2017

A Prototype for predicting real estate investment performance in Kenya

Moses Kimani Kihumba
Faculty of Information Technology (FIT)
Strathmore University

A Prototype for Predicting Real Estate Investment Performance in Kenya

Kihumba, Moses Kimani

Master of Science in Computer Based Information Systems
2017

A Prototype for Predicting Real Estate Investment Performance in Kenya

Kihumba, Moses Kimani

Submitted in partial fulfillment of the requirement of the Degree of Master of Science in Computer Based Information Systems at Strathmore University

Faculty of Information Technology
Strathmore University
Nairobi, Kenya

June, 2017

Declaration

I declare that this work has not been previously submitted and approved for the award of a degree by this or any other University. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made in the thesis itself.

Kihumba, Moses Kimani

....
June 2017

Approval
The dissertation of Kihumba, Moses Kimani was reviewed and approved by the following:

Dr. Vincent Omwenga,
Director Research and Postgraduate Programmes
Faculty of Information Technology,
Strathmore University.

Dr Joseph Orero

Dean, Faculty of Information Technology,

Strathmore University

Prof Ruth Kiraka

Dean, School of Graduate Studies,

Strathmore University

Abstract

Predicting investment performance is central to attracting investors in any property or business venture. Investors are keen to predict the future in order to protect their investments and choose assets with the best returns. All asset returns are predictable to some extent, with returns on real estate relatively easier to forecast due to the nature of assets. Forecasting is thus an important component in property investment decision-making. Currently, majority of investors in Kenyan real estate sector, rely on speculation and sales comparison to make investment decisions. Multiple regression models have been applied successfully in forecasting real estate investments in other markets. They incorporate socio economic variables, housing and proximity characteristics to estimate the value of real estate assets. The researcher applied a multiple regression model for predicting house prices by setting house price as the dependent variable (Y) while holding the Gross Domestic Product, income of households, cost of land and housing units developed as the predictor variables (X). This predicted house prices (Y) on the basis of the X variables and determined the influence of the variables on the price. Agile development methodology was applied in the development a web application that integrated the forecasting model, an analytical backend helps to present the forecasts to investors in terms of figures, charts, and graphs that are easy to interpret and compare. Various tests were also performed on the prototype including integration and system tests. User acceptance testing was also carried out where majority of the respondents found the interface easy to use, and indicated that the application met its stated objectives as outlined in the usability questionnaire.

Table of Contents

List of Figures

List of Tables

List of Equations

List of Abbreviations/Acronyms

ANN	-	Artificial neural Networks
ARFIMA	-	Autoregressive Fractionally Integrated Moving Average
ARIMA	-	Auto Regressive Integrated Moving Averages
APT	-	Arbitrage Pricing Theory
CAPM	-	Capital Asset Pricing Model
DCF	-	Discounted Cash Flows
EMH	-	Efficient market Hypothesis
GDP	-	Gross Domestic Product
GIS	-	Geographic Information System
MAD	-	Mean Absolute Deviation
MAE	-	Mean Absolute Error
KSHS	-	Kenya Shillings
RWA	-	Risk Weighted Assets
SARIMA	-	Seasonal Auto Regressive Integrated Moving Averages
SSE	-	Sum of Square Errors

Definition of Terms

Social Economic	-	Field of study that examines social and economic factors to understand how their combination influences something (Allysa, 2016).
Asset	-	An asset is something that is owned that has value (Brooks and Tsolacos, 2010).
Liquidity	-	Liquidity refers to the ability of assets that are readily convertible to cash (Brooks and Tsolacos, 2010).
HPI	-	Housing Price Index is a measure of movement of house prices, measuring average price changes in housing sector (Lynn & Wang, 2011).

Acknowledgements

I would like to acknowledge my supervisor Dr. Vincent Omwenga for his guidance, patience, and constant advice during the course of this research. My sincere gratitude to the Faculty of Information Technology at Strathmore University for the dedication shown in imparting knowledge and skills needed to succeed in this field.

Dedication

This research is dedicated to my family, especially my parents, and siblings who have offered me encouragement and support every step of the way. Special dedication to my son Aiden my constant source of motivation. Not to forget my friends and classmates who have always given a listening ear, thank you for walking this journey with me. All honor to the Almighty God for His sufficient grace.

Chapter 1: Introduction

1.1 Background of the Study

Forecasting real estate is a fundamental part of strategic decision making in the real estate sector. The ability to accurately predict the future outcomes of complex systems has been the goal of forecasters for decades. Larsen (2010) affirms that forecasting the performance of investments is a widely studied topic in many fields including trade, finance, and statistics. The motivation for which is naturally to predict the direction of future prices, such that assets can be bought and sold at profitable positions within defined time frames.

Krystalogianni, Matysiak, & Tsolacos (2014) state that forecasting total returns and its components, income returns and capital values, across commercial sectors is important in making investment allocation decisions, it is essential for investors and portfolio managers to forecast commercial investment return as it provides a prediction of the expected target return, which consequently assists in making accurate investment decisions. Most traditional models of investment prediction use statistical models or neural network models derived from price data. In addition to the conventional quantitative analysis, real estate analysts are increasingly monitoring several economic series in search of early signals for property market activity. This becomes more important at times of uncertainty arising from unsettling economic conditions and volatility in the wider investment markets.

Currently most real estate investors and developers in Kenya, rely on traditional approaches to make investment decisions. These approaches are mainly based on sales comparison and speculative assumptions. Taffese (2006) states, "An accurate and fast prediction of the real estate value is important to prospective homeowners, developers, investors, appraisers, tax assessors and other real estate market participants, such as, mortgage lenders and insurers. Real estate valuation based on traditional approaches such as cost and sale comparison approach lacks an accepted standard and a certification process."

According to Ma, Chen, and Zhang (2015), various statistical, data-mining, and machine-learning prediction algorithms can be used to forecast the prices of real estate assets. Basic algorithms such as linear regression have been applied using both intrinsic features of the real estate properties (living area, number of rooms, etc.), and additional geographical features such as socio demographical features (average income, population density, etc.). New machine learning algorithms have also been applied in predicting the temporal patterns of housing prices in different studies. In this study we predicted the price of real estate assets using multiple regression methods mainly stepwise regression and best subset regression.

Marwani (2014) supports the need for a real estate prediction system in his research on modelling and forecasting real estate, he observes that the availability of data makes it easier for real estate players to improve on their predictions and make more accurate investment decisions. He proposes the inclusion of Geographic Information Systems (GIS) data, in tradition models, to capture important valuation aspects of real estate such as location, transportation, and infrastructure development. The ability to predict can fill up the information gap in the real estate sector and improve efficiency of the real estate market players in Kenya.

1.2 Problem Statement

Purchasing a property is the largest single economic transaction made by Kenyans in their lifetime, and the decision should be based on proper analysis and forecasting to make an informed decision. Property investors would like to have a picture of what the future looks like in terms of cash flows and expected returns. Property prices in Kenya have been on an abnormal increase over the last decade, which has raised concerns that the market is unstable, and caused speculation that the "bubble" is about to burst. The result of such incredible asset price increases is severe financial crises, similar to what was witnessed in the Asian, global economic crunch and European debt crisis (Obere & Miregi, 2014; Muli, 2013).

There is a clear lack of studies on the factors that influence commercial real estate prices within Nairobi, and even fewer modern information systems tools that are being applied to aid investors in decision making. There is potential to utilize geographic information systems (GIS) and socio–economic factors to model causal relationships that can be used to forecast real estate prices (Marwani, 2014).

1.3 Objectives

The purpose of this study is to develop a model for predicting investment performance for real estate assets in Kenya to aid investors in making more informed investment decisions.

1.4 Specific Objectives

i. To analyze variables for predicting the performance of real estate investments in Kenya.
ii. To examine existing algorithms and models that can be applied in prediction of performance for real estate investments.
iii. To develop an application that can integrate the prediction model, geographic information systems data, and socio economic data to predict real estate investments.
iv. To carry out testing of the developed prototype.

1.5 Research Questions

i. What are the variables for predicting real estate investment performance in Kenya?
ii. What algorithms and methods have been applied in predicting the performance of real estate investments?
iii. How can the web application be developed to predict real estate investments?
iv. How can the developed prototype be tested?

1.6 Justification

The real estate sector in Kenya is a main contributor to economic development, and one of the most profitable. Real estate projects are characterized by high risk with many new investors susceptible to fraudulent land brokers, undiversified investment vehicles and limited ability to analyze the investments (Kibiru, Pokhariyal, & Obwocha, 2014). Real estate projects require heavy investments and as such detailed analysis should always precede commitment to these investments.

The proposed prediction application will enable project owners, developers and real estate entrepreneurs to demonstrate to investors that they are profit driven and that they will maximize returns for their financiers. This will give investors' confidence that they are investing in potentially profitable ventures. It will also aid Kenyans in making more informed real estate decisions that can help protect their hard earned cash. In addition this study will form a basis for future research to develop a better understanding of factors influencing real estate prices in Kenya.

1.7 Scope of the study

This study will be restricted to predicting performance of investments made in the real estate sector in Kenya. The focus of the study is to integrate a forecasting algorithm into a web technology to forecast real estate performance within the short to medium terms. The study limits itself to a multiple regression model to provide evidence that forecasting in the real estate sector can provide credible data to be used by the web application. This will guide investors to channel funds to attractive real estate projects, and help attract more funders to real estate projects. Data collection and prototyping will be conducted mainly in Nairobi and its outskirts.

Chapter 2: Literature Review

2.1 Introduction

This chapter reviews relevant literature and previous studies on predicting investment performance in Real Estate. It starts with a background on forecasting and modeling investments; it then attempts to define the commercial real estate categories, and the major theories that have been advanced in predicting the performance of investments in the real estate sector. It also reviews variables that affect prices of real estate assets, and models that have been applied in predicting those prices. It will conclude by highlighting any gaps and conclusions drawn by different authors, and summarize on the specific gap that this research study hopes to fill.

2.2 Prediction Methods

A prediction, or forecast, is a statement about an uncertain event. There is no universal agreement about the exact difference between the two terms and as such different authors have used the terms interchangeably in their research. According to Chambers, Mullick, and Smith (2001) there are three basic types of forecasting namely; qualitative techniques, time series analysis, and causal models. Qualitative techniques apply qualitative data such as the opinion of experts in predicting the future. They may or may not take the past into consideration. Time series analysis relies heavily on historical data to extrapolate variables into future circumstances. In time series analysis patterns and their changes are analyzed to inform what will happen in the future. Causal models rely on highly defined relationships established from analysis of trends to determine how different variables are correlated. Lizieri (2011) breaks down forecasting methods as in figure 2.1 below.

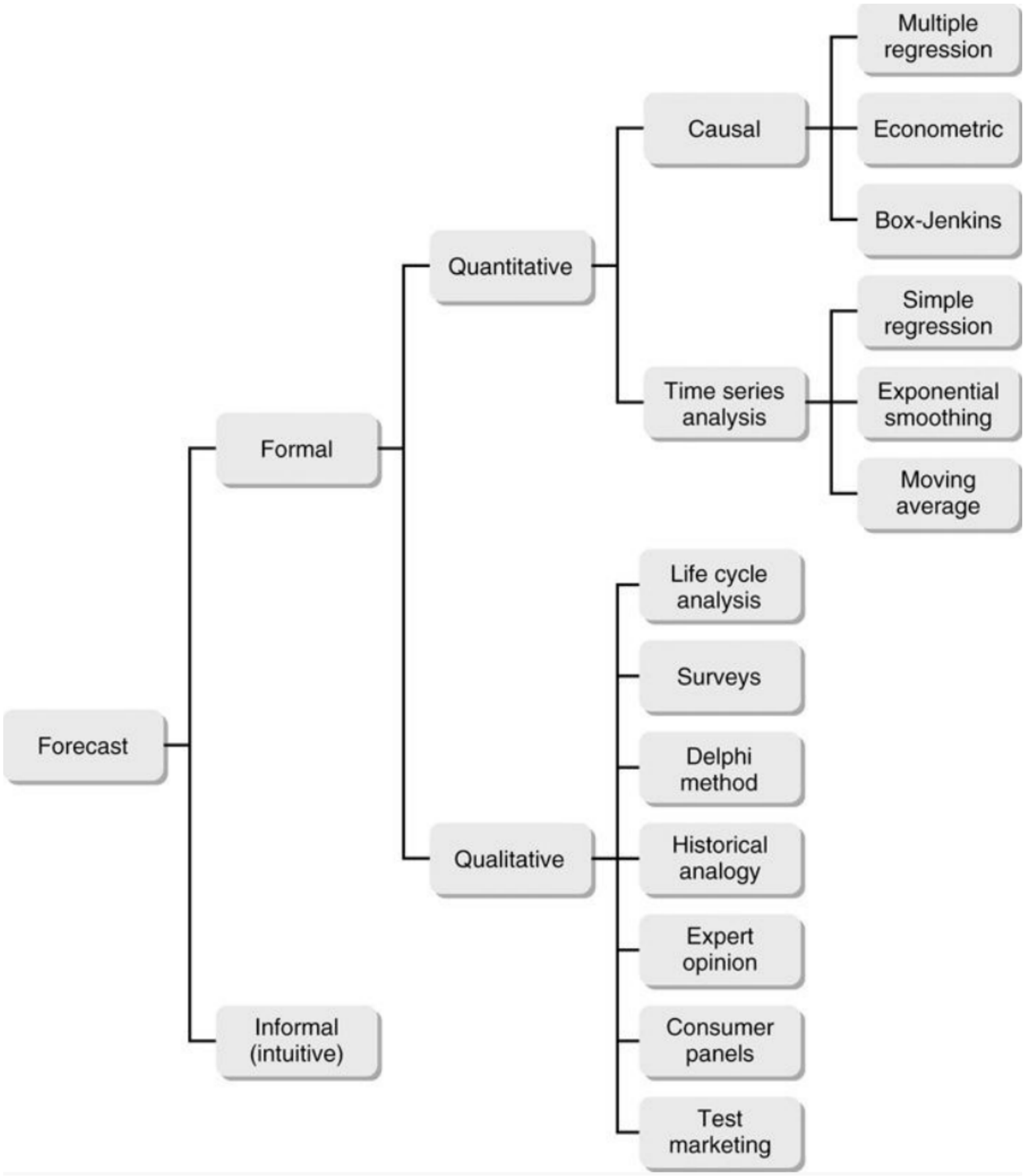


Figure 2.1 Forecasting Methods (Lizieri, 2011)

In selection of a forecasting model researchers have to determine several factors such as the variables in consideration, the accuracy of forecast and the quality and availability of historical data to be used, this influences how accurate the forecast will be (Chambers, Mullick, & Smith, 2001). There are two commonly used forecasting methods are discussed below.

2.2.1 Qualitative techniques

According to Stevenson (2011), quantitative forecasting techniques rely on personal judgment, opinions or personal experiences. They are often subjective and not based on any quantitative analysis. They permit inclusion of soft information such as human factors, personal opinions, hunches in the generating forecasts. When using quantitative techniques we so not consider human factors, as they cannot be quantified. The most common qualitative methods are identified in figure 2.2 below.

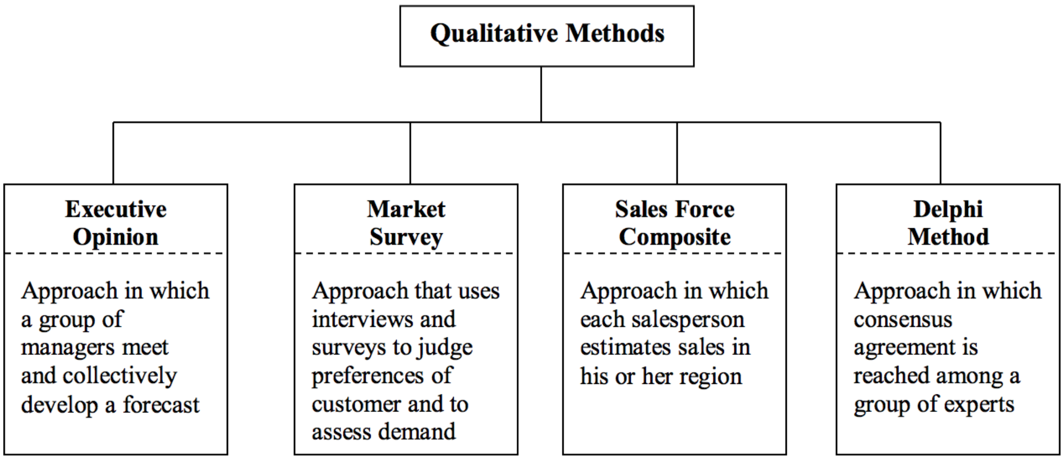


Figure 2.2 Qualitative Methods (Stevenson, 2011)

2.2.2 Quantitative techniques

Quantitative techniques mainly rely on statistical analysis by analyzing hard data. They tend to avoid personal biases unlike qualitative methods. They are seen to be more objective than qualitative methods. They can be classified into two main models, namely time series models & associative models as illustrated in figure 2.3 below (Stevenson, 2011).

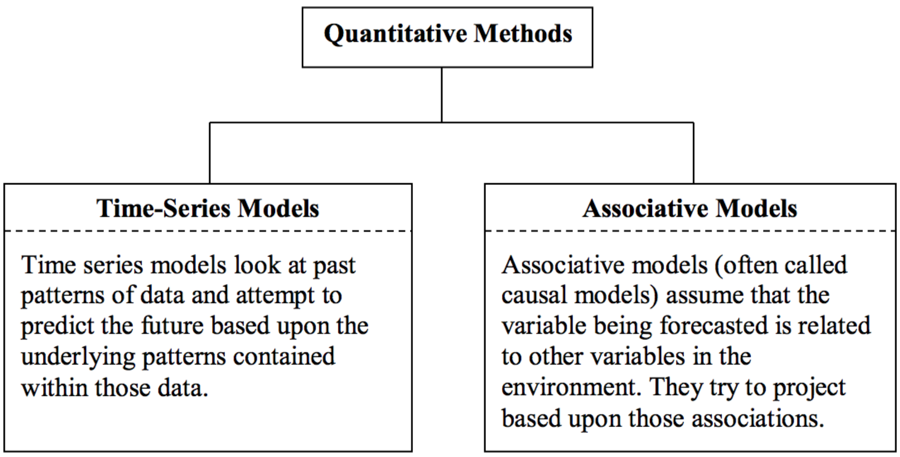


Figure 2.3 Quantitative Methods of Forecasting (Stevenson, 2011)

2.3 Forecasting and Modeling Investments - Investment Theory

Butler & Kazakov (2010), argue "one of the most interesting questions in investment theory is how to build robust models for making accurate and reliable predictions of the stock market and related financial assets." There are two main forms of analysis that are proposed in investment theory; these are fundamental analysis and technical analysis.

Fundamental analysis is concerned with a company's financial statements such as the balance sheet, income statement, and statement of cash flows. It also looks at stability in the organization, senior management, succession and how competitive an organization is within the industry. Technical analysis on the other hand focuses on historical price to inform investment decisions. Technical analysis is concerned with market patterns that can be observed and modelled. Butler further observes that investment theory concerns a decision making process where one intends to make investments to satisfy certain goals with regards to risk and return (Butler & Kazakov, 2010)

Ross (2010) emphasizes that the fundamental principle of finance is the trade-off between risk and return. This is seen as the basis for the commonly used statement, the higher the risk the higher the return. Portfolios with risky assets outperform the ones with less risky assets. A few theories that have been proposed to explain this trade-off between risk and return include: the capital asset pricing model (CAPM) and the portfolio theory and the efficient market hypothesis (EMH) and the arbitrage pricing theory (APT).

2.3.1 Capital Asset Pricing Model

Mukherjee and Metia (2001) in their discussions on Capital asset pricing models (CAPM) observe that they evolved out of the consumer's choice problem when they are faced with uncertainty. They assert that this model forms a major subject of research in finance theory. The model describes how the price of a claim to a future payoff is determined in the securities market. The capital asset pricing model (CAPM) describes the expected rate of return of financial assets like stocks, bonds, futures, options and other securities.

Baker (2001) concludes "To increase his wealth, an investor needs to buy property that has greater value to the investor than its purchase cost. To achieve this by design rather than by chance, the investor needs a formal theory of value to indicate to him what a property is worth."

2.3.2 Portfolio Theory

. Baker (2001) states that portfolio theory, developed from the notion that diversifying a portfolio reduces the risk of the portfolio. This is to mean that a portfolio with a mix of risky and risk free assets will have low risk due to diversification. He stipulated that completely random factors could determine the investment returns of an individual stock or investment, including the relationship it has with the market index.

To arrive at the CAPM, one needs to examine how imperfect correlation among asset returns affects the investor's tradeoff between risk and return. While risks combine nonlinearly because of the diversification effect, expected returns combine linearly. That is, the expected return on a portfolio of investments is just the weighted average of the expected returns of the underlying assets. Diversification thus leads to a reduction in risk without any sacrifice in expected return (Baker, 2001).

2.3.3 Arbitrage Pricing Theory

According to Ross (As Cited by Huberman and Wang, 2005), "The APT is a substitute for the Capital Asset Pricing Model (CAPM) in that both assert a linear relation between assets' expected returns and their covariance with other random variables." This means that the expected return of assets is linearly correlated to macro-economic factors such as the price indexes in that market.

2.3.4 Efficient Market Hypothesis

The efficient markets hypothesis (EMH), states that the price of assets in a market are a reflection of all available market information and as such investors cannot find opportunities to exploit information arbitrage in such a market. It tries to answer the question of why price changes in markets and how investors can exploit the information gaps to make abnormal returns for their portfolios.

Levišauskait (2010) concludes, "market efficiency means that the price which an investor is paying for a financial asset (stock, bond, other security), fully reflects fair or true information about the intrinsic value of this specific asset or fairly describe the value of the company – the issuer of this security."

2.4 Real estate types/ categories

According to Brueggeman and Fisher (2011), the term real in real estate comes from the term realty, which has, for centuries, meant land and all things permanently attached to the land, this may be buildings and anything else set up on the land Koenigsberg (2016) further breaks down real estate into multiple types. These are office, industrial, retail, residential, hotels & land.

2.4.1 Office

These include office buildings, mid-rise office blocks and converted residential offices. They range between of 20,000 – 400,000 square (Koenigsberg, 2016).

2.4.2 Industrial

These types of properties are heavily customized with machinery for the specific tenant such as industrial processing plants. Most large and small manufacturers' plants would also fall under this category (Koenigsberg, 2016).

2.4.3 Retail

These include community retail centers normally occupied by multiple tenants and anchoring businesses such as supermarkets and restaurants. Malls fall in this category. Malls range from 400,000 – 2 million squarefeet . (Koenigsberg, 2016).

2.4.4 Residential

These include apartments, which are properties around 3-9 stories, with between 20-100 units, often constructed in urban locations. It also includes high-rise properties found in larger settings, usually hosting 100+ units, which are professionally managed (Koenigsberg, 2016).

2.4.5 Hotels

These include full service hotels to guest houses and lodges. The category also holds boutique properties and extended stay hotels (Koenigsberg, 2016).

2.4.6 Land

This refers to undeveloped land such as agricultural land and pasture, developed land that is vacant and all parcels of land that was previously used for industrial or commercial properties in the past but are now available for re-use (Koenigsberg, 2016)

2.5 Real Estate Investments (Basic Theories)

2.5.1 Appraising commercial real estate

According to Schulz (2004), appraising commercial real estate involves estimating the value of the property. People appraise real estate for varied reasons including; buying and selling property, investment reasons, performance evaluations, loan applications to ascertain collateral value, and tax purposes. According to the RICS Valuation and Professional standards (2012), the market value of property is defined as "the estimated amount for which an asset or liability should exchange on the valuation date between a willing buyer and a willing seller in an arm's length transaction after proper marketing and where the parties had each acted knowledgeably, prudently and without compulsion."

A study by Koubkov´a (2015) suggests that there are three main valuation techniques, which are generally accepted as valuation standards. These are cost based approach, sales comparison approach and income approach. The cost based approach is mainly used in cases where data for comparison is lacking. This is mainly the case with construction of special use facilities such as museums and hospitals, which are rarely sold on open markets. The limitations of this approach are difficult determination of adjustments for depreciation and an absence of comparable land sales.

Koubkov´a (2015) additionally explains that the idea behind the sales comparison approach is that the informed buyer would never pay more than it has been recently paid for similar properties. The approach assumed that value of appraised property is closely related to the selling prices of comparable properties within the same market area. The value is therefore derived from recent sales data from comparable transactions that are subsequently adjusted to match the characteristics of appraised property. These adjustments should reflect the differences in size, scale, location, age, and quality of construction, selling date or in surrounding neighborhood. It should be taken into consideration that the more differences that must be adjusted for, the more different are the compared properties and the less reliable is the sales comparison approach (Brueggeman and Fisher, 2011).

The third technique for valuing real estate is the income approach. The income approach is based on the future flow of income relating to a particular property. This could be based on rental earnings, or capital appreciation (Schulz, 2004). This valuation approach is commonly known as the present value or discounted cash flow (DCF) approach.

2.5.1.1 Discounted cash flow analysis

Discounted cash flow analysis calculates the Net Present Value (NPV) of all future cash flows. In real estate valuation this includes expected rental earnings of a particular asset.

2.6 Predicting Performance for Real Estate

Ma, Chen, and Zhang (2015) in their study on prediction of real estate indexes based on neural networks establish that there are several methods of predicting the performance of real estate assets. They range from traditional forecasting methods such as time series models and Markovian models to modern day techniques such as neural networks and expert systems. Different methods produce different levels of accuracy over the short and long run depending on data availability and regularity.

2.6.1 Artificial Neural Networks

Zhao (2009), states that neural networks are mathematical models that were inspired by the structure of biological networks. Stergiou & Siganos (2012), add that they are inspired by the way biological nervous systems process information. They consist of a large number of highly interconnected processing elements working in unison to solve specific problems. Park (2011) identifies the most common type of artificial neural networks to consist of three layers. These are input units that are connected to a layer of hidden units, which are connected to a layer of output units. Figure 2.4 below shows the three layers of an artificial neural network.

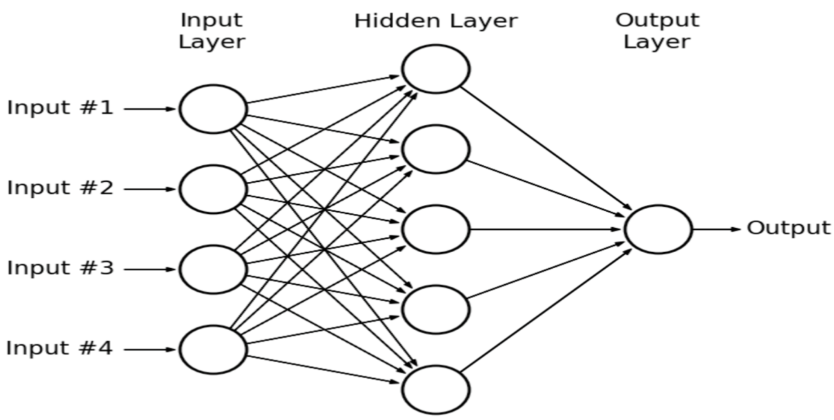


Figure 2.4 Neural Network Diagram (Park, 2011)

ANNs, like people, learn by example. They are configured for specific applications, such as pattern recognition or data classification, through a learning process. ANNs are also inherently non-linear, which makes them more practical and accurate in modeling complex data patterns, as opposed to various traditional linear approaches, such as autoregressive integrated moving average (ARIMA) methods. In many forecasting scenarios ANNs produced better analysis in the short run that time series models (Stergiou & Siganos, 2012). ANNs have been applied in many sectors to produce forecast, there has been a widespread application in solving business related problems and even weather prediction (Zhao (2009).

To evaluate the performance of ANNs, a lot empirical research has been undertaken comparing them with other traditional models. The performance measure being the accuracy of the forecasts produced. The results have of the comparisons been mixed and confounding.

Research by Kara, Boyacioglu and Baykan (2011), supports the view that ANNs show an improvement in forecasting accuracy in the short term, Zhao (2009) agrees with this view however point out that, "It is considerably harder to obtain substantial improvements in extrapolative forecasting with neural networks than might be assumed." He presents arguments to emphasize that ANNs are more suited to nonlinear and discontinuous series that are harder to forecast with time series techniques. However there is not enough evidence to show that they outperform them in the long term.

2.6.2 Time Series Forecasting Techniques for Property

A time series is a collation of observations of well-structured data, which has been obtained through repeated measurements over time (Marwani, 2014).

2.6.2.1 Mean Model and the Random Walk Model

Nau (2014) indicates that the simplest forms of time series models are the mean models and the random walk models. The mean models are based on the assumption that what happens next is an average of what has been happening until now. Random walk models are based on the assumptions that to predict what happens next we only need to look at what is happening now and ignore the past. The two models suggest that there is no significant change between one period and the next, and thus it is easy to tell what will happen if we know what has been happening.

Equation 2.1 Random Walk Function

$$\hat{Y} = Y_{t-1} + u_t \tag{2.1}$$

In the above equation u_t is the error term or white noise. In the simplest random walk process, future value of time series is given by its immediate previous value.

2.6.2.2 Simple Moving Average

Building on the mean model and simple Nau (2014) suggests that moving averages take advantage of what has happened in some window of the recent past. They use the average of the most recent data values in the time series as the forecast for the next period. Brooks and Tsolacos, (2010) state that moving averages represent the simplest form of time series models that a forecaster should consider. They present the function as follows. Let u_t (t = 1, 2, 3...) be a white noise process with E (u_t) = 0 and var (u_t) = σ^2. Then

Equation 2.2 Moving Average Function

$$y_t = \mu + u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \cdots + \theta_q u_{t-q} \tag{2.2}$$

Moving average models can be viewed as the combination of white noise processes, so that y_t depends on the current and previous values of white noise disturbance term.

2.6.2.3 Autoregressive Integrated Moving Average (ARIMA) models

Kumar, (2016) in his study of time series modeling and forecasting using stochastic models indicates that that they can take many forms. He points out two of the most popular models as Autoregressive (AR) and Moving Average (MA) models. He further asserts that a combination of these two has been seen to yield more accurate results that either model operating on its own. He presents two combinations of these models as yeilding better results in forecasting real estate properties. These are the Autoregressive Moving Average (ARMA) and the Autoregressive Integrated Moving Average (ARIMA). For seasonal time series forecasting, he proposes a variation of ARIMA, identified as the Seasonal Autoregressive Integrated Moving Average (SARIMA) model as more appropriate. The ARIMA models and the diffent variations of it are commonly reffered to as Box-Jenkins models. In these models the future value of a variable is assumed to be a linear function of several past observations and random errors. The ARIMA function takes the form of;

Equation 2.3 Autoregressive Integrated Moving Average Function

$$y_t = \theta_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots \phi_p y_{t-p}$$
$$+\mathcal{E}_t - \theta_1 \mathcal{E}_{t-1} - \theta_2 \mathcal{E}_{t-2} - \cdots - \theta_q \mathcal{E}_{t-q}, \tag{2.3}$$

Where y_t and $\mathcal{E}_t$ are the actual value and random error at time period t, respectively; θ_i (i = 1, 2,…,p) and θ_j (j = 0 1; 2,…,q) are model parameters. p and q are integers and often referred to as orders of the model. Random errors, $\mathcal{E}_t$, are assumed to be independently and identically distributed with a mean of zero and a constant variance of σ^2.

Vishwakarma (2013) in his study of the real estate market in Canada concludes that it is very difficult to produce accurate long-term forecasts using the Autoregressive Integrated Moving Average (ARIMA) family models. He noted that the models worked well in short term forecasting, almost at the same level with ANNs. He tested variations of the ARIMA model and presents the following findings. The ARIMA has constant mean and constant variance, the ARIMAX model includes exogenous variables to explain the data-generating process. The ARIMAX-GARCH model has all the features of the ARIMAX model, plus time-varying variance. Specifically, out of the three models tested, the ARIMA, ARIMAX and ARIMAX-GARCH models, the ARIMAX model works best in all circumstances by giving minimum mean absolute percentage error. Other models work best in one circumstance but fail drastically in other circumstances.

2.6.3 Regression Analysis

Brooks and Tsolacos, (2010) explain regression models as those that evaluate and describe the relationships between different variables and especially trying to explain how movement in the variables affect each other. These models have proved to be a reliable approach in forecasting real estate prices because of their ability to establish causal relationships between different fundamental variables. In the models the dependent variable is always the real estate price or rental income while the independent variables are any external factors that influence change in the dependent variable. For example income, inflation, GDP, location, infrastructure development etc. can all influence real estate prices.

Under regression models we have the simple and multiple regression models. Multiple regression models have been applied in many studies to assess the movement of real estate investments by setting the price as a function of different property characteristics (Marwani, 2014).

2.6.3.1 Simple Regression

Campbell & Campbell (2008) point out that in its simplest form, regression establishes the nature of a relationship between an Independent variable (X) and a dependent variable (Y), this can be represented as in the function below:

Equation 2.4 Simple Regression Function

$$Y = \beta_0 + \beta_1 X + u \qquad (2.4)$$

In the equation above (β_0) represents the status of the dependent variable, when the independent variable is absent the relationship can be defined by the direction of the slope represented by the parameter (β_1). An error term (u) captures the amount of variation not predicted by the slope and intercept terms. The regression coefficient shows how well the values fit the data.

2.6.4 Other Models

Ashram (2015) discusses other methods for forecasting real estate; these include the non-linear regression, which does not assume a linear relationship between variables-frequently used when time is the independent variable. Trend Analysis, which uses linear and nonlinear regression with time as the explanatory variable-used where pattern are observed over time. Birch and Suderman (2003) also introduced a two-way exponential smoothing system to estimate true movements in residential property prices.

Marwani (2014) discusses investments in real estate in his research. He suggests that the investments are of a heterogeneous nature and that market factors play an important role in explaining the variations in the prices. To generate forecasts one needs to consider many independent factors such as the type of tenant, age of property, location, socioeconomic parameters and other industry factors.

2.6.5 Comparison of forecasting algorithms

This section provides a brief review of the forecasting algorithms and techniques discussed like time series techniques and ANN's. It compares them on the basis of other studies that have applied the models and makes no attempt to be exhaustive or conclusive as to what algorithms are superior to the rest.

According to Zhao (2009), ANN's have gained enormous popularity in the recent years, especially in undertaking time series forecasting. Most applications, however, are in areas where data are abundant as NN are very data intensive. In real estate forecasting, due to the scarcity of large data samples, there exists only a few studies involving the use of NN that can be used to gauge their usefulness in the field. The results in these studies, based on out-of-sample forecasts, do not permit a demarcation between the linear models and NN as the latter is able to justify its theoretical superiority in only some of the cases. A considerable amount of research has been carried out in the recent years on NN. Vishwakarma (2013), states that despite their ability to capture nonlinear relationships, findings generally do not allow any discrimination between conventional linear statistical techniques and NN. One of the main reasons for this is that there are no well-defined guidelines to build NN for solving a particular task and their construction involves a lot of subjectivity on the part of researchers, thereby considerably restricting the power of NN and ultimately leading to the results of many previous studies being dubious.

The advantage of VAR models over ARIMA models is that they can incorporate more information in terms of other time series instead of just past observations and errors of the series to be forecast. In the absence of co-integration between the variables a common forecasting procedure would be to conduct a VAR on the first differences. However, if co-integrating relationships can be established between the variables, the VAR should also include the lagged co-integrating error term. A comparison of the results from the ARIMA and VAR forecasts suggest the multivariate models provide more accurate forecasts in the real estate and macro-economic sectors There is also growing evidence that socioeconomic series contain nonlinearities but linear models such as the ARIMA and VAR models are widely used for forecasting such series, despite the inability of linear models to cope with nonlinearities. New empirical evidence suggests there is higher relative forecasting performance of linear ARIMA and VAR models and the nonlinear NN (Brooks and Tsolacos, 2010).

Nguyen & Cripps in their study compared the predictive performance of artificial neural networks (ANN) and multiple regression analysis (MRA) for single family housing sales. In comparing the two models they used different functional models specifications and evaluation criteria to try and provide explanations as to why other studies that compared the two models had varied results. Some of the issues identified were methodological problems as well as data composition. They concluded that when a moderate to large data sample size is used, the ANN performs better than the MRA; however as the ANN model functional specification becomes more complex, the training sample size must also be increased in order for the ANN to perform better than the corresponding MRA model. They further assert that if one provides sufficient data training size and appropriate ANN parameters, then ANN performs better than MRA. Otherwise, the results vary.

For the review of literature and other practical purposes, the ANN is recommended in forecasting real estate performance and especially when there is sufficient sample data set and/or when there is no theoretical basis for the data model functional form. However where such limitations exist as the case in this study then the MRA is recommended.

2.6.6 Pooling Forecasting Methods

Wilson and Okunev (2001) in their research on enhancing information use to improve predictive performance in property markets studied the combination of forecasts from different property markets around the world. They observed that different forecasting techniques may produce different outcomes when applied to the same data sets. This presented a challenge for professionals in the property market around selecting the most appropriate forecasting method. As a means of dealing with this problem, they proposed combining or pooling the forecasts from different models, this resulted in better out of sample results. Pooling forecasts from different models tends to yield better outcomes for shorter-term forecasts than for component models. Further citing Winkler (1989) they discovered a similarity between forecast combination and diversification in portfolios, diversification reduced the risk of the portfolio; similarly a combination of forecasting methods yielded better results.

The majority of property forecasts are derived by combining the estimates of formal econometric models with more qualitative market overlays, the accuracy of the forecasts produced varies from institution to institution, across markets and over time.

2.6.7 Forecasting error

According to Ahmed and Nilsson (2011), all forecasting models are subject to implicit or explicit error structures. Forecasting error is interpreted as the deviation between the predicted values and the actual or true value. There are many methodologies that are applied in statistics to detect errors and to check the degree of accuracy. These include measures to access the quality of the models used, validation techniques applied to data, and other performance measures. Some of the common ones are the Mean Absolute Deviation (MAD) and Variance. For the error analysis purposes variance is preferred since variances of independent errors are additive; however, MAD is not additive.

2.7 Architecture of Web Based Prediction Systems

According to Kuang, Liu, Ma & Wang (2013), most web-based prediction systems are developed on a Client Server model with the prediction models being called through the web based systems and accessed through different web browsers. The figure 2.5 shows basic client server architecture.

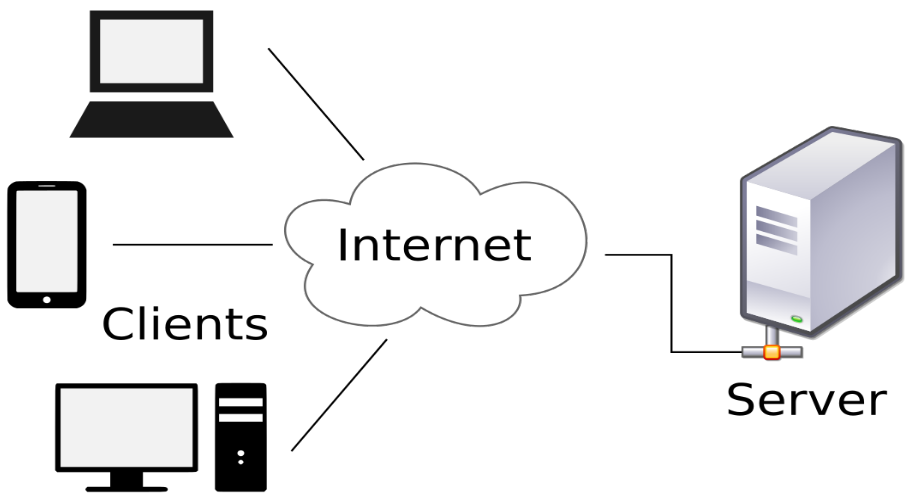


Figure 2.5: Client Server Model (Kuang, et al, 2013)

The models can be called and applied to prediction of different factors that are of importance to the predictor; prediction models can be stored in a database server such as SQL. The figure 2.6 below shows a web application deployed on a Web server and database in a separate server connected through LAN. Clients can access the Web application through browsers on laptops, desktops, mobile, etc. In this diagram, WWW is not just a component outside the client and server; it also standardizes what will flow through the wire at least with the application level protocol (in this case HTTP) between client and server.

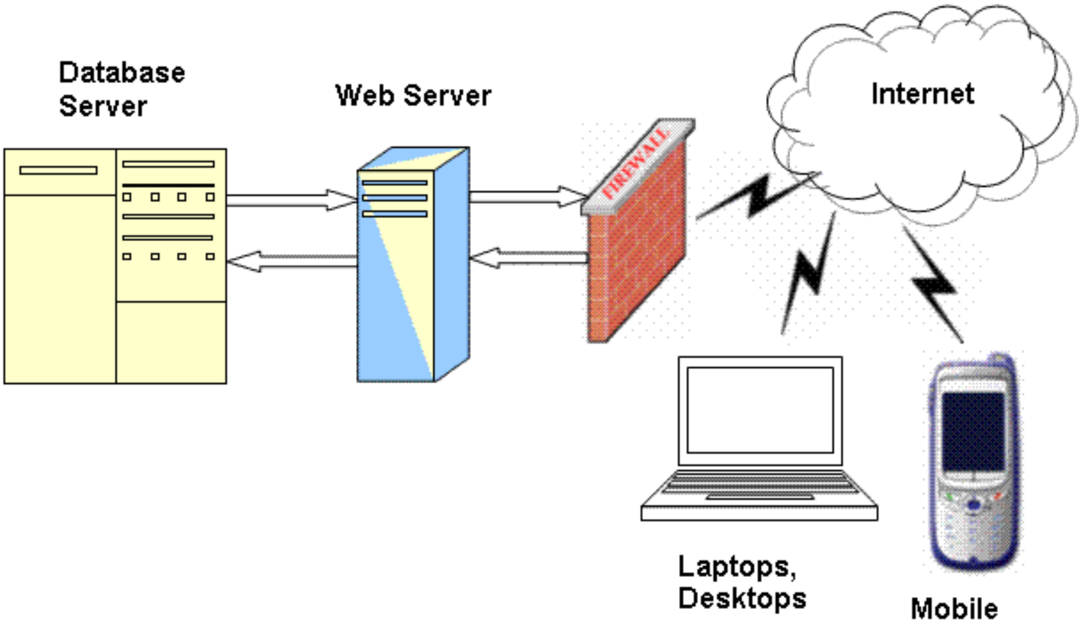


Figure 2.6 Web Application Architecture (Kuang, et al, 2013)

2.7.1 Development platforms

Some of the common development platforms for the web-based platforms are .Net, C# and JavaScript. Data is stored in SQL server, and Microsoft Visual Studio can be used as the development tool. The technology architecture of ASP.NET web-based applications can be demonstrated by separating the functionalities of the system into layers. The common layers are the data access layers the business logic layers and the Graphic User Interface layers. These layers can be demonstrated in the matrix below.

Table 2:1 Technology Matrix for ASP.NET Implementation

Layer	Functionality
GUI as Web Project (Common Layout)	ASP.NET 2.0
	Master Page for Common Layout

	Login Control
	Client Validation using validation controls
	The same images and CSS
	Data Grids for table view
BLL as Class Library Project	Checking for business roles like user privileges goes here. It also acts as bridge between GUI and the data access layer
DAL as class Library Project	SQLs are handled using SQL Data Adapter
	Datasets as mapping for DB schema
	User table as VB.NET class
	SQL Server 2016 as DBMS

2.7.2 Testing of web-based Applications

The aim of testing web-based applications is to ensure that they consistently perform when being used by multiple users concurrently. Applications that remain consistent during peak demand are seen as reliable and meet usability requirements. The tests applied are meant to ensure that the systems are able to scale and handle increased capacity during peak demand conditions. There are three main type of tests that can be used to establish the performance of web based systems they include the following

i. Load testing - Load tests exercise the code and try to find bottlenecks and bad code.
ii. Stress Testing - Stress tests try to break the system so it can be improved to fail gracefully or so that system activity can be monitored for signs of being overwhelmed.
iii. Capacity Testing - Capacity tests ensure that the web application can handle the number of users and transactions that it was designed to handle. They can also be used to help determine when the environment needs to be beefed up to meet demand.

2.8 Model Specification

Corsini (2009), in his study on statistical analysis of residential house prices observes that with most regression analysis studies that relate to real estate prices, the dependent variable is almost always the market price for a particular property, whether that is the sales price or rental price. All of the characteristics of that property (i.e. square footage, bedrooms, proximity factors, etc.) are the independent variables and are the predictors of that market price. The objective of the study was to integrate a forecasting model into a real estate prediction application. The model chosen for this study was a multiple regression model due to its ability to deal with limited data and multiple variables. Variable selection is crucial in the model as it can determine the accuracy of the forecast. In selecting the independent variables it's important to determine which characteristics are most meaningful in influencing the dependent variable and the nature of that influence. The multiple regression model selected for prediction of real estate performance will take the form of

Equation 2.5 Multiple Regression Function

$$\gamma = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots \beta_k X_k + e \qquad (2.5)$$

Where

γ = The predicted value of the dependent variable (In this study the Performance of Real Estate)

$X_1, X_2, X_3 \dots X_k$ are the independent variables selected for this study.

β_0 = The γ -intercept, or the value of γ when all of the predictor variables are 0

$\beta_0, \beta_1, \beta_2 \dots \beta_k$ are the population coefficients of the independent variables X_1, $X_2, X_3 \dots X_k$

e = The error term that is assumed to be associated with the Variable.

2.9 Proposed Model Framework

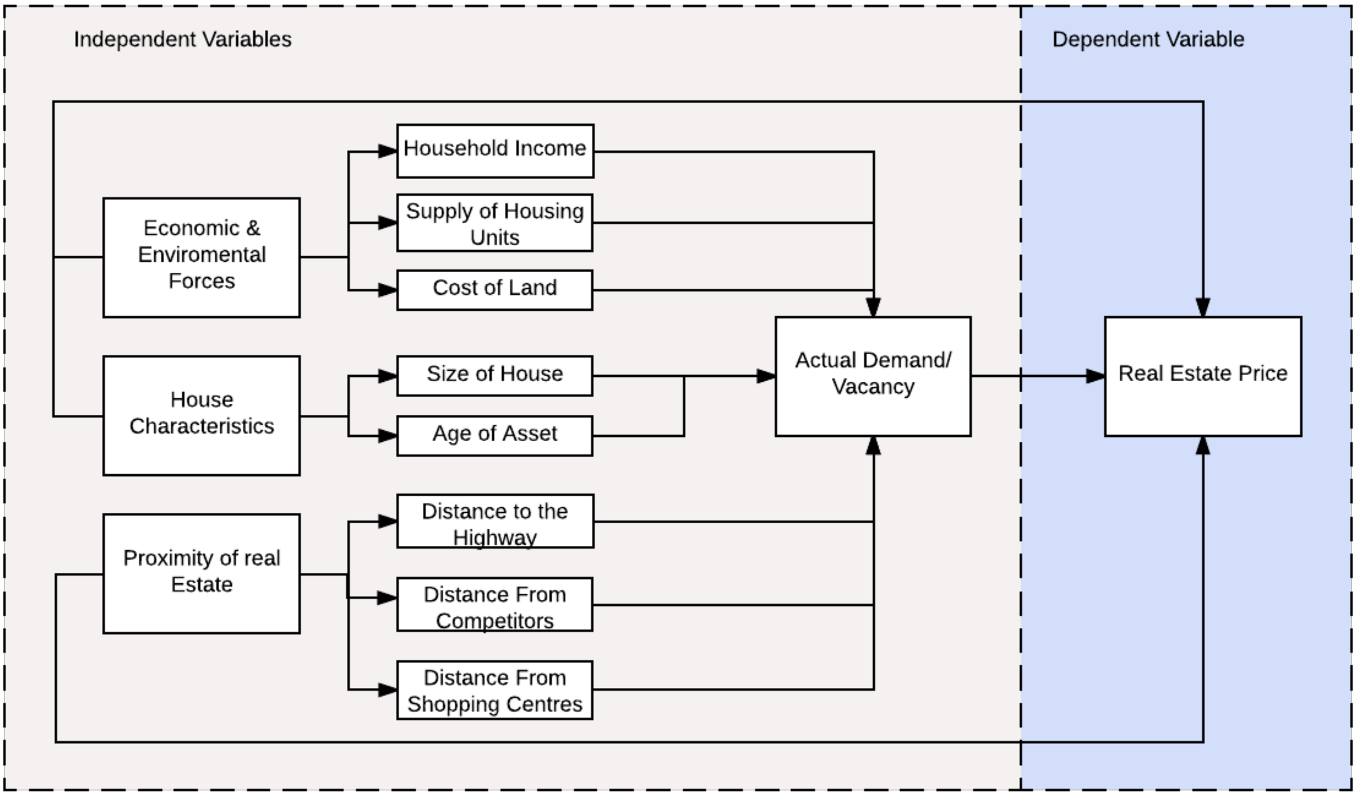


Figure 2.7 Proposed Regression Model Framework

Figure 2.7 shows the framework for the proposed regression model in this study. The model shows the variables used to develop the regression model for predicting house prices. The dependent variable is the market price for a particular property, which is the sales price or rental price. All of the characteristics of the property (i.e. square footage, proximity factors, etc.) are the independent variables and are the predictors of that market price. The real estate price is thus considered as the dependent variable while economic forces, house characteristics, and proximity factors are the independent variables. Through correlation analysis the researcher was able to identify factors that were most influential in house price movement and the correlation between different combinations of variables.

Chapter 3: Methodology

3.1 Introduction

This chapter describes the system development methodology, research design target population, research methods and the methodology used for this study. According to (Greener, 2008), research methods refer to the activities designed to produce research data e.g. interviews questionnaires, observations, and focus groups. On the other hand research methodology is about the researchers' attitude, and the strategy the researcher chooses to answer research questions.

This study employed qualitative and quantitative research design to identify the variables that influence performance of real estate investments, reviewed current performance prediction models and established requirements for the proposed model. This chapter also looks at system analysis, design and architecture of the web application to implement the designed model.

3.2 System Development Methodology

A system development methodology is a structure imposed on the development of a software product. The researcher applied agile software development methodology in this study. According to Li (2012) Agile software development is a capability that can rapidly respond to changing needs of the software, while focusing on rapid delivery of high quality software, and achieving customer satisfaction. The figure 3.1 shows the agile lifecycle as applied in this study.

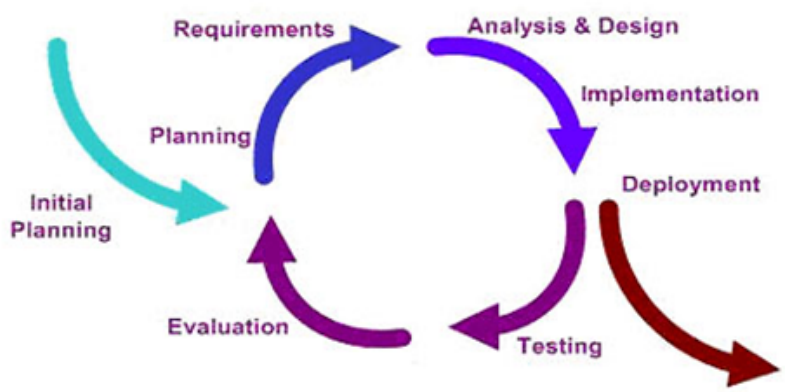


Figure 3.1 Agile Development Lifecycle

The first step in this process was planning and requirements gathering. Design and Development was then done based on the defined requirements in the previous process. After implementation the Quality assurance and testing was undertaken to improve on the product functionality and allow for feedback on any required changes. Issues identified during testing were then resolved and implementation was done.

3.3 Research Design

Kothari, (2012) states that the research design is the arrangement of conditions for collection and analysis of data in a manner that aims to combine relevance to the research purpose with economy in procedure. The research design sets out the blueprint for the collection measurement and analysis of data.

The study employed an exploratory research design in answering the research questions by examining relevant data in regards to real estate investment, and by identifying the effect the study variables had on real estate performance. The datasets in the study were then subjected to quantitative analysis. Through multiple regressions potential relationships between the variables proposed in the conceptual framework of the study were identified. The researcher then developed the architecture of the web based real estate prediction model proposed in the study and validated it. The application development lifecycle that this study used is as follows.

3.3.1 System Architecture

The system comprises of web based front end and a back end sub-system. The front end sub-system was developed using HTML and JavaScript while the backend system was based on the Asp.Net technology stack i.e. Asp.Net MVC, C# and SQL Server database.

3.3.2 Requirements Gathering

This phase involved determining the system needs and requirements. It involved fact-finding on different components of a web-based system including identifying the functional and non-functional requirements of the proposed system.

3.3.3 Analysis

During this phase the researcher answered questions about the system. The main aim was to review requirements to identify what the system would do and how users would interact with the system. Use case diagrams were developed to capture the different actors and their interaction with the systems. Denis, Wixom & Roth (2012) proposed the following steps to carry out in this phase.

i. Development of an analysis strategy to guide the project effort, in line with this the researcher reviewed the current systems and identified ways in which to design the proposed system.
ii. Requirements gathering, this was done through development of functional and non-functional requirements. Models to show the data and processes to support the data analysis were also reviewed in this step.
iii. Development of a system proposal, this captured system designs and prototyping.

3.3.4 Design

In the design phase requirements gathered during analysis phase above were used to create a blueprint of the real estate prediction system. This phase focuses on the technical system and how to satisfy the technical requirements gathered during analysis.

The researcher thus developed the data flow diagrams and case diagrams to show procedures that will be followed in interacting with the system. Denis et al (2012), state that user's interactions with the system should be considered in this phase. The system inputs and outputs are designed along with a plan or roadmap of the way the system's features will be navigated. Technical specifications and Data Flow Diagrams on how the system was built are illustrated in chapter 4 of this document.

3.3.5 Implementation

In this phase coding was undertaken and the model was run to populate the database with real estate predictions.

3.3.6 Testing

After coding the developed system was tested to ensure that the requirements gathered were met. Usability testing was undertaken in this phase. Usability is a measure of how easy an interface is to use. Several factors that were considered include ease of use, usefulness, satisfaction, and ease of learning (Lund, 2016). Feedback from users was also collected and user acceptance testing undertaken.

3.4 Target Population

According to Sekaran (2000), population refers to the entire group of people, events, or thing of interest that the researcher wishes to investigate. The target population refers to the entire group of individuals or objects to which researchers are interested in generalizing the conclusions. The research sample is the members of the target population from whom we collect our data.

To develop the model this study targeted the real estate sector in Kenya. The population size consisted of 54 years of economic data ranging from 1962 to 2016 available at the KNBS, Ministry of Housing, Hass Consult Real Estate, and Central Bank of Kenya.

3.4.1 Sample Population

The study employed convenience nonprobability sampling whereby data collection relied on data sets that were conveniently available to the researcher to include in the study. This technique is common with statistical research where forecasting is being done. Timmerman (2012) states that in this approach the data set is split into an in sample period, used for initial parameter estimation, and an out of sample period for evaluating the performance of the forecast. This study sample consisted of 15 years of macro-economic data available from the years 2000-2015 which was used to estimate the forecasting model from a total population of 54 years (1962-2016). The 15 years period was selected for the sample size because of data consistency from all the sources of macroeconomic data, this can be explained due to higher levels of digitization in this period.

3.5 Data Collection

The data for this study was obtained online from various publications of statistical abstracts and economic surveys by KNBS and industry data from real estate players. The data collected was time series in nature. The study focused only on secondary data for the period between 2000 and 2015, the main consideration being availability of data capturing annual figures for the preselected variables. Primary data was collected for testing the application through administering questionnaires to 21 respondents selected through purposive sampling of real estate market investors.

3.6 Data Analysis Approaches and Tools

The data collected was analyzed using SPSS and presented in various ways including graphs and charts. SPSS was selected for its extensive facilities for analyzing statistical data. The data was tested for accuracy, consistency and completeness. Through statistical analysis the researcher determined which characteristics were most important to the performance of the real estate investment. Regression analysis was used for the researcher to determine the impact of each variable on the house price. The following statistical elements were provided for each combination of variables: R-Square, Adjusted R-Square, and s statistic.

The analysis aimed to estimate the result of the correlation between the variables through cross-sectional forecasting. Through this the researcher was able to showcase the relationship between the price of real estate and some of the predictor variables. Changes in predictors were seen to have an effect in the output of the system in a predictable way, assuming that the relationship does not change this can be used to generate forecast values for different assets (Hyndman & Athanasopoulos, 2014). Through multiple regression analysis, the study was able to come up with a regression function that can be used to predict real estate prices using various characteristics of the house price identified in the conceptual framework. The analyzed data was then stored in the SQL Server database for prediction purposes through the web-based system.

3.7 Research Quality

Research quality is a measure to which the results of a study exhibit correctness and the extent which the researcher ensured that research quality was upheld. The researcher used the following measures to test the quality of this research

Validity is defined as the extent to which a concept is accurately measured in a quantitative study. Validity of the data used in this research was determined by ensuring only data from credible sources was used to run the model. The model was tested by backwards extrapolation to check how well the model fit into the historical data sets out of sample.

Reliability is the accuracy of the instrument. In other words, the extent to which a research instrument consistently has the same results if it is used in the same situation on repeated occasions (Heale & Twycross, 2015). To ensure reliability of the forecasts, measures of performance were computed as highlighted in data analysis section of the methodology.

Chapter 4: System Design and Architecture

4.1 Introduction

This study investigated factors influencing the performance of real investments and aimed to develop a web-based model for predicting the investments. In this chapter secondary data collected was analyzed and interpretation of hypothesis were summarized into; tables, graphs, charts, regression analysis outputs, ANOVA, F-test, t-test statistics, and Pearson's correlation. The chapter also looks at steps taken in design of the regression model that was used to answer the research questions. It concludes by looking at system analysis, design and architecture. System designs are illustrated including; use case diagrams, entity relationship, sequence and class diagrams.

4.2 Data & Data Sources

The source of the secondary data used in this study was economic surveys published by the Kenya National Bureau of Statistics (KNBS). Other secondary data sets were collected from Ministry of housing, Hass Consult Real Estate and the Central Bank of Kenya. The following are sample demographics from the data analysis. The data collected was analyzed to determine its influence on real estate performance.

4.2.1 Gross Domestic Product

The GDP represents the total value of goods produced within the economy. Kenya's GDP has maintained an overall upward trajectory within the sample period 2000 to 2015, as illustrated by figure 4.1, with the exception of disturbances caused mainly by political instability and natural disasters such as draught and famine. According to Cyttonn Real Estate survey (2016), the construction and real estate sector has shown the most growth in terms of contribution to GDP (14.1% growth) as compared to any other sector including agriculture and financial services. The growth in GDP can be expected to indicate a growth in the real estate sector.

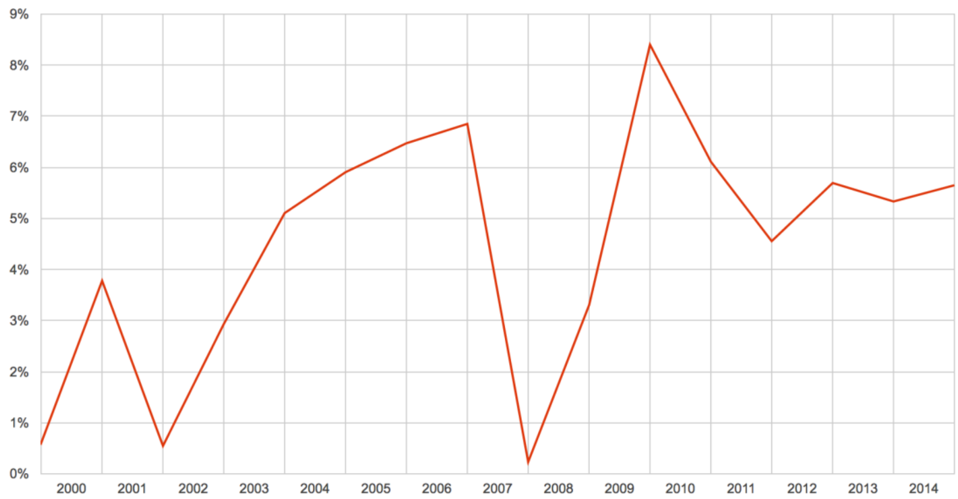


Figure 4.1: GDP growth 2000 – 2015 (KNBS Statistical Abstract, 2015)

4.2.2 Housing market Projections

According to the Ministry of Housing Kenya (2016) projections, the housing sector will continually see an increase in the number of units produced between the year 2008 and the year 2030. They estimate that about 4.3 million houses will be produced in this period. The figure 4.2 shows the expected house production from low-income houses to high-income houses. The low-low income houses are expected to be the majority of houses constructed due to the high population growth in the country as well as the high poverty rate in the country. The number of houses constructed was used to determine the average competitor distance for the regression model.

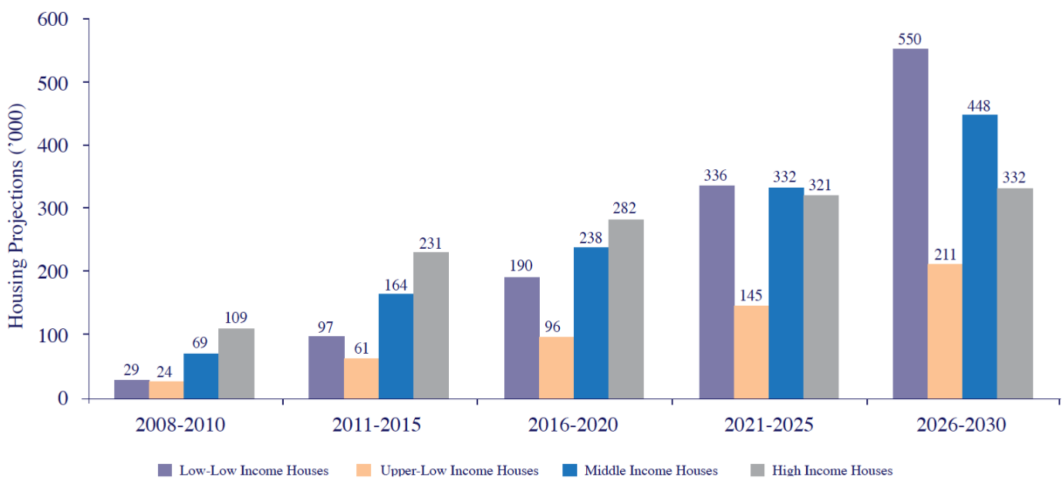


Figure 4.2 Housing projections, 2008-2030 (Ministry of Housing, 2016)

4.2.3 Land Prices Index

The Hass Property Index data as represented in figure 4.3 below is collected by Hass Real Estate Company, a leading real estate organization in Kenya. The index has become a valuable measure of Kenya's real estate performance. The information in the index is extracted from sold data from the company's' transactions. The index has shown a steady increase from the year 2011 to 2015 and this could in itself act as a predictor of house prices all other factors kept constant.

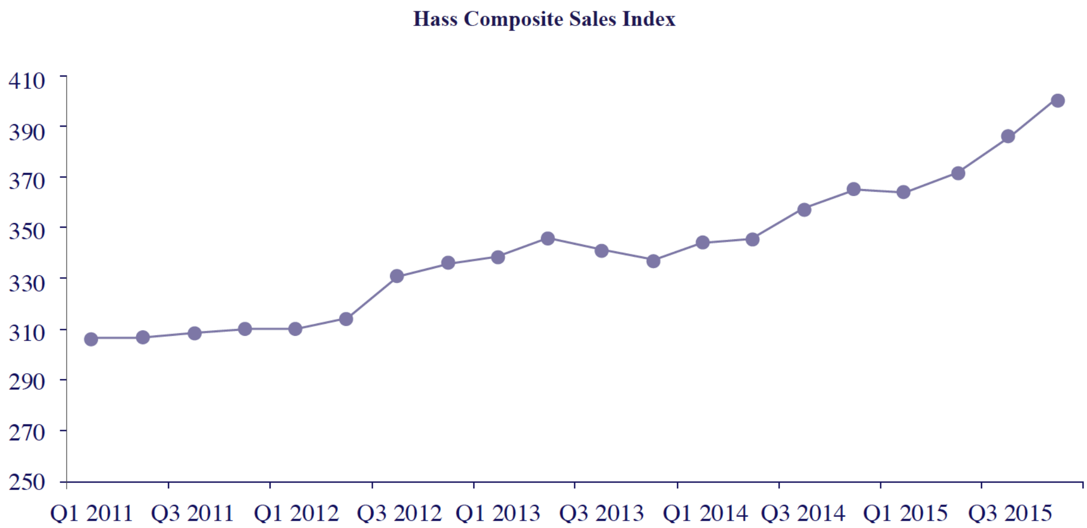


Figure 4.3 Hass Consult Property Index, 2011-2015 (Hass Consult Real Estate, 2016)

4.2.4 Income Per Capita

The income per capita is the average income of individual in an area for a certain period of time. The income per capita for Kenya shows the average income of Kenyans has been on the rise in the period between 2000 and 2015 indicating that more households have a higher quality of life. This can affects demand for real estate products. Income per capita in Kenya is represented in the figure 4.4 below.

Figure 4.4 Average Income Per Capita (KNBS Statistical Abstract, 2015)

4.2.5 Real Estate Sector Inputs V Outputs 2010 – 2015

The study also established that the real estate sector showed positive return on investment for the last five years according to the table 4.1 below. The outputs from the sector consistently exceeded the inputs and had a positive influence on the country's GDP. It was established that there was an 82.8% growth in the inputs to the real estate sector and a similar growth in the sector outputs over the period covered. This means that Kenyans are investing more money in the real estate sector and the sector has become a preference for many investors. The researcher also desired to establish the relationship between the income of real estate investors and the price of real estate. To do this a bivariate analysis was undertaken and correlation coefficient tested. The results obtained indicated a strong relationship (r = 0.783) in the return on investment and the prices of real estate assets.

Table 4:1 All Real Estate Firms Inputs & Outputs, 2011-2015

KES (Million)					
Year	**2011**	**2012**	**2013**	**2014**	**2015***
Input					
Special Trade Contractors	69,348	82,202	94,128	109,116	126,782
General Trade	121,065	143,506	164,324	190,491	221,332

Contractors					
All Other Contractors	81,690	96,832	110,879	128,535	149,345
Total	**272,103**	**322,540**	**369,331**	**428,143**	**497,459**
Output					
Special Trade Contractors	111,306	130,843	148,561	175,764	203,202
General Trade Contractors	194,314	228,420	259,352	306,842	354,743
All Other Contractors	131,115	154,128	174,999	207,044	239,365
Total	**436,734**	**513,391**	**582,912**	**689,650**	**797,310**

Source: Kenya National Bureau of Statistics (2016).

4.3 Requirements Analysis

This phase involved determining the requirements met based on the objectives identified in this study. Functional and non-functional requirements of the real estate prediction system were identified. The needs of different stakeholders using the system were an important input in determining the requirements.

4.3.1 Functional Requirements

These are specific functions the real estate prediction system performs. Functional requirements identified include

i. The System allows users to input a property type and the zone it is located.
ii. The system predicts the asset value of the selected property based on a regression function for that zone and property type.
iii. The application is hosted as a web service. It allows user to log in into the system through different web browsers.
iv. The system is able to consider multiple predictor variables in determining the appropriate function.
v. The system dynamically updates data used in the regression model.

4.3.2 Non-Functional Requirements

These are additional functions that improve the reliability of the prediction system and make it more robust. They include

i. The system is user friendly and easily navigable
ii. The system allows for multiple users to do prediction without downgrading on the performance
iii. The system is accessible from multiple devices and can scale to fit the screen size
iv. The application is secure to protect the integrity of the data uploaded by users. Other users should not see individual users personal information and data.
v. The system is available when needed by users and shall not have downtimes
vi. The application is responsive and efficient.

4.4 Proposed System Architecture

The purpose of this study is developing an application for predicting investment performance for real estate assets in Kenya. The application produces forecasts from the results of the model developed in the previous section. The following diagram represents the architecture of the proposed Real Estate Prediction System (REPS). The application has a user interface where investors will be able to enter property details; this will be crosschecked against the saved forecasts depending on geographical area and results produced from the price of the asset. The system architecture of the real estate prediction system is illustrated in the figure 4.5 below.

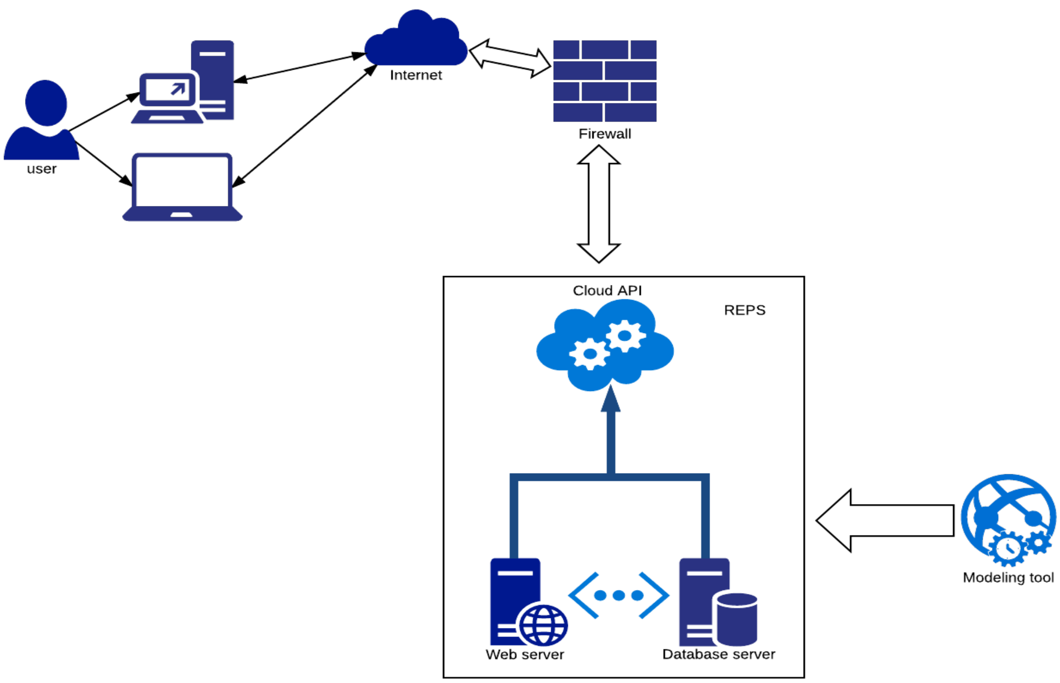


Figure 4.5 System Architecture

4.4.1 Use Case Diagram

UML diagrams were used in this study to represent system components, and to represent functional requirements of the system. Figure 4.6 shows the use case diagram indicating the major interactions taking place between the system and the various actors. The user is able to log in and pin property location and input property descriptions. The administrator and expert user manages property types and forecasts and ensure that updated forecast results are saved in the database. The system processes the data fetches the model and produces forecasts results when prompted by the user.

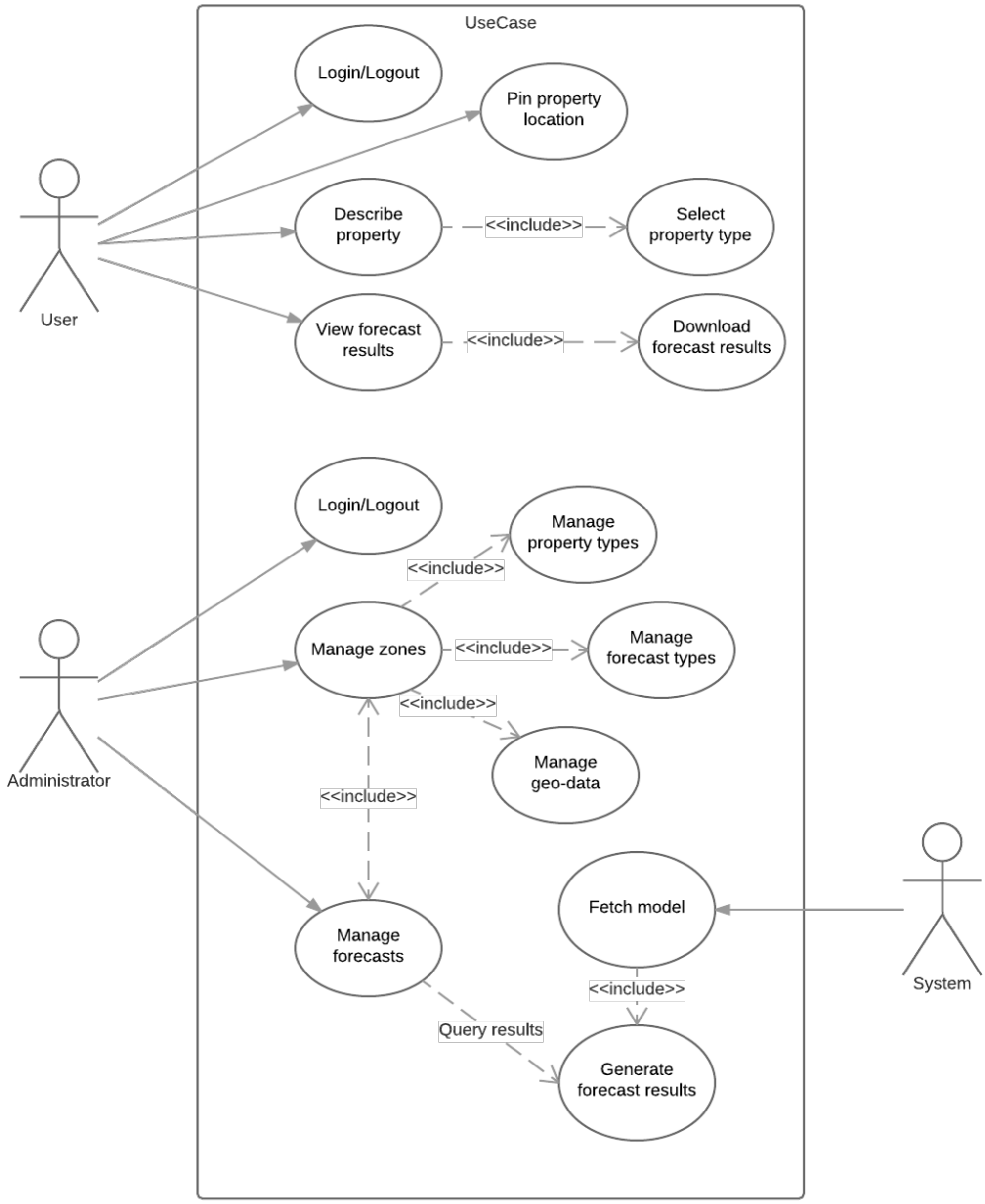


Figure 4.6 Use Case Diagram

4.5 Model Flow Chart

Figure 4.7 shows the flowchart or the model in this study. It captures the step-by-step flow of processes in the regression model used for prediction. Overall activities captured can be summarized in the following main processes

i. Input of data variables composed of economic and social economic data
ii. Comparison of variables based on their influence on real estate prices
iii. Identification of predictor variables
iv. Multiple regression analysis to produce regression function

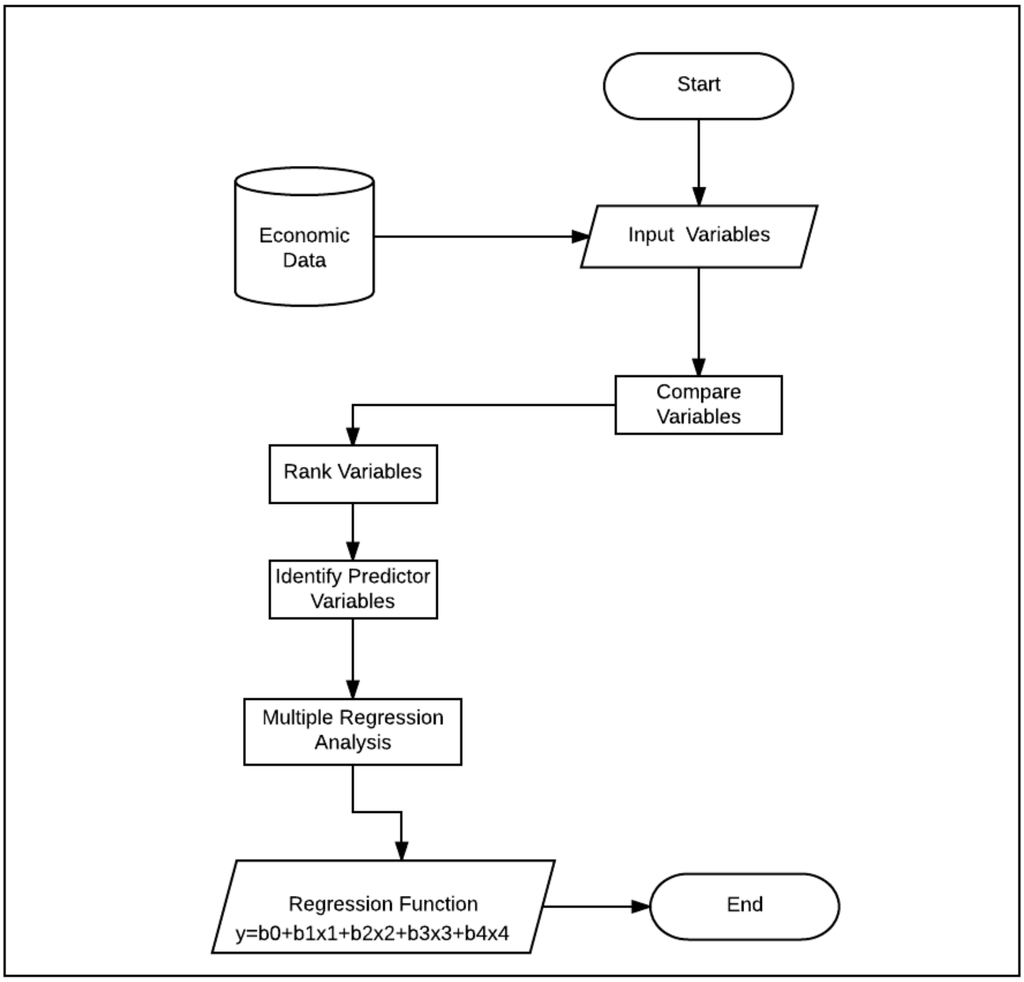


Figure 4.7 Model Flowchart

4.6 System Algorithm

i. Input collected data on economic and social economic variables. The variables should contain all relevant information about the real estate prediction model. Include time series data for house prices, GDP, incomes, and others discussed.

ii. Substitute missing values of any variables through averaging, this should be done by averaging the whole data set or the predecessor or successor of the missing value.

iii. Set the house price as the dependent variable (Y) while holding all others as variables (X1, X2, …, Xp) as the predictor variables or independent variables. This is to predict house price (Y) on the basis of the X's and to determine what is the "independent" influence of related variables.

iv. To determine the above run multiple linear regression analysis on the data this will give you the correlation between variables. The regression function will now take the form of $\gamma = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + e$.

v. Select predictor variables for particular zones and determine regression function for each zone. The model parameters $\beta_0, \beta_1, \beta_{2,} \beta_3$, & β_4 and the e will have values estimated from data. β_0 represents the intercept, $\beta_1 to\, \beta_4$ represent the regression coefficients, and e represents the residual standard deviation.

vi. Select the zone & type of property

vii. System fetches corresponding model for selected zone and property type.

viii. Output forecast results using regression function by replacing the function with estimated values of the predictors i.e by replacing values of X_1 with Average Income value, X_2 with housing units, X_3 with GDP data, X_4 with Cost of Land.

ix. Display estimated property values in form of charts, & graphs.

The system algorithm flowchart is represented in the figure 4.8 showing step by step the actions taken by the system in predicting real estate property values.

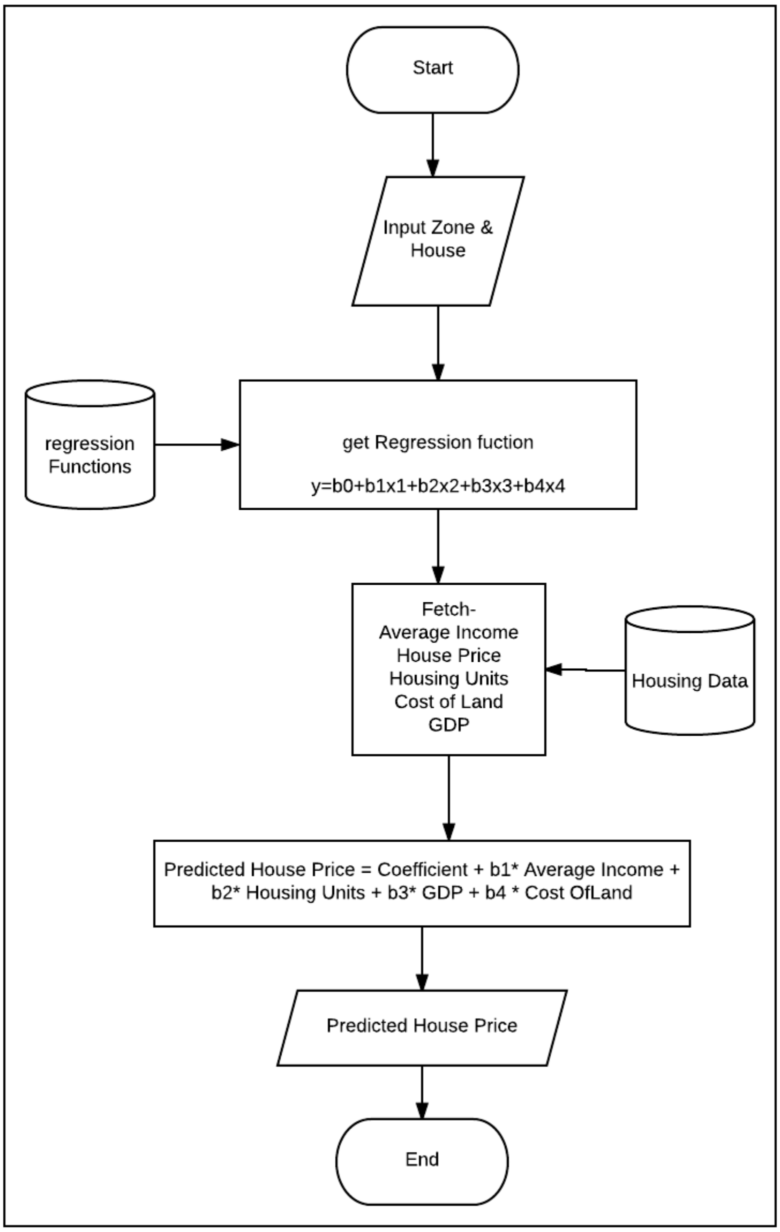


Figure 4.8 Algorithm Flow Chart

4.7 Sequence Diagram

Figure 4.9 shows the sequence diagram capturing the order of interaction in the real estate prediction system. It shows how objects operate with one another and in what order.

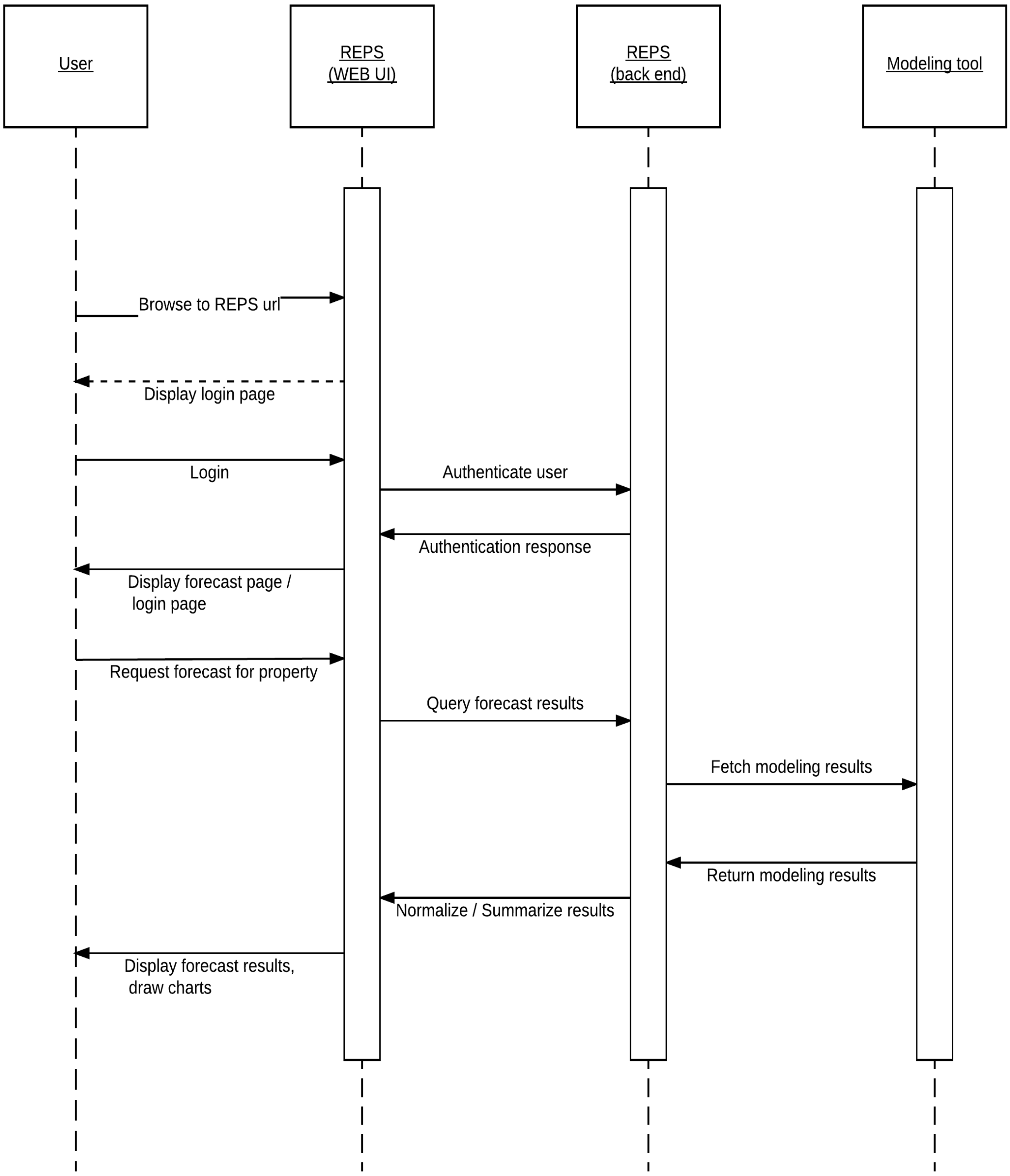


Figure 4.9 Sequence Diagram

4.8 Model Design

4.8.1 Context Diagram

The context diagram in the figure 4.10 below represents the flow of data from the main entities and the application. The user inputs information to login to the system, once logged in the user pins a property location and selects the type property. The system assigns a zone depending on user selection and selects the appropriate regression function for that zone. The system then generates forecasts based on selections. The system admin can manage zones and update databases.

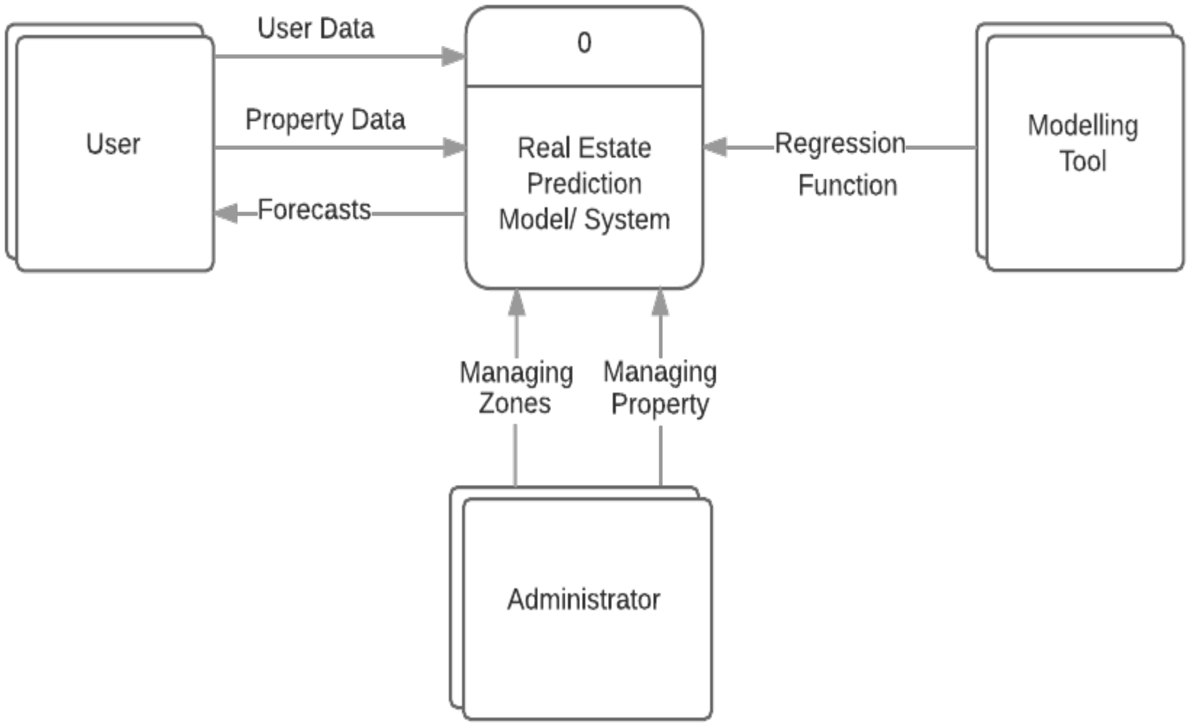


Figure 4.10 Context Level Diagram

4.8.2 Design Class Diagram

The class diagram in figure 4.11 describes the structure of a system by showing the prediction system classes, their attributes, operations and the relationships among objects.

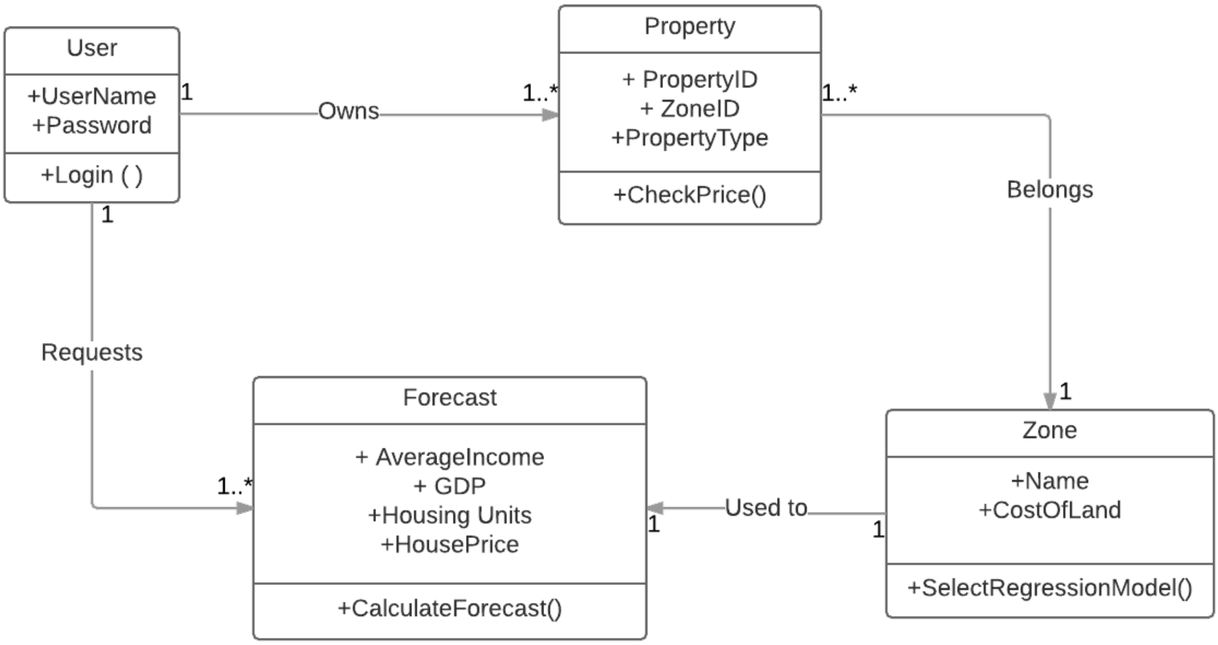


Figure 4.11 Design Class Diagram

Chapter 5: System Implementation and Testing

5.1 Introduction

The main objective of this study was to develop a model for predicting investment performance for real estate assets in Kenya. The model produces forecasts that are saved in the applications database and used to produce forecasts for different users. The system comprises of web based front end and a back end sub-system. The front end sub-system was developed using HTML and JavaScript while the backend system is based on the Asp.Net technology stack i.e. Asp.Net MVC, C# and SQL Server database.

5.2 System Implementation

The system is composed of the client side and the backend. It applies a multiple regression prediction model running on historical data. The regression model is based on the input historical data run through an analytical backend developed in C#. The library used is Math.NET which is an open source numerical library for .NET platform. The regression algorithm developed estimates the value of real estate prices as a function of the predictor variables for each case of the built data.

The relationships between the predictors and the dependent variables are summarized in the model and can be applied to different data sets such as different zones in Nairobi to produce forecasts. The predictions are displayed in the web map at the client side according to the type of property chosen and the zone selected. The backend is composed of user management module and model management module that are used to manage the users and to manage the SQL database, respectively.

The prediction model relies on different social economic variables affecting real estate prices include including demographic variables such as income levels and population growth, economic indicators such as changes in inflation GDP and Cost of Land.In developing the prototype the researcher adopted agile development methodology. The coding process was broken down into smaller iterative bits on the basis of system functionality, requirements for each bit were then collected, coded and implemented in weekly sprints over a two month period. At the end of each sprint the quality assurance and testing of functionalities was undertaken.

5.2.1 Front end

The web front end has been programmed using HTML5 and JavaScript. HTML markup is used to create the various interface elements while JavaScript is responsible for interaction with the back-end and manipulation of HTML elements. This sub-system comprises of interfaces allowing users to interact with and navigate through the system. After signing in the users are able to perform the following functions through web interfaces (Add data on open leaflet – to render map), mapbox (for map image), (Highcharts for the charts)

5.2.1.1 Admin Users

Admin users are able to: -

i. Manage property types - In this module administrative users define the various property types that are supported by the system. Property types include residential, commercial, retail, etc. This module is important as different types of properties are impacted differently by the model and as thus will have different future returns and values.

ii. Manage forecast types - It is envisaged that REPS will be able to run different prediction models. In this module administrative users are able to manage the supported forecasting as at any given time.

iii. Manage zones- The system is primarily geared towards generating forecasted returns for particular geographic zones and for different property types within those zones. To enable this core function of the system, administrative users have to distinguish zones by defining geographic boundaries on a map. In addition to defining boundaries social economic data about a region is specified as this has an effect on the future value and/or returns of properties within the zone.

iv. Manage forecasts - This module is used once zones, forecast types and property types have been set and the model generated. An expert user will come up with the various independent variable values for each zone and all available property types within a zone. The values provided will be for the current year and also forecasted for several years into the future. Once the independent variable values are set for each available property type in each zone, the model is run and the forecasts saved categorized by property type and zone. It is against these saved forecasts that users will query the future values and returns of their real estate properties.

5.2.1.2 Users

Based on the generated and saved forecasts investors are able to get predictions of the future returns and value of their real estate property as projected by the model. Investors are only required to pick the geographic zone of their property and specify the property type. With these details the system will return forecasted values in a time series format, return on investment ratios and generate charts that visualize the same.

5.2.2 Back End

The backend is built using the C# programming language with SQL Server serving as the storage database. IIS (Internet Information Services) webserver is used as the hosting server. This is the core sub-system responsible for running the model, generating forecast results against the model and presenting the forecasts to users, and storing user login data.

i. Integrate to the model - This module integrates to the modeling tool to fetch the model definition.
ii. Transform modeling results into forecasts - Using the model this module provides independent variable parameters for the entire forecasting period. The provided values are categorized into the various zones and property types already defined within the system. It is against these variable values that the model is run to generate forecast results.
iii. Store and retrieve forecasts - In this module the generated forecast zone and property types categorize results and stored in a time series format. The series is based on yearly steps. Once saved users are able to retrieve forecasted results of stored property types located in the stored zones.
iv. Drive the Web UI - This module is responsible for generating data displayed on the web user interface. It generates chart, graph and table data, while also calculating various Returns on Investment ratios to indicate the future value and profitability of a real estate property.

5.3 System Testing

The researcher tested the system to explore whether all the functionality proposed were working as required and to identify any bugs that may have been in they system. In light with this various tests were carried out in the development and implementation of the real estate prediction system. This included requirements testing from the functional requirements developed earlier, usability testing through the use of a questionnaire adopting System Usability Scale test, unit testing and system tests.

5.4 Unit Testing

Unit testing was conducted on the different modules in the application to check whether there were working properly. The unit testing was done to complement integration and system testing and to ensure bugs were identified and fixed early on in the coding process of the application. Some of the test cases used during unit testing are captured in table 5.1 .

Table 5:1 Unit Test Cases

Function	Test Cases	Results
Login Function	Enter Valid Login Details (Username & Password) Enter Invalid Login Details (Username or Password) Leave Empty Login Details and click Enter Leave Empty One Login Detail ((Username or Password) And click Enter	Passed. The login function was working as expected and no errors were found at this stage.

User Interface Functions	Test is all the buttons Menus, Navigation Pane details and Links are working in the expected way.	Passed. All the elements in the User Interface were functional
Database Connections	Test if connections to the SQL database are active, and that the application is able to connect, update, edit, delete and read database items	Passed. Database connection was successful. There was some errors in updating the database which were fixed and the test case repeated.
Back end Functions	Test for Model Integration and ability to download results into database Transformation of results into tables and charts Storage and retrieval of forecasts Generate data to be shown in the user Interface	Passed.

5.5 Usability Testing

Usability testing was also carried out on the application with questionnaires being issued to people who had interacted with the system. Out of 21 respondents who took part in the exercise 18 found the interface easy to use, and indicated that the application met its stated objectives as outlined in the questionnaire. Figure 5.1; 5.2; and 5.3 displays the results of the system usability testing.

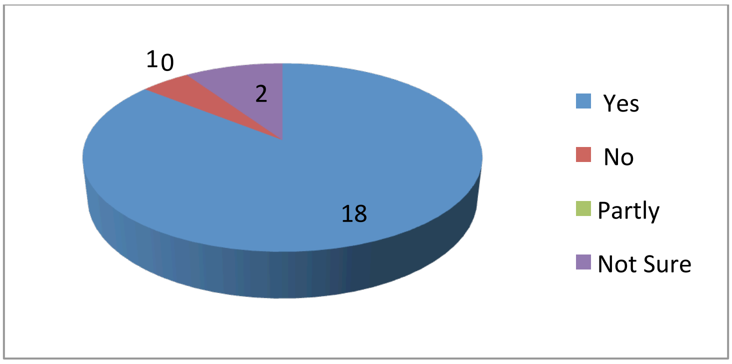


Figure 5.1 System Usability

All the respondents indicated that the menus were working properly and that they found it easy to navigate through the system.

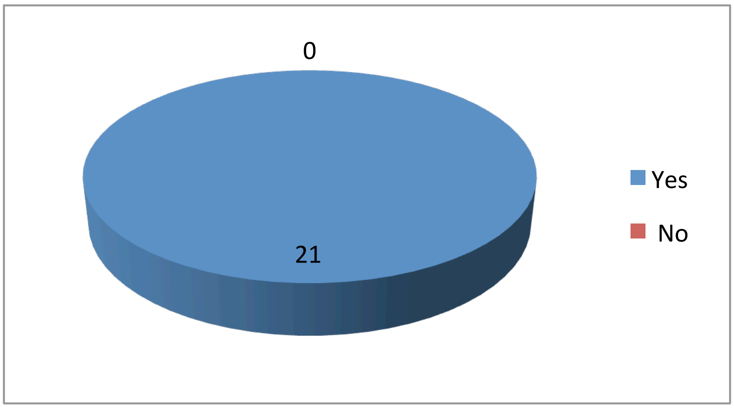


Figure 5.2 System reliability

19 of the respondents indicated they would like to use the system frequently once it was complete and more zones uploaded into the system

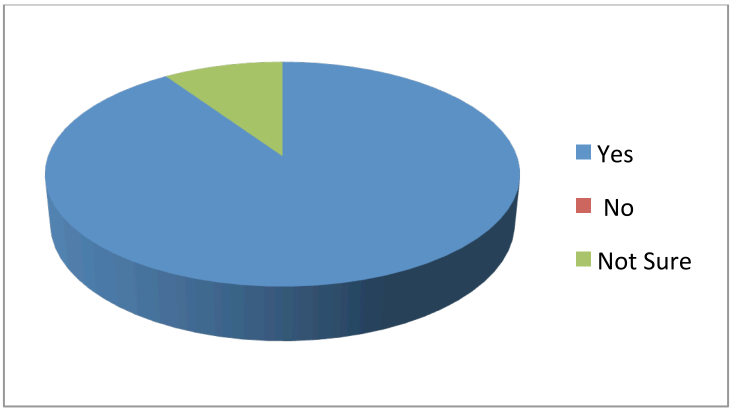


Figure 5.3 Respondents who would use the system for prediction

Chapter 6: Discussions

6.1 Introduction

This study set out to create a prediction model for predicting investment performance in the real estate sector in Kenya. The main objective was identifying the variables that affect the performance and to review existing models that have been applied in the prediction. The study also set out to develop a model and implement it in a web based application that could be used by home buyers, real estate investors, banks and other interested parties to analyze the performance of these investments going forward.

6.2 Analyzing variables that determine performance of real estate

Model development started with acquisition of secondary data from online sources. Data parameters were defined to ensure that only relevant data was captured. Data analysis was done through SPSS and predictor variables were narrowed down to the ones that showed the highest degree of correlation with the dependent variable. The factors that were used to develop the regression model are the average income of households, the GDP, the number on new housing units being developed, and the cost of land. It was determined that a regression function could be developed from these factors that can be used to predict real estate prices and that such a model can be tested for accuracy.

Through secondary data analysis the study was also able to identify economic variables that most significant to investment performance in the country. These variables included the GDP, government spending on infrastructure, the level of risk weighed assets held by real estate financiers and the housing price indexes. The results of these tests helped to identify variables that would be important in building a regression model for predicting house price movements and their characteristics.

6.3 Existing algorithms and models used to predict real estate assets

The second objective of this study was to examine existing algorithms and models that can be applied in prediction of performance for real estate investments. A number of methods and models were identified being able to give prediction of house prices. The literature review discussed several methods ranging from artificial neural networks, auto regressive moving averages and other time series forecasting techniques.

From review of literature the study could not identify any web-based tools available in Kenya that can help users predict the performance of their investments. Statistical analysis tools exist however this are for general analysis and expertise is needed to interpret results and make forecasts.

The researcher settled on a multiple regression model to test how these models produce results and how they can be integrated to a web application to help investors forecast the performance of real estate. The findings of the study show that it is possible to develop a model based on available social economic data, and to determine the variables that would be most useful in the model for accurate predictions. The model would then run on analysis software such as SPSS and produce forecasts for different zones in Nairobi. From this data individuals would be able to query their own property interests by simply choosing a location and narrowing down on property characteristics such as the size of the property, number of bedrooms, type of property and other characteristics that will be predefined in the system.

6.4 Developing a prototype for real estate prediction

The third objective of this study was to develop an application that can integrate prediction models, geographic information systems data, and socio economic data to predict real estate investments. A prototype was developed in line with study objectives. This was carried out through HTML5 and JavaScript. Agile development methodology was applied to the system development lifecycles through short iterative sprints that allowed for system testing when new features were implemented. The application allows for real estate investors to pin their location, enter property details and run their requirements through our models to generate forecasts. The application outputs estimated values for the asset they are interested in. A number of issues were faced during development of the application, including integration issues and usability concerns. These were all addressed in the testing and rectified to form the current prototype. To achieve the main aims of this study the following research objectives were developed in chapter 1, the following are the findings of this study in light of the set objectives.

6.5 Testing the Real Estate Prediction Application

The last objective was to carry out testing of the developed application. Unit tests and system testing was carried out and the application was seen to be functionally sound against the tests criteria. The application passed the tests on reliability, functionality and no major errors or bugs were identified. Usability testing was also carried out through the questionnaire attached in the appendix. Out of 21 respondents who participated in the study the majority 18 found the system easy to use and indicated their willingness to use the system regularly once more zones were included in the database.

Chapter 7: Conclusions and Recommendations

7.1 Conclusions

The goal of this research was to provide a means of predicting the performance of real estate assets for investors, homebuyers and other players in the sector. The research identified that there was a gap in the sector in terms of adoption of technology in analyzing these assets especially for people without a financial background. Whereas tools exists to do statistical analysis such as R, SPSS and Minitab most people are not able to use these tools and do not have the knowhow to interpret results into meaningful and actionable outputs.

In reviewing the literature many models for predicting real estate performance were identified. The research settled on Multiple Regression Analysis to identify which variables can be used in predicting prices and to demonstrate how the models can fit into the application. Multiple regression analysis was also preferred for its ability to carry out separate analysis of different variables and gauge the effect they have on the performance of real estate. That does not make the model more superior to other forecasting techniques and in building the application the researcher has made provisions for other models to be used to generate the forecasts.

7.2 Recommendations

From carrying out this research it was evident to the researcher that making predictions in real estate is complex and that non-skilled users would not be able to make predictions. It is thus the recommendation of the researcher that to advance the forecasting collaboration with real estate valuers and financial analysts to model and forecast different variables would be crucial in improving accuracy and producing credible forecasts.

There is need for more transparency in the real estate sector especially in terms of market data. Available data on sales and income from real estate showed a lot of inconsistencies and as such limit analysis and forecasting that can be applied. Future work could be done on data availability and artificial intelligence tools that can automatically collect and assess property listings data from real estate marketing websites, store this data in a database, analyze and sort through the data, to give more accurate assessments and predictions in the sector.

7.3 Future Work

The researcher plans to expand the application data to cover more zones in Nairobi County and other counties in Kenya. The researcher also plans to incorporate additional datasets such as satellite maps to help ease identification of property. The maps will also feature house addresses and estimates of house values based on recent sales approach data. The application will also be fitted with other tools such as a tool for predicting spending on real estate projects so as to budget accurately project costs of planned constructions. Finally the application algorithms will be developed to capture more accurately complex relationships between variables that affect the house price in a particular zone.

References

Ahmed, F., Nilsson, A. (2011). *Real Estate Discounted Cash Flow Model Development and Design* : The process of developing a new DCF model at a multinational real estate consultancy. Retrieved from http://www.diva-portal.org/smash/get/diva2:490753/fulltext01

Arsham, P. H. (2015). *Time-Critical Decision Making for Business Administration* . Retrieved 10 30, 2016, from University of Baltimore: http://home.ubalt.edu/ntsbarsh/stat-data/forecast.htm

Atzberger, P.J. (2010). An Introduction to Portfolio Theory, University of California Santa Barbara.

Baker, B (2001) *Residential Rental Real Estate:* An Investment In Need Of A Theory, Department of Accounting & Finance. Monash University. Retrieved from http://www.prres.net/papers/baker_residential_rental_real_estate_an_investment .pdf

Brooks, C., & Tsolacos, S. (2010). Real Estate Modeling and Forecasting. Cambridge: Cambridge University Press.

Brueggeman, W.B., Fisher, J.D. (2011). *Real Estate Finance and Investments (14th ed.)* McGraw-Hill

Butler, M., Kazakov, D. (2010). *Modeling the Behavior of the Stock Market with an Artificial Immune System*. Retrieved from http://www-users.cs.york.ac.uk/kazakov/papers/AIS.pdf

Can, A. (1998.). GIS and Spatial Analysis of Housing and Mortgage Markets. *Journal of Housing Research , Volume 9* (61).

Cave, W. C. (2014, April 23). *Prediction Theory For Control Systems*. Retrieved August 20, 2016, from predictsys: http://www.predictsys.com/PredictionTheory.pdf

CBK. (2015). Central Bank of Kenya. Kenya Economic Survey, Retrieved, from https://www.centralbank.go.ke

Chambers, J. C., Mullick, S. K., & Smith, D. D. (2001, July 9). How to Choose the Right Forecasting Technique. *Harvard Business Review* .

Corsini, R. K., (2009). Statistical Analysis Of Residential Housing Prices In An Up And Down RealEstate Market: A General Framework And Study Of Cobb

County, GA, MSc. Thesis, Building Construction, Georgia Institute of Technology, USA.

Cytonn Real Estate. (2016). Real Estate Market Survey. Retrieved from https://www.cytonn.com/research/

Denis, A., Wixom, H.B., Roth, M. R. (2012). Systems Analysis and Design, 5th Edition. John Wiley & Sons, Inc.

Greener, S. (2008). *Business Research Methods.* Ventus Publishing ApS.

Heale, R., & Twycross, A. (2015). *Validity and reliability in quantitative studies.* Retrieved November 15, 2016 , from http://ebn.bmj.com/content/18/3/66.full.pdf

Hyndman, R. J., & Athanasopoulos, G. (2014). *Forecasting: principles and practice.* Otexts.com.

Huberman, G., Wang, Z. (2005*). Arbitrage Pricing Theory*. The New Palgrave Dictionary of Economics, 2nd edition, London: Palgrave. Macmillan.

Kara, Y., Boyacioglu, M. A., & Baykan, O. K. (2011, May). *Predicting direction of stock price index movement using artificial neural networks and support vector machines*: The sample of the Istanbul Stock Exchange. Expert Systems with Applications.

KNBS. (2015). Kenya National Bureau of Statistics. Statistical Abstract. Retrieved, from http://www.knbs.or.ke/

Kibiru, C., Pokhariyal, R., & Obwocha, J. N. (2014). Determinants of Competitive Advantage in the Real Estate Industry in Kenya: A Case of Thika Greens Golf Estate in Muranga County. *International Journal of Social Sciences and Entrepreneurship, , 1*, 521-529.

Koenigsberg, R. (2016). *Commercial real estate types & categories*. Retrieved from https://cre.nyrej.com/commercial-real-estate-types-categories-part-ron-koenigsberg/

Kothari, C. R. (2012). *Research Methodology :: Methods and Techniques* (Vol. 3). New Age International Pvt Ltd Publishers.

Koubkov´a, K. (2015). *Risk model for real estate assets: Analysis and development.* Masters Theses, Charles University , Prague.

Krystalogianni, A., Matysiak, G., & Tsolacos, S. (2014, August). Forecasting UK Commercial Real Estate Cycle Phases With Leading Indicators: A Probit Approach. *Journal of Applied Economics*.

Kuang W., Liu W., Ma Z., Wang H. (2013) Development of a Web-Based Prediction System for Wheat Stripe Rust. In: Li D., Chen Y. (eds) Computer and Computing Technologies in Agriculture VI. CCTA 2012. IFIP Advances in Information and Communication Technology, vol 392. Springer, Berlin, Heidelberg

Kumar, V. (2016, December,). Time Series Modelling Using Stochastic Models: A Review . *International Journal of Engineering Sciences & Research Technology* , 585-589.

Larsen, J. I. (2010). *Predicting Stock Prices Using Technical Analysis and Machine Learning.* Norwegian University of Science and Technology Department of Computer and Information Science.

Levišauskait, K. (2010). *Investment Analysis and Portfolio Management.* Lifelong Learning Programmes. Leornado Da Vinci.

Lizieri, C.M. (2011). *Forecasting and Modeling Real Estate*. School of Real Estate & Planning, Henley Business School, University of Reading. Retrieved from http://eres.architexturez.net/system/files/pdf/eres2013_10.content.pdf

Lund M (2016). Measuring Usability with the USE Questionnaire. Usability Interface Retrieved from https://www.researchgate.net/profile/Arnold_Lund/publication/230786746_Measuring_usability_with_the_USE_questionnaire/links/56e5a90e08ae98445c21561c/Measuring-usability-with-the-USE-questionnaire.pdf

Lundström, S. & Gustafsson, C. (2016). Market Valuation Based On Discounted Cash Flow Analysis - Consistency In Assumptions. Retrieved from https://www.kth.se/polopoly_fs/1.167149!/Menu/general/column-content/attachment/WP60.pdf

Lynn, D., & Wang, T. (2011). *Real Estate Mathematics; Applied analytics and quantitative methods for private real estate investing.* PEI.

Ma, H., Chen, M., & Zhang, J. (2015). Study on the Prediction of Real estate Price Index based on HHGA-RBF Neural Network Algorithm. *International Journal of u- and e- Service, Science and Technology , 8* (7), 109-118.

Margret, J. (2005). *Why business models mater?* Harvard Business Review, R0205F. Retrieved from, http://info.psu.edu.sa/psu/fnm/asalleh/WhyBusModelMatter.pdf.

Marwani, H. A. (2014). *An Approach To Modeling And Forecasting Real Estate Residential Property Market.* London: Brunel Business School.

Miregi, M. O., & Obere, P. A. (2014). Effects of Market Fundamental Variables on Property Prices in Kenya: *A Case of Nairobi Residential Property Market*. IOSR Journal of Economics and Finance (IOSR-JEF) , 5 (5), 101-113.

Mukherjee, D and Metia,S (2001) Capital Asset Pricing Model When Data Is Skewed, Sankhyā: The Indian Journal of Statistics, Series B (1960-2002), Vol. 63, No. 1 (Apr., 2001), pp. 108-121. Retrieved from http://www.jstor.org/stable/25042383

Muli, N. F. (2013). *An Assessment of the Factors Affecting the Growth in Real Estate Investment in Kenya.* . University of Nairobi .

Nau, R. (2014, August). *Forecasting with moving averages*. (D. University, Editor, & F. S. Business, Producer) Retrieved September 17, 2016, from https://people.duke.edu/~rnau/Notes_on_forecasting_with_moving_averages--Robert_Nau.pdf

Park, H. (2011). *Study for Application of Artificial Neural Networks in Geotechnical Problems*. Korea Republic, Samsung C&TI

Pesaran, M., Schuermann, T., and Smith, L. (2009). *Forecasting Economic and Financial Variables with Global VARs*: International Journal of Forecasting 25, 642-675.

RICS Valuation Professional Standards. (2012, March). *Incorporating the International Valuation Standards*. Retrieved 11 5, 2016, from http://www.trigoncapital.com/: http://www.trigoncapital.com/upload/RICS_Red_Book_2012_-_Global_and_UK_edition.pdf

Ross, S. (2010). *A Practitioner's Guide to Arbitrage Pricing Theory.* Retrieved from http://www.cfapubs.org/doi/pdf/10.2470/rf.v1994.n4.4445.1

Sami, M. (2012). *Software Development Life Cycle Models and Methodologies.* Retrieved 11 18, 2016, from https://melsatar.wordpress.com/2012/03/15/software-development-life-cycle-models-and-methodologies/

Saunders, M., Lewis, P., & Thornhill, A. (2009). Research Methods for Business Students Fifth edition. Harlow, England: Prentice Hall / Pearsons.

Schulz, R. (2004, 2 3). *Valuation of Properties and Economic Models of Real Estate Markets.* Retrieved 12 7, 2016, from https://pdfs.semanticscholar.org/c5f4/ef67c7a849604730ced21ad7ed6361659f73.pdf

Sekaran, U. (2000). *Research methods for business: a skill-building approach* (Vols. 3, illustrated). Pennsylvania: Wiley, 2000.

Stergiou, C., & Siganos, D. (2012). *Neural Networks*. Retrieved 12 9, 2016, from Imperial College London : https://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/report.html#Contents

Stevenson, S., & McGarth, O. (2003). A comparison of alternative rental forecasting models: empirical tests on the London office market. *Journal Of Property Research , 20* (3), 235-260.

Stevenson, W. J. (2011). *Operations Management* (Vol. 11). Mcgraw Hill.

Taffese, Z. W. (2006). A survey of Artificial Intelligence in Real Estate Industry. Third International Conference on Articial Intelligence in Engineering & Technology. Nov 22-24 2006

Vishwakarma, V. K. (2013), "Forecasting Real Estate Business: Empirical Evidence From The Canadian Market", Global Journal of Business Research, Vol.7, No.3.

World Economic Forum. [WEF], (2014). Understanding the Commercial Real Estate Investment Ecosystem. Retrieved from http://www3.weforum.org/docs/WEF_IU_Understanding_the_Commercial_Real_Estate_Investment_Ecosystem.pdf

Young, D. C. (2012). *Software Development Methodologies* . Retrieved from researchgate: https://www.researchgate.net/publication/255710396_Software_Development_Methodologies

Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network mode. *Neurocomputing () , 50* , 159 – 175.

Zhao,L. (2009). Neural Networks In Business Time Series Forecasting: Benefits And Problems. *Business Information Systems Review Journal* –Vol.13, No 3. Retrieved from http://cluteinstitute.com/ojs/index.php/RBIS/article/viewFile/4324/4412.

Appendices

Appendix A: Usability Questionnaire

Real Estate prediction System Usability Questionnaire
The data captured in this questionnaire is for academic purposes and shall be treated as confidential. No personal information is required in this questionnaire.

Instructions
1. Access the application on the link provided
2. Select Location of property on map
3. Enter a name for the property
4. Enter estimated value of the property
5. Select property type from dropdown menu
6 Click Submit
7, System will display forecast value of property.

Section 1

1. I think the application met the above objectives as stated (Choose only one option)
   - ☐ Yes
   - ☐ No
   - ☐ Partly
   - ☐ Not Sure
2. All the functions and menus I used were working properly
   - ☐ Yes
   - ☐ No
3. I think I would use the system to predict performance for my real estate investments
   - ☐ Yes
   - ☐ No
   - ☐ Not Sure

4. Did the you experience any issues in trying to use the system?
   - ☐ Yes
   - ☐ No

   If you answered you answered Yes to the above question please list down the issues faced

5. Do you have any recommendations or comments about the system

Section 2: System Usability Scale

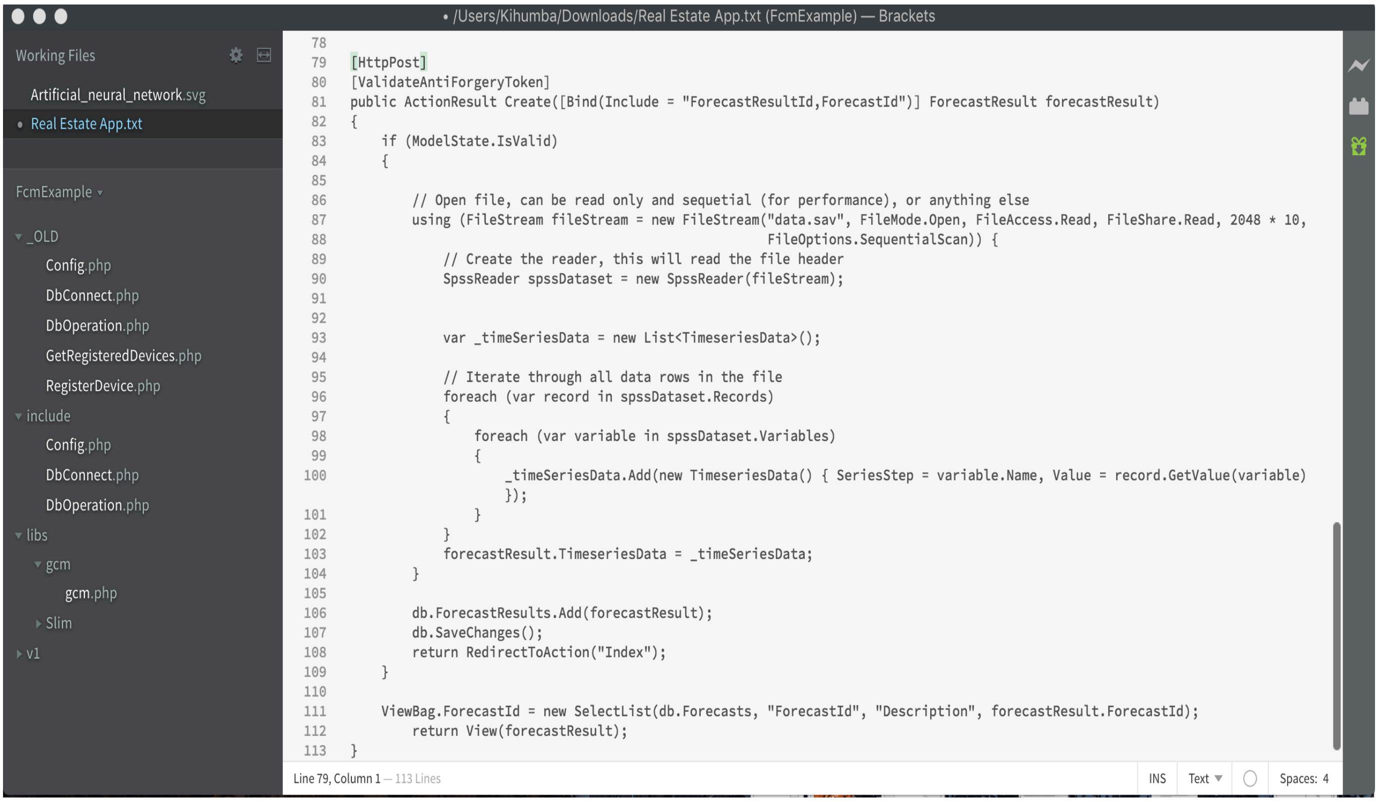
Strongly

Strongly

disagree agree

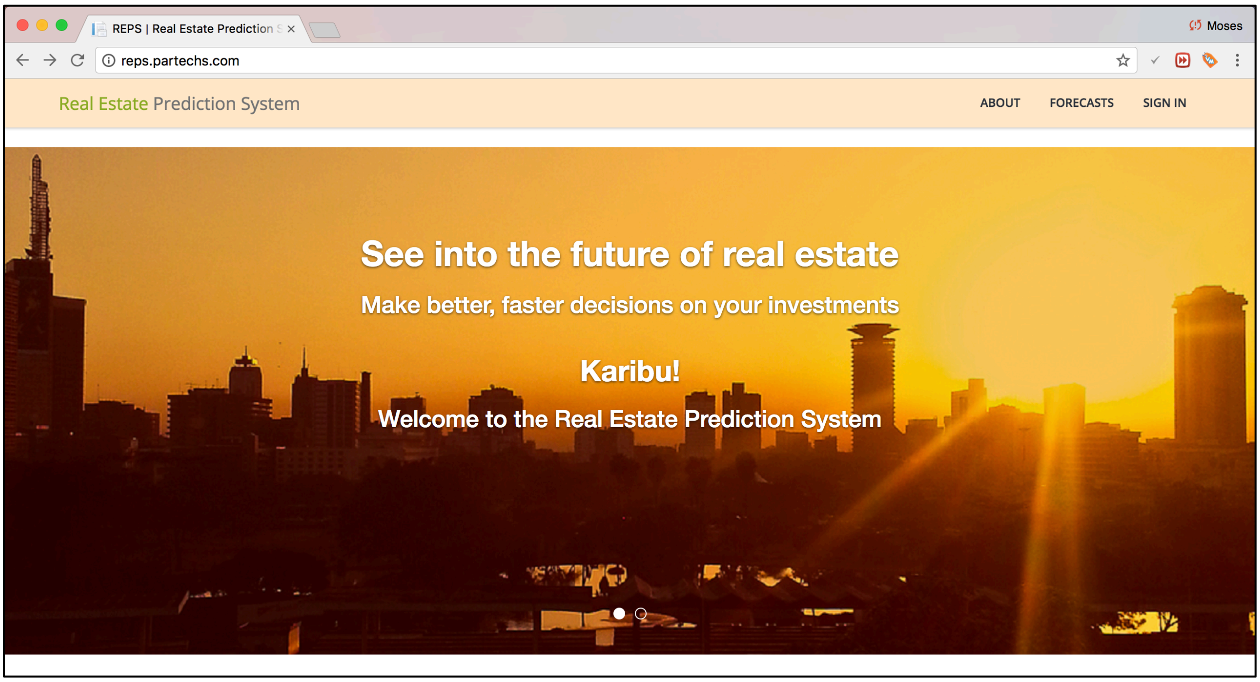
	1	2	3	4	5
1. I think that I would like to use this system frequently					
2. I found the system unnecessarily complex					
3. I thought the system was easy to use					
4. I think that I would need the support of a technical person to be able to use this system					
5. I found the various functions in this system were well integrated					
6. I thought there was too much inconsistency in this system					
7. I would imagine that most people would learn to use this system very quickly					
8. I found the system very cumbersome to use					
9. I felt very confident using the system					
10. I needed to learn a lot of things before I could get going with this system					

Appendix B: Sample Code

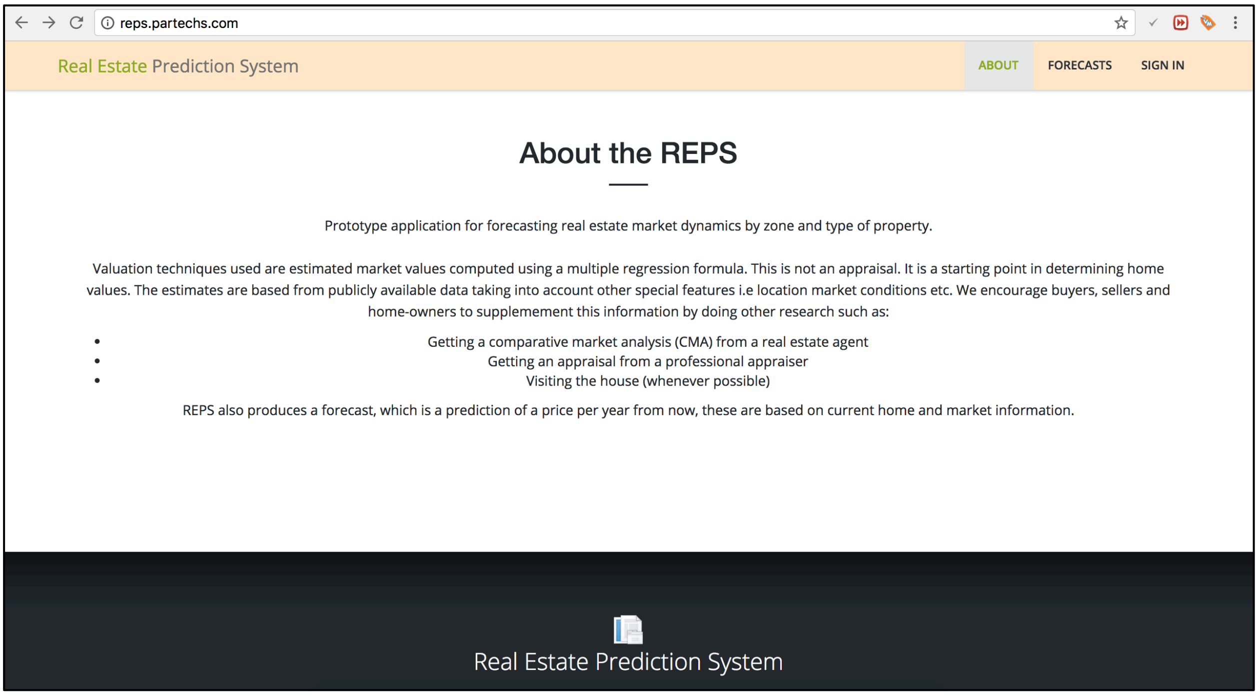


Appendix C: Web Application Screenshots

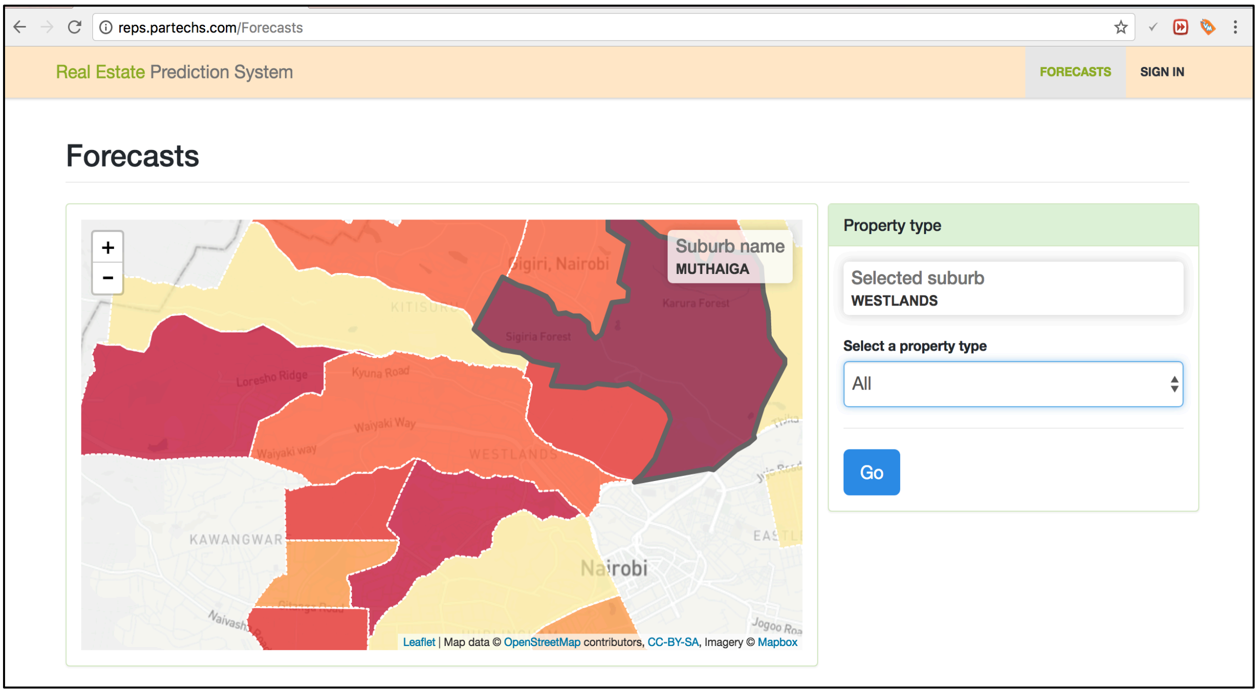
Home Page of the Real Estate Prediction System (REPS)



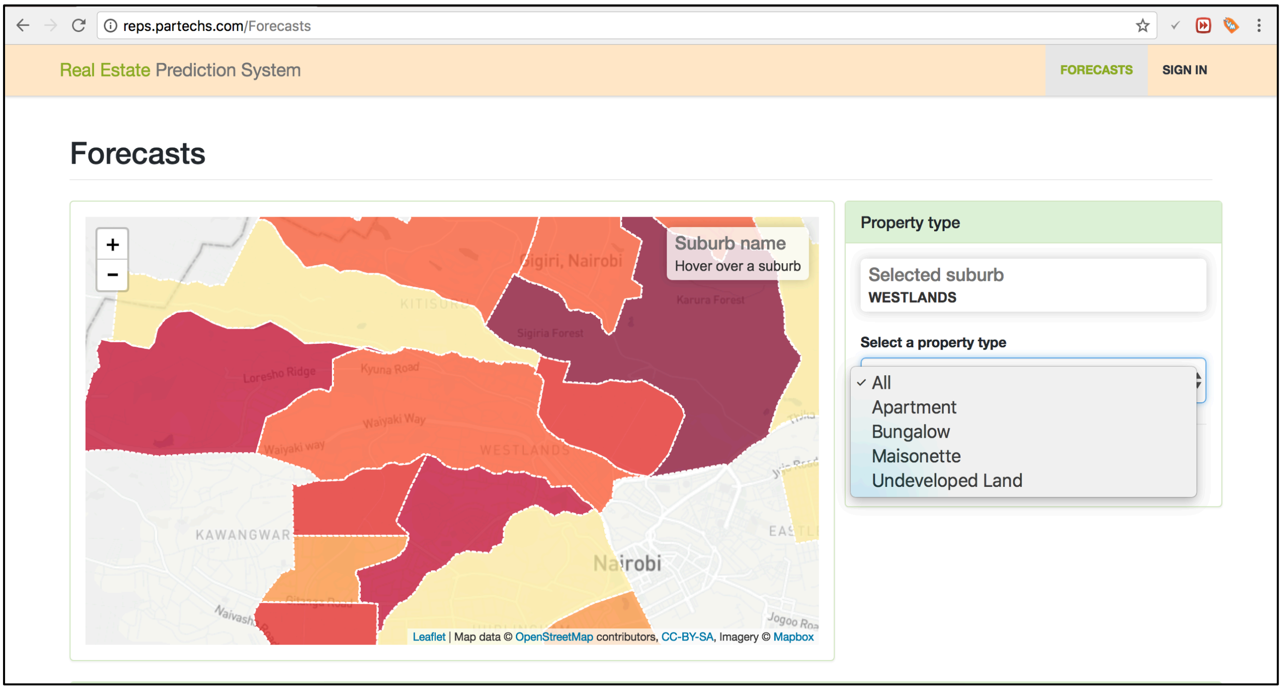
About REPS



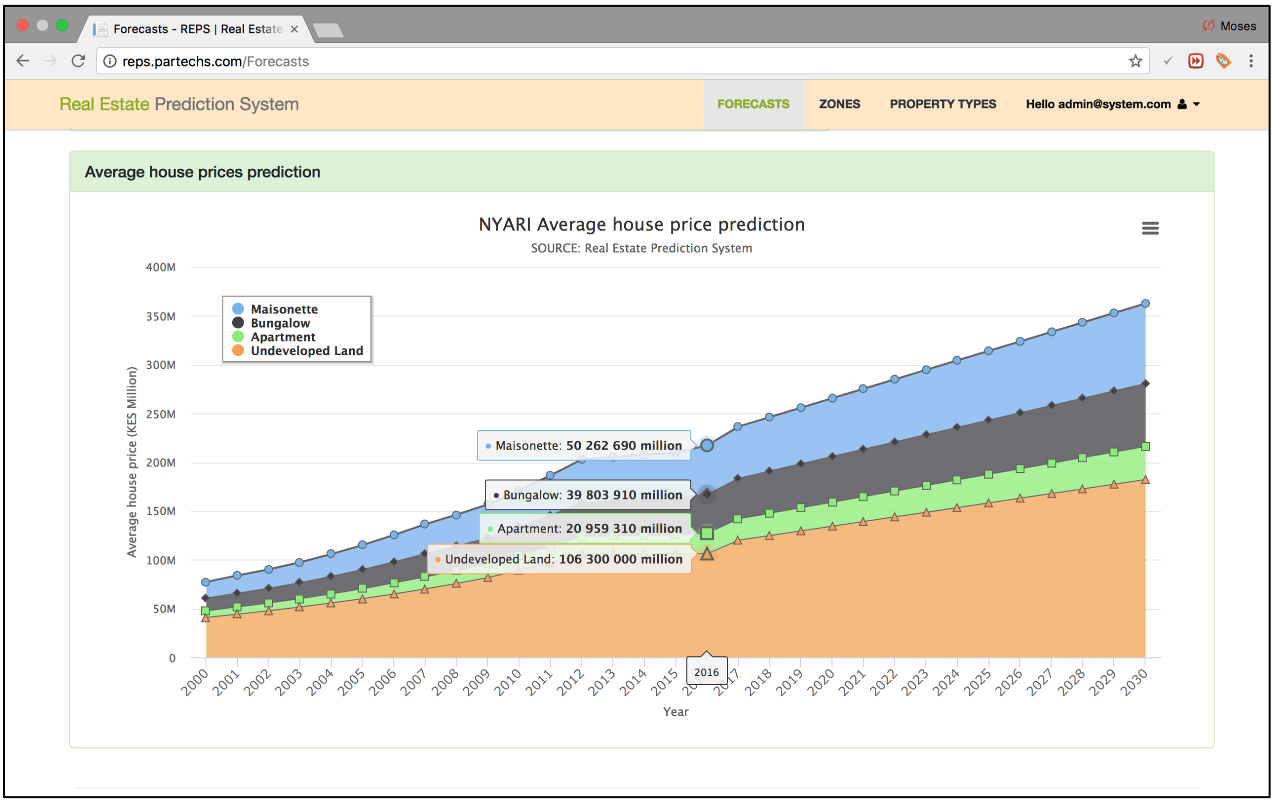
Forecasts (Selecting Zone)



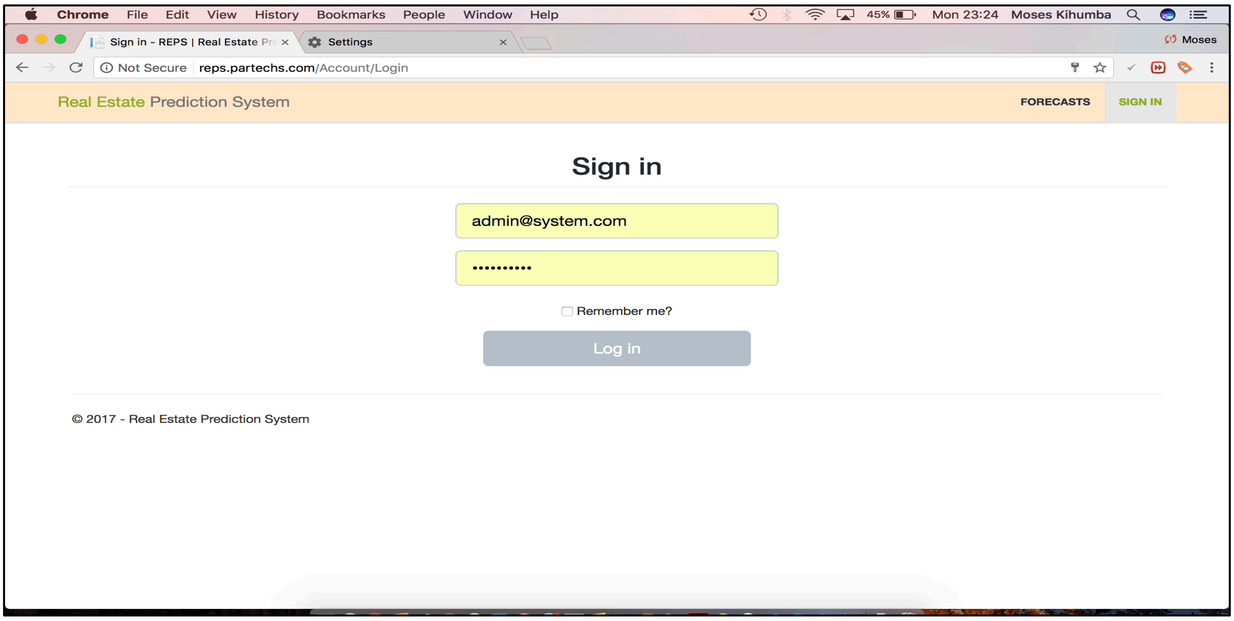
Forecasts (Selecting Property Type)



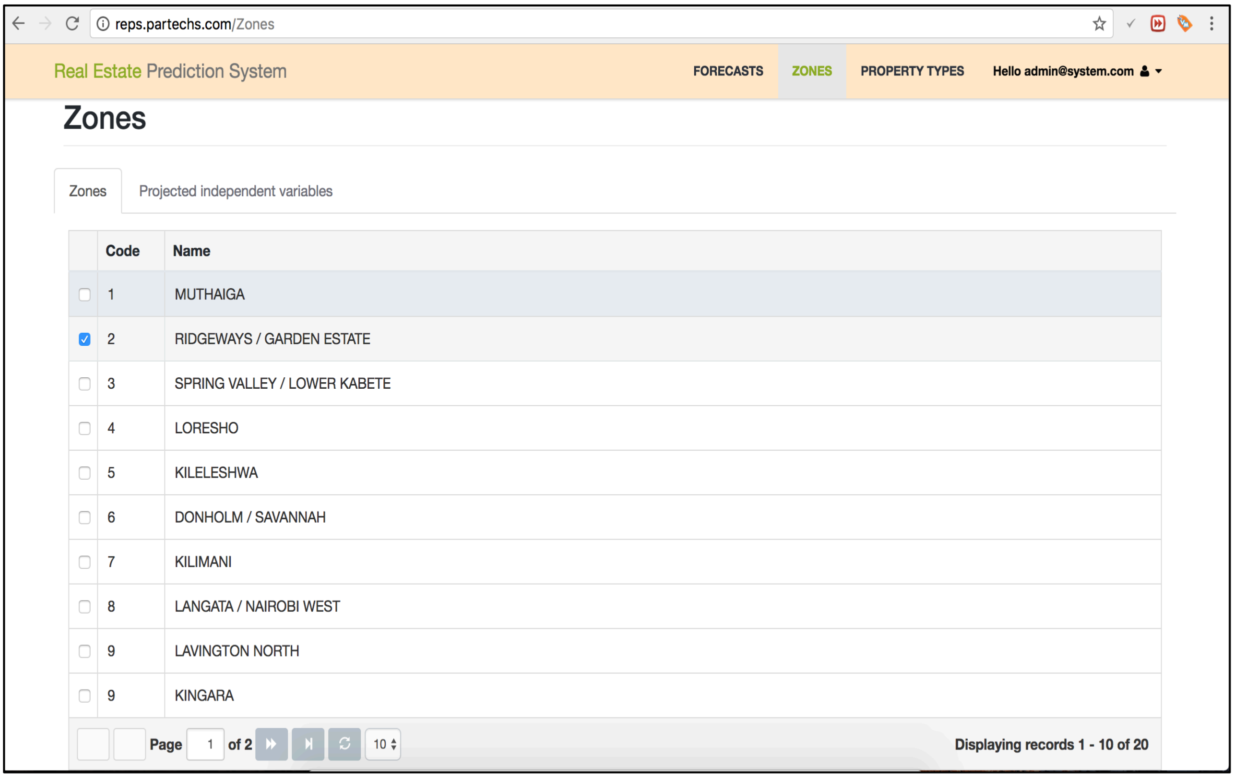
Viewing Predictions



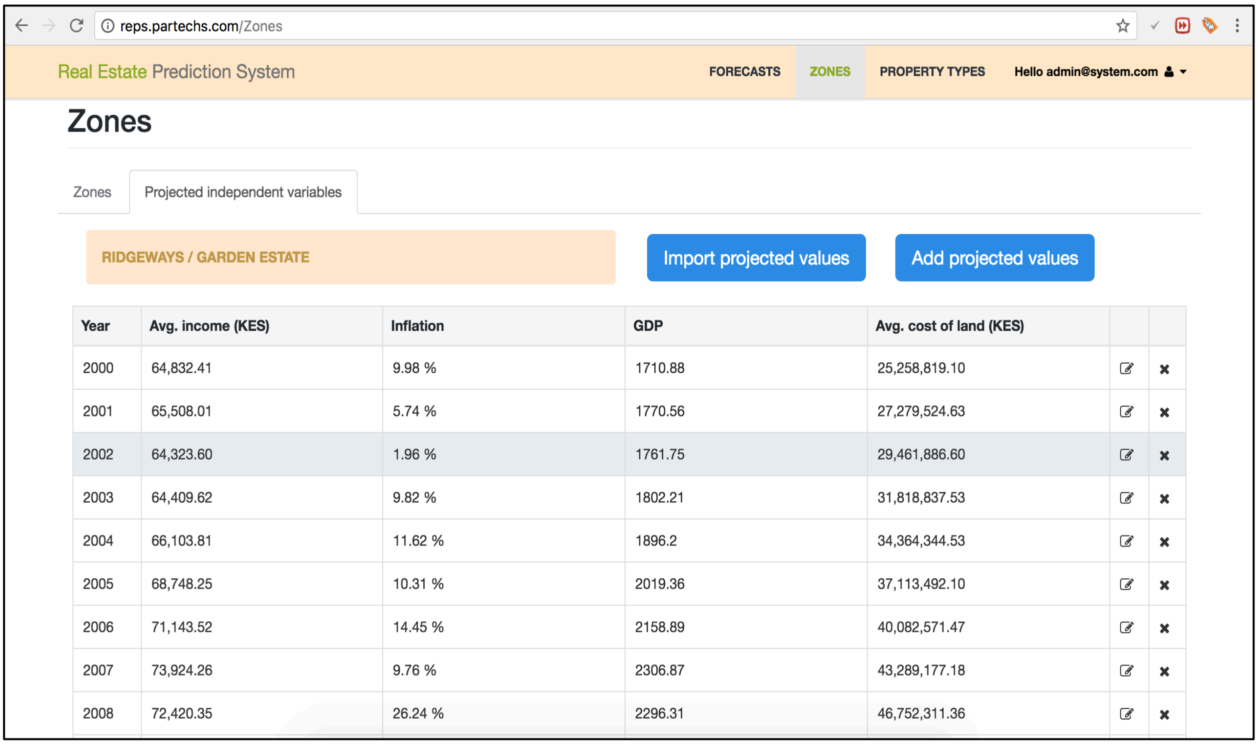
Login to Back-end



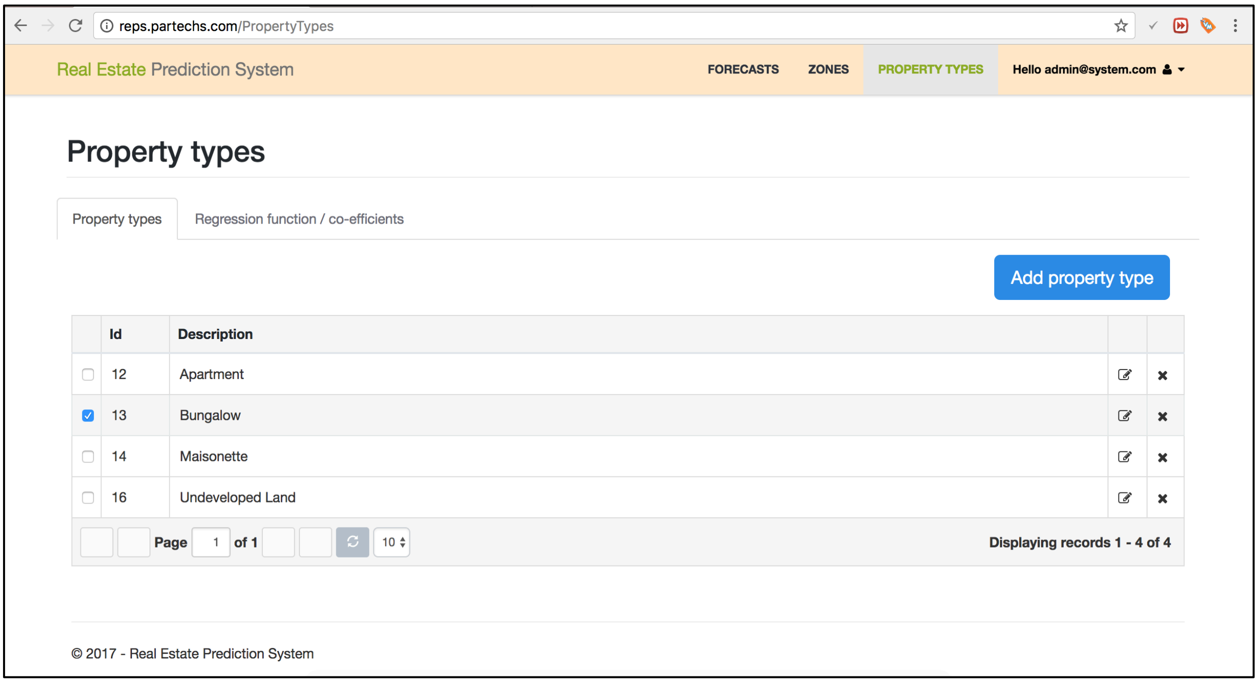
Managing Zones



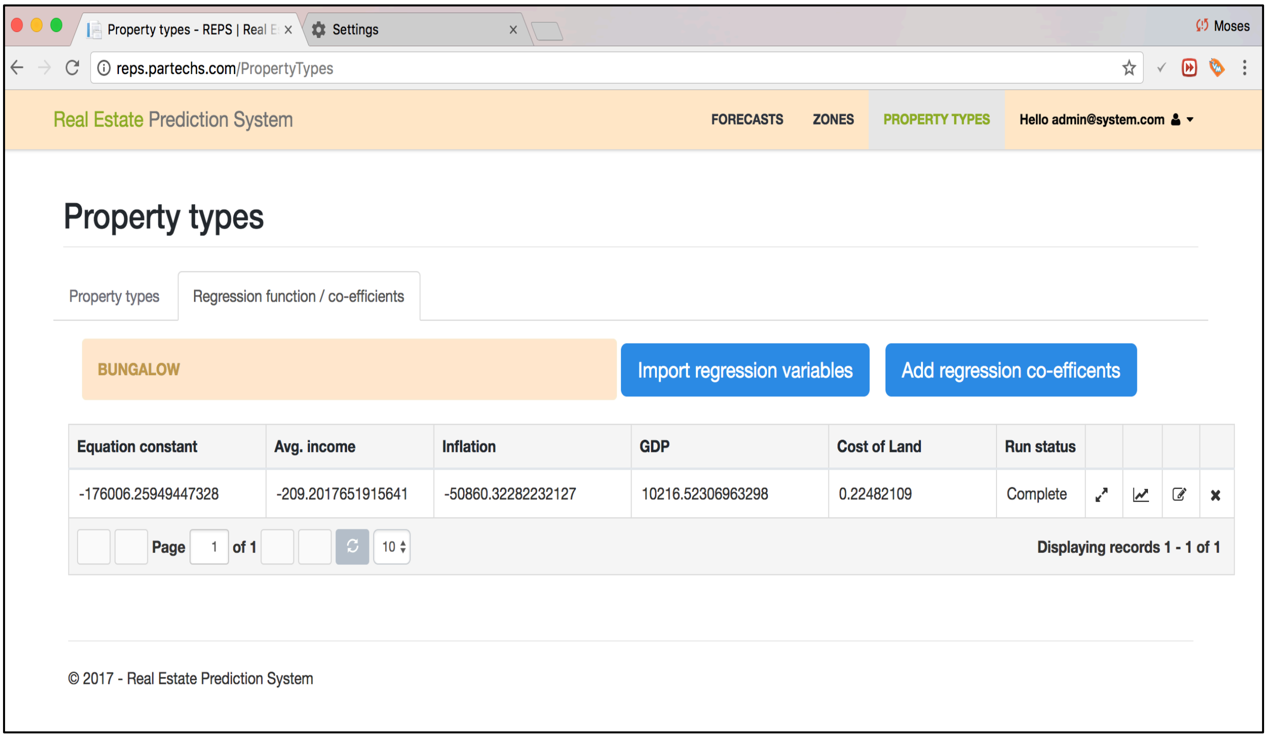
Managing Zone Data



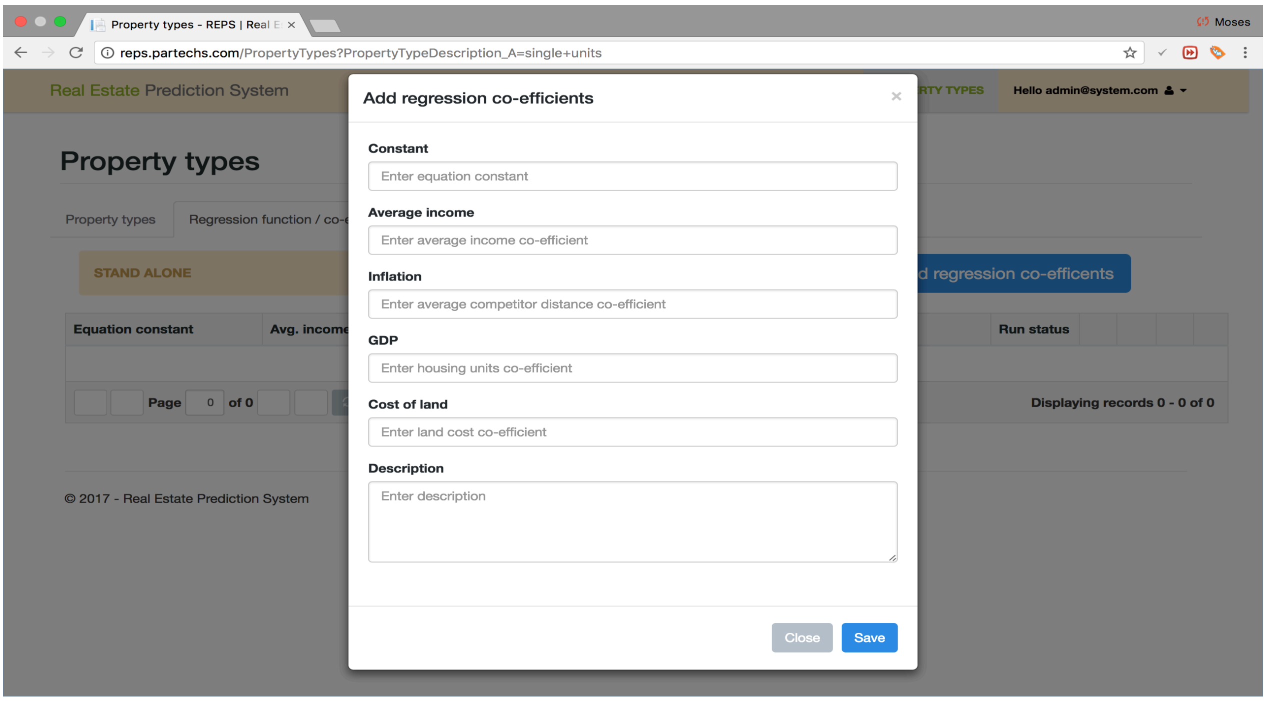
Managing Property Types



Managing Regression Functions



Managing Regression Functions



Appendix D: Turnitin Originality Report