve the effective and efficient development of an online web-based property listing and search platform, one of the open source Java web application frameworks, *Play Framework* [40] was selected in this dissertation work. *Java* programming language is used for the back-end development: multi-objective optimization tasks and *Scala* programming language is applied to the front-end development: online web-based property listing and search application. Play Framework follows the MVC software architectural design concept, and its server back end adopts the *Netty* server. Play Framework package version 2.5.10 is currently used in this dissertation work.

6.1.2. Database Management System

As for the data storage and management, the relational data model is preferred, and *MySQL* [41], one of the open source *Relational Database Management Systems* (RDBMS) was selected as a local database for this dissertation work. *SQL* language is used to store, manipulate, and retrieve the data back and forth to the database during run time. *MySQL Connector/J*, the JDBC driver for MySQL to connect with Java programming language, was set up for the connection between Java and MySQL. MySQL Workbench 6.3 Community version is currently used in this dissertation work.

6.1.3. Integrated Development Environment

Integrated Development Environment (IDE) was used to make the development life cycle of the web-based property listing and search platform more effective and efficient. One of the Java IDEs, *IntelliJ IDEA* [42] was selected for the developments of both back-end and front-end systems. IntelliJ IDEA Ultimate 2017.2 with the Student license is currently used in this dissertation work.

6.1.4. Google Maps APIs

Since the web-based property listing and search platform incorporates the map-based search services, the web services provided by Google Maps APIs [43] were applied in both back-end and front-end developments.

APIs for Back-end Development

During the development of the back-end multi-objective optimization tasks, *Distance Matrix API* [44] was used to search the real-world recommended route among the geolocation points and retrieve the estimated distance and duration which can be utilized in the evaluation of the candidate solutions. Furthermore, *Geocoding API* [45] was used for the conversion of the actual addresses of the real estate property into the geographic coordinates to position the property listings on the geographic map view and vice versa.

However, there was a usage limitation for each Google Map API service due to the pay-as-you-go pricing model currently adopted by Google. As for the *Distance Matrix API*, it costs 0.01 USD per element, which is calculated as the multiplication of the number of original addresses and destination addresses per query request. It makes the exhaustive search too expensive due to the linear comparison among all feasible solutions. Therefore, to achieve cost-effectiveness, the computation of the spherical distance among the geographic points was adopted in both exhaustive search method and the multi-objective optimization evolutionary algorithm search with the *distance* objective function during the experiments. Distance Matrix API was only adopted in the multi-objective optimization evolutionary algorithm search with the *duration* objective function. In order to achieve faster optimization performance, the results of API request calls were stored in the local database for the reusability in future API request calls which are similar.

As for *Geocoding API*, only 2,500 daily API request calls were allowed before the adaption to the pay-as-you-go model, which costs 0.005 USD per request call. Therefore, during the *4.1 Data Collection* process of *4.1.1 Singapore's Public Housing Estates*, 2,000 HDB flats were pre-processed daily and stored their geographical information in the local database.

APIs for Front-end Development

In the development of the front-end online web-based property listing and search platform, ***Maps JavaScript API*** [46] was used for the visualization of the geographical information on the map and display on the web pages and mobile devices. Property listings, which are the best-known optimal solutions provided by the multi-objective optimization evolutionary algorithm search, are displayed on the map to assist the users in decision-making. Similar to other APIs, APIs for the front-end development also adapts to the pay-as-you-go pricing model.

Places API [47] was applied for the implementation of an autocomplete service, which can search for information about the desired places. ***Directions API*** [48] was integrated into the map service to assist the users in finding the recommended driving routes among the property listings and places provided by the users. It calculates the distance and duration of each route. Additionally, ***Geolocation API*** [49] was embedded in the map for the detection of the user's current position when he/she is using a mobile device. In this dissertation work, API usages of the front-end online web-based property listing and search platform were kept within the standard limitation to achieve the cost-effectiveness.

6.1.5. MOEA Framework

To develop the multi-objective optimization evolutionary algorithm search in the property listing and search system, MOEA Framework [38], one of the open source Java libraries, was utilized to design the multi-objective optimization problem model as explained in the *5 System Design*. Fast *3.4 Non-Dominated Sorting Genetic Algorithm* (NSGA), which is provided by the MOEA Framework, was selected for solving the optimization tasks. The candidate evaluation function was designed and constructed according to the procedure mentioned in Algorithm 5.

6.1.6. User Interface

Interactive Map

User Interface (UI) of an interactive map page was designed and developed for the front-end online web-based property listing and search platform in which all relevant functionalities are provided to perform the optimization tasks conveniently within one page. In order to achieve the user-friendliness, the layout of the *Interactive Map* page was divided into four major sections, as shown in Figure 59:

1. Control Panel
2. Map Viewer
3. Travel Scheduler
4. Best Known Property Listings

Figure 59: User Interface of Interactive Map Page

Control Panel, as shown in Figure 60, provides the user with the ability to make the property listings and search in three different ways: 1) the exhaustive search, 2) the multi-objective optimization with the distance traveled and 3) the multi-objective optimization with the duration taken. Adjustments can be made on three objective functions with the use of price range, facility filter, distance meter, and time setter. Moreover, the functionality to rank the property listings according to various preference priorities is provided.

Figure 60: Control Panel of Interactive Map Page

In order to visualize the property listings on the geographic map, *Map Viewer* was designed and embedded, as shown in Figure 61. Results of the best known optimal solutions can be visualized conveniently on the geographic map, and the rank of the solutions can be easily analyzed based on the choice of preference priority set by the user as displayed in Figure 62.

Figure 61: Map Viewer of Interactive Map Page

Visualization of Best-Known Optimal Solutions without Ranking

Ranking of Best-Known Optimal Solutions based on Price Expense

Figure 62: Ranking of Best-Known Optimal Solutions according to Various Preference Priorities

Moreover, the recommended routes among the property listings and locations specified by the user can be found with the use of *Travel Scheduler* as shown in Figure 63 which allows the user to define more than one location and provides the recommended routes with the information of distance in kilometre and duration in minute. In Figure 64, the recommended routes are visualized in the Map Viewer, and the information of distance and duration are provided in the ascending order for the user to observe the property listings which are the nearest to the specified locations.

Travel Schedule

Destinations

Enter a location

Route

View

Route Results

Distance (km)	Duration (min)

Figure 63: Travel Scheduler of Interactive Map Page

Figure 64: Visualization of Recommended Routes among Property Listings and Locations specified by the user

Best Known Property Listings section displays the real estate property listings, which are the optimal solutions found by the search algorithms. Relevant information such as the full address of the real estate property, price, living facilities and distance/duration between the real estate property and the specified locations specified are shown in Figure 65 and Figure 66.

Figure 65: Best Known Property Listings of Interactive Map Page

Figure 66: Property Listings displayed in Best Known Property Listings section with relevant information

6.2. Price Estimation Model

6.2.1. Integrated Development Environment

Integrated Development Environment (IDE) was used for the development of the price estimation model more effective and efficient. Python programming language was used for the training of artificial neural networks. One of the Python IDEs, *PyCharm* [50] was selected for the developments of the standalone offline system. PyCharm 2018.3 Professional Edition is currently used in this dissertation work.

6.2.2. Keras

In order to perform the training of the neural networks and the price estimation using the neural networks, Keras, the Python deep learning library [51], which is a high-level neural networks API, was applied in the development. Architecture of the neural networks was designed and set up with Keras.

6.2.3. Training and Validation of Neural Networks

Neural networks model was trained under the different settings of the number of epochs. Epoch is defined as the number of times the whole dataset is fed to the neural network. One epoch means that the whole dataset is fed to the neural network model only once. For the experiment, the epoch was set as 25, 50, 75, and 100 in each training cycle, and Figure 67 shows the results of the neural network model's loss function. Based on the training results, it can be observed that the results of loss function are achieved around 0.005 in all four values of epoch parameter. Therefore, it can be concluded that a small number of epoch value can be selected for the training of the neural network model to achieve the similar result of loss function and faster training processing time.

Epoch = 25

Epoch = 50

Epoch = 75

Epoch = 100

Figure 67: Training Results of Neural Networks in Different Epoch Settings

7. SYSTEM TESTING

7.1. Experimental Results

A brief experiment was done to analyze the performance of the multi-objective optimization in the property listing and search system. The back-end system, which can perform the multi-objective optimization tasks, was initially developed and assessed for the performance of the property search.

7.1.1. Experimental Setup

The back-end system was developed with IntelliJ IDEA, and the initial performance assessment was done in the local environment. Table 8 provides the local environment settings prepared for the experiments, and Table 9 describes the parameters setting of the multi-objective optimization evolutionary algorithm, which was used for the experiments.

Local Environment Setting

Windows Edition	Windows 7 Professional
System	Dell Precision T3600
Processor	Intel Xeon CPU E5-1650 0 @ 3.20 GHz
Memory	16.0 GB
System Type	64-bit Operating System

Table 8: Local Environment Setting for Performance Assessment

Parameters Setting of MOEA

MOEA Technique	Non-dominated Sorting Genetic Algorithm (NSGA)
Population Size	100
Simulated Binary Crossover Rate	1.0
SBX Offspring Distribution Index	15.0
Polynomial Mutation Rate	$\frac{1}{\# \text{ of decision variables}}$
PM Offspring Distribution Index	20.0
Half Uniform Crossover Rate	1.0
Bit Flip Mutation Rate	0.01
Maximum Objective Function Evaluations	5000

Table 9: Parameters Setting for Multi-Objective Evolutionary Algorithm

For the initial performance analysis, 20 geographic coordinate points were manually selected for the distance objective function evaluation. As shown in Figure 68, various location points were chosen on the Singapore map, and the multi-objective optimization tasks were done independently for each location point.

Figure 68: 20 Handpicked Geographic Coordinate Points on Map for Performance Analysis

7.1.2. Initial Performance Assessment

In this experiment, both exhaustive search and multi-objective optimization-based search were run in order to observe the global Pareto optimal solutions and the best-known Pareto optimal solutions, respectively. Afterward, the best-known Pareto optimal solutions were compared with the global Pareto optimal solutions for the performance assessment of the optimization tasks. Simple performance measurement was done by using a Confusion Matrix as described in Table 10 in which search performance results were recorded in detail and various components of the Confusion Matrix: Recall, Precision, F-Score, Accuracy, Misclassification Rate, and False Positive Rate were computed.

Based on the results, it is found that the *accuracy*: $\frac{TP+TN}{Total\ Solutions} = 99\%$ is achieved in a single run of the multi-objective optimization search. According to the *precision* value, which determines how much the optimization model can search the Pareto optimal solutions out of the global Pareto optimal solutions, it is discovered that the multi-objective optimization tasks achieve more than 0.65 precision value in most cases. However, the *recall* value which can

provide information about the true positive rate is found to be low, around 0.46 on average. Based on this performance analysis, it is essential to improve the performance of multi-objective optimization. Various improvements were considered to assist in the optimization tasks such as the adjustment of parameters setting of MOEA, initialization of the population and decoding the candidate individuals into the actual solutions (selection of the nearest solution points in Algorithm 5).

In terms of the search space in optimization, it can be observed that the multi-objective optimization evolutionary algorithm evaluates around 2000 solutions (i.e., property listings) instead of the total 8463 solutions which achieves the fast online processing time, less than 10 seconds, while the exhaustive search takes around 30 seconds for a single run. It gives the optimization-based search superiority in the web-based online property listing and search system where the users can provide various inputs (i.e., locations) dynamically during the search.

Location Points	Search Method	# of Pareto Optimal Solutions	# of Comparisons / Search Space	Processing Time (seconds)	Confusion Matrix									
					TP	TN	FP	FN	Recall	Precision	F-Score	Accuracy	Misclassification Rate	False Positive Rate
Nanyang Child Care Centre	Baseline	40	436013	32	16	8402	21	24	0.400	0.432	0.416	0.995	0.005	0.002
	MOEA	37	1803	8										
11 Kent Ridge Rd, SG 119220	Baseline	49	340963	31	27	8407	7	22	0.551	0.794	0.651	0.997	0.003	0.001
	MOEA	34	1502	6										
Mount Faber Rd, SG 099205	Baseline	30	236440	31	13	8431	2	17	0.433	0.867	0.578	0.998	0.002	0.000
	MOEA	15	1388	5										
4 Lor M Telok Kurau, SG 425283	Baseline	37	248679	32	14	8419	7	23	0.378	0.667	0.483	0.996	0.004	0.001
	MOEA	21	1532	7										
80 Airport Blvd, SG 819642	Baseline	54	380766	32	29	8396	13	25	0.537	0.690	0.604	0.996	0.004	0.002
	MOEA	42	1949	8										
601 Island Club Rd, SG 578775	Baseline	39	269389	31	19	8415	9	20	0.487	0.679	0.567	0.997	0.003	0.001
	MOEA	28	1645	6										
3 Jln Mata Ayer, SG 759150	Baseline	54	340789	31	28	8398	11	26	0.519	0.718	0.602	0.996	0.004	0.001
	MOEA	39	1525	6										
100-110 Lim Chu Kang Lane 3	Baseline	42	383782	40	16	8384	37	26	0.381	0.302	0.337	0.993	0.007	0.004
	MOEA	53	1942	8										
257 Jln Endut Senin, SG 508352	Baseline	56	382113	30	22	8385	22	34	0.393	0.500	0.440	0.993	0.007	0.003
	MOEA	44	1892	7										
Raffles Place Station	Baseline	44	305680	31	23	8400	19	21	0.523	0.548	0.535	0.995	0.005	0.002
	MOEA	42	1266	5										
70 Airport Boulevard SG 819661	Baseline	68	460910	32	34	8383	12	34	0.500	0.739	0.596	0.995	0.005	0.001
	MOEA	46	1843	7										
118 Rivervale Dr, SG 540118	Baseline	53	376906	31	21	8390	20	32	0.396	0.512	0.447	0.994	0.006	0.002
	MOEA	41	2087	8										
24 Neram Cres, SG 807829	Baseline	40	332929	31	24	8421	2	16	0.600	0.923	0.727	0.998	0.002	0.000
	MOEA	26	1820	7										
11 Woodlands Street 83, SG 738489	Baseline	59	424830	31	33	8395	9	26	0.559	0.786	0.653	0.996	0.004	0.001
	MOEA	42	1350	6										
31 Yishun Central, SG 768827	Baseline	45	307482	31	22	8412	6	23	0.489	0.786	0.603	0.997	0.003	0.001
	MOEA	28	1302	5										
671A Choa Chu Kang Cres, SG 681671	Baseline	48	371866	31	21	8403	12	27	0.438	0.636	0.519	0.995	0.005	0.001
	MOEA	33	1677	7										
202 Ang Mo Kio Ave 3, SG 560202	Baseline	40	289339	31	20	8410	13	20	0.500	0.606	0.548	0.996	0.004	0.002
	MOEA	33	1767	8										
296 Lor Ah Soo, SG 536742	Baseline	52	349636	30	18	8392	19	34	0.346	0.486	0.404	0.994	0.006	0.002
	MOEA	37	2015	8										
75 Marine Dr, SG 440075	Baseline	45	292176	31	21	8409	9	24	0.467	0.700	0.560	0.996	0.004	0.001
	MOEA	30	1606	7										
Bef Telok Blangah Hts	Baseline	31	271953	30	13	8427	5	18	0.419	0.722	0.531	0.997	0.003	0.001
	MOEA	18	1422	5										

Table 10: Performance Assessment of Multi-Objective Optimization using Confusing Metrix

7.1.3. Improvement in Performance Assessment

Based on the 7.1.2 *Initial Performance Assessment* conducted previously, various improvements were made to assist in the optimization tasks. One of the improvements was the addition of the weightage on the computation of the living facilities objective function. Associated weights were pre-processed based on the frequency distribution of the real estate property data set, as shown in Figure 34 and Table 6 provides the value of weight for each living facilities.

Moreover, decoding the candidate individual into the actual solution was revised (i.e., the selection of the nearest solution point in **Algorithm 5**). In the selection of the actual solution point which is the nearest to the candidate solution produced by the NSGA-II algorithm, the computation of distance was improved by the addition of constant weights to each decision variable and the dominance comparison check for the solutions which have the same distance. **Algorithm 6** provides the procedure to find the nearest actual solution to the candidate solution.

Performance assessment was conducted on the same 20 data points, and the results of the Confusion Matrix were recorded, as shown in Table 11. It can be observed that the average precision value is improved from 0.65 to 0.71. The most significant improvement is the search space of the optimization. It is found that the multi-objective optimization algorithm evaluates less than 1,000 solutions, which reduces around 50% compared to the previous performance assessment. Moreover, the search time is observed to be less than 5 seconds, which achieves faster processing time.

Location Points	Search Method	# of Pareto Optimal Solutions	# of Comparisons / Search Space	Processing Time (seconds)	Confusion Matrix									
					TP	TN	FP	FN	Recall	Precision	F-Score	Accuracy	Misclassification Rate	False Positive Rate
Nanyang Child Care Centre	Baseline	58	540243	37	34	8396	9	24	0.586	0.791	0.673	0.996	0.004	0.001
	MOEA	43	867	5										
11 Kent Ridge Rd, SG 119220	Baseline	54	349991	33	21	8398	11	33	0.389	0.656	0.488	0.995	0.005	0.001
	MOEA	32	834	5										
Mount Faber Rd, SG 099205	Baseline	35	263049	33	18	8425	3	17	0.514	0.857	0.643	0.998	0.002	0.000
	MOEA	21	759	4										
4 Lor M Telok Kurau, SG 425283	Baseline	50	328955	32	26	8411	2	24	0.520	0.929	0.667	0.997	0.003	0.000
	MOEA	28	874	5										
80 Airport Blvd, SG 819642	Baseline	64	443354	33	32	8389	10	32	0.500	0.762	0.604	0.995	0.005	0.001
	MOEA	42	908	5										
601 Island Club Rd, SG 578775	Baseline	46	320174	32	25	8412	5	21	0.543	0.833	0.658	0.997	0.003	0.001
	MOEA	30	905	5										
3 Jln Mata Ayer, SG 759150	Baseline	61	377487	32	27	8393	9	34	0.443	0.750	0.557	0.995	0.005	0.001
	MOEA	36	797	5										
100-110 Lim Chu Kang Lane 3	Baseline	63	471196	32	33	8389	11	30	0.524	0.750	0.617	0.995	0.005	0.001
	MOEA	44	882	5										
257 Jln Endut Senin, SG 508352	Baseline	53	395559	34	21	8394	16	32	0.396	0.568	0.467	0.994	0.006	0.002
	MOEA	37	841	5										
Raffles Place Station	Baseline	54	357193	32	25	8404	5	29	0.463	0.833	0.595	0.996	0.004	0.001
	MOEA	30	629	4										
70 Airport Boulevard SG 819661	Baseline	74	480148	32	28	8363	26	46	0.378	0.519	0.438	0.991	0.009	0.003
	MOEA	54	963	5										
118 Rivervale Dr, SG 540118	Baseline	55	390619	31	22	8395	13	33	0.400	0.629	0.489	0.995	0.005	0.002
	MOEA	35	952	5										
24 Neram Cres, SG 807829	Baseline	43	357979	30	18	8408	12	25	0.419	0.600	0.493	0.996	0.004	0.001
	MOEA	30	838	4										
11 Woodlands Street 83, SG 738489	Baseline	66	500930	31	32	8382	15	34	0.485	0.681	0.566	0.994	0.006	0.002
	MOEA	47	817	5										
31 Yishun Central, SG 768827	Baseline	50	348497	31	23	8397	16	27	0.460	0.590	0.517	0.995	0.005	0.002
	MOEA	39	810	4										
671A Choa Chu Kang Cres, SG 681671	Baseline	63	451631	31	34	8390	10	29	0.540	0.773	0.636	0.995	0.005	0.001
	MOEA	44	850	5										
202 Ang Mo Kio Ave 3, SG 560202	Baseline	43	311555	31	20	8413	7	23	0.465	0.741	0.571	0.996	0.004	0.001
	MOEA	27	936	5										
296 Lor Ah Soo, SG 536742	Baseline	62	389734	30	21	8386	15	41	0.339	0.583	0.429	0.993	0.007	0.002
	MOEA	36	1008	5										
75 Marine Dr, SG 440075	Baseline	58	366947	30	29	8396	9	29	0.500	0.763	0.604	0.996	0.004	0.001
	MOEA	38	826	5										
Bef Telok Blangah Hts	Baseline	35	281198	30	15	8418	10	20	0.429	0.722	0.531	0.996	0.004	0.001
	MOEA	25	858	4										

Table 11: Improved Performance Assessment of Multi-Objective Optimization using Confusing Metrix

7.2. Web-based Property Listing and Search Demonstration

System testing of the online web-based property listing and search platform was prepared to observe the real-time performance of the multi-objective optimization-based search. Different test cases were prepared to represent various real-world case scenarios which are commonly occurred during the real estate property search.

7.2.1. Local Environment Setup

Web-based property listing and search platform was developed in IntelliJ IDEA, and it is running in the local environment with the use of the Integrated HTTP Server (Netty) which is supported by Play Framework. The local environment settings for the performance assessment can be referred to Table 8. A web browser which was used to run the web-based property listing and search platform is provided in Table 12.

Web Browser Setting for Performance Assessment

Web Browser	Mozilla Firefox
Version	Firefox Quantum 67.0.4 (64-bit)
Network Connection	Local Area Connection (LAN)
Domain Network	main.ntu.edu.sg

Table 12: Web Browser Setting for Performance Assessment of Web-based Property Listing and Search Platform

7.2.2. Test Cases

Test cases were designed to comply with real-world case scenarios. There were three major groups into which all test cases are categorized.

- 1) Bi-objective based test case (*price* and *living facilities*)
- 2) Multi-objective based test case (*price*, *living facilities*, and *distance/duration*)
- 3) Multi-objective based test case with the user's preference (*price*, *living facilities* and *distance/duration* with *constraints*)

Case Scenario 1: Bi-Objective Based Test Case

Case Study: A customer wanted a long-term stay in Singapore, and he/she does not know any information about the place to stay. He/she wanted a house with a low rental price and sufficient living facilities included.

Control Panel: Search for the price and living facilities.

The control panel includes the following sections:

- Price:** A slider ranging from S\$500 to S\$2,000, with a S\$10,000 marker.
- Facility Filter:** A dropdown menu labeled "Facilities".
- Travel Distance:** A slider ranging from 0 to 10,000m, with a 100m marker.
- Travel Time:** A slider ranging from 0 to 120min, with a 10min marker.
- Search Buttons:** A row of buttons including "Exhaustive Search", "Optimizer with Travel Distance" (with a red circle around the house icon), and "Optimizer with Travel Time".
- Ranking:** A row of buttons: "Default" (highlighted), "Price", "Living", and "Distance".

Figure 69: Search for Price and Living Facilities Operation Button in Control Panel

Number of Property Listings found: 11

Rank by: the price

Figure 70: Result of Bi-Objective Based Test Case with Price Ranking on Map

Rank by: the living facilities

Figure 71: Result of Bi-Objective Based Test Case with Living Facilities Ranking on Map

Figure 72: Good HDB Flat recommended for Bi-Objective Based Test Case

Result Analysis: Among the ranking of 11 property listings based on the price and living facilities, it is discovered that a 3-Room HDB flat in 13 Telok Blangah Cres, Block 13, Singapore 090013, is a good option for the customer since it is in the lowest price group and the highest living facilities group. The criteria rank indicator will be defined for this HDB flat.

Case Scenario 2: Multi-Objective Based Test Case

Case Study: A customer wanted a long-term stay in Singapore, and his/her workplace is near Raffles Place Station. He/she wanted a house with a low rental price, which is nearby the station. Sufficient living facilities included would be favorable.

Control Panel and Map Viewer: Select the location on the map (at Raffles Place Station) and search for the price, living facilities, and distance.

Figure 73: Search for Price, Living Facilities and Distance Operation Button in Control Panel

Number of Property Listings found: 43

Rank by: the price

Figure 74: Result of Multi-Objective Based Test Case with Price Ranking on Map

Rank by: the living facilities

Figure 75: Result of Multi-Objective Based Test Case with Living Facilities Ranking on Map

Rank by: the distance to the specified location

Figure 76: Result of Multi-Objective Based Test Case with Distance Ranking on Map

Result Analysis: Among the ranking of 43 property listings based on three criteria, it is discovered that there are a few good options for the customer. Detailed analysis can be done using the table of Best-Known Property Listings, which can efficiently perform the ranking adjustment. Based on the result analysis, for the customer who prefers to stay near his/her workplace, two 3-Room HDB flats in 4 Sago Ln, Singapore 050004, with the criteria indicator:

, are the good options due to the nearest group to the Raffles Place station , the

highest living facilities group 📍, and the medium price group 📍. For the customer who prefers to find the lower priced HDB flats, a 3-Room HDB flat in 16 Taman Ho Swee, Block 16, Singapore 163016, with the criteria indicator: 📍📍📍, is a good choice since it is nearby Tiong Bahru station which is three stations away from the Raffles Place station, the highest living facilities group, and the medium price group.

Best Known Property Listings

Show 10 entries

Search:

PropertyID	Address	Price (SGD)	Total Facilities	Facilities List	Dist (m) / Dur (s)
4915	532 Upper Cross St, Singapore 050532	2600.0	7	[aircon, bed, fridge, stove, tv, washer]	602.82
853	538 Upper Cross St, Singapore 050538	2450.0	4	[aircon, fridge, washer]	654.85
8046	536 Upper Cross St, Singapore 050536	2700.0	0	[]	666.19
2289	533 Upper Cross St, Singapore 050533	2300.0	16	[aircon, bed, closet, dining room furniture, dryer, fridge, oven, sofa, stove, tv, walk in closet, washer, corner unit, high floor, renovated]	679.01
6472	533 Upper Cross St, Singapore 050533	2600.0	0	[]	679.01
590	4 Sago Ln, Singapore 050004	2200.0	9	[aircon, bed, closet, dining room furniture, fridge, sofa, stove, washer]	972.21
913	4 Sago Ln, Singapore 050004	2350.0	11	[aircon, bed, dining room furniture, fridge, oven, sofa, stove, tv, washer, high floor]	972.21
3620	4 Sago Ln, Singapore 050004	2400.0	1	[]	972.21
2631	335B Smith St, Singapore 052335	3200.0	2	[aircon]	972.46
3447	32 New Market Rd, Singapore 050032	1900.0	7	[aircon, bed, fridge, stove, tv, washer]	998.28

Showing 1 to 10 of 43 entries

Previous

1

2

3

4

5

Next

Figure 77: Good Options for the Customers who prioritize the Location

Best Known Property Listings

Show 10 entries

Search:

PropertyID	Address	Price (SGD)	Total Facilities	Facilities List	Dist (m) / Dur (s)
6275	225 Lor 8 Toa Payoh, Block 225, Singapore 310225	1600.0	2	[aircon]	6179.71
2224	12 Telok Blangah Cres, Block 12, Singapore 090012	1600.0	1	[]	3623.80
1324	49 Circuit Rd, Block 49, Singapore 370049	1650.0	2	[aircon, renovated]	5950.60
6869	28 Tanglin Halt Rd, Block 28, Singapore 140028	1650.0	2	[aircon]	5960.71
6151	83 Lor 2 Toa Payoh, Block 83, Singapore 310083	1680.0	4	[aircon, fridge, washer]	5771.46
2792	85 Circuit Rd, Singapore 370085	1700.0	9	[aircon, bed, closet, dining room furniture, fridge, sofa, stove, washer]	5745.06
1644	271 Queen St, Singapore 180271	1700.0	6	[aircon, bed, dining room furniture, fridge, tv, washer]	1907.98
1780	1 Beach Rd, Singapore 189673	1700.0	0	[]	1205.21
4150	16 Taman Ho Swee, Block 16, Singapore 163016	1800.0	13	[aircon, bed, closet, dining room furniture, fridge, sofa, stove, tv, washer, corner unit, high floor]	2254.38
2767	10 Jln Kukoh, Singapore 162010	1800.0	6	[aircon, bed, fridge, tv, washer]	1442.37

Showing 11 to 20 of 43 entries

Previous

1

2

3

4

5

Next

Figure 78: A Good Option for the Customers who are Price conscious

Case Scenario with Duration Objective Function

The same case scenario was conducted one more time with the duration criteria instead of the distance criteria.

Control Panel and Map Viewer: Select the *location* on the map (at Raffles Place Station) and search for the *price, living facilities, and duration*.

Figure 79: Search for Price, Living Facilities and Duration Operation Button in Control Panel

Number of Property Listings found: 35

Rank by: the duration to the specified location

Figure 80: Result of Multi-Objective Based Test Case with Duration Ranking on Map

Result Analysis: It is observed that with the duration criteria, different property listings are recommended by the web-based property listing and search platform. It is due to the consideration of the real-time duration between the house and the specified location. Among the ranking of 35 property listings based on three criteria, it is discovered that there are a few good options found for the customer. Based on the detailed analysis, for the customer who prefers to stay near his/her workplace, a 3-Room HDB flat in 32 New Market Rd, Singapore 050032, and a 2-Room HDB flat in 10 Jln Kukoh, Singapore 162010, with the criteria indicator:

, are the good options due to the nearest group to the Raffles Place station, the medium living facilities group, and the medium price group. For the customer who prefers to find the lower priced HDB flats, a 3-Room HDB flat in 13 Telok Blangah Cres, Block 13, Singapore 090013, with the criteria indicator: , is a good choice since it is in the lowest price group, the highest living facilities group, and it is nearby Tiong Bahru station which is three stations away from Raffles Place station.

Best Known Property Listings

Show 10 entries

Search:

PropertyID	Address	Price (SGD)	Total Facilities	Facilities List	Dist (m) / Dur (s)
21	34 Upper Cross St, Block 34, Singapore 050034	2500.0	12	[aircon, bed, closet, dining room furniture, fridge, sofa, stove, tv, washer, high floor, utility room]	225.00
1006	34 Upper Cross St, Block 34, Singapore 050034	3000.0	3	[aircon, stove]	232.00
6898	34 Upper Cross St, Block 34, Singapore 050034	1900.0	0	[]	232.00
853	538 Upper Cross St, Singapore 050538	2450.0	4	[aircon, fridge, washer]	251.00
2289	533 Upper Cross St, Singapore 050533	2300.0	16	[aircon, bed, closet, dining room furniture, dryer, fridge, oven, sofa, stove, tv, walk in closet, washer, corner unit, high floor, renovated]	265.00
3447	32 New Market Rd, Singapore 050032	1900.0	7	[aircon, bed, fridge, stove, tv, washer]	277.00
2767	10 Jln Kukoh, Singapore 162010	1800.0	6	[aircon, bed, fridge, tv, washer]	287.00
143	8 Jln Kukoh, Singapore 162008	2300.0	12	[aircon, bed, closet, dining room furniture, fridge, sofa, stove, tv, washer, renovated]	367.00
7010	140 Jalan Bukit Merah, Singapore 160140	3000.0	6	[aircon, bed, fridge, stove, washer]	411.00
2488	87 Zion Rd, Singapore 160087	3100.0	14	[aircon, bed, closet, dining room furniture, dryer, fridge, sofa, stove, tv, washer, corner unit, low floor, renovated]	462.00

Showing 1 to 10 of 35 entries

Previous 1 2 3 4 Next

Figure 81: Good Options for the Customers who prioritize the Location nearby Workplace

Best Known Property Listings

Show 10 entries

Search:

PropertyID	Address	Price (SGD)	Total Facilities	Facilities List	Dist (m) / Dur (s)
1718	692A Choa Chu Kang Cres, Block 692A, Singapore 681692	600.0	0		2558.00
2242	20 Eunos Cres, Block 20, Singapore 400020	700.0	0		916.00
1829	59 Commonwealth Dr, Block 59, Singapore 140059	1400.0	0		902.00
2238	413 Bukit Batok West Ave 4, Singapore 650413	1400.0	3	[aircon, fridge, washer]	1828.00
7383	206 Marsiling Dr, Singapore 730206	1400.0	9	[aircon, bed, dining room furniture, fridge, sofa, tv, washer, renovated]	2086.00
3830	28 Tanglin Halt Rd, Block 28, Singapore 140028	1500.0	0		790.00
1845	13 Telok Blangah Cres, Block 13, Singapore 090013	1500.0	18	[aircon, bathtub, bed, closet, dining room furniture, dryer, fridge, oven, sofa, stove, tv, walk in closet, washer, corner unit, high floor, renovated]	809.00
3673	34 Tanglin Halt Rd, Block 34, Singapore 142034	1500.0	2	[aircon]	894.00
3609	12 Telok Blangah Cres, Block 12, Singapore 090012	1500.0	8	[aircon, bed, closet, dining room furniture, fridge, sofa, tv]	959.00
449	43 Bendemeer Rd, Singapore	1600.0	4	[aircon, high floor, renovated]	771.00

Showing 1 to 10 of 35 entries

Previous
1
2
3
4
Next

Figure 82: A Good Option for the Customers who prefers Lower Price

Case Scenario 3: Multi-Objective Based Test Case with the User's Preference

Case Study: A customer wants a long-term stay in Singapore, and his workplace is at Nanyang Technological University, and his child attends to Yuan Ching Secondary School, which is near Lakeside station. Moreover, he owns a car which can fetch his child to the school before he goes to his workplace. He wanted a house with a low rental price, with the budget less than S\$2,500. Sufficient living facilities included would be favorable; however, he preferred a house that has aircon, bed, dining room furniture, fridge, sofa, and TV for the living convenience.

Control Panel and Travel Scheduler: Use the travel scheduler for the two location inputs (Yuan Ching Secondary School and Nanyang Technological University), set the price range (between S\$500 and S\$2,500), specify the preferred living facilities in the facility filter (Aircon, Bed, Dining Room Furniture, Fridge, Sofa and TV), set the distance meter (less than 1,500m), and search for the *price, living facilities* and *distance* with the user's preference.

In the current web-based property listings and search platform, the distance meter was set as the constraint for the *first* specified location point on the map.

Figure 83: Setting of Price Range and Location Distance Range, and Search for Price, Living Facilities and Distance Operation Button in Control Panel

Figure 84: Setting of Facilities in Control Panel and Location Points in Travel Scheduler

Number of Property Listings found: 14

Figure 85: Result of Property Listings based on the Case Study

Rank by: the price

Figure 86: Result of Multi-Objective Based Test Case with User's Preference in Price Ranking on Map

Rank by: the living facilities

Figure 87: Result of Multi-Objective Based Test Case with User's Preference in Living Facilities Ranking on Map

Rank by: the distance to the specified locations

Figure 88: Result of Multi-Objective Based Test Case with User's Preference in Distance Ranking on Map

Result Analysis: Among the ranking of 14 property listings based on three criteria and the user's preferences, it is discovered that the property listing and search system recommends a few good options near the first location which is Yuan Ching Secondary School for the customer. Based on the result analysis, a 3-Room HDB flat in 480 Jurong West Street 41, Block 480, Singapore 640480, with the criteria indicator: would be a good option for the customer. It is in the lowest price group and within the specified budget (i.e., S\$2,500), and the highest living facilities group and provides all preferred facilities (i.e., aircon, bed, dining room furniture, fridge, sofa, and tv). Moreover, it is in the medium distance range group to Yuan Ching Secondary School (i.e., an estimated distance of 650 meters).

Figure 89: Good HDB Flat recommended for Multi-Objective Based Test Case with User's Preference

PropertyID	Address	Price (SGD)	Total Facilities	Facilities List	Dist (m) / Dur (s)
6650	516 Jurong West Street 52, Singapore 640516	1500.0	0	[]	402.72
2791	524 Jurong West Street 52, Block 524, Singapore 640524	1700.0	1	[]	722.85
2402	480 Jurong West Street 41, Block 480, Singapore 640480	1800.0	10	[aircon, bed, closet, dining room furniture, fridge, sofa, stove, tv, washer]	655.96
446	346 Kang Ching Rd, Block 346, Singapore 610346	1900.0	4	[aircon, low floor, renovated]	214.13
1181	176 Boon Lay Dr, Block 176, Singapore 640176	1900.0	9	[aircon, bed, closet, dining room furniture, fridge, sofa, stove, washer]	1374.30
1020	554 Jurong West Street 42, Block 554, Singapore 640554	2000.0	10	[aircon, bed, closet, dining room furniture, fridge, sofa, stove, tv, washer]	1311.58
462	335 Kang Ching Rd, Block 335, Singapore 610335	2200.0	9	[aircon, bed, dining room furniture, fridge, sofa, stove, tv, washer]	393.85
480	196B Boon Lay Dr, Singapore	2200.0	11	[aircon, bed, closet, dining room furniture, fridge, sofa, stove, washer, high floor, renovated]	865.36
451	337B Tah Ching Rd, Singapore 612337	2400.0	4	[aircon, high floor, renovated]	513.24
2358	335 Kang Ching Rd, Block 335, Singapore 610335	2400.0	10	[aircon, bed, closet, dining room furniture, fridge, sofa, stove, tv, washer]	393.85
3462	488 Jurong West Ave 1, Block 488, Singapore 640488	2450.0	11	[aircon, bed, dining room furniture, fridge, oven, sofa, stove, tv, washer, corner unit]	1000.37
2447	337A Tah Ching Rd, Singapore 611337	2500.0	15	[aircon, bed, dining room furniture, fridge, oven, sofa, stove, tv, washer, bomb shelter, corner unit, high floor, renovated, utility room]	503.54
6595	353 Kang Ching Rd, Block 353, Singapore 610353	2500.0	9	[aircon, bed, closet, dining room furniture, fridge, sofa, stove, washer]	351.76
1894	345 Kang Ching Rd, Block 345, Singapore 610345	2750.0	11	[aircon, bed, dining room furniture, fridge, stove, walk in closet, washer, corner unit, high floor, utility room]	268.63

Figure 90: Result of Property Listings in the table of Best-Known Property Listings ranked by Price

Travel Scheduler: Add two location inputs and perform the routing to find the driving directions from 14 property listings.

Figure 91: Search of Driving Directions from Property Listings to the specified Locations in Travel Scheduler

Result Analysis: Driving directions from all 14 property listings to two specified locations are searched and listed according to the ascending order of the distance. Moreover, all driving directions are visualized on the map, as shown in Figure 92.

Figure 92: Visualization of Driving Directions from Property Listings to the specified Locations on the Map

Figure 93: Driving Direction from a selected Property Listing to the specified Locations

Figure 93 visualizes the driving direction from the house, which was previously selected as a good option for two specified locations. According to the real-time driving directions listed in Figure 91, it takes around 19.10 minutes to drive 8.05 kilometres from the house to Yuan Ching Secondary School, and then to Nanyang Technological University. Moreover, it is observed that the estimated spherical distance can be used for the search process due to its compatibility with the real-time distance computed online, i.e., the selected house is ranked 9th in both estimated spherical distance and the real-time distance among the property listings as shown in Figure 94 and Figure 91 respectively.

Best Known Property Listings						
<div> <div>Show 25 entries</div> <div> <div>Search:</div> <input type="text"/> </div> </div>						
PropertyID	Address	Price (SGD)	Total Facilities	Facilities List	Dist (m) / Dur (s)	
446	346 Kang Ching Rd, Block 346, Singapore 610346	1900.0	4	[aircon, low floor, renovated]	214.13	
1894	345 Kang Ching Rd, Block 345, Singapore 610345	2750.0	11	[aircon, bed, dining room furniture, fridge, stove, walk in closet, washer, corner unit, high floor, utility room]	268.63	
6595	353 Kang Ching Rd, Block 353, Singapore 610353	2500.0	9	[aircon, bed, closet, dining room furniture, fridge, sofa, stove, washer]	351.76	
462	335 Kang Ching Rd, Block 335, Singapore 610335	2200.0	9	[aircon, bed, dining room furniture, fridge, sofa, stove, tv, washer]	393.85	
2358	335 Kang Ching Rd, Block 335, Singapore 610335	2400.0	10	[aircon, bed, closet, dining room furniture, fridge, sofa, stove, tv, washer]	393.85	
6650	516 Jurong West Street 52, Singapore 640516	1600.0	0	[]	402.72	
2447	337A Tah Ching Rd, Singapore 611337	2500.0	15	[aircon, bed, dining room furniture, fridge, oven, sofa, stove, tv, washer, bomb shelter, corner unit, high floor, renovated, utility room]	603.54	
451	337B Tah Ching Rd, Singapore 612337	2400.0	4	[aircon, high floor, renovated]	513.24	
2402	480 Jurong West Street 41, Block 480, Singapore 640480	1800.0	10	[aircon, bed, closet, dining room furniture, fridge, sofa, stove, tv, washer]	655.96	
2791	524 Jurong West Street 52, Block 524, Singapore 640524	1700.0	1	[]	722.85	
480	196B Boon Lay Dr, Singapore	2200.0	11	[aircon, bed, closet, dining room furniture, fridge, sofa, stove, washer, high floor, renovated]	865.36	
3462	488 Jurong West Ave 1, Block 488, Singapore 640488	2450.0	11	[aircon, bed, dining room furniture, fridge, oven, sofa, stove, tv, washer, corner unit]	1000.37	
1020	554 Jurong West Street 42, Block 554, Singapore 640564	2000.0	10	[aircon, bed, closet, dining room furniture, fridge, sofa, stove, tv, washer]	1311.58	
1181	176 Boon Lay Dr, Block 176, Singapore 640176	1900.0	9	[aircon, bed, closet, dining room furniture, fridge, sofa, stove, washer]	1374.30	

Case Scenario with Duration Objective Function

The same case scenario was conducted one more time with the duration criteria instead of the distance criteria.

Control Panel and Map Viewer: Select the *location* on the map (at Raffles Place Station) and search for the *price*, *living facilities*, and *duration*.

Control Panel and Travel Scheduler: Use the travel scheduler for two location inputs (Yuan Ching Secondary School and Nanyang Technological University), set the price range (between S\$500 and S\$2,500), specify preferred living facilities in the facility filter (Aircon, Bed, Dining Room Furniture, Fridge, Sofa and TV), define the time setter (less than 10min), and search for the *price*, *living facilities* and *duration* with the user's *preference*.

In the current web-based property listing and search platform, the time setter was set as the constraint for the *first* specified location point on the map.

room furniture, fridge, sofa, and tv). Moreover, it is in the medium duration range group to Yuan Ching Secondary School (i.e., an estimated driving time of 4 minutes).

<div> Show 10 entries <div>Search:</div> </div>						
PropertyID	Address	Price (SGD)	Total Facilities	Facilities List	Dist (m) / Dur (s)	
1807	121 Yuan Ching Rd, Block 121, Singapore 610121	1588.0	0	[]	216.00	
2791	524 Jurong West Street 52, Block 524, Singapore 640524	1700.0	1	[]	475.00	
2399	326 Tah Ching Rd, Block 326, Singapore 610326	1800.0	10	[aircon, bed, closet, dining room furniture, fridge, sofa, stove, tv, washer]	232.00	
8136	475 Jurong West Street 41, Block 475, Singapore 640475	1800.0	3	[aircon, stove]	439.00	
446	346 Kang Ching Rd, Block 346, Singapore 610346	1900.0	4	[aircon, low floor, renovated]	162.00	
6167	339B Kang Ching Rd, Singapore 612339	2100.0	7	[aircon, bed, fridge, stove, tv, washer]	128.00	
451	337B Tah Ching Rd, Singapore 612337	2400.0	4	[aircon, high floor, renovated]	186.00	
2447	337A Tah Ching Rd, Singapore 611337	2500.0	15	[aircon, bed, dining room furniture, fridge, oven, sofa, stove, tv, washer, bomb shelter, corner unit, high floor, renovated, utility room]	167.00	
6676	333 Kang Ching Rd, Singapore 610333	3000.0	9	[aircon, bed, closet, dining room furniture, fridge, sofa, stove, washer]	241.00	
Showing 1 to 9 of 9 entries					<div> <div>Previous</div> <div>1</div> <div>Next</div> </div>	

Figure 98: Result of Property Listings in the table of Best-Known Property Listings ranked by Price

7.3. Summary

In this chapter, the experimental assessments were done to analyze the performance of the proposed property listing and search system. 20 geographic coordinate points were selected for the distance objective function evaluation. Both exhaustive search and multi-objective optimization-based search were run to observe the global Pareto optimal solutions and the best-known Pareto optimal solutions, respectively. Afterward, the comparison between the best-known Pareto optimal solutions and the global Pareto optimal solutions were conducted for the performance assessment of the optimization tasks with the Confusion Matrix.

Based on the performance results, it is found that the *accuracy*: 99% was achieved in a single run of the multi-objective optimization search. It is discovered that the optimization tasks achieved more than 0.65 *precision value* in most cases. However, the *recall* value was found to be low, around 0.46 on average. Based on this performance analysis, multi-objective optimization tasks were improved in various areas: adjustment of parameters setting of MOEA, initialization of the population, addition of weightage on the computation of objective function and decoding the candidate individuals into the actual solutions. In terms of the search space, the proposed search system evaluated around 2000 solutions instead of the total 8463 solutions which achieved the fast online processing time, less than 10 seconds, while the exhaustive search took around 30 seconds. It gives the optimization-based search superiority in the web-based online property listing and search system where the customers provide various inputs dynamically during the search.

After the improvements, it can be observed that the average *precision value* was improved from 0.65 to 0.71. The most significant improvement was the search space. It was found that the proposed search system evaluated less than 1,000 solutions, which reduced around 50% compared to the previous assessment. Moreover, the search time was observed to be less than 5 seconds, which achieved the faster processing time.

Furthermore, system testing of the online web-based property listing and search platform was conducted to observe the real-time performance of the multi-objective optimization-based search. Different test cases were prepared to represent various real-world case scenarios: 1) bi-objective based test case, 2) multi-objective based test case, and 3) multi-objective based test case with the user's preference. Detailed demonstrations were made to provide the step-by-step procedures on the property listing search and results analysis was done on each test case.

8. CONCLUSION

In this dissertation, a new kind of property search system was proposed and designed as a decision support system, which can be differentiated from existing property search methods. With an adoption of the multi-objective optimization techniques, an online web-based property listing and search system was designed to consider multiple criteria in the search with the minimum preference input from the customers and recommend the property listings which are the ideal possible options for the customers to make an intelligent decision in the property selection. Moreover, in order to achieve the goal of a convenient transition from the selection of a dream home to a successful business contract between the customer and house owner, a price negotiation model was cooperated in the decision support system to perform the appropriate price estimation of the real estate property.

The whole dissertation work was mainly organized into three types of data analytics: descriptive analytics which were used for understanding the data with various data visualization techniques, predictive analytics which were applied in the data to perform data cleansing and data transformation to achieve the knowledgeable discovery, and prescriptive analytics which explained the step by step procedures of the design and development of an online web-based property listing and search system. According to the performance assessment, it was discovered that the property listing and search system can perform a good recommendation of the property listings considering three multiple criteria in the search performance: 1) *minimizing* the price expense, 2) *maximizing* the facilities offered in the real estate property, and 3) *minimizing* the distance/duration it takes to go to the specified locations. Various real-world case scenarios were conducted in order to evaluate the performance of the online web-based property listing and search system in which it can perform the intelligent recommendation based on multiple criteria and suggest the property listings to the customers to make the criteria adjustments by themselves.

Moreover, this dissertation work encourages the research community to contribute or apply multi-objective optimization techniques and innovative technologies in various PropTech areas such as *Investment/Crowd Financing* (MOO problem: risk analysis and estimation), *Mortgage and Lending* (MOO problem: debt financing), *Agent Matching* (MOO

problem: agent finder) and *Property Management* (MOO problem: sustainable planning and development). With its nature of adaptability and generalization, multi-objective optimization problem models can be constructed based on various problem scenarios by defining 1) *decision variables*, 2) *constraints*, and 3) *objective functions*. The design and development framework of this dissertation work are believed to guide the potential researchers and developers in the development of multi-objective optimization model in various PropTech areas as the starting point of their optimization models.

9. FUTURE WORKS

The research works on this dissertation can be extended to various problem models. One of the research works will be the problem formulation with more than three objectives or criteria, which will be represented as a ***many-objective optimization problem***. This dissertation work can be the crucial stating point of research works focused on the optimization-based search methods applied in the online real estate property search in which the challenges of time and space complexity of the search performance are required to be tackled.

Furthermore, the advanced components such as efficient routing method with the use of ***vehicle routing problem*** can be adopted in the online web-based property listing and search platform in order to save the expensive cost consumption of the request calls to Google APIs. It can be applied in search of the property listings, which can accommodate many specified locations at once with low cost and high speed.

CHAPTER 9

REFERENCES

- [1] A. Baum, “PropTech 3.0: the future of real estate,” Saïd Business School, University of Oxford, 2017.
- [2] “Early-Stage Real Estate Tech: 120+ Companies Building The Industry’s Future,” CB Insights, 15 June 2017. [Online]. Available: <https://www.cbinsights.com/research/real-estate-tech-startup-market-map-early-stage/>.
- [3] “Home, Sweet Home: 96 Tech Startups Reshaping Residential Real Estate,” CB Insights, 4 May 2016. [Online]. Available: <https://www.cbinsights.com/research/residential-real-estate-tech-market-map-company-list/>.
- [4] M. Wong, “The rise of real estate tech,” CB Insights, 2018.
- [5] A. Couse, “The Growing Influence of PropTech,” JLL, 2018.
- [6] N. Wedlake and B. Crist, “Market Map, Part Two: 170+ Technology Companies Reshaping Commercial Real Estate,” Thomvest Ventures, 8 August 2018. [Online]. Available: <https://blog.thomvest.com/market-map-part-two-170-technology-companies-reshaping-commercial-real-estate-45e0a3ed5040>.
- [7] N. Wedlake and B. Crist, “Market Map: 140+ Real Estate Tech Companies Transforming the \$32 Trillion Housing Market,” Thomvest Ventures, 11 July 2018. [Online]. Available: <https://blog.thomvest.com/market-map-140-real-estate-technology-companies-transforming-the-32-trillion-housing-market-103ec015a54c>.
- [8] “Zillow,” [Online]. Available: <https://www.zillow.com/>.
- [9] “Zoopla,” [Online]. Available: <https://www.zoopla.co.uk/>.
- [10] “Disrupt Property,” [Online]. Available: <http://disruptproperty.com/>.
- [11] “PropertyGuru,” [Online]. Available: <https://www.propertyguru.com.sg/>.
- [12] “99.co,” [Online]. Available: <https://www.99.co/>.
- [13] “EdgeProp,” [Online]. Available: <https://www.edgeprop.sg/>.
- [14] “keylocation.sg,” [Online]. Available: <https://keylocation.sg/>.

- [15] “Google Scholar,” Google, [Online]. Available: <https://scholar.google.com.sg/>.
- [16] “IEEE Xplore Digital Library,” IEEE, [Online]. Available: <https://ieeexplore.ieee.org/Xplore/home.jsp>.
- [17] J. Branke, K. Deb, K. Miettinen and R. Slowinski, Multiobjective Optimization: Interactive and Evolutionary Approaches, Springer-Verlag Berlin Heidelberg, 2008.
- [18] C. A. C. Coello, G. B. Lamont and D. A. V. Van Veldhuizen, Evolutionary Algorithms for Solving Multi-Objective Problems, Springer Science+Business Media, LLC, 2007.
- [19] S. D. Sudhoff, “Lecture 9: Multi-Objective Optimization,” 2007. [Online]. Available: <https://engineering.purdue.edu/~sudhoff/ee630/Lecture09.pdf>.
- [20] A. E. Eiben and M. Schoenauer, “Evolutionary computing,” *Information Processing Letters*, vol. 82, no. 1, pp. 1-6, 2002.
- [21] R. Atencia, “Nature-inspired Computation: Timeline,” [Online]. Available: <https://www.timetoast.com/timelines/evolutionary-computation>.
- [22] A. E. Eiben and J. E. Smith, “Evolutionary Computing: The Origins,” in *Introduction to Evolutionary Computing*, Springer, 2015, pp. 13-24.
- [23] A. E. Eiben and J. E. Smith, “What Is an Evolutionary Algorithm?,” in *Introduction to Evolutionary Computing*, Springer, 2015, pp. 25-48.
- [24] N. Siddique and H. Adeli, Computational Intelligence: Synergies of Fuzzy Logic, Neural Networks and Evolutionary Computing, 2013.
- [25] E. Zitzler, K. Deb, L. Thiele, C. A. C. Coello and D. Corne, “A Short Tutorial on Evolutionary Multiobjective Optimization,” in *Evolutionary Multi-Criterion Optimization*, 2001.
- [26] A. Ghosh and S. Dehuri, “Evolutionary Algorithms for Multi-Criterion Optimization: A Survey,” *International Journal of Computing & Information Sciences*, vol. 2, no. 1, p. 20, 2004.
- [27] A. Konak, D. W. Coit and A. E. Smith, “Multi-objective optimization using genetic algorithms: A tutorial,” *Reliability Engineering & System Safety*, vol. 91, no. 9, p. 16, 2006.
- [28] K. Deb, A. Pratap, S. Agarwal and T. Meyarivan, “A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II,” *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182-197, 2002.

- [29] K. Deb, "Multi-Objective Optimization Using Evolutionary Algorithms: An Introduction," in *Multi-objective Evolutionary Optimisation for Product Design and Manufacturing*, Springer, 2011, pp. 3-34.
- [30] A. A, "First neural network for beginners explained (with code)," 13 January 2019. [Online]. Available: <https://towardsdatascience.com/first-neural-network-for-beginners-explained-with-code-4cfd37e06eaf>.
- [31] K. Jewmaidang and P. Tunchanok, "Interns Explain Basic Neural Network," 30 January 2018. [Online]. Available: <https://blog.datawow.io/interns-explain-basic-neural-network-ebc555708c9>.
- [32] F. v. Veen, "The Neural Network Zoo," 14 September 2016. [Online]. Available: <https://www.asimovinstitute.org/author/fjodorvanveen/>.
- [33] "Housing & Development Board," [Online]. Available: <https://www.hdb.gov.sg>.
- [34] "Rental Statistics," Housing & Development Board, [Online]. Available: <https://www.hdb.gov.sg/cs/infoweb/residential/renting-a-flat/renting-from-the-open-market/rental-statistics>.
- [35] "GADM maps and data," GADM, [Online]. Available: <https://gadm.org/>.
- [36] "Spatial Convenience Functions," MySQL, [Online]. Available: <https://dev.mysql.com/doc/refman/5.7/en/spatial-convenience-functions.html>.
- [37] "Google Maps Distance Matrix API," Google, [Online]. Available: <https://developers.google.com/maps/documentation/distance-matrix/>.
- [38] "MOEA Framework," [Online]. Available: <http://moeaframework.org/>.
- [39] B. Xue, "Flowchart of NSGAIL," [Online]. Available: https://ecs.victoria.ac.nz/foswiki/pub/Groups/ECRG/Talks/EMONSGAIL_and_SPEA2.pdf.
- [40] "Play Framework," [Online]. Available: <https://www.playframework.com>.
- [41] "MySQL," [Online]. Available: <https://www.mysql.com>.
- [42] "IntelliJ IDEA," JetBrains, [Online]. Available: <https://www.jetbrains.com/idea/>.
- [43] "Google Maps APIs," Google, [Online]. Available: <https://developers.google.com/maps/web-services/>.
- [44] "Distance Matrix API," Google, [Online]. Available: <https://developers.google.com/maps/documentation/distance-matrix/start>.

- [45] “Geocoding API,” Google, [Online]. Available:
<https://developers.google.com/maps/documentation/geocoding/start>.
- [46] “Maps JavaScript API,” Google, [Online]. Available:
<https://developers.google.com/maps/documentation/javascript/tutorial>.
- [47] “Places API,” Google, [Online]. Available: <https://developers.google.com/places/web-service/intro>.
- [48] “Directions API,” Google, [Online]. Available:
<https://developers.google.com/maps/documentation/directions/intro>.
- [49] “Geolocation API,” Google, [Online]. Available:
<https://developers.google.com/maps/documentation/geolocation/intro>.
- [50] “PyCharm,” [Online]. Available: <https://www.jetbrains.com/pycharm/>.
- [51] “Keras: The Python Deep Learning library,” [Online]. Available: <https://keras.io/>.
- [52] D. Hadka, “MOEA Framework - A Free and Open Source Java Framework for Multiobjective Optimization Version 2.12,” [Online]. Available:
<http://www.moeaframework.org/>.

APPENDIX

Author's Publications

1. Fuzzy Aggregated Topology Evolution

Iti Chaturvedi and Chit Lin Su

Journal: Cognitive Computation

Status: *under review*