REAL ESTATE VALUE FORECASTING SYSTEM USING DATA MINING AND NEURAL NETWORK APPROACH (REVFOS)

by

AZIZUL BIN ABDULLAH

Dissertation submitted in partial fulfillment of
the requirements for the
Bachelor of Technology (Hons.)
Business Information Systems

JANUARY 2008

Universiti Teknologi PETRONAS
Bandar Seri Iskandar
31750 Tronoh
Perak Darul Ridzuan

CERTIFICATION OF APPROVAL

REAL ESTATE VALUE FORECASTING SYSTEM USING DATA MINING AND NEURAL NETWORK APPROACH (REVFOS)

by

Azizul Abdullah

Final Report submitted in partial fulfillment of
the requirements for the
Bachelor of Technology (Hons.)
Business Information Systems

Approved by,

Ahmad Izuddin

Mr. Ahmad Izuddin bin Zainal Abidin
Supervisor

UNIVERSITI TEKNOLOGI PETRONAS

TRONOH, PERAK

JANUARY 2008

CERTIFICATION OF ORIGINALITY

This is to certify that I am responsible for the work submitted in this project that the original work is my own except as specified in the references and acknowledges and that the original work contained herein have not been undertaken or done unspecified sources or persons.

Azizul bin Abdullah

ABSTRACT

Modern science and engineering are based on applying first-principle models in order to describe physical, biological and social systems. It is an approach starts with a basic scientific model such as Newton's Law of Motion or Maxwell's equations in electromagnetism and it leads to building various applications in mechanical and electrical engineering. However, in many domains that underlying first principle is unknown and unable to elaborate or the systems developed are too complex to be mathematically formalized. Thus, there is currently a paradigm shift from classical modeling and analyses based on first principles to developing models and the corresponding analyses directly from data.

The necessitate to understand large, complex, information-rich data sets in widespread to virtually all fields of business, science and engineering as in the business world, corporate and customer data are treated as a strategic assets. The ability to extract useful knowledge hidden in these sets of data and to act on that knowledge is becoming increasingly important in today's competitive world. The entire process of applying computer-based methodologies including new techniques for discovering knowledge from raw data is known as data mining.

Data mining is the process of identifying and analyzing data from diverse perspectives and summarizing it into constructive and useful information which it can be utilized as revenue increments, costs reductions and input productions. Technically, data mining application or software is been treated as one of analytical tools for analyzing data and input gathered from various sources. Furthermore, it also allows users to scrutinize data from different dimensions and angels in order to categorize it before summarize the possible relationship identified. In addition, data mining is the process of identifying correlations or patterns among dozens of fields in large relational databases.

As from business perspectives, data mining is used by business practitioners with strong consumer focus such as retail, communication, financial and marketing. It is believed to assist these companies to determine relationship among internal factors for instance; price, product positioning or workforce skills and external factors like economic indicators, competition and consumer demographics. It enables them to decide on the impact on company's sales, customer satisfactions and corporate profits. It also facilitates them to drill down into concluded information in order to view transactional data.

Based on stated definition on data mining, the report performs as the detailed provisional report for Final Year Project Part I namely Real Estate Value Forecasting System using Data Mining and Neural Network Approach (REVFOS). The project will be developed in data mining environment and neural network (NN) will be treated as major platform in dealing with several crucial issues identified from the affected industries. Several models will be developed in order to test and train gathered data from the industry practitioner whereby all these data plays crucial impacts in determining precise possible results as final outcome.

Historical data will be gathered from the industrial practitioners mainly property agents or analyst on their manual prediction on property future value. To test the extent of identified datasets, neural network models will be constructed to predict accurate results on possible increase on value using only numbers of bedrooms, present value estimations, interest rates and locations as major inputs. All these data captured will be segregated into calibration, verification and test subsets after going through several important phases acquired in data-mining environment and model development. Furthermore, various mathematical calculations will be used in developing neural models in order to support any justifications made from the findings on the final predictive results.

As the result, all the predictions and models involved will be presented and demonstrated visually with some justifications in order to confirm accurate predictive results. The system is believed to provide precise value estimation on real estate price for its user based on the historical data used in the neural network models.

ACKNOWLEDGEMENT

First and foremost, I would like to take this opportunity to express my gratitude to The One and The Almighty, Allah SWT because of His permission; I am able to face all the challenges and difficulty during complete the project. Also, I would like to wish the deepest and sincere thanks to Mr. Ahmad Izuddin bin Zainal Abidin for his endless support and guidance in helping me to complete my project. His constructive comments, ideas and suggestions have guided me throughout this project.

Secondly, I would like to acknowledge Universiti Teknologi PETRONAS and CIS department; exclusively the final year project committees for proper organization and management on this course. I also would like to thank all the external parties such as real estate developers, property agents and bank officers in providing me all the information needed in order to accomplish the project's objectives.

Last but not least, all the endless gratitude to my family for their great morale and financial supports during conducting this project. Furthermore, to all my friends and colleagues who have given me strength, motivations and determination in all my undertakings.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS AND NOMENCLATURES

Acronym	Significance
ANN	Artificial Neural Network
DM	Data Mining
FYP	Final Year Project
IR	Interest Rates
NN	Neural Network
REVFOS	Real Estate Value Forecasting System
WWW	World Wide Web

CHAPTER 1

INTRODUCTION

1. Introduction

Data mining is an iterative process within which progress is defined by discovery through either automatic or manual processes. It also defined as a process of discovering various models, summaries and derived values from a given collection of datasets. Data mining is most useful in an exploratory analysis scenario in which there are no predetermined concepts about what will represent an interesting outcome. It is the search for new, valuable and nontrivial information in large volumes of data which is a cooperative effort of humans and machines.

In practice, the two primary goals of data mining (DM) tend to be prediction and description. Prediction involves using some variables or fields in the dataset to predict unknown or future values of other variables of interests. Description focused on finding patterns describing the dataset that can be interpreted by humans. The goals of prediction and description can be achieved after completing the primary DM tasks such as classification, regression, clustering, summarization, dependency modeling and change or deviation detection.

The successes of data mining engagement into certain projects depend fundamentally on the amount of energy, knowledge and creativity that the designer or developer puts into it and one of its greatest strengths which is wide range of methodologies and techniques plays the crucial parts in determining the triumph of managing host of problem sets.

The system will apply most of the techniques in data mining in order to achieve its goals to offer most reliable and accurate predictive results from historical data gathered from the industry for industrial practitioners and property buyers.

All the knowledge in data mining and its concept will help the application to accomplish its missions by providing further assistance to real estate buyer. Thus, the system is believed can be successfully implemented for massive usages.

1.1 Project Background

REVFOS is an abbreviation for Real Estate Value Forecasting System using Data Mining and Neural Network Approach whereby it will be developed using Artificial Neural Network (ANN) methodologies. It will be served as solutions for property buyers to predict their property possible value in future without consulting the real estate agents.

REVFOS is developed by creating network modules or diagram as shown below in order to test and train all possible data collected before the author decided on default weighted values and possible functions to be used throughout the system and its hidden layer or neural network (NN).

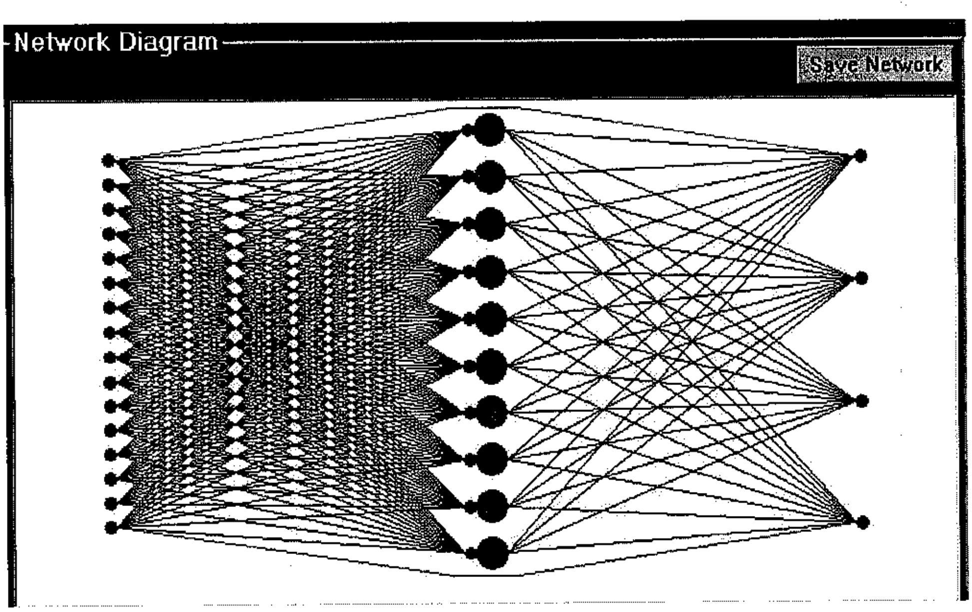


Figure 1.1: Example of Neural Network Diagram

All the predictions will be estimated by comparing historical data gathered from several property analysts from various locations nationwide. All these data will be tested using mathematical models created by understanding project's algorithms and the outcomes will be displayed in graphical representations.

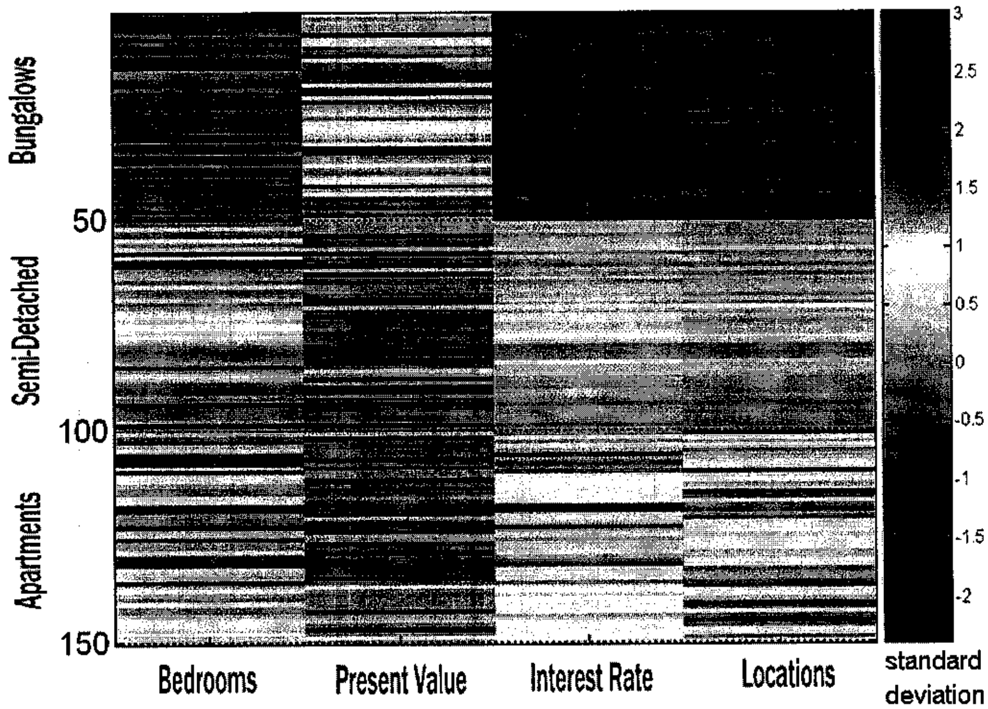


Figure 1.2 : Sample of Data Visualization of the System

1.2 Problem Statement

After several observations on current consumers' issues particular in property transaction, the author decided to develop REVFOS which can help not only real estate developers and property agents but also it will assist consumers in determining possible predictive values on respective properties. The selection of the topic as major focus in the project was based on several factors which are:

a) ***Manual calculation of property rate which can be lead to potential data redundancy***

The author was informed by industrial practitioners in real estate industries that the forecast values on certain land properties are been calculated manually by agents or banks based on figures captured on interest rates (IR) and other variables which have not been standardized throughout the country. Each agent may acquire diverse values based on dissimilar calculation methods. Therefore, REVFOS will serve as crucial solutions for industrial practitioners to avoid potential data redundancy on identical properties which share same decisive factors.

b) ***Inaccurate estimation derived from numerous crucial factors such as human errors, incorrect data gathered and erroneous value evaluation by property analysts***

The likelihood of inaccurate value estimations on real estate property to happen is moderately high due to several factors as mentioned previously. Thus, it will be an advantage to property analysts to acquire REVFOS that will provide them standardized value estimations nationwide since it will obtain data from major development places such as Klang Valley, Penang and Ipoh as its main preliminary target audiences.

c) ***Lack of awareness on real estate value estimation by real estate's proprietors***

It is believed that most of the land owner and property buyers are not fully aware with the potential market increase on their real estate possessions and they are solemnly dependent to property agents and bank land analysts to do the estimations on prospective analytical value of their property. Due to this attitude, it will straightforwardly increase the number of data redundancy on potential value estimation and the author do believe that REVFOS will not only provide better understanding on how to forecast property value but also avoid data redundancy from happening.

1.3 Objectives and Scope of Study

1.3.1. Objectives

The project is believed to achieve several objectives and goals based on potential implementation in the real estate industry such as:

a) ***To assist real estate analysts in measuring and forecasting the property's value more efficiently and effectively***

By applying REVFOS in the industry, the author do believe that it will be able to assist property analysts in computing and predicting real estate's value in effective methods. Furthermore, it will directly avoid result and data redundancy in determining forecast market rate on respective properties. REVFOS also will served as alternative approach for real estate analysts in calculation potential increase on price rate for preferred property without any clashes on possible values decided.

b) ***To create awareness to property developers and potential buyers on prospective value on the property based on respective criteria assigned***

REVFOS is capable in forming understanding and awareness to industry practitioners so that they will not be cheated or use inaccurate prediction for any properties since all the forecast values will be done based on specific characteristics and features.

1.3.2. Scope of Study

There are various knowledge and expertise required for further understanding before developing the system so that it will meet the functional and non-functional requirements as expected by the industrial practitioners and it fulfills the business needs of the organizations.

a) Identifying processes and procedures in data mining environment

It is crucial for the author to master the concept of developing the application in DM environment since REVFOS will solemnly developed by using processes and procedures in DM practices. Therefore, all the requirements on the system will be followed as proposed by DM experts and author's FYP supervisor.

b) Gathering potential data and additional information from the industry

As part of business requirements in predicting accurate potential property values, the author will gather all the data from the real estate industry and actual interest rate from banking sectors. All these raw data will be analyzed and classified into several clusters before the test and train phase take place. REFVOS will scrutinize these data before determining possible weighting value on selected nodes in DM models and the predictive result will be shown in graphical representations.

c) Selecting possible mechanisms and tools to be used

DM tools will be used in testing and training the data before the author continue in developing the application. Tools such as NeuroLab, NuMap & NuClass, Weka Projects will be applied as supportive software for REVFOS's development and as for system creation, the system will be developed on Java platform collaborated with Matlab for result representation.

d) Determining system requirements based on business needs

The purpose of the system to be developed is to be main alternative for the industry in predicting property future value. Thus, most of the requirements will be based on feedbacks gathered from the industrial practitioners such as property agents & buyers, consumers and mortgage personnel in banking sectors. Nevertheless, as for educational purposes in achieving FYP's objectives, REVFOS will be developed based on theories learned in class and platform used in developing group projects in respective courses.

CHAPTER 2
LITERATURE REVIEW AND THEORY

2.1. Literature Review

Mehmed Kantardzic (2003) claimed that data mining is one of the fastest growing fields in computer industry. Once a small interest area within computer science and statistics, it has quickly expanded into a field of its own. One of the greatest strengths of data mining is reflected in its wide range of methodologies and techniques that can be applied to a host of problem sets. In business community, DM can be used to discover new purchase trends, plan investment strategies and detect unauthorized expenditures in the accounting system. It can improve marketing campaign and the outcomes can be used to provide customers with more focused support and attention. DM techniques can be applied to problems of business process reengineering, in which the goal is to understand interactions and relationships among business practices and organizations. (p.3).

Witten and Frank (2000) are very definite: "Data mining is defined as the process of discovering patterns in data. The process must be automatic or (more usually) semi-automatic. The patterns discovered must be meaningful in that they lead to some advantage, usually an economic advantage. The data is invariably present in substantial quantities. DM is about solving problems by analyzing data already present in databases. Suppose, to take a well-worn example, the problem is fickle customer loyalty in a highly competitive marketplace. Behavior patterns of former customers can be analyzed to identify distinguishing characteristics of those likely to switch products and those likely to remain loyal. As the world grows in complexity, overwhelming us with the data it generates, DM becomes our only hope for elucidating the patterns that underlie it as intelligently analyzed data is a valuable resource which can lead to new insights and competitive advantages". (p.3).

Sushmita Mitra and Tinku Acharya (2003) stated that data mining tasks can be descriptive, (i.e. discovering interesting patterns or relationship describing the data), and predictive (i.e. predicting or classifying the behaviors of the model based on available data). In other words, it is an interdisciplinary field with a general goal of predicting outcomes and uncovering relationship in data. It uses automated tools that (a) employ sophisticated algorithms to discover mainly hidden patterns, associations, anomalies, and / or structure from large amount of data stored in data warehouses or other information repositories and (b) filter necessary information from this big dataset. (p.4)

Ian H. Witten and Eibe Frank (2005) said that data mining is a practical topic and involves learning in a practical, not a theoretical, sense. We are interested in techniques for finding and describing structural patterns in data as a tool for helping to explain that data and make predictions from it. The data will take the form of a set of examples - examples of customers who have switched loyalties, for instance, or situations in which certain kinds of contact lenses can be prescribed. The output takes the form of predictions about new examples – a prediction of whether a particular customer will switch or a prediction of what kind of lens will be prescribe under given circumstances. (p.7)

2.2. Theory

As one of major features in data mining, ANN has been treated as important technique in developing prediction models for forecasting accurate results. Hinton (1992) points out that ANN "is a mathematical structure designed to mimic the information processing functions of a network of neurons in the brain. ANN is particularly well suited for problems in which large datasets contain complicated nonlinear relations among many different inputs. ANN is able to find and identify complex patterns in datasets that may not be well described by a set of known processes or simple mathematical formulas."

In addition, "artificial neural networks, with their remarkable ability to derive meaning from complicated or imprecise data, can be used to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques. A trained neural network can be thought of as an "expert" in the category of information it has been given to analyze. This expert can then be used to provide projections given new situations of interest and answer "what if" questions." (Stergiou & Siganos, 2001, p.16).

CHAPTER 3
METHODOLOGY / PROJECT WORK

3.0. Methodology / Project Work

3.1. Data Mining Process

As stated before, REVFOS will fully developed using data mining as it main backbone or field of studies, therefore it is important to realize that the predicament of discovering or predicting dependencies from datasets or discovering totally new data is only one part of the general experimental procedure used by scientists, engineers and industry practitioners who apply standard steps in drawing conclusions from the raw data itself. It will follow several processes and procedures stated in DM references as guidelines for plan, design and analyze the proposed system. The general experimental procedure adapted to DM issues in developing REVFOS involves the following steps which are :

3.1.1. State the problem and formulate the hypothesis

Most data-based modeling studies are performed in a particular application domain. Domain-specific knowledge and experience are usually necessary in order to happen with a meaningful problem statement. A modeler usually specifies a set of variable for the unknown dependency and a general form of this dependency as an initial hypothesis and in practice; it generally means a close interaction between the DM expert and the application expert.

In REVFOS's perspective, real estate industry will be treated as core domain-specific model in order to solve any issues related to the industry. As for its problem statement, data redundancy and inaccurate value predictions are major drawbacks to be solved by the author in providing the proposed system. Hypothesis soon to be built up to support the findings from the data gathered within the industry.

3.1.2. Collect the data

This particular step is concerned with how the data are generated and collected as it will be divided into two distinct possibilities which are *designed experiment* and *observational approach.* Typically, the sampling distribution is completely unknown after the data are gathered or it is partially and implicitly given in the data-collection procedure since it has been treated as crucial factor for determining priori knowledge and final interpretation of the results.

Data will be gathered from several locations nationwide in order to acquire precise value and interest rate before determining ANN models to be used. All these crucial data will be collected from interviews and observations with property valuers such as Henry Butcher and real estate developers like YTL, Worldwide Holding, Platinum Victory and others.

3.1.3. Preprocessing the data

In the observational environment, data are generally collected from the existing databases, data warehouses and data marts. Data preprocessing includes at least two common tasks which area *outlier detection* and *scaling or encoding.*

Outlier detection and removal is a method of classifying unusual data values resulted from measurement errors, coding and recording errors and natural or abnormal values. There are strategies in dealing with outliers such as detect and remove outliers as a part of the preprocessing phase or develop robust modeling methods that are insensitive to outliers.

Scaling, encoding and selecting features are used in scaling data or variables and bring both features to the same weight for further analysis. Furthermore, application-specific encoding methods usually achieve dimensionality reduction by providing smaller number of informative attributes for subsequent data modeling.

Data congregated from these sources will be processed and analyzed using methods as stated above. Data with missing values will be either eliminated or filled with default value using outlier detection and removal in order to obtain meaningful data prediction. In the other hand, some of the data will be scaled and encoded to attain dimensionality reduction in data preprocessing.

3.1.4. Estimate the model

The selection and implementation of the appropriate DM technique is the main task on estimating the particular model as implementation will be based on several models and selecting the best model is an additional goal to be achieved.

As for REVFOS, neural models will be created based on the data gathered so that accurate predictive result can be generated. The model will obtain some mathematical calculations and statistic functions in order to provide accurate predictive results on property value. The sample of the model will be shown in Chapter 4 : Results and Discussion.

3.1.5. Interpret the model and draw conclusion

From the reported developed by experts, it is believed that DM models should help organizations in making decisions. Hence, such models need to be interpretable in order to be useful throughout the organization. The problem of interpreting these models is considered as separate task with specific techniques to validate the possible results.

Based on the model(s) created for REVFOS, several justifications supported with mathematical equations will be made to support the findings so that any reasonable doubts can be reduced and the goals of the system can be accomplished. The ANN model developed will be interpreted via graphical representations.

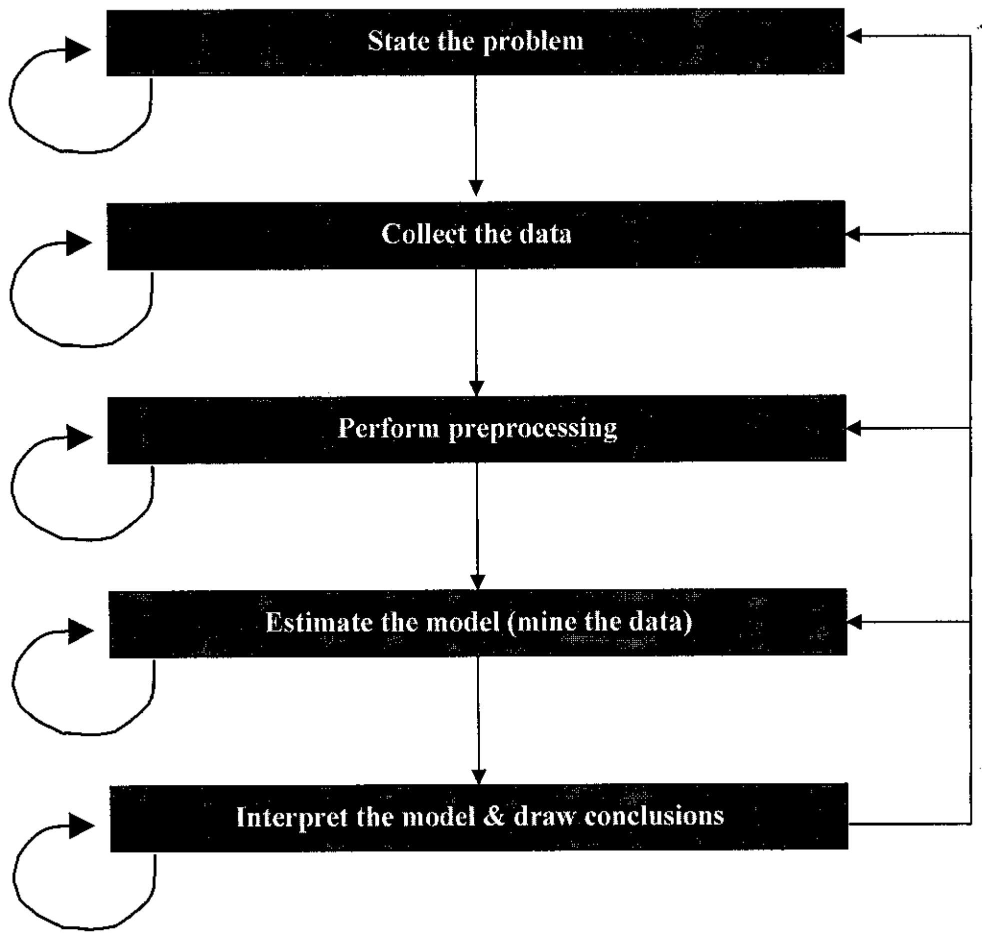


Figure 3.1 : The data-mining process

3.2 Data Preparation Process

When gathering the data samples from multiple sources such as databases or data warehouses, the author needs to categorize them into groups before further analysis to be conducted. Basically, the data will be divided into two common types which are numeric and categorical. In REVFOS' data preparation, numeric values include real-value variables or integer value such as present value, interest rate and number of bedrooms and it contains two crucial properties line an order relation and a distance relation.

In contrast, categorical or symbolic variables have neither of these properties because it only supports value either equal or not equal in its result interpretations; for instance of symbolic variables in REVFOS will be location of the property. Habitually in DM models, categorical variables can be converted to a numeric binary variable with two values: 0 and 1.

Most of the data-mining problems arise because there are large amounts of samples with different types of features. Moreover, these data samples are very often high dimensional which means they acquired extremely large number of measurable features and it causes the problem known as the curse of dimensionality. The properties of high dimensional spaces appear counterintuitive to DM practitioners and DM experts have developed guidelines for these issues in the interpretation of the input data. The guidelines provided are if a dataset yielding the same density of data points in an n-dimensional space increases exponentially with its dimensions, a larger radius is needed to enclose a fraction of the data points in a high dimensional space, almost every point is closer to an edge than to another sample point in a high-dimensional space and lastly, almost every point is a outlier.

3.2.1 Transformation of Raw Data

In the REVFOS project, all the data gathered will be analyzed and transformed into groups or clusters based on various obtainable techniques depending on types of data, amounts of data and general characteristics of the DM task. These possible schemes will help the author in obtaining problem-dependent data and improving DM results.

3.2.2 Normalization

Some methods in DM process may require normalized data for better and accurate outcomes especially techniques based on distance computation between points. It is because if the values are not normalized and standardized, the distance measures between two points of data gathered for REVFOS will overweight those features that obtained average or larger values.

a) Decimal scaling

This normalization method moves the decimal point - for instance present value and interest rate - but it still preserves most of the original digit value. The typical scale will be used in the data normalization is in range of -1 to 1 based on the following equation.

$$v'(i) = v(i) / 10^k$$

The above equation describes decimal scaling where $v(i)$ is the value of the feature v for the case i and v'(i) is a scaled value for the smallest k such that max $(|v'(i)|)$ < 1. For instance, if the largest value in the set for present value of the property is 455,000 and the smallest value is -38,000, therefore the maximum absolute value of the feature becomes 0.38 and the divisor for all $v(i)$ is 100,000 (since k=6).

b) Min-max normalization

This technique will be used for normalizing selective data in REVFOS particularly those data with interest rate as provided by banking industry on property loans. The automatic-computation of min and max values requires an additional search through the entire dataset but computationally, it will caused unintentional accumulation of normalized values based on below equation to be used :

$$v'(i) = (v(i) - min(v(i))) \,/\, (max(v(i)) - min\ (v(i)))$$

Suppose that the data for a feature v are in range between 85,000 and 1,500,000. In that case, the previous method of normalization will give all normalized data between .0085 and .15 but it will accumulate the value on a small subinterval of the entire range. In order for us to obtain better distribution of values on a whole normalized interval, min-max formula works best.

c) Standard deviation normalization

Normalization by standard deviation often works well with distance measures but it might be possibly transforms the data into an unrecognizable from original data. The value for this normalization will be computed using stated equation :

$$v'(i) = (v(i) - mean\ (v)) \,/\, sd\ (v)$$

For example, if the initial set of values of the attribute is $v = \{1, 2, 3\}$, then $mean\ (v) = 2$, $sd(v) = 1$ and the new set of normalized value is $v^* = \{-1, 0, 1\}$. This type of normalization can works best for interest rate but since the possibility for it to transform the data into an unrecognizable format, the author decided to disregard this technique for result accuracy issue.

Normalizations are very useful for several diverse methods of data mining especially for REVFOS which will be dealing with enormous number of raw data. If a method requires data to be normalized, available data will be transformed and prepared for the selected DM technique but an identical normalization must be applied in other phases with new and future data will applicable solution(s) in order to circumvent incorrect predictive results for the final outcome.

3.2.3 Data smoothing

The method used to standardize random variations of same underlying values in order to avoid insignificant values or performance degradation on methods and final results. For instance, data captured in F with its own real value is {0.93, 1.01, 1.014, 3.02, 2.99, 4.95, 5.11, 5.0}, then it can be smoothed to be $F_{smoothed}$ {1.0, 1.0, 1.0, 3.0, 3.0, 5.0, 5.0, 5.0}. As for the system, some of the data captured will be undergo this simple transformation so that it will not losing any quality in its dataset and it is believe to reduce the number of different real values for the feature to four variables.

There are DM experts quoted that some of these smoothing algorithms are more complex when they are used in reducing the number of distinct values for a feature which lead to reduction of the dimensionality of the data space at the same time. Nevertheless, they also believed that smoothers can be used in the DM systems in order to discretize continuous data or features into a set of features with binary true-false values.

3.2.4 Differences and ratios

It is believed that even small changes in data / features can produce significant improvement in DM performance especially when using two types of simple transformations such as differences and ratios for output features application. These transformations – sometimes – produce better results than the simple, initial goal of predicting an output as for many data-mining methods, a smaller number of alternatives will or can improve the efficiency of the algorithms and will be giving better outcomes.

As for REVFOS, ratios transformation will be used for some of the data gathered which it is a method of using $s(t+1) / s(t)$ as the output of the DM process instead of absolute value $s(t+1)$. It means that the level of increase or decrease in the value may also improve the REVFOS's performance of the entire data mining process for accurate results.

3.3 Missing Data

For many knowledge-based applications particularly in DM areas, although there are massive amounts of data, the subset of cases with complete data may be relatively small. It is believed that available samples and future cases may obtain missing values in its datasets. Some of the DM methods do accept missing values and satisfactorily process data in order to reach final conclusion but there are some methods require all values to be available.

Thus, it is very crucial for the author as system developer to treat this issue as major drawback when gathering all important data for REVFOS's development. It is true which some of the data obtained from the industry have missing values in them. As recommended by experts, the simplest solution for arise problem is the reduction of datasets or the elimination of the affected samples with missing values because if the author does not eradicate the samples with missing values, then the author has to locate default value for these missing value.

Firstly, a data miner – the author - need to manually examine data samples gathered which have no values and the author need to put reasonable or expected value based on domain experience of the system developed. Then, as a second approach, the author need to provide simpler solution for eliminating of missing values which is based on a formal and automatic replacement of missing values with some possible constants such as replace all missing values with single global constant (a selection of a global constant is highly application-dependent), replace a missing value with its feature mean and replace a missing value with its feature mean for given class which this approach is only possible for classification problems where samples area classified in advance.

Although these solutions area tempting, their main flaw that needs to be considered by applying these elucidations in REVFOS's development is the substituted value is not the exact or accurate value. By replacing the missing value with a constant or changing the value for a few different features, the data are biased. The replaced values will homogenize the cases with missing values into a uniform subset directed towards the class with most missing value or artificial class.

Based on the observations made and clarification from DM experts, one possible interpretation of missing values is they are "don't care" values. In other words, DM practitioner suppose that these missing values do not have any influence on the final result. For example, if one three-dimensional sample X is given $X = \{1, ?, 3\}$, where the second feature's value is missing, the process will generate five artificial classes or samples for the feature domain [0, 1, 2, 3, 4].

$X_1 = \{1, 0, 3\}$, $X_2 = \{1, 1, 3\}$, $X_3 = \{1, 2, 3\}$. $X_4 = \{1, 3, 3\}$ and $X_5 = \{1, 4, 3\}$

Finally, the data miner will generate a predictive model to predict each of the missing value. For instance, if three features A, B and C are given for each sample, then based on samples that have all three values as a training set, the author can generate a model of correlation between these features. Different techniques such as regression, Bayesian formalism, and clustering or decision-tree induction will be used in the system depending on data types. Once the author has a trained model, the author can present a new sample that has missing values and generate a "predictive" value.

In general, it is speculative and often misleading to replace missing values using simple and artificial schema of data preparation. It is best to generate multiple solutions of data mining with and without features that acquire missing values before analyze and interpret them.

3.4 Outlier Analysis

In large datasets, there are samples that do not comply with the general behavior of the data model. Such samples, which are significantly different or inconsistent with the remaining set of data, are called as outliers. Outliers can be caused by measurement error or result of inherent data variability. The value captured could be typographical error or it could be correct and represent real variability for the given attribute.

Many data-mining algorithms try to minimize the influence of outliers of the final model or to eliminate them in the preprocessing phase. Some of the data-mining applications are focused on outlier detection due to it is an essential result of data analysis. For instance in the real world, while detecting fraudulent credit card transactions in a bank, the outliers are typical examples that may indicate fraudulent activity and the entire data-mining process is concentrated on their detection.

Outlier detection and potential removal from a dataset can be described as a process of the selection of k out of n samples that are considerably dissimilar, exceptional or inconsistent with respect to the remaining data. Thus, the problem of defining outliers is nontrivial particularly in multidimensional samples. Data visualization methods that are useful in outlier detection for one to three dimensions are weaker in multidimensional samples or data because of lack of adequate visualization methodologies for these spaces.

The simple approach to outlier detection for one-dimensional samples is based on statistics. Assuming that the distribution of values is given, it is necessary to find basic statistical parameters such as mean value and variance. Based on these values and the expected number of outliers, it is possible to establish the threshold value as a function of variance but the main problem with this methodology is a priori assumption about data distribution. For example, if the given dataset represents the feature `Age` with twenty different value :

$$Age = \{3,56,23,39,156,52,41,22,9,28,139,31,51,20,-67,37,11,55,45,37\}$$

Then, the corresponding statistical parameters are :

Mean	=	39.9
Standard deviation	=	45.65

If we select the threshold value for normal distribution of data as

$$Threshold = Mean \pm 2 \times Standard\ deviation$$

Then, all the data that are out of range [-54.1, 131.2] will be treated as potential outliers.

Distance-based outlier detection is a second method that eliminates some of the limitations imposed by the statistical approach. The most important difference is that this method is applicable to multidimensional samples while statistical descriptors analyze only a single dimension or several dimensions separately. In other words, distance-based outliers' area those sample which do not have enough neighbors whereby neighbors are defined through the multidimensional distance between samples. The table of Euclidian distance, $d = [(x1 - x2)^2 + (y1 - y2)^2]^{1/2}$, for the set S is given in Table 3.1 and based on this table, the author can calculate a value for a parameter p with the given threshold distance (d = 3) for each sample.

	S_1	S_2	S_3	S_4	S_5	S_6	S_7
S_1		2.236	3.162	2.236	2.236	3.162	2.828
S_2			2.236	1.414	4.472	2.236	1.000
S_3				3.605	5.000	4.472	3.162
S_4					4.242	1.000	1.000
S_5						5.000	5.000
S_6							1.414

Table 3.1 : Table of distance for dataset S

3.5 Data Reduction

For large datasets, there is likelihood that an intermediate and additional step which is data reduction should be performed prior to applying the data-mining techniques. Whereas large datasets have the potential for better mining results, there is no guarantee that it will yield better knowledge than small datasets. More commonly, a general solution is deducted from a subset of available features or samples and it will remain the same even when the search space is enlarged.

The choice of data representation and selection, reduction or transformation of features is probably the most important issue that determines the quality of DM solutions. Besides influencing the nature of DM algorithms, it can determine whether the problem is solvable at all or how powerful the resulting model of DM is. Performing standard data-reduction operation operations such as deleting rows, columns or values as a preparation for data mining, the author needs to be acquainted with what the author gains and lose with the activities performed. The overall comparison involved the following parameters for analysis which are:

a) *Computing time*

It has been stated that simpler data – a result from data reduction process – can hopefully lead to a reduction in the time consumed for data mining. In most cases, DM practitioners cannot afford to spend too much time on the data-preprocessing phases including a reduction of data dimensions although the more time we spend in preparation, the better the predicted outcome.

b) *Predictive or descriptive accuracy*

This is a dominant measure for most DM models since it measures how well the data is summarized and generalized into the DM model. DM experts generally expect that by using only relevant features, a DM algorithm cannot only learn faster but also with higher accuracy. Irrelevant data may mislead a learning process and a final model, while redundant data may complicate the task of learning which cause unexpected DM results.

c) *Representation of DM models*

The simplicity of representation obtained with data reduction, often implies that a model can be better understood. The simplicity of the induced model and other results depends on its representation. The need for a balanced view between accuracy and simplicity is necessary and dimensionality reduction is one of the mechanisms for obtaining the balance.

Algorithms that perform all basic operations for data reduction are not simple especially when these algorithms are applied to large datasets. Recommended characteristics of data-reduction algorithms that may be guidelines for DM model designers of these techniques are measurable quality, recognizable quality, monotonicity, consistency, diminishing returns, interruptability and preemptablity.

In feature reduction process, the author should obtain the result in terms of less data so that the DM algorithm can learn faster, higher accuracy of DM process so that the model can generalize better from the data, simple results of DM process so that they are easier to understand and use; and fewer features so that in the next phase of data collection, a saving can be made by removing redundant or irrelevant features or samples of data.

3.5.1 Entropy Measure for Ranking Feature

A method for unsupervised feature selection or ranking based on entropy measure is a relatively simple technique but with large number of features, its complexity increases significantly. The approach is based on the observation that removing any irrelevant features from REVFOS's gathered data but it may not change the basic characteristics of the dataset. The basic idea is to remove as many features as possible but yet maintain the level of distinction between the samples in the dataset as if no features had been removed.

All the data gathered for testing and training in REVFOS will be analyzed using several entropy measure algorithms. For instance, the algorithm is based in a similarity measure S that is inverse proportion to the distance D between two n-dimensional samples of data. The distance measure D is small for close samples – close to 0 – and large for distinct pairs (close to 1). When the feature are numeric, the similarity measure S of two samples can be defined as

$$S_{ij} = e^{-\alpha D_{ij}}$$

Where D_{ij} is the distance between samples x_i and x_j and α is a parameter mathematically expressed as

$$\alpha = -(\ln 0.5) / D$$

If the features area not numeric, the similarity for nominal variables is measured directly using Hamming distance :

$$S_{ij} = \left(\sum_{k=1}^{n} |X_{ik} = X_{jk}|\right) / n$$

Where $|X_{ik} = X_{jk}|$ is 1 if $X_{ik} = X_{jk}$, and otherwise. The total number of variables is equal to *n*. For mixed data, we can discreatize numeric values and transform numeric features into nominal features before applying this similarity measure.

Entropy is a global measure and it is less for ordered configurations and higher for disorder configurations. The proposed technique compares the entropy measure for a given dataset before and after removal of feature as for dataset of N samples, the entropy measure is

$$E = - \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} (S_{ij} \times \log S_{ij} + (1 - S_{ij}) \times \log (1 - S_{ij})$$

where S_{ij} is the similarity between samples x_i and x_j. The measure is computed in each of the iterations as a basis for deciding the ranking of features. The steps of the algorithm area based on sequential backward ranking and they have been successfully tested on several real-world applications and it also will be implemented in REVFOS development in pseudocodes as stated below :

1. Start with the initial full set of features F.
2. For each feature f € F, remove one feature f from F and obtain a subset F_f. Find the difference between entropy fro F and entropy for all F_f. as in REVFOS environment, the author will compare the differences ($E_F - E_{F-F1}$), ($E_F - E_{F-F2}$) and ($E_F - E_{F-F3}$).
3. Let f_k be a feature such that the difference between entropy for F and entropy for F_{fk} is minimum.
4. Update the set of features F = F – { f_k}, where – is a difference operation on sets.
5. Repeat steps 2-4 until there is only one feature in F.

3.5.2 Feature Discretization : Chimerge Technique

ChiMerge is one automated discretization algorithm that analyzed the quality of multiple intervals for a given feature by using x^2 statistics. The algorithm determines similarities between distributions of data in two adjacent intervals based on output classification of samples. It the conclusion of the x^2 test is that the output class is independent of the feature's intervals, then the intervals should be merged; otherwise it indicates that the difference between intervals statistically significant.

Chimerge algorithm consists the three basic steps for discretization :

1. Sort the data for the given feature in ascending order.
2. Define initial intervals so that every value of the feature is in a separate interval.
3. Repeat until no x^2 of any two adjacent intervals is less than threshold value.

The x^2 test or contingency-table test is used in the methodology for determining the independence of two adjacent intervals. When the data are summarized in a contingency table, the x^2 test is given by the formula :

$$x^2 = \sum_{i=1}^{2} \sum_{j=1}^{k} (A_{ij} - E_{ij})^2 / E_{ij}$$

Where

k	=	number of classes
A_{ij}	=	the number of instances in the i-th interval, j-th class
E_{ij}	=	the expected frequency of A_{ij} which is computed as $(R_i . C_j) / N$
R_i	=	the number of instances in the i-th interval = ΣA_{ij} , j = 1, ... k,
C_j	=	the number of instances in the j-th class = ΣA_{ij} , i = 1,2 ..
N	=	the total number of instances = ΣR_i , I = 1,2.

CHAPTER 4
RESULTS AND DISCUSSIONS

4.0 Results and Discussions

After all data gathered have been analyzed and transformed using data-mining methods and techniques, a model will be developed in order to train and test all the data. The purpose of the model is to identify any possible error and determine weighted for each neuron for accurate prediction of results. All the neuron nodes in respective model will be interacted with each other in order to predict accurate and reliable value for potential result as requested by users. In order to test and train the data, several applications available from the WWW such as Tiberius, NNClass datasets and NeuroLab will be used so that the finding of the REVFOS predictive result can be accomplished.

As for the project, the entire datasets gathered will be tested using below models with unique criteria specified. The derived NN model is developed based on observations of data gathered from the industry and it will be used throughout the system widely. All clustered data will be saved in .dll format which compatible with DM applications before it will be analyzed using available learning tasks in data mining environment.

After the gathered data undergo several analytical phases, all the analyzed data will be compiled and transform in graphical representation so that the comparison on result accuracy with some mathematical justifications can be made. The predictive result generated from the REVFOS will be an excellent solution for property analysts in determining potential property value in the market, plus it will help consumers in deciding which property is worth to their money.

4.1 Project Updates

4.1.1 Alteration of Project Title

Prior to the responses received from internal examiners – Dr. Dominic and Mdm. Rohaiza – on irrelevant title for the project during FYP 1 Oral Presentation, the author had decided to change the project title from Real Estate Value Forecasting System (REFVOS) to Real Estate Value Forecasting System using Data Mining and Neural Network Approach. It will represent the area covered to develop the system and to ensure the users aware of the application's platform.

4.1.2 Data Gathering

As part of project requirements, data has been gathered from several industrial practitioners through interviews, sales pamphlets, and informal surveys. All data has been analyzed and categorized using data mining techniques as mentioned in previous reports. The finding of the data gathered is illustrated in Appendix A.

During the interviews with several real estate agents and valuers, they clarified that there are some key factors need to be considered in determining the weighted value of the nodes in the project's neural network architecture. The author has been advised to consider various locations in Klang Valley as benchmark in deciding possible increment of forecast value on respective property. The survey has been conducted to 100 respondents in order to monitor customers' preferences on locations and the result is displayed in table below.

Location	Predilections
City Center	0.78
Damansara – Penchala	0.88
Wangsa Maju – Maluri	0.61
Bukit Jalil – Seputeh	0.53
Bandar Tun Razak – Sg. Besi	0.49
Others	0.11

* predilections is measured from scale 0 to 1

Table 4.1 : Location branding for weighted value

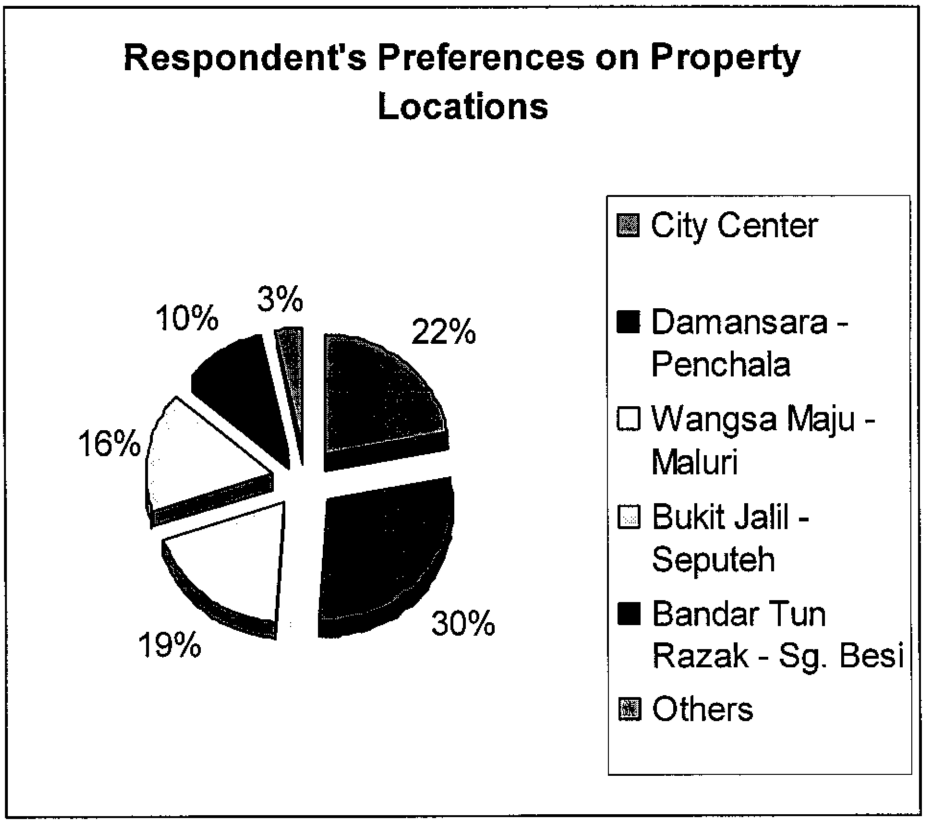


Figure 4.1 : Respondent's preferences on property locations

4.1.3 Data Testing and Training

After all the data has been gathered from various sources, it will be trained and tested in order to minimize errors so that accurate preditions can be generated. As for testing and training the data, the author decided to use NNClass and the findings are as stated below.

a) Neural Network Architecture Patterns

Network Architecture Options

Number of Inputs *(bewtween 2 and 50)*	**4**		
Number of Hidden Layers *(1 or 2)*	**2**	Hidden Layer sizes *(Maximum 20)*	**Hidden 1** **Hidden 2** 5 5
Learning parameter *(between 0 and 1)*	**0.5**	Initial Wt Range (0 +/- w): w =	**0.5**
Momentum *(between 0 and 1)*	**0.5**		

Training Options

Total #rows in your data *(Minimum 10)*	**40**	No. of Training cycles *(Maximum 500)*	**150**
Present Inputs in Random order while Training ?	**NO**	Training Mode *(Batch or Sequential)*	**Sequential**
Saving Network Weights	**With least Training Error**		
Training / Validation Set	**Partition data into Training / Validation set**		Build Model

If you want to partition, how do you want to select the Validation set ?

Please choose one option	**1**	**Option 1 : Randomly select**	25%	**of data as Validation set** (between 1% and 50%)
Please fill up the input necessary for the selected option		**Option 2: Use last**	10	**rows of the data as validation set**

Save model in a separate workbook? **NO**

b) Sample of data gathered

Enter your Data in this sheet

Instructions:

Start Entering your data from cell **AC105**. Specify variable name in row 103.
Make sure that the row 104 is blank.

Specify variable type in row 102.

Cont - for continuous Input,
Cat - for Categorical Input,
Output -for Output var.
Omit - if you don't want to use the variable in tl Specify variable names in row 103

For each continuous Input, there will be 1 neuron in Input Layer.
For Each categorical Input with *K* levels, there will be *K* neurons in Input Layer
Please make sure that there are no more than 50 neurons in Input Layer.
There should be exactly 1 Output variable - application will treat it as Categorical
There should be no more than 40 Categorical Variables.

Var Type	Omit	Omit	Output	Cont	Cat	Cont	Cat	Omit	Omi	Or
Var Name	**Species_ID**	**Project name**	**Type**	**SquareFeet**	**Location**	**Min Price**	**Amenities**	**JUNK**		
	1	Magna Ville Condominium	Condominium	1261	City Center	180000	Transportation	A		
	2	Putra Avenue	Link	796	Nilai	398888	Education	B		
	3	Taman Desa Mas	Semi-D	1800	Subang Jaya	149800	Commercial	C		
	4	Amansiara	Duplex Terrace	1749	Damansara	169880	Commercial	C		
	5	Parklane Heights	Duplex Terrace	1313	Subang Jaya	276870	Commercial	C		
	6	Jelutong Heights	Semi-D	1650	City Center	1098144	Transportation	A		
	7	Taman Sutera	Apartment	3600	Damansara	87420	Commercial	C		
	8	Data Suria	Semi-D	850	Nilai	988800	Education	B		
	9	Unipark Condominium	Condominium	3200	Nilai	193800	Education	B		
	10	Taman Putra Prima	Semi-D	1088	City Center	265500	Transportation	A		
	11	Cahaya Permai Apartment	Apartment	2110	Nilai	100000	Education	B		
	12	Changkat View Condominium	Condominium	930	City Center	225556	Commercial	C		
	13	Timur Enstek	Link	1365	Nilai	272600	Education	B		
	14	Taman Tasik Prima	Link	1580	Damansara	359290	Commercial	C		
	15	Andari Townvilla	Link	2065	Nilai	160600	Education	B		
	16	Bayu Permai Acadia	Link	1679	City Center	72000	Commercial	C		
	17	Anggunpuri Condominium	Condominium	1190	City Center	194800	Commercial	C		
	18	Taman Cheras Idaman	Link	1945	City Center	238000	Transportation	A		
	19	Taman Pinggiran Mahkota	Duplex Terrace	1400	City Center	228900	Transportation	A		
	20	Riana Green East	Condominium	850	City Center	187600	Commercial	C		
	21	Asmara Condominium	Condominium	1100	City Center	180000	Transportation	A		
	22	Marine Height	Semi-D	1500	Nilai	265000	Education	B		
	23	Taman Perak Setia	Semi-D	1640	Subang Jaya	149800	Commercial	C		
	24	Damansara Suria	Duplex Terrace	1533	Damansara	169880	Commercial	C		
	25	Parklane Avenue	Duplex Terrace	1200	Subang Jaya	276870	Commercial	C		
	26	Segambut Heights	Semi-D	1700	City Center	1098144	Transportation	A		
	27	Taman Perdana Mewah	Apartment	850	Damansara	87420	Commercial	C		
	28	Puchong Suria	Semi-D	1250	Nilai	988800	Education	B		
	29	Plerish Condominium	Condominium	1130	Nilai	193800	Education	B		
	30	Taman Putra Perdana	Semi-D	1390	City Center	265500	Transportation	A		
	31	Damai Permai Apartment	Apartment	930	Nilai	100000	Education	B		
	32	Setapak Lake Condominium	Condominium	1109	City Center	225556	Commercial	C		
	33	Timur Enstek	Semi-D	1655	Nilai	272600	Education	B		
	34	Taman Prima	Duplex Terrace	2065	Damansara	359290	Commercial	C		
	35	Asmarinda Townvilla	Link	2065	Nilai	160800	Education	B		
	36	Bayu Permai Sutera	Condominium	1679	City Center	72000	Commercial	C		
	37	Taman Desa Condominium	Condominium	1050	Nilai	194800	Commercial	C		
	38	Taman Cheras Perdana	Link	1945	City Center	238000	Transportation	A		
	39	Taman Pinggiran Bolton	Duplex Terrace	1750	Damansara	228900	Transportation	A		
	40	Wangsa Melati Heights	Condominium	990	City Center	187600	Commercial	C		

c) Neural Network Model for Classification

Neural Network Model for Classification Created On : 24-Mar-08

% MissClass.(Training) 6.67% % MissClass.(Validation) 40.00%

Number of Hidden Layers 2
Layer Sizes 9 5 5 5

True Output (if available)
Model Output link

	Bias	SquareFee	Location	Min Price	Amenities
		Cont	Cat	Cont	Cat
Raw Input	1	3319.6001	city center	289095.1875	transportation

Enter your Inputs in the range AG112:AJ112 - the cells marked in green.

	Bias	SquareFeet	Location.city center	Location.nilai	Location.subang jaya	L a		nsportation	Amenities.education	Amenities.commercial	
Transformed Input	1	0.8999	1.0000	0.0000	0.0000	0.0000	0.2116	1.0000	0.0000	0.0000	
Hdn1_bias	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Hdn1_Nrn1	-3.0616	28.1738	-1.0211	-6.5454	5.2583	-1.0132	14.3285	-3.0011	0.6271	-0.1089	21.3025
Hdn1_Nrn2	4.0479	-3.5673	-8.1480	-1.1956	10.2019	4.6329	-31.6712	2.4586	4.2782	-2.3644	-11.5541
Hdn1_Nrn3	-3.6152	-2.1283	2.3530	1.8497	5.8860	-14.3159	27.5417	-7.7567	-1.4197	5.2273	-5.1058
Hdn1_Nrn4	-3.5966	-2.1267	-6.8529	-1.7315	11.5387	-6.1676	45.6809	1.2705	1.6863	-6.6253	-1.4258
Hdn1_Nrn5	-2.9537	35.7479	5.2142	8.4137	-3.5764	-13.2445	-20.8244	-10.0880	5.7438	1.4131	19.9357
	1.0000	1.0000	0.0000	0.0060	0.1938	1.0000					
Hdn2_bias	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000				
Hdn2_Nrn1	-1.5286	-1.0295	-4.2307	3.1947	-1.7335	-0.1942	-3.0690				
Hdn2_Nrn2	1.0287	-4.7501	-3.2818	2.0625	-0.1894	1.4735	-2.2723				
Hdn2_Nrn3	1.6267	-7.2186	1.7498	-1.7324	-6.9082	4.4258	-2.5150				
Hdn2_Nrn4	-2.2622	0.0865	8.3508	-2.0087	-1.7571	-6.2534	-8.7817				
Hdn2_Nrn5	0.1441	-1.6567	1.3161	5.8030	5.7998	-6.9063	-7.2601				
	1.0000	0.0444	0.0934	0.0748	0.0002	0.0007					
Op_bias	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000				
Op_Nrn1	-3.3485	1.9150	3.1632	0.9312	-2.9081	0.0422	-2.8986				
Op_Nrn2	2.3039	-0.7521	-1.7695	-2.7300	-3.0357	-5.4152	1.8966				
Op_Nrn3	-1.1510	0.4064	-1.3380	-3.2438	-6.7536	3.8393	-1.4990				
Op_Nrn4	-2.8223	-2.2570	-2.0290	-5.1442	5.4640	0.3731	-3.4959				
Op_Nrn5	-3.4250	-4.1848	-1.8800	7.1158	2.1592	-4.3172	-3.2569				
	1.0000	0.0522	0.8695	0.1826	0.0294	0.0371					
	2										

Category Table

	Type	**Location**	**Amenities**
	5	4	3
1	condominiu	city center	transportation
2	link	nilai	education
3	semi-d	subang jaya	commercial
4	duplex terra	damansara	
5	apartment		

Confusion Matrix - Training set

TRUE	Predicted: condominiu	link	semi-d	duplex terrace	apartment
condominium	8	1	0	0	0
link	0	6	0	0	0
semi-d	0	0	7	0	0
duplex terrace	0	1	0	4	0
apartment	0	0	0	0	3

Confusion Matrix - Validation

TRUE	Predicted: condominiu	link	semi-d	duplex terra	apartment
condominiu	1	1	0	0	0
link	0	0	2	0	0
semi-d	0	0	3	0	0
duplex terra	0	1	0	1	0
apartment	0	0	0	0	1

d) Profile plots for fitted model

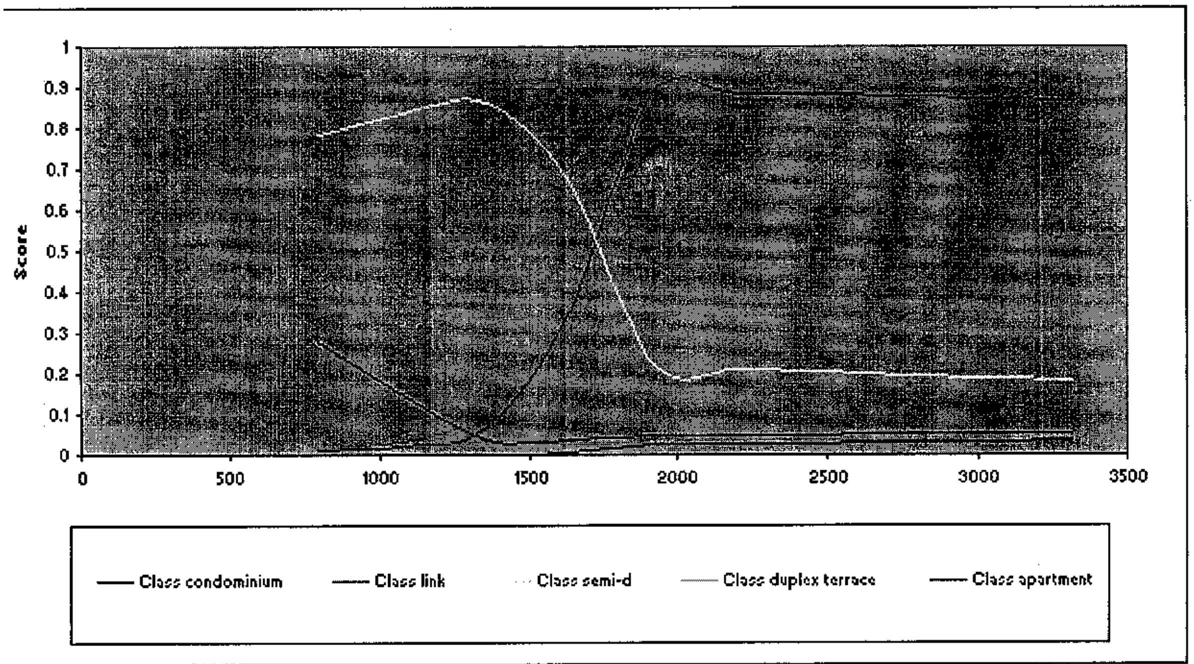
Profile plot for the fitted model Create Profile

Generate profile for each of the Class categories
Generate 10 data points
by varying **SquareFeet** between 796 and 3600
keeping the other predictors fixed at the specified values
Notes:
For each of the Class category one profile is created
X axis is the predictor you have chosen to vary
Y axis is the Score - the scaled output of the network for that category
At any value of X - the category having the highest score is the precicted Class category.

Predictor	SquareFeet	Location	Min Price	Amenities
Fixed Value	1523.800	city center	289035.188	transportation

Min / Max in Original Data (for user's reference only)

Min	796.00	72000.00
Max	3600.00	1098144.00

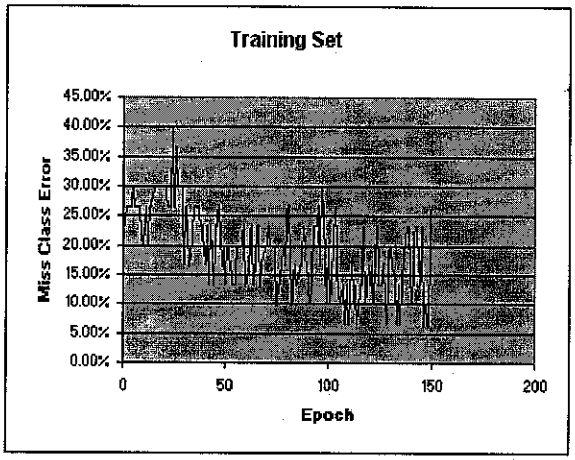


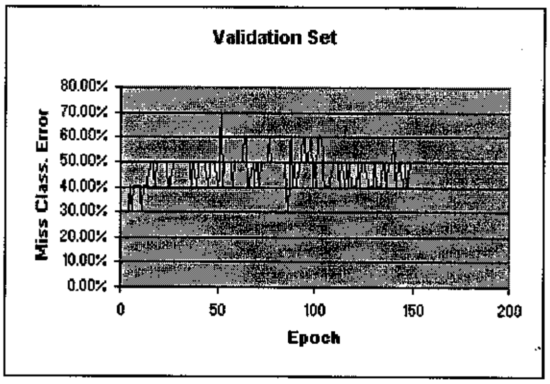
Profile Data
Number of Class Categories 5

SquareFeet	Class condominiur	Class link	Class semi-c	Class duplex	Class apartment
796	0.27973724	0.01022547	0.78602557	0.01498202	0.00014186
1076.400024	0.147701171	0.01863885	0.83668409	0.0297557	0.00019303
1356.800049	0.039392519	0.06742375	0.86478332	0.07770348	0.00065152
1637.199951	0.038957056	0.35035058	0.66755746	0.06355521	0.00295626
1917.599976	0.046288437	0.88876306	0.22130853	0.04273964	0.02511769
2198	0.048810505	0.88474598	0.20887128	0.03879134	0.02719035
2478.399902	0.049562054	0.88197837	0.20337439	0.03670868	0.02881554
2758.800049	0.050356547	0.87855095	0.19719486	0.03443049	0.03093411
3039.199951	0.051250308	0.87436278	0.19017611	0.03196294	0.03367577
3319.600098	0.052222398	0.86950714	0.18257698	0.02943048	0.03708091

e) Output from Training Set / Validation Set

Epoch	% Missclassified (Training Set)	% MissClassified (Validation Set)
1	23.33%	30.00%
2	26.67%	30.00%
3	26.67%	30.00%
4	26.67%	30.00%
5	30.00%	40.00%
6	26.67%	30.00%
7	26.67%	40.00%
8	26.67%	40.00%
9	23.33%	40.00%
10	20.00%	40.00%
11	26.67%	30.00%
12	20.00%	40.00%
13	26.67%	40.00%
14	26.67%	40.00%
15	30.00%	50.00%
16	30.00%	50.00%
17	30.00%	50.00%
18	30.00%	40.00%
19	30.00%	50.00%
20	30.00%	50.00%
21	30.00%	50.00%
22	26.67%	50.00%
23	26.67%	50.00%
24	40.00%	50.00%
25	26.67%	40.00%
26	36.67%	40.00%
27	30.00%	50.00%
28	30.00%	50.00%
29	30.00%	50.00%
30	16.67%	50.00%
31	26.67%	50.00%
32	16.67%	50.00%
33	20.00%	50.00%
34	20.00%	50.00%
35	26.67%	50.00%
36	26.67%	50.00%
37	26.67%	40.00%
38	23.33%	50.00%
39	23.33%	40.00%
40	16.67%	50.00%
41	23.33%	50.00%
42	13.33%	50.00%
43	23.33%	50.00%
44	13.33%	50.00%
45	23.33%	40.00%
46	23.33%	50.00%
47	26.67%	50.00%
48	16.67%	50.00%
49	20.00%	50.00%
50	13.33%	40.00%
51	20.00%	50.00%
52	20.00%	70.00%
53	13.33%	40.00%
54	13.33%	50.00%



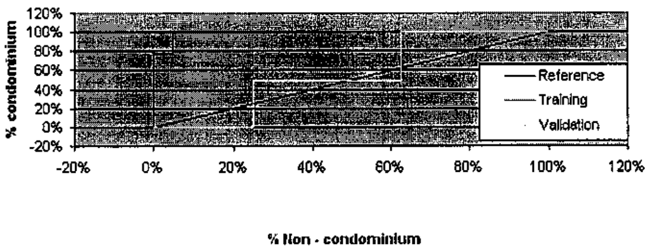


f) Lift Chart for the Fitted Model

Lift Chart for the Fitted Model

Generate Lift Chart for the Class Category condominium

Create Lift Chart



Class categories
condominium
link
semi-d
duplex terrace
apartment

	Training		Validation	
Total +ve's	9		Total +ve's	2
Total -ve's	21		Total -ve's	8

Train / Validation	Class	Score	Training %-ve	Training %+ve	Validation %-ve	Validation %+ve	Reference %-ve	Reference %+ve
1	1	0.902953	0.00%	0.00%	0.00%	0.00%	0%	0%
1	1	0.898284	0.00%	11.11%	12.50%	0.00%	100%	100%
1	1	0.867912	0.00%	22.22%	25.00%	0.00%		
1	1	0.866353	0.00%	33.33%	25.00%	50.00%		
1	1	0.613793	0.00%	44.44%	37.50%	50.00%		
1	1	0.542171	0.00%	55.56%	50.00%	50.00%		
1	1	0.373635	0.00%	66.67%	62.50%	50.00%		
1	0	0.317809	0.00%	77.78%	62.50%	100.00%		
1	1	0.317809	4.76%	77.78%	75.00%	100.00%		
1	1	0.248635	4.76%	88.89%	87.50%	100.00%		
1	0	0.180847	4.76%	100.00%	100.00%	100.00%		
1	0	0.085013	9.52%	100.00%				
1	0	0.082862	14.29%	100.00%				
1	0	0.080146	19.05%	100.00%				
1	0	0.05645	23.81%	100.00%				
1	0	0.05645	28.57%	100.00%				
1	0	0.05512	33.33%	100.00%				
1	0	0.040992	38.10%	100.00%				
1	0	0.038385	42.86%	100.00%				
1	0	0.036178	47.62%	100.00%				
1	0	0.03592	52.38%	100.00%				
1	0	0.029312	57.14%	100.00%				
1	0	0.022556	61.90%	100.00%				
1	0	0.022556	66.67%	100.00%				
1	0	0.017402	71.43%	100.00%				
1	0	0.006352	76.19%	100.00%				
1	0	0.002706	80.95%	100.00%				
1	0	0.002028	85.71%	100.00%				
1	0	0.002024	90.48%	100.00%				
1	0	0.002008	95.24%	100.00%				
0	0	0.644742	100.00%	100.00%				
0	0	0.210051						
0	1	0.161877						
0	0	0.090266						
0	0	0.082862						
0	0	0.075274						
0	1	0.036149						
0	0	0.031028						
0	0	0.025792						
0	0	0.002956						

4.1.4 Neural Network Variables

Below is the model developed using NN environment that will be the backbone of the REVFOS system which it will be supported in several ways of DM techniques.

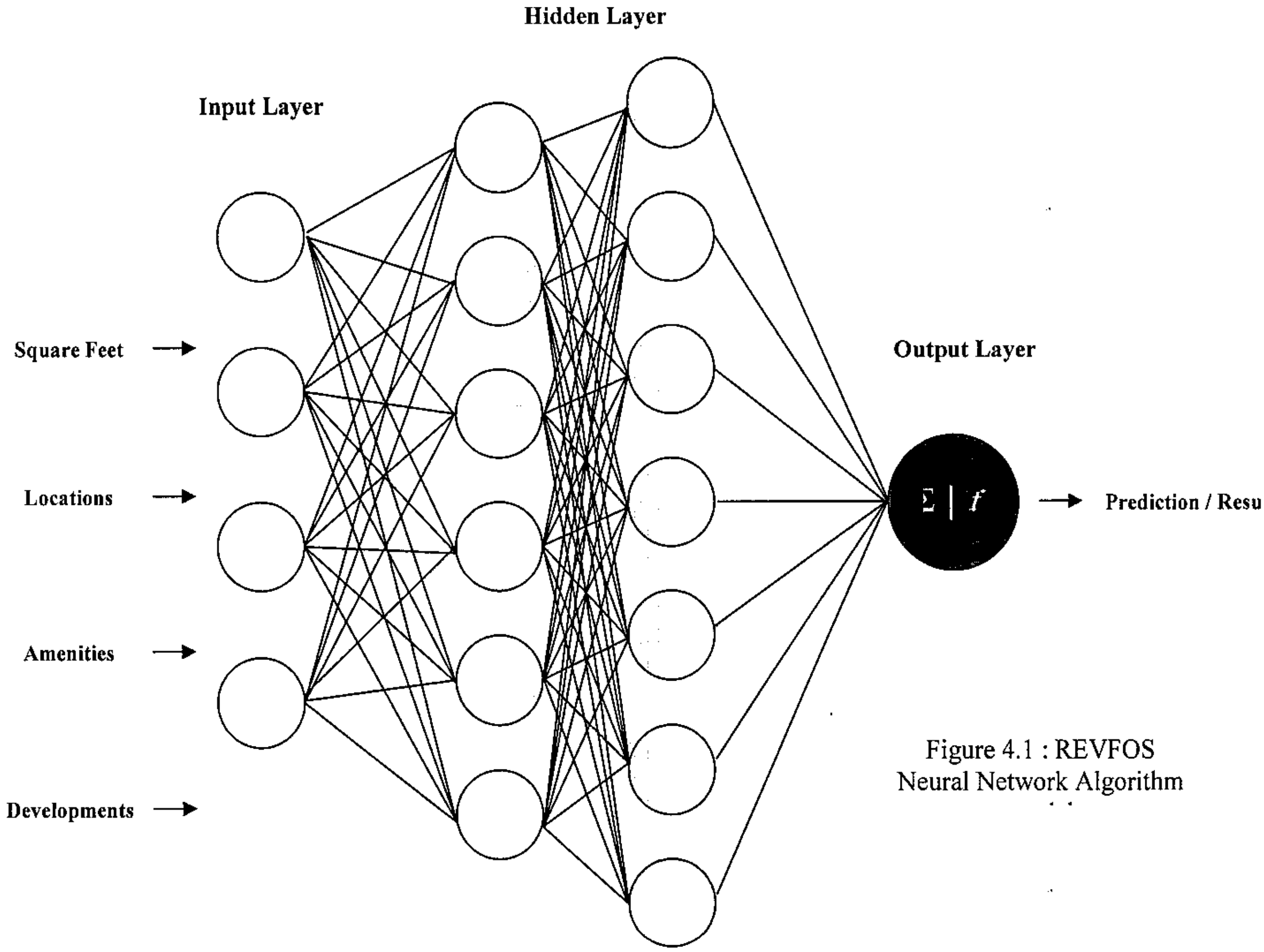


Figure 4.1 : REVFOS Neural Network Algorithm

Features of nodes in NN hidden layer :

a) **Accessibility ***
b) **Features**
c) **Future Development ***
d) **Tenure of Land**
e) **Property Type**
f) **Location Branding**

CHAPTER 5
CONCLUSION AND RECOMMENDATIONS

5.0 Conclusion

Data mining has emerged to be very important research area that it is believed to help organizations make use of the tremendous amount of available data. As we know, in the past, it was almost impossible to gather information from large databases, datasets or data marts but due to technology advancement and efficient manpower, data mining has unlocked these limitations as with the combination of many other research disciplines, data mining has turns raw and unprocessed data into useful and valuable information.

Data-mining application will be developed slightly different with common software development process which it will be solemnly based on data gathering and preprocessing before proceed with result prediction. For instance, REVFOS is developed using techniques available in DM environment since it required the author to gather all historical data before developing NN model for testing and training the samples obtained. By using several crucial elements from the industry such as number of bedrooms, interest rate, present value and locations, REVFOS will be able to predict potential value in property market based on some mathematical justifications that will directly support the findings. Therefore, it will be crucial to the author in ensuring that REVFOS will be ultimate solutions for real estate industry in predicting property value.

REVFOS will be implied several techniques from SDLC processes as for minimal requirements for system development and all the weighted values assigned to respective criteria will be stored, updated and generated from the databases.

5.1 Recommendations

Since the project serves as a solution of property agents and consumers in forecasting the market price in real estate industry, thus there exist many other possible patches or extension to this project in order to make it reliable and powerful application.

Firstly, due to time constraint in gathering data from various sources, the project had limited its scope to Klang Valley areas whereby it can be covered larger areas with additional features to be added in future patches. It will help the author to acquire larger market and generate more accurate predictive value. Secondly, more artificial intelligence interaction between the system and the databases in order to update and capture values in computing mathematical formulas in the system.

Lastly, more preprocessing features such as scaling data, data regression using various data mining tools can be implemented to allow for higher quality data that believed can contribute more accurate weighted value in predicting real estate price.

REFERENCES

Mehmed Kantardzid, 2003, *Data Mining : Concept, Models, Methods and Algorithms,* IEEE Press, John Wiley & Sons

Witten, Ian H. and Frank, Eibe, 2000, *Data Mining : Practical Machine Learning Tools and Techniques with Java Implementations,* Waikaito University, Morgan Kaufmann Publishers, ISBN 0 23 088407 0, pg. 3-7

Fausett L., *Fundamentals of Neural Networks*, Prentice-Hall, 1994. ISBN 0 13 042250 9, pg. 63-73

Haykin S., *Neural Networks* , 2nd Edition, Prentice Hall, 1999, ISBN 0 13 273350 1 pg. 144-149

Mitra, Sushmita and Acharya, Tinku, *Data Mining: Multimedia, Soft Computing and Bioinformatics*, John Wiley and Sons, ISBN 0 47 146054 0, pg. 4

Nurazzah Abd Rahman, Zainab Abu Bakar & Normaly Kamal Ismail, 2005, *Cluster-Based Hadith Retrieval System,* Ph.D. Thesis, MARA University of Technology, Malaysia

Margeret H. Dunham, 2003, *Data Mining Introductory and Advanced Topics*, New Jersey, Prentice Hall

David Hand, Keikki Mannila, Padhraic Smuth, 2002, *Principles of Data Mining*, The MIT Press

Michalski R. et al, 1998, *Machine Learning and Data Mining: Methods and Applications*, John Wiley & Sons

Khairudin Omar, Sales Executive. Platinum Victory, Kuala Lumpur. Personal Interview. August. 14. 2007.

Nordiana Abd. Rani, Marketing Executive. Worldwide Holding Berhad, Shah Alam. Personal Interview, January 31, 2008

Christos Stergiou and Dimitrios Siganos, 17 March 2001
<http://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/report.html>

Kuala Lumpur City Hall, 9 February 2008
< http://www.dbkl.gov.my/pskl2020/english/housing/index.htm>

APPENDICES

Appendix 1. Kuala Lumpur Residential Area Distribution

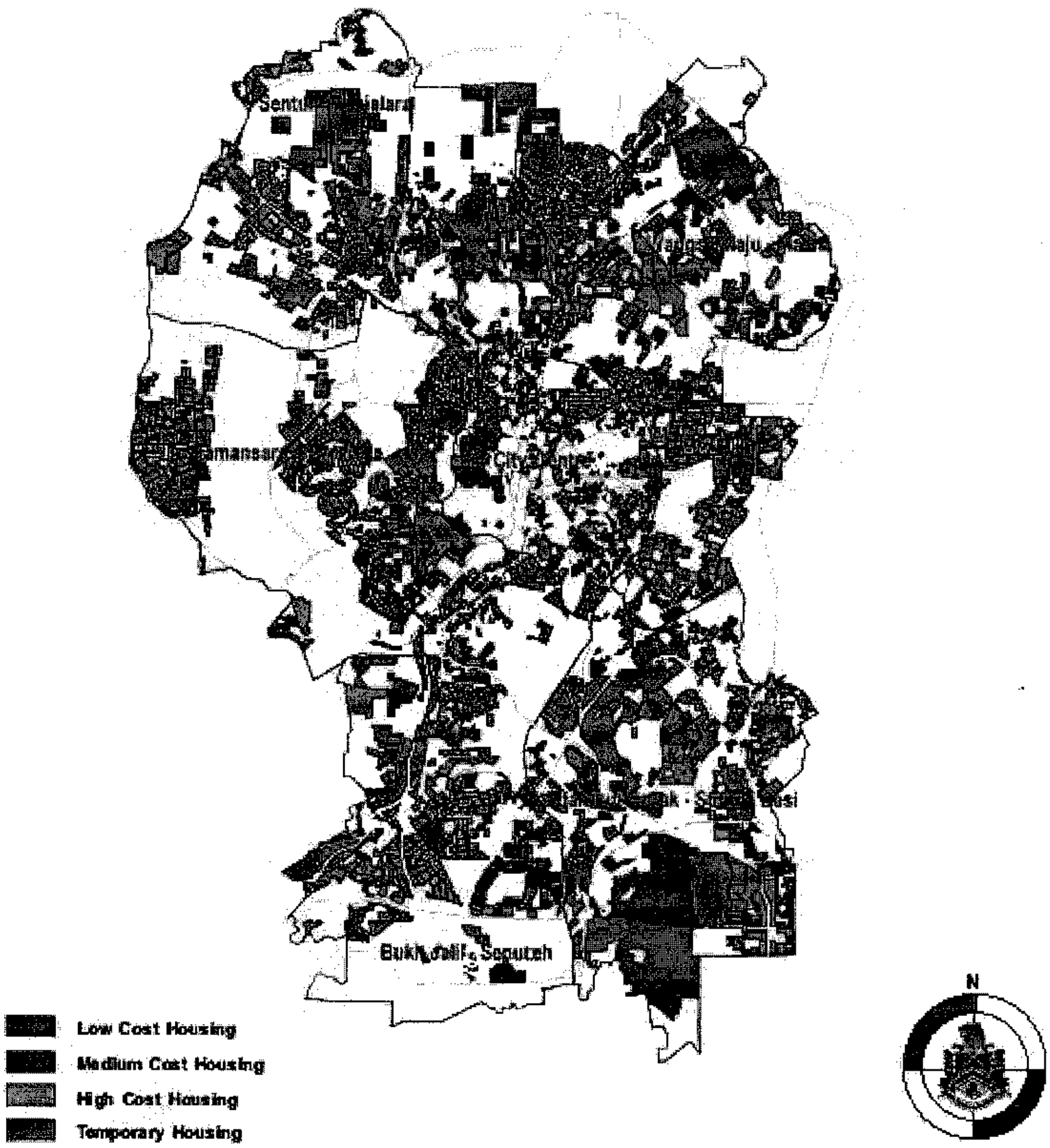
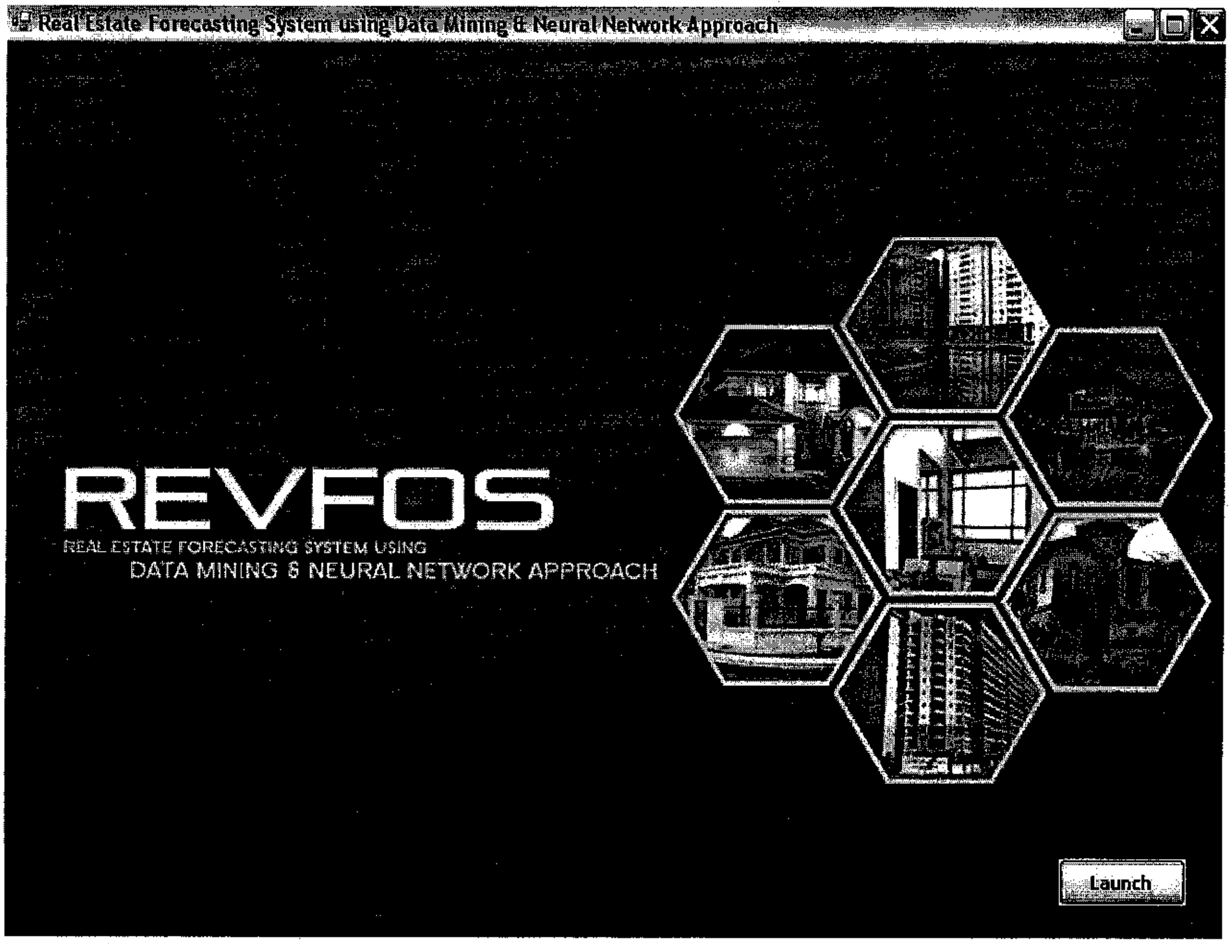
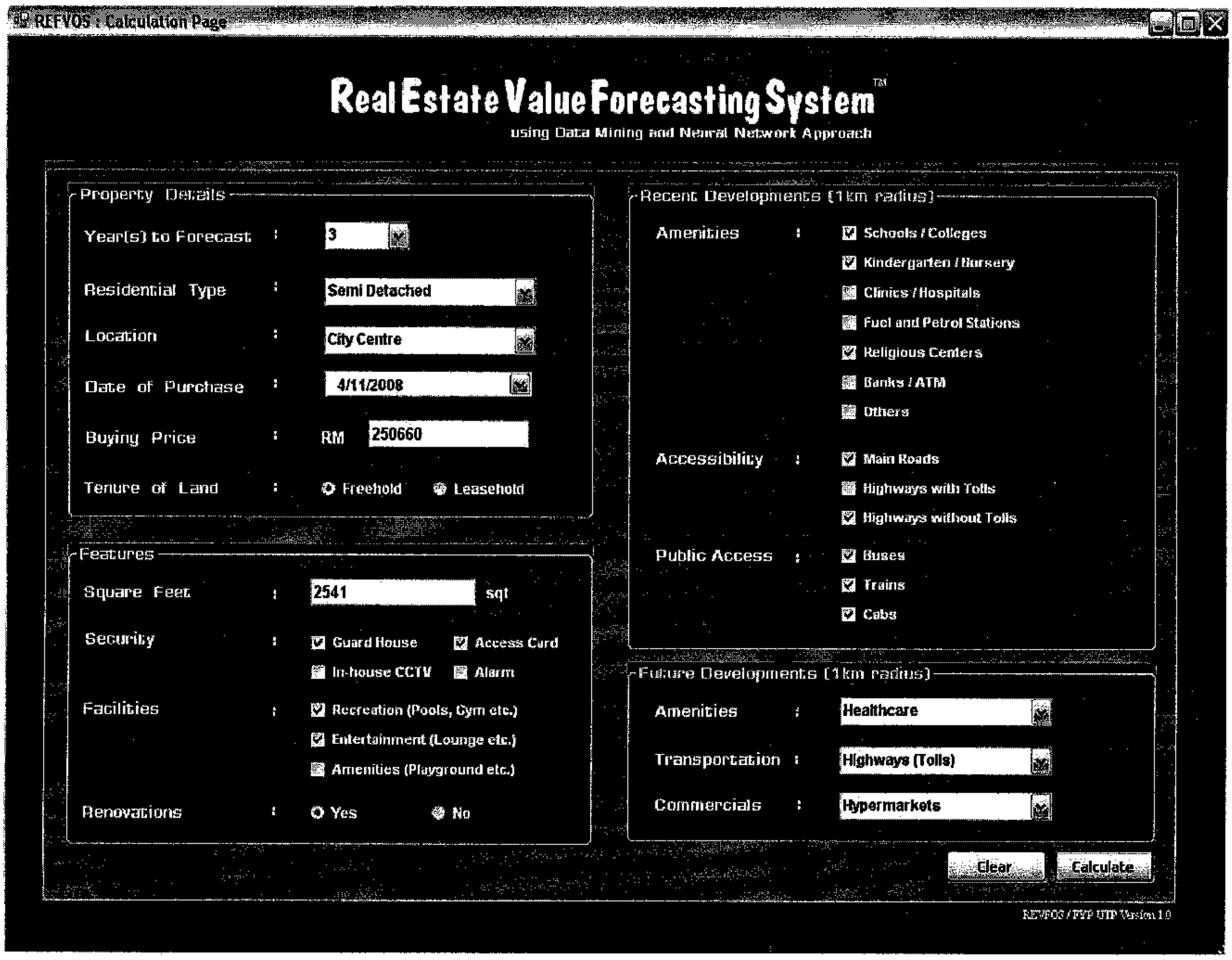


Figure 12.3 : Distribution of housing by type, 2000

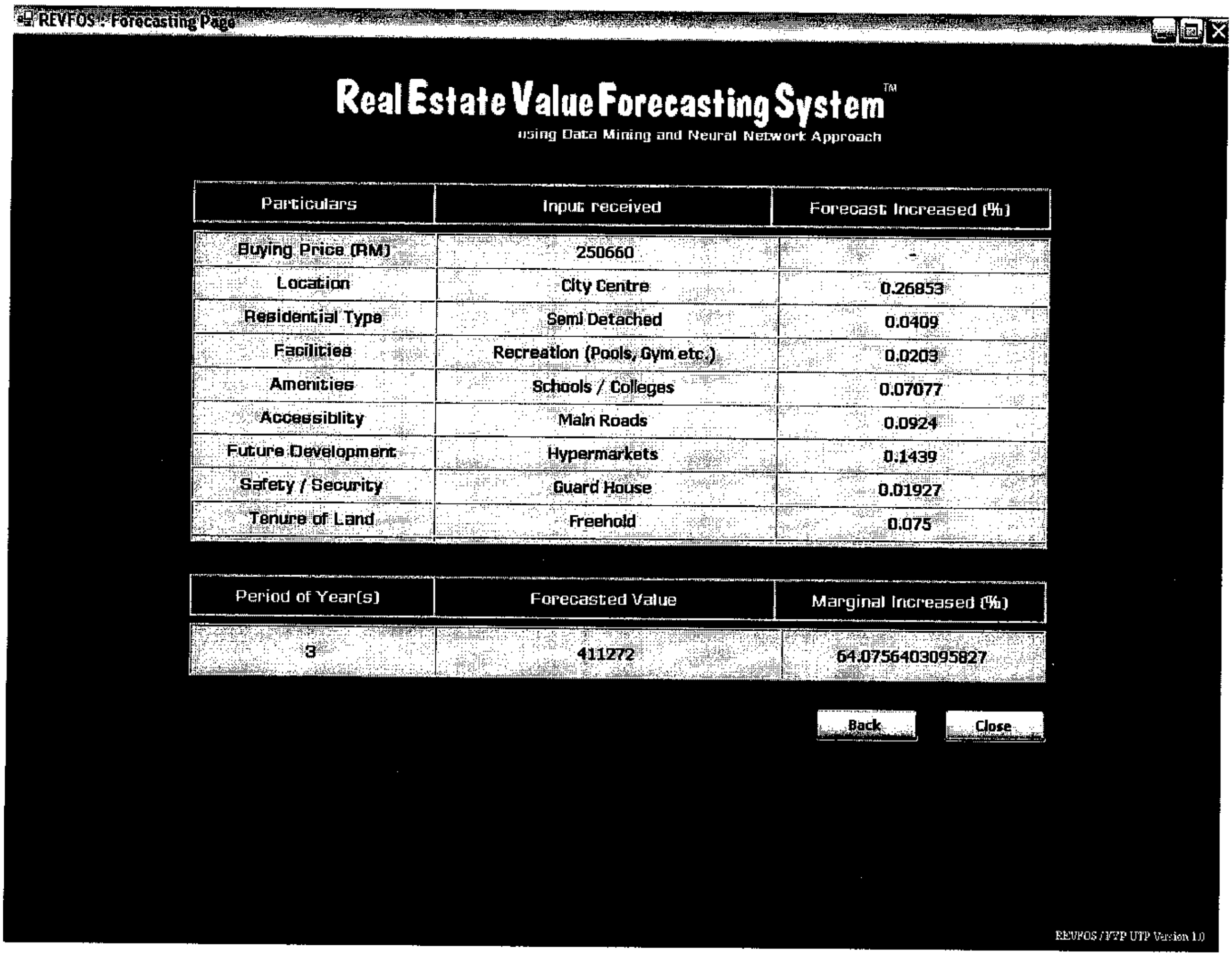
Appendix 2. System Interfaces



Landing Page


REFVOS : Calculation Page
Real Estate Value Forecasting System
using Data Mining and Neural Network Approach
Property Details
Year(s) to Forecast :
3
Residential Type :
Semi Detached
Location :
City Centre
Date of Purchase :
4/11/2008
Buying Price :
RM
250660
Tenure of Land :
Freehold
Leasehold
Features
Square Feet :
2541
sqt
Security :
Guard House
Access Card
In-house CCTV
Alarm
Facilities :
Recreation (Pools, Gym etc.)
Entertainment (Lounge etc.)
Amenities (Playground etc.)
Renovations :
Yes
No
Recent Developments (1km radius)
Amenities :
Schools / Colleges
Kindergarten / Nursery
Clinics / Hospitals
Fuel and Petrol Stations
Religious Centers
Banks / ATM
Others
Accessibility :
Main Roads
Highways with Tolls
Highways without Tolls
Public Access :
Buses
Trains
Cabs
Future Developments (1km radius)
Amenities :
Healthcare
Transportation :
Highways (Tolls)
Commercials :
Hypermarkets
Clear
Calculate

Main Page



Forecasting page

Appendix 3. Data Sample

House Price list (pg 1)							
Project's Name	Location	Area	Type	Tenure of Land	Square Feet	Min. Price	Max. Price
Magna Ville Condominium	Selayang	Selayang	Condominium	Leasehold	1,261	$180,000.00	$267,000.00
Acacia Avenue	Nilai	Nilai	Shoplots	Freehold	1,506	$231,000.00	$330,000.00
Putra Avenue	Subang Jaya	Subang Jaya	Double-storey Link	Freehold	1,800	$398,888.00	$740,888.00
Taman Desa Mas	Rawang	Selayang	Semi-D	Leasehold	1,749	$149,800.00	$306,000.00
Impian Meridian	USJ	Subang Jaya	Residence Suite	Freehold	1,302	$305,800.00	$458,226.00
Amansiara	Jalan Ipoh	Selayang	Duplex Terrace	Leasehold	1,313	$169,880.00	$428,000.00
Parklane Heights	Rawang	Selayang	Duplex Terrace	Leasehold	1,650	$276,870.00	$409,645.00
Jelutong Heights	Bukit Jelutong	Shah Alam	Semi-D	Freehold	3,600	$1,098,144.00	$1,238,800.00
Taman Sutera	Kajang	Kajang	Apartment	Freehold	850	$87,420.00	$100,440.00
Data Suria	Ampang	Ampang Jaya	3-half Semi D	Leasehold	3,200	$888,800.00	$2,488,800.00
Unipark Condominium	Putrajaya	Sepang	Condominium	Freehold	1,088	$193,800.00	$250,800.00
Taman Putra Prima	Puchong	Sepang	Double Storey Terrace	Freehold	2,110	$265,500.00	$408,100.00
Cahaya Permai Apartment	Seri Kembangan	Subang Jaya	Apartment	Leasehold	930	$100,000.00	$100,000.00
Changkat View Condominium	Kota Damansara	Damansara	Condominium	Freehold	1,365	$225,556.00	$716,000.00
Timur Enstek	Sepang	Nilai	Single Storey Link	Freehold	1,580	$272,600.00	$431,377.00
Taman Tasik Prima	Puchong	Subang Jaya	Double-storey Link	Leasehold	2,065	$359,290.00	$651,090.00
Andari Townvilla	Selayang	Selayang	Double-storey Link	Leasehold	1,679	$160,800.00	$242,547.00
Bayu Permai Acadia	Rawang	Selayang	Single Storey Link	Leasehold	796	$72,000.00	$132,000.00
Ansara Impian Condominium	Batu Caves	Selayang	Condominium	Freehold	1,021	$210,500.00	$275,800.00
Brickmall Avenue	Dengkil	Nilai	Shoplots	Freehold	1,200	$189,000.00	$251,600.00
Idaman Mahligai Avenue	USJ	Subang Jaya	Double-storey Link	Freehold	1,680	$425,000.00	$862,000.00
Taman Desa Perak	Rawang	Selayang	Semi-D	Freehold	2,011	$250,669.00	$459,000.00
Impian Meridian II	Kelana Jaya	Subang Jaya	Residence Suite	Freehold	1,750	$550,900.00	$1,025,650.00
Damansiara Height	Jalan Ipoh	City Center	Duplex Terrace	Leasehold	1,200	$250,750.00	$501,500.00
Melati Impian Heights	Sg. Buloh	Rawang	Duplex Terrace	Freehold	1,560	$290,665.00	$500,400.00
Jelutong Perdana	Batu Tiga	Shah Alam	Semi-D	Freehold	3,760	$1,250,950.00	$1,560,995.00
Taman Bayu Senja	Kajang	Kajang	Apartment	Freehold	850	$87,420.00	$100,440.00
Data Palma	Ampang	Ampang Jaya	3-half Semi D	Leasehold	3,150	$1,026,441.00	$2,788,690.00
Unipark Apartment	Nilai	Sepang	Condominium	Freehold	1,204	$178,004.00	$205,630.00
Taman Putra Perdana	Bandar Baru Puchong	Puchong	Double Storey Terrace	Freehold	2,110	$350,115.00	$465,210.00
Cahaya Kasih Apartment	Seri Kembangan	Subang Jaya	Apartment	Leasehold	930	$95,600.00	$125,000.00
Exselview Condominium	Sri Hartamas	Damansara	Condominium	Freehold	2,056	$568,009.00	$1,085,991.00
Timur Enstek II	Sepang	Nilai	Single Storey Link	Freehold	1,580	$326,544.00	$531,778.00
Taman Emas Prima	Bandar Baru Puchong	Puchong	Double-storey Link	Leasehold	2,165	$409,100.00	$751,600.00
Amandari Townvilla	Selayang	Selayang	Double-storey Link	Leasehold	1,799	$175,800.00	$292,457.00
Bayu Permai Height	Sg. Buloh	Rawang	Single Storey Link	Leasehold	796	$72,000.00	$114,000.00
Melati Impian Condominium	Taman Melati	Setapak	Condominium	Freehold	1,021	$180,650.00	$250,667.00
Dayu Suria Condominum	Taman Desa	Kerinchi	Condominium	Freehold	1,240	$265,001.00	$362,500.00
Desa Delima Avenue	Kelana Jaya	Kelana Jaya	Double-storey Link	Freehold	1,680	$375,000.00	$567,800.00
Taman Suria Enigma	Bandar Tun Razak	Bandar Tun Razak	Semi-D	Freehold	2,011	$356,000.00	$615,290.00
Impiana Cheras Height	Cheras	Seputeh	Residence Suite	Freehold	1,750	$758,800.00	$988,988.00
Medan Idaman	Gombak	Setapak	Duplex Terrace	Leasehold	1,350	$360,500.00	$780,441.00
Teratai Mewah Apartment	Taman Melati	Setapak	Apartment	Leasehold	950	$96,000.00	$132,560.00
Paya Jaras Avenue	Sg. Buloh	Rawang	Semi-D	Freehold	2,050	$650,700.00	$765,250.00
Semarak Api Apartment	Balakong	Kajang	Apartment	Freehold	755	$82,500.00	$112,650.00
Palma Indah Terrace	Taman Kosas	Ampang	3-half Semi D	Leasehold	2,855	$1,250,660.00	$1,875,005.00
Zihan Avenue	Brickfield	Bangsar	Condominium	Freehold	1,520	$365,000.00	$654,000.00
Taman Putra Perdana	Putrajaya	Sepang	Double Storey Terrace	Freehold	1,980	$350,115.00	$455,625.00
Merak Apartment	Seri Kembangan	Sg. Besi	Apartment	Leasehold	930	$88,001.00	$100,500.00
Valencia Height	TTDI	Damansara	Suite	Freehold	3,570	$2,500,600.00	$5,002,359.00
Prima Perdana I	Dengkil	Nilai	Single Storey Link	Freehold	1,260	$355,900.00	$478,620.00
Taman Mewah	Kg. Pandan	Datuk Keramat	Terrace (Single)	Leasehold	1,005	$85,000.00	$165,000.00

House Pricelist (pg 2)

Security	Landscape	Others	Facilities: Recreations	Facilities: Entertainment	Facilities: Family
Guard House	No	NS	Swimming Pool, Courts, Gym	Wi-fi, Mini Theater	BBQ, Playground, Nursery, Laundry
CCTV	No	NS	No	No	No
Guard House	Yes	Air-cond, Autogate	Jogging Paths	No	Playground
Access Card, Guard House	Yes	Full tile	Jogging Paths	No	BBQ, Playground, Café
Home Alarm System	No	NS	Courts, Gymnasium, Pools, Sauna	Wi-fi, Broadband	BBQ, Playground, Café
Access Card, Guard House	Yes	NS	Clubhouse	Broadband	BBQ, Playground
Alarm, Access Card	Yes	NS	Parks, Jogging Tracks	Broadband, Bistro	Playground
Smart Home System	Yes	Warranty (36months)	Jogging Paths	No	Playground
Nil	No	NS	No	No	Playground
Guard House	Yes	Sky Garden	Parks, Jogging Tracks	No	BBQ, Playground, Café
CCTV, Security Guard, Access Card	No	NS	Courts, Gymnasium, Pools	Broadband, Bistro	MPH
Guard House	Yes	Ceramic Tile	Parks, Jogging Tracks	No	BBQ, Playground, Café
Guard House	No	NS	Pools, Courts	Lounges	Kindergarten, Playground, Shops
CCTV, Security Guard, Access Card	No	Intercom, Air-cond	Courts, Gymnasium, Pools	Wi-fi, Broadband, Bistro	BBQ, Playground, Nursery, MPH
CCTV, Security Guard	Yes	NS	Parks, Jogging Tracks	No	Kindergarten, Playground, Shops
Guard House	Yes	Air-cond	Jogging Paths	No	Nursery, Laundry
Guard House	No	NS	Swimming Pool, Courts, Gym	No	Nursery, Laundry
CCTV, Security Guard	No	NS	No	No	No
Guard House, Access Card	No	NS	Swimming Pool, Courts, Gym	Wi-fi, Broadband, Bistro	BBQ, Playground, Nursery, Laundry
CCTV	No	NS	No	No	Playground
Guard House	Yes	Air-cond, Autogate	Jogging Paths	No	Playground
Guard House	Yes	Full tile	No	No	No
Home Alarm System	No	NS	Courts, Gymnasium, Pools, Sauna	Wi-fi, Broadband	BBQ, Playground, Café
Access Card, Guard House	Yes	NS	Clubhouse	Broadband	BBQ, Playground
Access Card, Guard House	Yes	NS	Parks, Jogging Tracks	Broadband, Bistro	Playground
Smart Home System	Yes	Warranty (36months)	Parks, Jogging Tracks	No	No
Nil	No	NS	Courts, Gymnasium, Pools, Sauna	No	Playground
Guard House	Yes	Sky Garden	No	No	BBQ, Playground
CCTV, Security Guard, Access Card	No	NS	Courts, Gymnasium, Pools	Broadband, Bistro	BBQ, Playground, Nursery, Laundry
Guard House	Yes	Ceramic Tile	Jogging Paths	No	No
Guard House	No	NS	Pools, Courts	Lounges	Kindergarten, Playground, Shops
CCTV, Security Guard, Access Card	No	Intercom, Air-cond	Courts, Gymnasium, Pools	Lounges	BBQ, Playground, Nursery, MPH
Guard House, Alarm	Yes	NS	Jogging Paths	No	Nursery, Laundry
Guard House	Yes	Air-cond	Jogging Paths	No	Kindergarten, Playground, Shops
Guard House	Yes	NS	Swimming Pool, Courts, Gym	No	Nursery, Laundry
Alarm, Guard House	Yes	NS	No	No	No
Access Card, Guard House	No	NS	Swimming Pool, Courts, Gym	Wi-fi, Broadband	BBQ, Playground, Nursery, Laundry
CCTV	No	NS	Courts, Gymnasium, Pools, Sauna	Wi-fi, Broadband	BBQ, Playground, Café
CCTV, Guard House	Yes	Air-cond, Autogate	Jogging Paths	No	Playground
Smart Home System	Yes	Full tile	NS	Lounges	Playground
Home Alarm System, Access Card	No	NS	Courts, Gymnasium, Pools, Sauna	Wi-fi, Broadband	BBQ, Playground, Café
Access Card, Guard House	Yes	NS	Clubhouse	Broadband	BBQ, Playground
Nil	No	NS	Parks, Jogging Tracks	Broadband, Bistro	Playground
Smart Home System	Yes	Warranty (36months)	Jogging Paths	No	Playground
Nil	No	NS	Parks, Jogging Tracks	No	Playground
Guard House	Yes	Sky Garden	Jogging Paths	No	No
CCTV, Security Guard, Access Card	No	NS	Courts, Gymnasium, Pools	Wi-fi, Broadband	BBQ, Playground, Café
Smart Home System	Yes	Ceramic Tile	No	No	BBQ, Playground, Café
Guard House	No	NS	Pools, Courts	Lounges	Kindergarten, Playground, Shops
CCTV, Security Guard, Access Card	Yes	Intercom, Air-cond	Courts, Gymnasium, Pools	Wi-fi, Broadband, Bistro	BBQ, Playground, Nursery, MPH
Alarm, Guard House	Yes	NS	Jogging Paths	No	No
Alarm	Yes	Air-cond	No	No	Playground

House Pricelist (pg 3)

Projects Name	Amenities (radius 1 km)					
	Health Care	Educations	Banks	Fuel / Gas	Religious	Others
Magra Ville Condominium	Hospital	Schools, Kindergartens	Yes	Yes	Yes	NS
Acacia Avenue	Hospital	USIM,UIAM,INTI	NS	NS	Mosque	Indoor Stadium
Putra Avenue	Clinics	Yes (NS)	Yes	NS	NS	NS
Taman Desa Mas	Clinics	Schools, Kindergartens	Yes	Yes	NS	Police Station
Impian Meridian	Clinics	Colleges, Schools	NS	NS	NS	NS
Amansiara	Hospital	Chinese Primary Schools	NS	NS	NS	Police Station
Parklane Heights	Hospital, Clinic	UNISEL, Schools	NS	NS	NS	Post Office, 7-11
Jelutong Heights	Clinics	Primary & Secondary	NS	NS	Mosque	NS
Taman Sutera	Hospital, Clinic	NS	NS	NS	NS	NS
Data Suria	Hospital	International Schools	Yes	Yes	Yes	NS
Unipark Condominium	Hospital	KLIUC, UNITEN	NS	NS	NS	NS
Taman Putra Prima	Hospital	Primary Schools, University, College	Yes	Yes	Yes	NS
Cahaya Permai Apartment	Clinics	NS	Yes	Yes	Yes	NS
Changkat View Condominium	Clinics	International Schools	Yes	Yes	Yes	Hentian Duta
Timur Enstek	Clinics	NS	NS	Yes	Yes	F1 Circuit
Taman Tasik Prima	Hospital, Clinic	Yes (NS)	Yes	Yes	Yes	NS
Andari Townvilla	Hospital, Clinic	Yes (NS)	Yes	Yes	Yes	Stadium
Bayu Permai Acadia	Hospital, Clinic	Schools	Yes	Yes	Yes	Police Station
Ansara Impian Condominium	Hospital, Clinic	Schools, Colleges	NS	Yes	NS	NS
Brickmall Avenue	Hospital, Clinic	Schools, Colleges	NS	Yes	NS	Indoor Stadium
Idaman Mahligai Avenue	Clinics	Yes (NS)	Yes	NS	NS	NS
Taman Desa Perak	Clinics	Schools, Kindergartens	Yes	Yes	NS	Police Station
Impian Meridian II	Clinics	Colleges, Schools	NS	NS	NS	NS
Damansiara Height	Hospital	Chinese Primary Schools	NS	NS	NS	Police Station
Melati Impian Heights	Hospital, Clinic	UNISEL, Schools	NS	NS	NS	Post Office, 7-11
Jelutong Perdana	Hospital, Clinic	Primary & Secondary	NS	NS	Mosque	NS
Taman Bayu Senja	Hospital, Clinic	NS	NS	NS	NS	NS
Data Palma	Hospital	International Schools	Yes	Yes	Yes	NS
Unipark Apartment	Hospital	KLIUC, UNITEN	NS	NS	NS	NS
Taman Putra Perdana	Clinics	Primary Schools, University, College	Yes	Yes	Yes	NS
Cahaya Kasih Apartment	Clinics	NS	Yes	Yes	Yes	NS
Excelview Condominium	Hospital, Clinic	International Schools	Yes	Yes	Yes	Hentian Duta
Timur Enstek II	Hospital, Clinic	NS	NS	Yes	Yes	F1 Circuit
Taman Emas Prima	Clinics	Yes (NS)	Yes	Yes	Yes	NS
Amandari Townvilla	Hospital, Clinic	Yes (NS)	Yes	Yes	Yes	Stadium
Bayu Permai Height	Hospital, Clinic	Schools	Yes	Yes	Yes	Police Station
Melati Impian Condominium	Hospital, Clinic	Schools, Colleges	NS	Yes	NS	NS
Dayu Suria Condominum	Hospital, Clinic	Schools, Colleges	NS	Yes	NS	Indoor Stadium
Desa Delima Avenue	Clinics	Yes (NS)	Yes	NS	NS	NS
Taman Suria Enigma	Clinics	Schools, Kindergartens	Yes	Yes	NS	Police Station
Impiana Cheras Height	Clinics	Colleges, Schools	NS	NS	NS	NS
Medan Idaman	Hospital, Clinic	Chinese Primary Schools	NS	NS	NS	Police Station
Teratai Mewah Apartment	Hospital, Clinic	UNISEL, Schools	NS	NS	NS	Post Office, 7-11
Paya Jaras Avenue	Hospital, Clinic	Primary & Secondary	NS	NS	Mosque	NS
Semarak Api Apartment	Hospital, Clinic	NS	NS	NS	NS	NS
Palma Indah Terrace	Hospital, Clinic	International Schools	Yes	Yes	Yes	NS
Zihan Avenue	Hospital, Clinic	KLIUC, UNITEN	NS	NS	NS	NS
Taman Putra Perdana	Clinics	Primary Schools, University, College	Yes	Yes	Yes	NS
Merak Apartment	Clinics	NS	Yes	Yes	Yes	NS
Valencia Height	Hospital, Clinic	International Schools	Yes	Yes	Yes	Hentian Duta
Prima Perdana I	Hospital, Clinic	NS	NS	Yes	Yes	F1 Circuit
Taman Mewah	Clinics	Yes (NS)	Yes	Yes	Yes	NS

House Pricelist (pg 4)

Project's Name	Transportation (radius 1 km)		Others (radius 1 km)	Room(s)	Carpark(s)
	Highways	Public Access	Amenities		
Magna Ville Condominium	MRR2	Buses & Taxis	NS	NS	Yes (1)
Acacia Avenue	NS	NS	Jati Park	NS	NS
Putra Avenue	Elite,NKVE,LDP	NS	NS	NS	Yes (2)
Taman Desa Mas	PLUS, Guthrie Link, LDP	Buses & Taxis	NS	4	Yes (2)
Impian Meridian	KESAS	LRT	NS	NS	Yes (3)
Amansiara	MRR2	NS	Waterfalls, Forest	3	Yes (1)
Parklane Heights	PLUS, NKVE, Shah Alam Klang Expressway	KTM	Tropicana GCC	6	Yes (1)
Jelutong Heights	Federal Highway, NKVE, Guthrie Corridor	NS	Bukit Cahaya Park	5	Yes (3)
Taman Sutera	SILK Highway, PLUS	NS	Saujana Impian GCC	3	NS
Data Suria	MRR2, Elevated KLCC Highway	Taxis	NS	NS	Yes (3)
Unipark Condominium	LDP, SILK Highway, PLUS	ERL	IOI Golf Club	3	Yes (1)
Taman Putra Prima	KESAS, NS Central Link, LDP	Buses & Taxis	NS	4	Yes (2)
Cahaya Permai Apartment	LDP, PLUS, Besraya, Elite	All	Technology Park	3	Yes (1)
Changkat View Condominium	NKVE	Taxis	Mahkamah	3	Yes (1)
Timur Enstek	PLUS	NS	KLIA	4	Yes (2)
Taman Tasik Prima	LDP	NS	Kirrara Golf	5	Yes (2)
Andari Townvilla	Rawang Highway	NS	NS	5	NS
Bayu Permai Acadia	PLUS, MRR2	KTM, Taxis	NS	3	Yes (1)
Ansara Impian Condominium	MRR2, Mainroad	Buses & Taxis	NS	NS	Yes (1)
Brickmall Avenue	PLUS	NS	Park	NS	NS
Idaman Mahligai Avenue	Elite,NKVE,LDP	NS	NS	NS	Yes (2)
Taman Desa Perak	PLUS, Guthrie Link, LDP	Buses & Taxis	NS	4	Yes (2)
Impian Meridian II	KESAS	LRT	NS	NS	Yes (3)
Damansiara Height	MRR2	NS	Waterfalls, Forest	3	Yes (1)
Melati Impian Heights	PLUS, NKVE, Shah Alam Klang Expressway	KTM	Tropicana GCC	6	Yes (1)
Jelutong Perdana	Federal Highway, NKVE, Guthrie Corridor	NS	Bukit Cahaya Park	5	Yes (3)
Taman Bayu Senja	SILK Highway, PLUS	NS	Saujana Impian GCC	3	NS
Data Palma	MRR2, Elevated KLCC Highway	Taxis	NS	NS	Yes (3)
Unipark Apartment	LDP, SILK Highway, PLUS	ERL	IOI Golf Club	3	Yes (1)
Taman Putra Perdana	KESAS, NS Central Link, LDP	Buses & Taxis	NS	4	Yes (2)
Cahaya Kasih Apartment	LDP, PLUS, Besraya, Elite	All	Technology Park	3	Yes (1)
Excelview Condominium	NKVE	Taxis	Mahkamah	3	Yes (1)
Timur Enstek II	PLUS	NS	KLIA	4	Yes (2)
Taman Emas Prima	LDP	NS	Kirrara Golf	5	Yes (2)
Amandari Townvilla	Rawang Highway	NS	NS	5	NS
Bayu Permai Height	PLUS, MRR2	KTM, Taxis	NS	3	Yes (1)
Melati Impian Condominium	MRR2, Mainroad	Buses & Taxis	NS	NS	Yes (1)
Dayu Suria Condominum	PLUS	NS	Park	NS	NS
Desa Delima Avenue	Elite,NKVE,LDP	NS	NS	NS	Yes (2)
Taman Suria Enigma	PLUS, Guthrie Link, LDP	Buses & Taxis	NS	4	Yes (2)
Impiana Cheras Height	KESAS	LRT	NS	NS	Yes (3)
Medan Idaman	MRR2	NS	Waterfalls, Forest	3	Yes (1)
Teratai Mewah Apartment	PLUS, NKVE, Shah Alam Klang Expressway	KTM	Tropicana GCC	6	Yes (1)
Paya Jaras Avenue	Federal Highway, NKVE, Guthrie Corridor	NS	Bukit Cahaya Park	5	Yes (3)
Semarak Api Apartment	SILK Highway, PLUS	NS	Saujana Impian GCC	3	NS
Palma Indah Terrace	MRR2, Elevated KLCC Highway	Taxis	NS	NS	Yes (3)
Zihan Avenue	LDP, SILK Highway, PLUS	ERL	IOI Golf Club	3	Yes (1)
Taman Putra Perdana	KESAS, NS Central Link, LDP	Buses & Taxis	NS	4	Yes (2)
Merak Apartment	LDP, PLUS, Besraya, Elite	All	Technology Park	3	Yes (1)
Valencia Height	NKVE	Taxis	Mahkamah	3	Yes (1)
Prima Perdana I	PLUS	NS	KLIA	4	Yes (2)
Taman Mewah	LDP	NS	Kirrara Golf	5	Yes (2)

House Pricelist (pg 5)

Projects Name	Future Development (radius 1 km)		
	Amenities	Transportation	Commercial
Magna Ville Condominium	UiTM (Medic)	NS	NS
Acacia Avenue	NS	NS	NS
Putra Avenue	NS	Putra Height Interchange	NS
Taman Desa Mas	NS	NS	Shoplots
Impian Meridian	NS	LRT Route	NS
Amansiara	NS	NS	NS
Parklane Heights	NS	Latar Highway	NS
Jelutong Heights	Post offices	NS	NS
Taman Sutera	NS	NS	NS
Data Suria	NS	NS	NS
Unipark Condominium	International Convention Center	NS	Shop Offices
Taman Putra Prima	NS	South KV Expressway	NS
Cahaya Permai Apartment	NS	NS	NS
Changkat View Condominium	NS	Duke Expressway	NS
Timur Enstek	NS	NS	NS
Taman Tasik Prima	NS	NS	NS
Andari Townvilla	Clubhouse	NS	NS
Bayu Permai Acadia	NS	Selayang-Rawang Expressway	NS
Ansara Impian Condominium	IPTA	NS	Hypermarkets
Brickmall Avenue	NS	NS	Hypermarkets
Idaman Mahligai Avenue	NS	Putra Height Interchange	NS
Taman Desa Perak	NS	NS	Shoplots
Impian Meridian II	NS	LRT Route	NS
Damansiara Height	NS	NS	NS
Melati Impian Heights	NS	Latar Highway	NS
Jelutong Perdana	NS	NS	NS
Taman Bayu Senja	IPTA	NS	NS
Data Palma	NS	NS	NS
Unipark Apartment	Convention Center	NS	Shop Offices
Taman Putra Perdana	NS	South KV Expressway	NS
Cahaya Kasih Apartment	NS	NS	NS
Excelview Condominium	NS	Duke Expressway	NS
Timur Enstek II	NS	NS	NS
Taman Emas Prima	NS	NS	NS
Amandari Townvilla	Clubhouse	NS	NS
Bayu Permai Height	NS	Selayang-Rawang Expressway	NS
Melati Impian Condominium	IPTA	NS	Hypermarkets
Dayu Suria Condominum	NS	NS	Hypermarkets
Desa Delima Avenue	NS	Putra Height Interchange	NS
Taman Suria Enigma	NS	NS	Shoplots
Impiana Cheras Height	NS	LRT Route	NS
Medan Idaman	NS	NS	NS
Teratai Mewah Apartment	NS	Highway	NS
Paya Jaras Avenue	Schools & Colleges	NS	NS
Semarak Api Apartment	Clinics	NS	NS
Palma Indah Terrace	NS	NS	NS
Zihan Avenue	International Convention Center	NS	Shop Offices
Taman Putra Perdana	NS	South KV Expressway	NS
Merak Apartment	NS	NS	NS
Valencia Height	NS	Duke Expressway	NS
Prima Perdana I	NS	NS	NS
Taman Mewah	NS	NS	NS