BINDURA UNIVERSITY OF SCIENCE EDUCATION

FACULTY OF SCIENCE AND ENGINEERING

DEPARTMENT OF COMPUTER SCIENCE

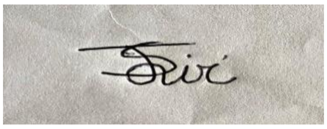
Property Evaluation in Real Estate using Machine Learning.

NAME:	GODWIN S JIRI
REG NUMBER:	B213599B
LEVEL:	4.2
PROGRAM:	COMPUTER SCIENCE
SUPERVISOR:	Mr. P. Chaka

APPROVAL FORM

The undersigned certify that they have supervised the student Godwin Jiri's dissertation entitled, "**Property Evaluation in Real Estate Using Machine Learning**" submitted in partial fulfilment of the requirements for a Bachelor of Science Honors Degree in Computer Science at Bindura University of Science Education.

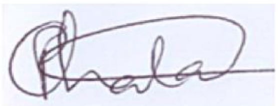
STUDENT: **DATE:**



17/06/25

………………………………………………… ……………………………...

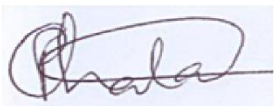
SUPERVISOR: **DATE:**



17/06/25

………………………………………………… ……………………………...

CHAIRPERSON: **DATE**



17/06/25

………………………………………………… ……………………………...

EXTERNAL EXAMINER: **DATE:**

………………………………………………… ……………………………...

DEDICATION

This project, "Property Evaluation in Real Estate Using Machine Learning," is dedicated with heartfelt appreciation to my parent, Mrs. Jiri. Your unwavering support, love, and sacrifices have been the foundation of my academic journey. Thank you for nurturing my curiosity, encouraging my ambitions, and standing by me through every challenge. Your belief in me has given me the strength to pursue my goals and bring this research to completion. This achievement is a reflection of your endless devotion and guidance.

ACKNOWLEDGEMENTS

First of all, I want to thank God for giving me the strength, health, and grace to complete this journey. This project wouldn't have been possible without His guidance. A big thank you to my supervisor, Mr. P. Chaka, for his continuous support, honest feedback, and encouragement from the very beginning. Your guidance made this process a lot clearer and more manageable.

To my amazing family my parent, my siblings thank you for always believing in me even when things got tough. Your love, prayers, and sacrifices mean everything to me. To my friends Themba and Panashe, you guys have been my rock from the late-night talks to the endless motivation, I genuinely appreciate it all. This research journey had its ups and downs, but because of each one of you, I made it through. Thank you all, from the bottom of my heart.

ABSTRACT

This research investigates the application of supervised machine learning techniques in the prediction of property prices, with a specific focus on residential real estate in Bulawayo, Zimbabwe. The study addresses the limitations of traditional manual valuation methods, such as human error, delays, and inconsistencies, which often lead to inaccurate pricing and affect key stakeholders including buyers, sellers, and real estate agents. To provide a data-driven solution, a machine learning-based system was developed using the Random Forest algorithm due to its robustness and high accuracy in predictive tasks. The model was trained on a real estate dataset comprising various property attributes such as location, number of bedrooms, land size, and property type. Performance evaluation metrics of the model produced the following results: A Mean Squared Error (MSE) of 4,669,713,572.27, Root Mean Squared Error (RMSE) of 68,335.30, R-squared (R^2) of 0.871, Mean Absolute Error (MAE) of 43,224.97, and Mean Absolute Percentage Error (MAPE) of 0.25%. These values indicate strong predictive performance, validating the model's effectiveness in estimating property values. The developed system allows users to input property features to obtain estimated values and also provides functionality to load your property in the market after evaluation. This study contributes to enhancing property valuation practices by introducing a reliable, automated, and user-friendly approach to real estate pricing using machine learning.

.

TABLE OF CONTENTS

APPENDICES

AI	Artificial Intelligence
ML	Machine Learning
ZIM	Zimbabwe
CRISP-DM	Cross-Industry Standard Process for Data Mining
R^2	R-squared
MAE	Mean Absolute Error
RMSE	Root Mean Squared Error
MAPE	Mean Absolute Percentage Error
CMA	Comparative Market Analysis
RE	Real Estate
ANNs	Artificial Neural Networks
DFD	Data Flow Diagram

CHAPTER 1

PROBLEM IDENTIFICATION

1.1 Introduction

Property price forecasting is a critical but difficult process requiring full comprehension of many factor that affect the price of houses. Housing is a most critical factor as not only does it act as a home, but it is also an income generating source (Bowen, 2018). The current housing market rests primarily on manual evaluation systems, and these are fraught with numerous problems for the sellers, buyers, and real estate agents. Traditional property valuation methods are usually subjective, time-consuming, and highly dependent on human expertise. Such methods never capture the dynamic, ever-changing essence of real estate markets accurately.

There are many things that impact real estate, including the location of the property, material used to construct, legal ownership information, and so much more. These factors may vary significantly from one property to another, and it may be challenging to create a uniform, reliable standard for judging property values. It is thus possible that current techniques may be lacking in delivering precise or fair valuations and that there may be room for improvement in real estate valuations. This study aims to create a Property Assessment System using machine learning techniques to automatically improve the accuracy of property prices estimates in order to address these problems effectively.

The proposed system includes essential properties of property such as location, total square meters, bathrooms and bedrooms, floors (single story or multi-story), condition, year built, and year refurbished in forecasting property values. By analysing the factors, the system seeks to offer a more accurate and effective alternative for conventional hand methods. Validity of a machine learning system should take into account the inputs used, the selection and utilization process, and the system's output (ROOT, et al., 2023). The present study also investigates the use of machine learning to property valuation process in determining the best model in forecasting property value based on its attributes (Ja'afar, 2021) . Through the implementation of advanced machine learning technique, this research aims to transform the conduct of property valuation. The aim is to create a system more transparent, data-driven, and scalable and therefore easier for every interested party in the property market e.g., buyers, sellers, agents to make informed decisions.

1.2 Background to the Study

Any item’s price is likely to alter based on conditions like advertising and market strength. Homes are made up of a variety of characteristics that add value to their asking prices (Begum, 2022). Valuers must physically visit the property site in order to anticipate home prices in the present home market. They accomplish this by applying conventional methods of property valuation, such as the cost approach and comparative market analysis (CMA). This physical site attendance sometimes makes it hard for Valuers to attend all the clients within a day, leading to queues and increasing work pressure. This forces them to deal with the clients the following day or postpone the service. It also takes a long time for real estate valuers to determine the right price for a property due to factors such as the lack of comprehensive and up-to-date property data, the influence of informal settlements on property values, and the impact of economic factors like inflation and currency devaluation. Homeowners then have to wait for weeks or months to get the service and feedback, which delays the selling and buying process. Due to this long waiting, some home owners end up losing interest and motivation to sell, considering competitors, or cancelling the deal.

The extended waiting period causes a devaluation in the property's value, due to market price volatility. Opportunities lost occur when home sellers do not catch the attention of prospective buyers who are specifically looking for homes within that range. Applying machine learning methods in property price prediction has attracted significant attention during the past few years (José-Luis Alfaro-Navarro, 2020). Lack of data causes inadequate support for important contributions concerning the factors affecting property value (Goh, 2024). The process of property valuation in real estate is essential for the economy because it affects different systems such as taxation, insurance markets, infrastructure development, and accounting (Gružauskas, 2020).

1.3 Statement of the Problem

The property market is vulnerable to extreme risks as a result of market price fluctuations, which may influence property prices during long valuation procedures. Manually driven systems used conventionally for property price prediction are susceptible to human mistakes, time-consuming, and raise work pressure, thereby leading to inefficiency. Such limitations are likely to overestimate or underestimate properties, to the disadvantage of both buyers and sellers. Moreover, loss of interest by property sellers, cancellation of transactions, and lower profits for real estate agents are consequences of delay and inaccuracy in valuation. To overcome these challenges, there is a need to develop and apply a model that utilizes machine learning in a precise, efficient, and reliable manner in property price measurement and thereby eliminating issues resulting from the traditional approach.

1.4 Research Objectives

The aim is to design a property evaluation system. The objectives: -

1. To analyse the different artificial intelligence methods that can be used in property evaluation.
2. To design and implement a Machine learning model (Random forest) for property evaluation system.
3. To evaluate the effectiveness of using Random forest in property evaluation for Real Estate.

1.5 Research Questions

1. What are the advantages and disadvantages of present Artificial Intelligence techniques used to evaluate real estate?
2. What key property attributes should be considered while developing a machine learning-based property evaluation system to ensure reliable property evaluation?
3. What are the expected outcomes of using the designed machine learning-based property evaluation system in terms of accuracy, consistency, and reliability on the real estate market?

1.6 Research Propositions/ Hypothesis.

Hypothesis 1: Existing AI property evaluation techniques have useful advantages like enhanced accuracy, reduced cost and time for appraisal, and greater consistency and scalability than the traditional techniques. Their performance, however, also depends equally on the algorithms employed, e.g., Random Forest, and are prone to being adversely affected by data quality and availability issues. Both indicate the promise as well as the limitations of AI in present-day property evaluation.

Hypothesis 2: The incorporation of the most determinant property attributes in a machine learning algorithm greatly enhances the performance of the property evaluation system.

Hypothesis 3: There will be fewer failed transactions of property due to delayed evaluation when a machine learning-based model is used to value property.

1.7 Justification/ Significance of the Research

This research is important as real estate markets become more sophisticated and rational property appraisals are critical to the well-informed investment decision-making of investors, buyers, sellers, and legislators. Traditional methods rely too often on subjective judgment and outdated data, which can lead to unsound and inconsistent results. The research employs machine learning to create a system which can manage a huge amount of data—like market trends, location data, and economic factors to provide adequate, impartial, and responsive evaluations of properties.

The actual value of this system is that it possesses the capability to transform the property sector. It can make assessment more transparent, non-discretionary, and give cost-effective and scalable solutions. It can make the market efficient, enable better financial and policy decision-making, and property management and assessment innovation and sustainability by giving stakeholders data-driven insights.

1.8 Assumptions

1. **Availability of Data**: There is sufficient, correct, and representative historical property data (e.g., location, size, age, trends) available to train and test the model.
2. **Applicability of Model**: Machine learning models can make valuation projections correctly from pattern identification in property data.
3. **Accuracy of User Input**: Users will be entering truthful, relevant, and accurate parameter values while running the system.

4. **Market Stability**: General economic and property market conditions are relatively stable during the time of system deployment.
5. **Stakeholder Involvement**: It is assumed during the research that the stakeholders, like property agents, vendors, and buyers of property, will utilize the system and provide feedback.
6. **Appropriate Features**: Identified input features completely represent the key drivers of property prices.
7. **Technical Infrastructure**: Adequate computational resources exist for development, deployment, and up-scaling of models.

1.9 Limitations/ Challenges

Data Quality and Availability: Completeness and quality of the training data are an important contributor to the efficiency of the system. The use of incomplete, outdated, or even biased data sets can lead to incorrect evaluation.

Feature Selection: Identification of the most suitable features for property valuations might be difficult. Irrelevant or redundant features may have a negative impact on the interpretability and performance of the model.

Model Complexity: Over complexity of machine learning models also causes overfitting. Developing a model that is accurate and generalizable is the central issue.

Interpretability: Uptake and acceptance may be hindered by the fact that most machine learning algorithms, including deep learning models, are "black boxes" and it is difficult for users to see how the decisions are built.

Regulatory Enforcement: There are numerous regulations controlling the real estate sector. The system needs to be complying with statutory rules and moral codes but can be bureaucratic.

Market Dynamics: Real estate markets are subject to numerous outside variables (local policies, economic climate, etc.) open to unpredictable changes and make the model unusable in a short time horizon.

Integration with Existing Systems: Integration into existing property evaluation processes and systems may be technical and operational challenging.

User Acceptance: User acceptance among buyers, sellers, and agents will be the system breaker or maker.

Computational Resources: Certain research environments are constrained by the high computational resources and ability needed to train complex machine learning models.

Ethical Issues: It should not reflect amplifying bias which has been present in past data so it will not render discriminatory race, gender, or income judgments.

1.10 Scope/ Delimitations of the Research

Target Residential Properties: The research will be looking into the residential property valuation and not industrial and commercial property so that there is a clear strategy.

Geographic Limitation: The research will be looking at just a given geographical location so that the data becomes significant and manageable, taking note of the fact that property prices vary geographically.

Data Sources: The research will take into account publically available datasets, existing real estate listings, and past sales data. Proprietary data sources or specialized databases will not be taken into account since they are not accessible.

Machine Learning Techniques: The research will take into account particular machine learning algorithms and rule out other practices such as deep learning or reinforcement learning for simplicity and readability.

User Interface Design: Even though the research will make a user interface design of the system an observation, designing and implementing it are beyond project scope.

Temporal Scope: The research will employ data in a time frame that the model mimics the current market trend and scenario.

Exogenous Variable Reduction: The study will eliminate certain of the exogenous variables like economic downturns, natural disasters that would impact the value of property but utilize innate property traits.

1.11 Definition of terms

Real Estate

Real estate is property and anything physically affixed to it, such as buildings, homes, and natural resources. Physical property as well as legal rights to the property of ownership or possession are encompassed in real estate.

Property Evaluation

Property appraisal refers to the determination of the value of a property, whether residential, commercial, or industrial property. This is normally undertaken based on reviewing various aspects such as location, size, condition, market trends, and comparable sales.

Machine Learning

Machine learning is an artificial intelligence discipline that deals with creating algorithms which enable machines to learn from and make predictions or make decisions based on data. It enables systems to learn automatically over time rather than being clearly programmed for a task.

Dataset

A systematic collection of facts (e.g., real estate records, sale prices, location features, and market trends) used for model training, model validation, or model evaluation. In real estate valuation, it usually includes characteristics such as size, age, features, neighbourhood, and past prices so that algorithms can learn patterns and make predictions.

Data Features

Features are individual measurable properties or attributes used in machine learning algorithms. In property assessment, features may be location of the property, total square meters, bathrooms, bedrooms, price of the property, floors (single floor or multiple floors), condition, year built, and year renovated.

Training Data

A training dataset is a collection of data used to train a machine learning algorithm. The model is made up of input-output pairs that learns to map inputs (features) to outputs (property values) based on historical data.

Validation Data

Validation data is another data that is used to calculate the accuracy of the machine learning algorithm during training. It is used to improve the model parameters as well as to prevent overfitting.

Test Data

The final performance of a trained machine learning model is assessed with test data. It gives an unbiased measurement of the predictive performance of the model and isn't used for training or validation.

Comparative Market Analysis (CMA)

A traditional appraisal method that involves comparing a property with recently sold similar properties ("comps") in order to reach an estimate of what it would fetch in the marketplace.

Predictive Modelling

A statistical method called predictive modelling builds a model that can forecast future occurrences based on past data. In real estate appraisal, it involves the application of machine learning algorithms to forecast property values based on input features.

Real Estate Market Trends

Time series of property demand, supply, and price, designed to input predictive models (e.g., housing price impact of interest rate).

Property Valuers

Professionals (also referred to as appraisers or assessors) estimating the market value of property from experience, practice, and consideration of such factors as location, comparable sales, market conditions, and property condition.

Random Forest

Is a machine learning algorithm that makes predictions using a combination of numerous decision trees. It improves the accuracy by combining the output of various trees to eliminate errors and overfitting.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

The real estate industry has warmly embraced technological advancements, particularly in property valuation. Traditional approaches to valuation, while common, have been criticized for subjectivity, inaccuracy, and inability to capture price volatility. This has made automated valuation systems, particularly those based on artificial intelligence (AI) and expert systems (ES), increasingly popular (Kusuma, 2018).

This literature review aims to critique earlier related research findings and integrate the most valid evidence, which will subsequently be combined in order to advance present knowledge of intervention with an amalgamation of various types of sources existing literature on property price prediction, emphasizing the traditional valuation methods and aggregating algorithms of machine learning. In order to develop authentic evidence, the review attempts to collect, analyse, and synthesize data of different published reviews of literature associated with the system for property appraisal, forecasting property or real estate prices using machine learning. By synthesizing empirical and theoretical research, the review is going to establish the premise for the proposed system for property evaluation in Real Estate using Machine learning.

2.2 Real Estate

Development of the Zimbabwean real estate market begins from colonial times, where land ownership was defined by racially discriminatory methods to accommodate settler societies. Land occupation based on race was reconfirmed in the Land Apportionment Act of 1930 that reserved good-quality land for cultivation by white settlers, thus establishing a double property regime that has affected land ownership patterns up to modern days (Wikipedia, 2024). Since the nation became independent in 1980, urbanization and land reform began to change the industry so that there would be increased involvement in property ownership and development. In 2024, the Zimbabwe government introduced a land title system that would allow black farmers to own leased land in full for collateral use and financing property (APNews, 2024).

Zimbabwe real estate is currently an impetus of the economy since it allows employment and investment. The four major categories are residential, commercial, industrial, and agricultural, and the most sought-after among them is urban property due to population growth and rural-urban migration. From 33% in 2012 to 38.6% in 2022, the urban population in Zimbabwe grew, as reported by the 2022 national census, and it exerted pressure on the housing stock of the urban areas (ZIMSTAT, 2022). The capital for real estate comes from institutions and private investors, and increasingly there is focus on high-density residential as a means of addressing the country's housing shortage. However, the industry is characterised by usual problems like ill-specified regimes of land holding, restricted mortgage capital, and economically volatile property prices (ZIMPropertyDigest, 2025).

The trends are through government intervention into computerizing land records, openness, and legalization of informal settlements. The dominant firms in the real estate market are Old Mutual Property, Dawn Properties, and Knight Frank Zimbabwe and they deal in property development, management, and investment in capital cities (ZimAdvocate, 2024).

2.3 Challenges in Real Estate

Zimbabwean property industry is faced with sophisticated challenges to its development and survival. Included among them is the lack of established property and land title systems. The majority of the urban land, as well as the squatter settlements, do not have legally enforceable property records restricting individuals and institutions from leveraging property as a security for investment mobilization or obtaining rights of long-term occupation. Property insecurity causes investment aversion and discourages property transfer (Tsabora, 2016). Financial insecurity in the form of exchange rate volatility and inflation has negatively affected the property market. They depreciate the value of property and rental income, and developers and investors cannot make long-term plans (Zhanda, 2020). Mortgage finance is also limited, with tight lender terms and high interest rates, thus low affordability of home ownership.

Urbanization has come with the growth in infrastructure, and there has been no provision for amenities and houses in urban areas like Harare. The increase in population has overtaken house provision, and therefore sprawl and clustering of unplanned urban settlements without amenities have resulted (Average, 2019). Corruption and inefficiency in land management and building permits then decelerate the industry. Politicized allocations and politicized land sales have deterred genuine developers and investors (TransparencyInternational, 2024). Inability to digitize land records leads to delays and litigation. The efforts to develop digital land registries have been slow, which makes property transactions not be transparent and efficient (Gambe, 2015). The delays in the process deter the overall public from having faith in the housing system due to previous land scandals and house scheme failures. This kind of distrust discourages entry and discourages positive growth of the housing market (Kwangwama, 2024).

It calls for an even blend of land administration reforms, economic stabilization, infrastructure building, and leveraging technology in making efforts to correct the property market and promote inclusive urbanization in Zimbabwe.

2.4 Machine Learning in Real Estate Valuation

Machine learning has transformed real estate appraisal since it can handle enormous data and recognize intricate patterns, which are difficult to detect with the human eye. Wei et al. (2020) also add that ML models, i.e., ANNs, outperformed traditional hedonic models in estimating property value. ANNs, which model the neurostructure of the human brain, can detect non-linear relationships between property attributes and prices without specification of functional forms. The non-normalization ability of ANNs to easily learn advanced dynamics in the property market makes them very appropriate for large-scale property valuation issues. Researchers do admit that there are some challenges, though, and that these are that it requires humongous, high-quality data sets and sheer processing capacity to actually train them.

Machine learning methods used in real estate appraisal also have some disadvantages despite their benefits. One of the biggest limitations is the "black box" nature of most machine learning models, such as artificial neural networks (ANNs), which will make it challenging to interpret why their outputs are what they are (Wei, 2020). Since stakeholders expect clear reports of value implications, incompleteness of this kind can actually be a huge liability for the property sector. Second, the quality and quantity of training data also affect the performance of ML models. The performance of the models may be compromised by biased or inaccurate predictions with changing or incomplete data. Other than that, the expense of compute power to implement and deploy sophisticated machine learning models could be exorbitantly expensive, especially for a small firm or in countries with underdeveloped technological infrastructure.

The advantages of machine learning in real estate appraisal are enormous even though these challenges abound. Machine learning operations, by automation and removal of redundancy in the process of appraisal, can cut down time and labor efforts by manual analysis. They can process huge amounts of data and therefore can prove to be useful for mass appraisal application in high-transaction cities. Wei et al. (2020) note that ML models, and ANNs in particular, can be applied to make better predictions by finding nuanced patterns between the characteristics of a property that cannot be found in normal models. This is particularly relevant in new markets where property values are resolved by anything from social trends to the quality of the neighborhood and economic status. In addition, valuations of property can be made more accurate and sophisticated by the integration of machine learning (ML) with other technology like geographic information systems (GIS).

2.5 Property value estimation review

With respect to literature analysed above, the researcher learns that the majority of authors made use of machine learning techniques only, although some of them applied them in conjunction with robotic process automation, artificial neural networks, and hedonic regression. I want to recognize the work of different studies, for example, (Mora-Garcia, 2022), in forecasting real estate prices across the different nations mentioned above using machine learning techniques. (Khan, 2023) describe that, by the efficient use of the enormous amount of information available for houses, machine learning algorithms can offer an improved and more accurate means of house value prediction (Satish, 2019). Utilizing measurements such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE) value, the authors critiqued the performance of seven machine learning algorithms. Gradient boosting, Ada boost, random forest, ridge regression, elastic networks, and neural networks are some of the machine learning models that have been experimented. House area, property area, bathrooms, bedrooms, house price, location, floors, and property title are some of the characteristics or features of the dataset used. Random Forest recorded the maximum value of R square among all other models according to the study findings. The authors recommend using this model to construct an expert system in property valuation.

According to findings of some researchers, such (Quang, 2020), (Khan, 2023), (Satish, 2019) and (Mora-Garcia, 2022), who were comparing the performance of various machine learning models, the random forest performed better. It is these findings that have guided the use of the random forest approach in the current study.

There was also research conducted by (Shah, 2020). The authors describe that location, number of rooms, carpet area, and age of property are just a few of the numerous variables which have to be taken into account while estimating the value of a home. Real-time data extraction will be made possible using Robotic Process Automation (RPA) along with the CatBoost algorithm. In robotic process automation, software robots are employed to automatically execute data extraction tasks and machine learning procedures employed to forecast home prices from datasets. It is thus possible to conclude that this study aimed at reducing errors exposed by the manual techniques employed in the current industry and ensuring the problems faced by customers are mitigated. Upon analysing this research, the new system to be established is anticipated to respond to customers' needs and assist them in making their decisions.

Out of all the literature that were used within this study, two authors (Khan, 2023) & (Ake, 2022) developed graphic user interface for their model. The end-user is only able to interact with the GUI to comprehend the price of a house. The system forecasts the price based on parameters that are input by the user (Ake, 2022). It should be provided with an interface that is user-friendly for users to engage with the house price prediction model. Users may submit important details about a residential house, including its size, the number of bedrooms, bathrooms, and region, through a GUI-based system. These were submitted and passed through the Random Forest model, which gave an estimated value of the house according to the information available. These findings inspired the development of appealing interfaces to establish an even user experience.

2.5.1 Machine learning

In order to improve computer capabilities and give computers the ability to learn and make decisions like humans do, modern machine learning research focuses on teaching computers to learn new things by processing different types of data, including text, images, and numbers (Mao, 2023). Machine learning is the process of giving computers structured data so they can use various contemporary technologies to learn and make better decisions. Consequently, the creation of a precise prediction model for estimating home prices required a deep comprehension of and commitment to putting into practice suitable machine learning techniques (Khan, 2023). A subfield of artificial intelligence called machine learning uses various technical instruments and algorithmic procedures to glean valuable knowledge from data. It can be used in the big data field because it is very hard to manually examine such massive amounts of data (Ake, 2022).

Predictive models use data for training, which yields somewhat precise outcomes. Machine learning involves developing these models from data and using the data to predict new data (Liu, 2023). It has improved automation in its ways to make work easier as the world pushes ahead with using modern technologies. At that time, we provide a housing cost prediction model to help a real estate agent or home seller get better information based on the house's valuation. (Satish, 2019). Artificial intelligence includes machine learning as a subset. It provides the system with the ability to learn, allowing it to perform better on its own without needing to be explicitly programmed (Shah, 2020).

2.5.2 Hedonic regression

The impact of particular attributes on house values has been ascertained, and variations in dwelling prices have been explained using the hedonic pricing models (Wei, 2022). Hedonic pricing has limitations when it comes to identifying the non-linear correlations between a house's attributes and price. Hedonic model performance can also be hindered by variables like outliers and model specification processes. Hedonic processes have defined functional forms and rigorous presumptions (Yazdani, 2021).

2.5.3 Robotic Process Automation.

The robotic process is the Automation is the process of automating tasks and services that replicate human labour. (Ribeiro, 2021). Machine learning algorithms are utilized to forecast home values based on the given dataset, and software robots are employed in robotic process automation to automate data extraction chores. Using software robots to automate business processes and minimize human labour is known as robotic process automation (Shah, 2020).

2.5.4 Artificial neural networks

The structure of the human brain served as the inspiration for artificial neural networks, a deep learning technique. ANNs are designed to simulate the way the human brain learns. In order to determine the associated output for any new inputs, the specific objective of developing a neural network is to ascertain the relationship between inputs and outputs in the given data set (Abdolrasol, 2021). It is not necessary to assume explicit functions between the study's input and output when using ANNs. An artificial neural network (ANN) models complex and non-linear relationships by directly observing data. It requires a number of processing neurons to be connected together. There are layers of hidden input and output in an artificial neural network. There are one or more neurons in each layer. Every neuron in a layer is linked to every node in the layer next to it. The weight given to each connection determines how strongly each neuron influences the others. The number of neurons and layers in a neural network can vary. The number of neurons in the hidden layers determines the model's width, while the number of layers determines the model's depth (Yazdani, 2021).

2.6 Application of Random forest in Real Estate

Random Forest, a machine learning ensemble method, has become increasingly used across the real estate sector to apply to tasks from valuation of property to price estimation and market analysis. It is well adapted to the complex, non-linear relationships inherent in real estate data because it can handle large data sets.

Used Random Forest amongst all other machine learning models to predict house rental rates in Ghana (Adzanoukpe, 2025). The article brought out how the model can identify complex patterns in the real estate market as well as how location, number of bedrooms, and furnishing status of a home are some of the key determinants in determining the prices of rentals. Thus, in anticipating San Francisco Bay Area rentals per square foot, Random Forest regression has been compared to customary multiple regression models (Waddell, 2020) They concluded from their study that Random Forest performed more accurately than customary models and thus validated the excellence of Random Forest to handle dissimilar real estate data. Random Forest was utilized in predicting rentals and selling price indicators of 1.7 million properties in the Swiss real estate market (Moosavi, 2017). The study achieved a median absolute relative error of 6.57%, reflecting the model's accuracy in mass property valuation exercises.

These also identify Random Forest's power and flexibility in a range of applications of real estate from country-level property appraisals to neighbourhood-level rental rate forecasts. Random Forest is a convenient tool for real estate practitioners and policy makers who would want to improve decision-making processes because it can deal with a number of markets and complexities in the data.

2.7 Advantages of Random forest in Real Estate

Random Forest ensemble learning technique has grown in popularity within the real estate industry due to its high reliability and accuracy in the analysis. It may be used for most of the applications such as market prediction, price estimation, and property appraisals because it is capable of handling complex sets of data.

1. **Handling Complex Interactions**

Interdependencies and correlations among variables like location, size, and facilities are of non-linear forms in real estate data. Random Forest can manage such complexity without appeal to carrying out monstrous pre-data processing. For instance, the author demonstrated Random Forest performing better than other models for forecasting the price of houses in the Indian real estate market, portraying its capability to learn complex patterns (Putatunda, 2019).

2. **Resistance to Overfitting**

Through averaging the prediction of multiple decision trees, Random Forest minimizes the risk of overfitting and hence improves the model to be more generalizable to new data. Random Forest worked well using various sample sizes in predicting rent for an apartment and was also stable (Yoshida, 2024).

3. **Missing Data and Outliers**

Random Forest is resistant to missing values and outliers, which are characteristics of the data. It does contain some algorithmic framework within it so that it can take reasonably mediocre quality data and turn it into extremely high quality predictions.

4. **Feature Importance and Interpretability**

The model allows the stakeholders to identify important drivers of house price by looking at relative importance of each feature. Interpretability directly facilitates making informed decisions regarding property development and investments.

5. **Performance and Scalability**

Random Forest is able to handle huge amounts of data with ease and is therefore well equipped to handle complex real estate markets. Random Forest was employed by a research scholar in a database of over 62,000 records with very good house price prediction accuracy, proving to be a scalability test for it (Jha, 2020).

6. **Ease of Use**

Random Forest flexibility can make it applicable to a variety of real estate uses such as price prediction, trend in the market, and investing risk evaluation. Dynamic locality-enabled Random Forest models with exact house price prediction, demonstrating its flexibility (Coleman, 2022).

2.8 Integration of ML

Application of machine learning (ML) in Zimbabwe in the property sector has boundless possibilities for property valuation to become more precise, the market to become more affordable, and operations to become enhanced (Yazdani, 2023). The conventional property valuation models may not have the potential to cope with the volatility of real estate markets because they are static models and manual valuations. ML is data-driven and hence can give more accurate property prices and be sensitive to changes in market trends.

Some of the features of house prices that can be established through machine learning algorithms include property age, address, area, number of bedrooms and bathrooms, and closeness to amenities like shopping malls, hospitals, and schools. Machine learning algorithms are able to process extremely large data sets and therefore can identify patterns and relationships that would be impossible to do with other algorithms. For example, due to their ability of modelling nonlinear composite relations between variables, algorithms such as Random Forests and Gradient Boosting have been able to achieve enormously successful performances in terms of property value prediction.

In the same way in Africa, ML has also been applied in property valuation as seen from research. A study utilized CatBoost to forecast rental value of houses with an R^2 of 0.876, meaning that it performed amazingly well to forecast (Adzanoukpe, 2025). A web application in South Africa utilized an ML algorithm to forecast listing price of houses to help sellers with competitive pricing and reduced days on market (Bax, 2019).

ML use in the Zimbabwean property market can lead to several benefits:

1. **Improved Valuations**: ML algorithms are able to forecast property value more precisely based on more variables and their complex inter-relationships.
2. **Market Transparency**: Value-based valuations are able to generate information asymmetry in such a manner that buyers and sellers make sound decisions.
3. **Operational Efficiency:** Automation of the valuation process can be time and energy-conserving for property agents.
4. **Minimizing Risks:** Bank valuations can result in the better estimation of lending risk of real estate transactions.

In order to realize these gains, Zimbabwean stakeholders, including government agencies, banks, and real estate associations, need to put in place data collection infrastructure, facilitate data sharing, and conduct training in real estate applications of ML. Joint ventures with universities and technology firms can also be utilized to adapt and implement ML models to the local situation.

2.9 challenges of implementing ML in Real Estate in Zimbabwe

Implementation of machine learning (ML) in Zimbabwe's property industry has certain challenges, spurred by the unique socio-economic and infrastructural circumstances in Zimbabwe. They consist of data availability and quality, technology infrastructure, regulatory frameworks, human capability, finance, and organizational preparedness.

1. **Availability and Quality of Data**

One of the fundamental issues is a deficiency of high-quality and detailed real estate information. The majority of the real estate deals in Zimbabwe are not reported or reported outside of the official system, producing incomplete datasets. Incomplete data prevents it from being easy to build robust ML models of property value and market assessment (Bhanye, 2024). The entire African real property market has a low transparency rate and rare presence of robust sources of information (Mupondo, 2023).

2. **Technological Infrastructure**

The infrastructural context in Zimbabwe also poses concerns in terms of the deployment of ML. Non-existence of high-speed internet connectivity and suitable computing resources confines the application of complex ML algorithms. Erratic and poor power supply, as well as non-existent digital infrastructure, are also to be held responsible for such constraints (Chigwada, 2022).

3. **Regulatory Frameworks**

Lack transparent and nurturing regulation of AI and ML technology in Zimbabwe generates ambiguity in stakeholders. Lack of data privacy, security, and ethics legislation can discourage real estate firms from embracing ML solutions. Such regulatory ambiguity can discourage innovation and implementation of ML in the real estate sector (Chigwada, 2022).

4. **Human Resource Constraints**

There is lack of data science, ML, and real estate analysis professionals in Zimbabwe. Skills in programming, analysis, and domain are the ones needed to be able to use ML effectively. In the absence of such skills, industry-specific implementations of ML models that can be used in the real estate industry cannot be created and deployed (Hlongwane, 2024).

5. **Cost Constraints**

ML technology adoption is costly, and this is a major barrier. Funds are needed to acquire data, infrastructure, software, and training. For most property companies with very low profit margins, the investment will be too expensive (Chigwada, 2022).

6. **Organizational Readiness**

There may be resistance to change in organizations when employing ML. Historic use of property can be established and there may be resistance to algorithmic decision-making. Cultural lag needs to be addressed by utilizing more change management and by demonstrating material benefits of ML implementations (Chigwada, 2022).

2.10 Gap in Literature

The lack of adequate literature within the Zimbabwean real estate market portrays a serious deficiency in the usage of high-end, data-demanding methodologies on the valuation of property. The conventional methods of property valuation are usually based on subjective inference and narrow datasets and it makes property price unstable and incompatible (Paradza, 2015). The incompatibility portrays the necessity to embrace machine learning (ML) techniques in an attempt to make property evaluations more compatible and dependable.

Regression, decision tree, random forest, and ensemble methods like XGBoost are some of the machine learning algorithms that have outperformed in estimating real estate values globally (Ho, 2021). As an example, research has established that ensemble methods specifically, XGBoost and Random Forest are more accurate and reliable than the previous models in the case of mixed housing markets (Pastukh, 2025) (Sharma, 2024).

In Zimbabwe, application of such ML techniques will be able to address current property valuation issues using available data to come up with more precise and consistent pricing models (Özalp, 2024). Application of such models will put stakeholders sellers, buyers, investors, and policymakers in a better position to make sound decisions and create a more transparent and effective real estate industry (Coleman, 2022).

2.11 Chapter Summary

This chapter provided a synoptic literature review of the subject, both theoretical and empirical insights. It commenced with defining the purpose and significance of the literature review, prior to delving into discussions on the theoretical background that proffered the relations between variables of interest. Terms utilized in the study were explained with caution and defined. In addition to that, the chapter experimented with the issues encountered in the private insurance sector and concluded how Machine Learning techniques can be employed against such issues. The various model options available for use are also presented. These include data gathering procedures, data analytical techniques, study design, and ethical considerations, which will be elucidated in detail in the next chapter.

CHAPTER 3

METHODOLOGY

3.1 Introduction

This chapter presents the machine learning method in creating a Property Evaluation in Real Estate, a system to assist stakeholders in making accurate property valuations with respect to application of machine learning approach, which rely on data-driven models of property evaluation. The chapter provides an easy overview of the research project, presenting the strategies, tools, and procedures required to develop, create, and implement the system. By blending the Agile philosophy of iterative and elastic development and CRISP-DM (Cross-Industry Standard Process for Data Mining) rigid framework, this method guarantees a right and effective process.

3.2 System Development

System implementation of Property Evaluation in Real Estate with Machine Learning harnesses the synergy between Python, Jupyter Notebook, and Streamlit to offer powerful data analysis, model development, and user interface deployable that is ready to be fine-tuned. Jupyter Notebook is ideal for the initial half of system development, namely data analysis, pre-processing, and building the machine learning model. The graphical interactive interface of the Jupyter Notebook helps scientists properly analyse real estate data in a bid to assess the property. Jupyter supports libraries like Matplotlib or Seaborn in making informative plots, Pandas for data manipulation, and Scikit-learn for applying machine learning. Properties of statistical models are applied using these libraries. Streamlit transforms the system into an interactive web-based application once it has developed and tested its predictive machine learning (ML) model in Jupyter Notebook.

Streamlit is employed to create an interactive graphical user interface where the users, i.e., the prospective purchasers or the valuers, can enter respective parameters of the property (e.g., square footage, area, bedrooms, etc.) and receive instant price quotes. These are generated by the ML model upon the prediction from data.

The combination of Jupyter Notebook and Streamlit provides an end-to-end development cycle from data preprocessing, model training, to deployment without losing usability or technicality. Streamlit turns model deployment into a complete application as a complete functional end-to-end web application with features like real-time prediction view, visual analytics, and end-user usability without having to be technical. By undergoing this process of development, the researcher can create a system for property evaluation closing the gap between top-level data analysis and on-the-ground usability through efficient, useful, and intelligent means. It also allows for rapid iteration, testing, and refinement of the system based on user input and changing trends in the data so that the model may be utilized in the high-speed real estate market environment.

3.3 Methodology Selection

In the next section, we discuss the rationale behind why methodologies used in system development of Property Evaluation in Real Estate using Machine Learning were selected. For a productive, flexible, and iterative process of system development, two methodologies were selected: CRISP-DM and Agile. The two-methodology approach provides the structured process of data-mining and the flexibility of agile development so that precise model development can be done along with adaptive system development.

3.3.1 Cross-Industry Standard Process for Data Mining (CRISP-DM)

Business understanding, data understanding, data preparation, modelling, evaluation, and project implementation are the six broader phases of the model. Following is a diagram showing CRISP-DM processes:

Figure 3.1: CRISP-DM processes

3.3.1.1 Business understanding

The goal of this CRISP-DM step is to clearly define the project goals, requirements, and limitations. The primary goal of this study is to develop a property price prediction model so that the problems of conventional practices in the real estate sector are minimized (Verma, 2023).

Specifically, the requirements are to determine the factors that affect property prices, use a machine learning technique (random forest) to predict property prices, and visualize and report the model output. The system will provide evidence-based forecasts and assist decision-makers with interpretable model outputs and visual insights.

3.3.1.2 Data understanding

The data scientist must decide on the data to utilize that serves the objectives set down in the business understanding step in the first step of this phase, data gathering. The data could be obtained through research specialists, extracted from data warehouses, or web scraped.

The process of obtaining and analysing the data pertinent to the analysis is known as the data-understanding phase. Data used in this study are historic real estate information, square foot, bedrooms, bathrooms, floors, location, year built, and year renovated.

The downloaded United States dataset from Kaggle will be properly pre-processed in a manner to render it worthwhile and suitable for Zimbabwean standards before models are created.

3.3.1.3 Data Preparation

The data set would have already been chosen and processed by now. The second step is data selection, which involves choosing the data relevant to the project; for our example, it is choosing the features of significance that are useful in predicting the price of the property sold with a minimum of errors. The parameters on which the data (features) are accepted or rejected must be well defined in the documentation of the process of data selection.

Data preparation stage aims to convert the data into a shape in which it will become analysable. Data cleaning and feature engineering are some of the processes that are carried out during this stage. Data cleaning is a name given to the process of dealing with missing numbers, data inconsistencies, and outliers. Feature engineering is the discovery of appropriate features and creation of new ones that may improve the model's predictive power.

3.3.1.4 Modelling

The modelling process starts with the choice of technique to be employed. A random forest model will be employed in this project for constructing the house price prediction model, and no missing variables will be accounted for by the modelling approach. To avoid overfitting in our model, the dataset is split into halves of training and testing (Angrick, 2022).

Machine learning techniques were used in the development of the property estimation model. The machine learning package, employing the random forest algorithm, will be used for model development. Once the data supplied is divided into training and test sets, the models are trained on the training set and tested on the test set.

3.3.1.5 Evaluation

This phase is about reviewing or testing the model's output and evaluating how well it achieves the business goals. To verify whether our model meets the project goals, it has to be tested using either the test dataset or a new dataset.

This stage evaluates the performance of the model created. Depending on the analysis of the advantages, disadvantages, and weaknesses of the model, any enhancements or changes that are needed will be implemented. The ML model will be evaluated using measures like R^2 Score and RMSE (Root Mean Squared Error).

3.3.1.6 Deployment

During this stage, a deployment plan is established where the project will be implemented to the client or any user. It should be indicated if the project will be released in the form of an online or website application or a running application. The model is then monitored and updated to see if it's doing its job and being beneficial. Property price prediction model is utilized in a real environment in deployment. A Python library called Streamlit will be used in order to integrate the model into an internet application.

3. 3.1.7 Justification for the methodology used

The writer learnt that CRISP-DM methodology has been widely utilized for many writers for nearly equally studied works and provides a scientific and systematic way of researching easily in the form of organization, i.e., the probability of more business-oriented approach and better output in shortest possible time; thus, the writer found it quite suitable for this forecasting study.

CRISP-DM is particularly well-suited to data-driven approaches like machine learning, where structured exploration, development, and testing of models according to business goals is supported. This approach supports efficient iteration and decision-making, which are most important in rapidly evolving domains like real estate valuation.

3.3.1.8 Advantages of CRISP-DM Methodology:

- It is an iterative process; the method allows for creating a long-term strategy through the use of short iterations at the beginning of working on the project.

- It is a successful and orderly methodology for all data science projects.

- CRISP-DM can be applied in projects of varying sizes and levels of complexity since it is scalable.

- It provides a standard framework for project planning and management.

- CRISP-DM is a tried and tested methodology; it minimizes the possibility of errors and delays, ultimately saving time and money.

- The CRISP-DM methodology places an emphasis on the business issue that the machine learning solution ought to resolve, which counts as one of its strong points.

- The CRISP-DM methodology is systematic in nature, but it is also flexible enough to be adapted by researchers to suit their particular requirements.

- Structured Approach: CRISP-DM methodology offers a structured way of developing and deploying a machine learning solution. It is a proven and tested framework that takes researchers through all the processes of the project, from problem determination to solution implementation (Wang, 2019).

3.3.2 Agile Methodology

The Agile Software Model is a dynamic and iterative system development process that emphasizes flexibility, continuous collaboration, and responsiveness to stakeholder input during the development process. This makes it especially appropriate for applications like the Property Evaluation in Real Estate Using Machine Learning, where both user needs as well as technical requirements may evolve as understanding is created.

In contrast to linear sequential approaches like the Waterfall model, Agile emphasizes the early and continuous delivery of functional system components. Agile is designed to deal with changing requirements, which is a must in the integration of machine learning into a system.

One of the key features of Agile is its iterative development cycle, which is commonly referred to as sprints. Incremental working parts of the ML model can be delivered regularly due to these time-boxed short iterations, usually lasting two to four weeks. Possibly deployable features—such as the user interface, or prediction module—are delivered, tested, and inspected at the end of each sprint.

Agile's flexibility makes it suitable for work where data-driven results (ML models) reveal themselves incrementally. For instance, new trends can be discovered during the analysis of data, where frequent updating of model features, pre-processing methods, or evaluation metrics is necessary. Agile provides the framework to incorporate such alterations effectively and quickly.

One of the primary advantages of Agile is its collaborative nature. Throughout the development of the machine learning-based system, cross-functional groups of developers, data scientists, domain experts (e.g., real estate agents), and stakeholders work in close coordination with each other. Regular sprint meetings, planning, and reviews ensure that technical development aligns with domain-specific expectations and user requirements.

Stakeholder and customer involvement is one of the cornerstones of Agile. End-users such as real estate evaluators or property investors are kept actively engaged through feedback sessions, user testing, and demo presentations. This keeps the evolving solution of real value and in alignment with real-world requirements.

The Agile methodology also values individuals and interactions more than processes and tools and promotes open communication, trust, and continuous knowledge transfer. The cross-disciplinary nature of this project—combining software engineering, machine learning, and real estate expertise—benefits immensely from this interactive process.

Agile promotes repetitive delivery of functional software increments. On this project, that would mean releasing functional versions of the system periodically for instance, a prototype with basic prediction functionality, then incorporating features for geographic filtering or interactive dashboards.

Agile also embraces continuous improvement by holding retrospective meetings at the conclusion of each sprint. Via retrospectives, the team may review its progress, recognize problems, and tune its methodology to be more productive and effective in future iterations.

Finally, cross-functionality of Agile teams ensures that development, data pre-processing, model training, interface design, and testing occur in parallel and in synchronization with one another. This reduces development silos and promotes faster, more integrated system development.

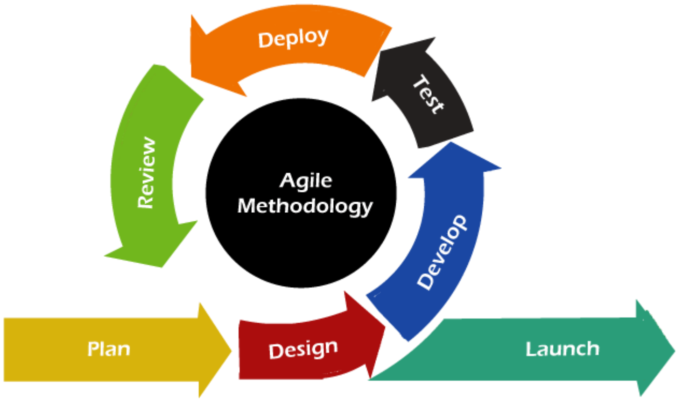


Figure 3.2: Agile Methodology

Agile has been utilized in this case because it is adaptable, iterative, and collaborative. These are required in a complex machine learning system based on predictive analytics. Agile offers ongoing feedback, agile planning, and ongoing delivery to gain a capable and user-centric Property Evaluation in Real Estate with the help of Machine Learning, and robust against changing project goals and knowledge gained from data.

3.4 Requirements Analysis

Requirements analysis phase is a critical role in defining the functionalities, performance needs, and restrictions of the property valuation system. Throughout the process, the stakeholders' requirements such as the sellers', buyers', real estate agents', and lawyers' are defined.

3.4.1 Functional Requirements

- Properties must be valued with regard to relevant factors, e.g., location, size, amenities, market conditions, and legality.
- It should have a good justification of why the property is valued.

- Users should input the data of the property and receive an automatic valuation.
- In order to predict the prices of properties from past facts and market trends, machine learning operations should be integrated into the system.
- The system should update itself with new data every day to provide more accurate valuations.

3.4.2 Non-Functional Requirements

- The system should be scalable to handle large numbers of property valuations in parallel.
- The system needs to be easy to use and offer an easy-to-use interface for technical and non-technical personnel.
- The system should offer proper data security and confidentiality, particularly for confidential property data.
- The system should offer integration with external real estate APIs and databases for fetching the market data in real time.

3.4.3 Hardware Requirements

- Intel Core i5 laptop/desktop processor or higher with a minimum of 8GB RAM

3.4.4 Software Requirements

- **Operating System:** Windows 10 or higher.
- **Programming Language:** Python3.13
- **Machine Learning Frameworks:** Scikit-Learn
- **Database:** SQLite Studio
- **Integrated development environment:** Jupyter Notebook
- **Deployment tools:** Streamlit, Visual Studio code, Browser

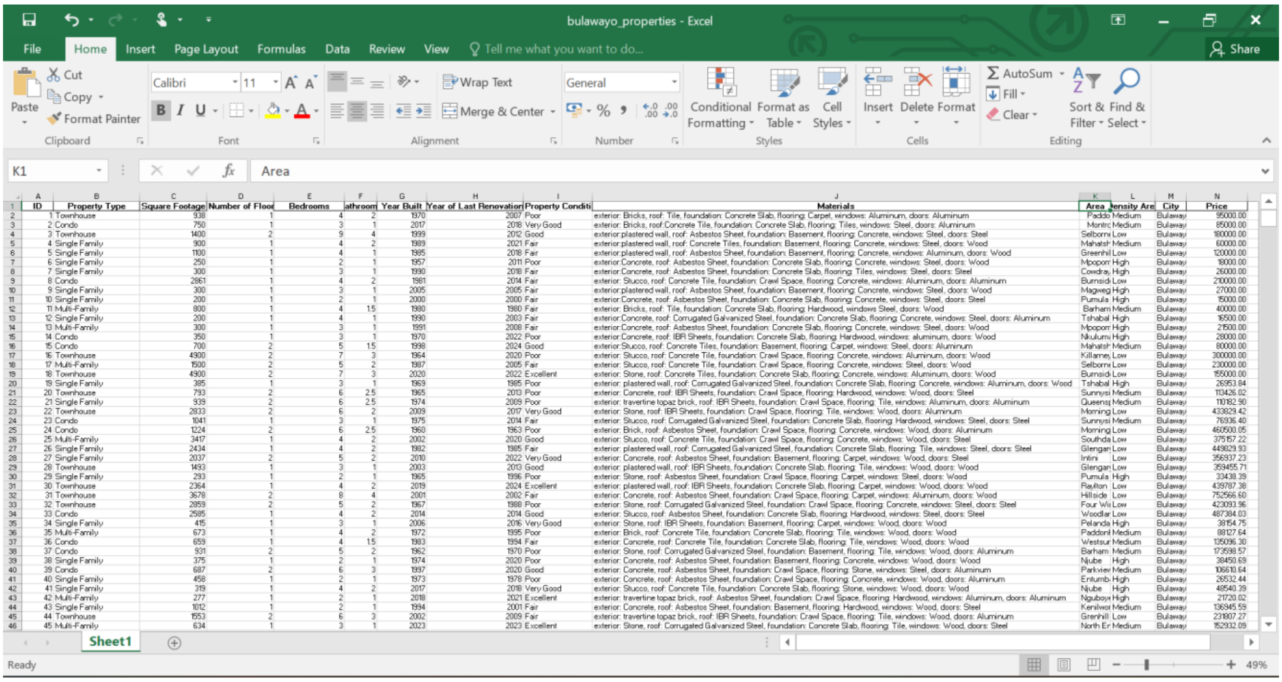
3.5 Data Understanding and Preparation

Data understanding phase is of utmost importance in defining the performance and effectiveness of property price predictive models with machine learning. In this phase, one collects, defines, explores, and validates data with regard to defining how suitable it will be to use in predictive model development. In this phase, focus is put on the attainment of knowledge of different aspects of the dataset utilized throughout the property price forecasting in Zimbabwe, being Bulawayo.

3.5.1 Dataset

Information used in this project involve major variables that have been found to affect the price of real estate. The variables are location and physical characteristics, which play a critical role in determining the value of properties in real estate. The data involve the following variables:

1. **Location:** Geographical location of where the property is situated, playing a central part in its market value on the basis of accessibility and facilities in the surroundings. The data include the below variables.
2. **Size (Square Footage):** Usually, size of property, with costlier properties being larger.
3. **Number of Bedrooms**: While deciding accommodation capacity, influencing demand and price by buyers.
4. **Number of Bathrooms**: Additional bathrooms are value-increasing, especially in big families.
5. **Year Built:** Older houses require renovation, something that will influence their market value.
6. **Year Renovated**: Newly renovated houses contribute to the value of property through new additions.
7. **Number of Floors**: Multi-story houses have excess space, something that influences their valuation.



ID	Property Type	Square Footage	Number of Floor	Bedrooms	athroom	Year Built	Year of Last Renovatior	Property Conditi	Materials	Area	ensity Are	City	Price
1	Townhouse	938	1	4	2	1970	2007	Poor	exterior: Bricks, roof: Tile, foundation: Concrete Slab, flooring: Carpet, windows: Aluminum, doors: Aluminum	Paddo	Medium	Bulaway	95000.00
2	Condo	750	1	3	2	2017	2018	Very Good	exterior: Bricks, roof:Concrete Tile, foundation: Concrete Slab, flooring: Tiles, windows: Steel, doors: Aluminum	Montrc	Medium	Bulaway	85000.00
3	Townhouse	1400	2	9	4	1999	2012	Good	exterior:plastered wall, roof: Asbestos Sheet, foundation: Basement, flooring: Concrete, windows: Steel, doors: Steel	Selborn	Low	Bulaway	180000.00
4	Single Family	900	1	4	2	1989	2021	Fair	exterior:plastered wall, roof: Concrete Tiles, foundation: Basement, flooring: Concrete, windows: Steel, doors: Wood	Mahatsh	Medium	Bulaway	60000.00
5	Single Family	1100	1	4	1	1985	2018	Fair	exterior:plastered wall, roof: Asbestos Sheet, foundation: Basement, flooring: Concrete, windows: Aluminum, doors: Wood	Greenhil	Low	Bulaway	120000.00
6	Single Family	250	1	2	1	1957	2011	Poor	exterior:Concrete, roof: Asbestos Sheet, foundation: Concrete Slab, flooring: Concrete, windows: Steel, doors: Wood	Mpopom	High	Bulaway	18000.00
7	Single Family	300	1	3	1	1990	2018	Fair	exterior:Concrete, roof: Asbestos Sheet, foundation: Concrete Slab, flooring: Tiles, windows: Steel, doors: Steel	Cowdray	High	Bulaway	26000.00
8	Condo	2861	1	4	2	1981	2014	Fair	exterior: Stucco, roof: Concrete Tile, foundation: Crawl Space, flooring: Concrete, windows: Aluminum, doors: Aluminum	Burnsid	Low	Bulaway	210000.00
9	Single Family	300	1	3	1	2005	2005	Fair	exterior:plastered wall, roof: Asbestos Sheet, foundation: Basement, flooring: Concrete, windows: Steel, doors: Wood	Magweg	High	Bulaway	27000.00
10	Single Family	200	1	2	1	2000	2000	Fair	exterior:Concrete, roof: Asbestos Sheet, foundation: Concrete Slab, flooring: Concrete, windows: Steel, doors: Steel	Pumula	High	Bulaway	15000.00
11	Multi-Family	800	1	4	1.5	1980	1980	Fair	exterior: Bricks, roof: Tile, foundation: Concrete Slab, flooring: Hardwood, windows Steel, doors: Wood	Barham	Medium	Bulaway	40000.00
12	Single Family	200	1	4	1	1990	2003	Fair	exterior:Concrete, roof: Corrugated Galvanized Steel, foundation: Concrete Slab, flooring: Concrete, windows: Steel, doors: Aluminum	Tshabal	High	Bulaway	16500.00
13	Multi-Family	300	1	3	1	1991	2008	Fair	exterior:Concrete, roof: Asbestos Sheet, foundation: Concrete Slab, flooring: Concrete, windows: Steel, doors: Wood	Mpopom	High	Bulaway	21500.00
14	Condo	350	1	3	1	1970	2022	Poor	exterior:Concrete, roof: IBR Sheets, foundation: Concrete Slab, flooring: Hardwood, windows: aluminum, doors: Wood	Nkulum	High	Bulaway	28000.00
15	Condo	700	2	5	1.5	1998	2024	Good	exterior:Stucco, roof: Concrete Tiles, foundation: Basement, flooring: Carpet, windows: Steel, doors: Aluminum	Mahatsh	Medium	Bulaway	80000.00
16	Townhouse	4900	2	7	3	1964	2020	Poor	exterior: Stucco, roof: Concrete Tile, foundation: Crawl Space, flooring: Concrete, windows: Aluminum, doors: Wood	Killarney	Low	Bulaway	300000.00
17	Multi-Family	1500	2	5	2	1987	2005	Fair	exterior: Stucco, roof: Concrete Tile, foundation: Crawl Space, flooring: Concrete, windows: Steel, doors: Wood	Selborn	Low	Bulaway	230000.00
18	Townhouse	4900	2	7	3	2020	2022	Excellent	exterior: Stone, roof: Concrete Tiles, foundation: Concrete Slab, flooring: Concrete, windows: Aluminum, doors: Wood	Burnsid	Low	Bulaway	155000.00
19	Single Family	385	1	3	1	1969	1985	Poor	exterior: plastered wall, roof: Corrugated Galvanized Steel, foundation: Concrete Slab, flooring: Concrete, windows: Aluminum, doors: Wood	Tshabal	High	Bulaway	26953.84
20	Townhouse	793	2	6	2.5	1965	2013	Poor	exterior: Concrete, roof: IBR Sheets, foundation: Crawl Space, flooring: Hardwood, windows: Wood, doors: Steel	Sunnysi	Medium	Bulaway	113426.02
21	Single Family	939	2	6	2.5	1974	2009	Poor	exterior: travertine topaz brick, roof: IBR Sheets, foundation: Crawl Space, flooring: Tile, windows: Aluminum, doors: Aluminum	Queensp	Medium	Bulaway	110182.90
22	Townhouse	2833	2	6	2	2009	2017	Very Good	exterior: Stone, roof: IBR Sheets, foundation: Crawl Space, flooring: Tile, windows: Wood, doors: Aluminum	Morning	Low	Bulaway	433829.42
23	Condo	1041	1	3	1	1975	2014	Fair	exterior: Stucco, roof: Corrugated Galvanized Steel, foundation: Concrete Slab, flooring: Hardwood, windows: Steel, doors: Steel	Sunnysi	Medium	Bulaway	76936.40
24	Condo	1224	2	6	2.5	1960	1963	Poor	exterior: Brick, roof: Asbestos Sheet, foundation: Crawl Space, flooring: Concrete, windows: Wood, doors: Aluminum	Morning	Low	Bulaway	460500.05
25	Multi-Family	3417	2	8	4	2002	2020	Good	exterior: Stucco, roof: Concrete Tile, foundation: Crawl Space, flooring: Concrete, windows: Wood, doors: Steel	Southda	Low	Bulaway	375157.22
26	Single Family	2434	1	4	2	1982	1985	Fair	exterior: plastered wall, roof: Corrugated Galvanized Steel, foundation: Concrete Slab, flooring: Tile, windows: Steel, doors: Steel	Glengar	Low	Bulaway	449829.93
27	Single Family	2037	2	5	2	2010	2022	Very Good	exterior: Concrete, roof: Asbestos Sheet, foundation: Basement, flooring: Carpet, windows: Wood, doors: Steel	Intini	Low	Bulaway	356937.23
28	Townhouse	1493	1	3	1	2003	2013	Good	exterior: plastered wall, roof: IBR Sheets, foundation: Concrete Slab, flooring: Tile, windows: Wood, doors: Wood	Glengar	Low	Bulaway	359455.71
29	Single Family	293	1	2	1	1965	1996	Poor	exterior: Stone, roof: Asbestos Sheet, foundation: Crawl Space, flooring: Carpet, windows: Steel, doors: Wood	Pumula	High	Bulaway	33438.39
30	Townhouse	2364	1	4	2	2019	2024	Excellent	exterior: plastered wall, roof: IBR Sheets, foundation: Concrete Slab, flooring: Carpet, windows: Wood, doors: Wood	Raylton	Low	Bulaway	439787.38
31	Townhouse	3678	2	8	4	2001	2002	Fair	exterior: Concrete, roof: Asbestos Sheet, foundation: Crawl Space, flooring: Carpet, windows: Aluminum, doors: Wood	Hillside	Low	Bulaway	752566.60
32	Townhouse	2859	2	5	2	1967	1988	Poor	exterior: Stone, roof: Corrugated Galvanized Steel, foundation: Crawl Space, flooring: Concrete, windows: Steel, doors: Steel	Four Wi	Low	Bulaway	423093.96
33	Condo	2585	1	4	2	2014	2014	Good	exterior: Stucco, roof: Asbestos Sheet, foundation: Concrete Slab, flooring: Hardwood, windows: Steel, doors: Steel	Woodlar	Low	Bulaway	487384.03
34	Single Family	415	1	3	1	2006	2016	Very Good	exterior: Stone, roof: IBR Sheets, foundation: Basement, flooring: Carpet, windows: Wood, doors: Wood	Pelanda	High	Bulaway	38154.75
35	Multi-Family	673	1	2	1	1972	1995	Poor	exterior: Brick, roof: Concrete Tile, foundation: Concrete Slab, flooring: Tile, windows: Wood, doors: Wood	Paddonl	Medium	Bulaway	88127.64
36	Condo	659	1	4	1.5	1983	1994	Fair	exterior: Concrete, roof: Concrete Tile, foundation: Concrete Slab, flooring: Tile, windows: Wood, doors: Wood	Westsur	Medium	Bulaway	135096.30
37	Condo	931	2	5	2	1962	1970	Poor	exterior: Stone, roof: Corrugated Galvanized Steel, foundation: Basement, flooring: Tile, windows: Wood, doors: Aluminum	Barham	Medium	Bulaway	173598.57
38	Single Family	375	1	2	1	1974	2020	Poor	exterior: Concrete, roof: Asbestos Sheet, foundation: Basement, flooring: Concrete, windows: Wood, doors: Wood	Njube	High	Bulaway	38450.69
39	Condo	687	2	6	3	1997	2020	Good	exterior: Concrete, roof: Asbestos Sheet, foundation: Crawl Space, flooring: Stone, windows: Steel, doors: Aluminum	Parkviev	Medium	Bulaway	106610.64
40	Single Family	458	1	2	1	1973	1978	Poor	exterior: Concrete, roof: Asbestos Sheet, foundation: Crawl Space, flooring: Concrete, windows: Wood, doors: Aluminum	Entumb	High	Bulaway	26532.44
41	Single Family	319	1	4	2	2017	2018	Very Good	exterior: Stucco, roof: Concrete Tile, foundation: Concrete Slab, flooring: Stone, windows: Wood, doors: Wood	Njube	High	Bulaway	48540.39
42	Multi-Family	277	1	2	1	2018	2021	Excellent	exterior: travertine topaz brick, roof: Asbestos Sheet, foundation: Crawl Space, flooring: Hardwood, windows: Aluminum, doors: Aluminum	Nguboy	High	Bulaway	21720.02
43	Single Family	1012	1	3	1	1994	2001	Fair	exterior: Concrete, roof: Asbestos Sheet, foundation: Basement, flooring: Hardwood, windows: Wood, doors: Steel	Kenilwo	Medium	Bulaway	136945.59
44	Townhouse	1553	2	6	3	2002	2009	Fair	exterior: travertine topaz brick, roof: IBR Sheets, foundation: Crawl Space, flooring: Tile, windows: Wood, doors: Aluminum	Grenhill	Low	Bulaway	231807.27
45	Multi-Family	634	1	3	1	2023	2023	Excellent	exterior: Stone, roof: Corrugated Galvanized Steel, foundation: Concrete Slab, flooring: Tile, windows: Wood, doors: Steel	North Er	Medium	Bulaway	152932.09

Figure 3.3: Dataset

3.5.2 Justification for the Effectiveness of Variables in Determining Property Prices

When developing a model for property price prediction through machine learning, e.g., the application of algorithms like Random Forest, it is necessary to select good variables that are able to capture the determinants of property prices. The variables chosen are yours for your data set—location, size (square footage), bedrooms, type of property, bathrooms, year built, and year renovated are all reasonable because they are pertinent and beneficial in determining the values of real estate. Below is a full explanation of each variable along with supporting quotes from relevant studies.

1. **Location**:

Location is an important determinant of property prices as it includes elements of quality surroundings, amenity proximity, and general desirability. Quality areas have houses that sell at a premium. A study highlights the issue that block and area location factors play a tremendous role in house prices (Zhong, 2024).

2. **Size (Square Footage)**:

The total area of a property is directly proportional to its market value. Bigger properties hold bigger space, and that is a more appealing aspect for potential buyers. Size is one of the key drivers of the prices of properties in a house price estimation study based on machine learning (Sharma, 2023).

3. **Number of Bedrooms**:

The number of bedrooms in a home dictates its viability and appeal to potential clients, especially household units. Research also discovered a reverse relationship between the worth of a home and the number of bathrooms that it contains, such that each extra bedroom can make a substantial contribution to the value of the property (Zhong, 2024).

4. **Number of Bathrooms**:

Similar to with bedrooms, more bathrooms make a property more functional and convenient. According to research that was carried out by the Federal Reserve Bank of Chicago, more bathrooms contribute to house prices, hence making houses more valuable (Toussaint-Comeau, 2018).

5. **Year Built**:

An age of a property can affect its market value; newer properties are more valuable because they save energy and meet the newest building codes. Newer properties in the same location, according to studies by Mosaik Real Estate, will typically have higher prices per square foot than older properties (Mamre, 2024) .

6. **Year Renovated**:

Upgrades can increase the beauty and usability of a property, which can increase its market value. Upgraded houses can demand a sought-after premium, based on a Journal of Real Estate Finance and Economics report, while downgraded houses can demand a negative premium (Mamre, 2024).

3.5.3 Data Collection and Description

Data was gathered from publicly available lists of actual property and Bulawayo's property records. Every effort was made to make the data representative of prevailing market conditions and for a vast majority of different types of properties and locations.

3.5.4 Data Exploration

To comprehend the distribution and connections between variables, exploratory data analysis, or EDA, was carried out. To find trends and possible outliers, methods like scatter plots and correlation matrices were used. This stage is essential for identifying underlying patterns and guiding the predictive model's feature selection.

3.5.5 Data Preparation

Data preparation involved several steps for preparing the dataset for modelling:

- **Missing Value Handling**: Imputation techniques were used to replace any points of missing values in the data to make it complete.
- **Categorical Variable Encoding**: Effective encoding techniques suitable for use by machine learning algorithms were used to convert categorical data like location into a numeric format.
- **Scaling Features:** Numerical features were scaled to a similar range of values to allow for model convergence and performance.
- **Train-Test Split**: In order to evaluate the predictive capability of the model in terms of new data, the data was split into training and test partitions.

3.5.3 Data Quality Verification

Verification of data quality is of paramount significance while creating a sound model for property valuation. It consists of a number of major steps:

- **Missing Value Check**: Identification and treatment of missing values through imputation or deletion are necessary to prevent model bias. In real estate, where every minor piece of information is a deciding factor for property value, missing data would make values significantly off.

- **Consistency and Accuracy**: It is essential to have the data constantly and accurately reflect actual situations in the world. This entails maintaining consistency in data types, unit of measurement, and categorical variables. Correct data accurately reflects actual property attributes, and it is crucial during model training as well as prediction accuracy.
- **Detection of Outliers and Treatment**: The data must be scanned for anomalies for outliers as they cause the process of learning by the model to become biased. For instance, the houses which are very high or very low compared to the average market price can be instances of wrong data entry or special cases to be handled separately.
- **Feature Relevance Evaluation:** Verification of the relevance of each feature for the purpose of forecasting the property prices help to enhance the model. Irrelevant and redundant features can be removed to enhance model performance and simplify it.

3.6 Modelling

This phase outlines building and validation of the predictive model used to predict property prices using the Random Forest algorithm. Each step is defined as follows based on the Cross-Industry Standard Process for Data Mining(CRISP-DM) approach:

3.6.1 UML diagrams

UML stands for Unified Modelling Language and is used to model and represent the product or end system abstractly. The use of diagrams is of utmost importance in the system modelling and designing phase of system development life cycle (Bhatt, 2021).

Following are the UML diagrams of this system.

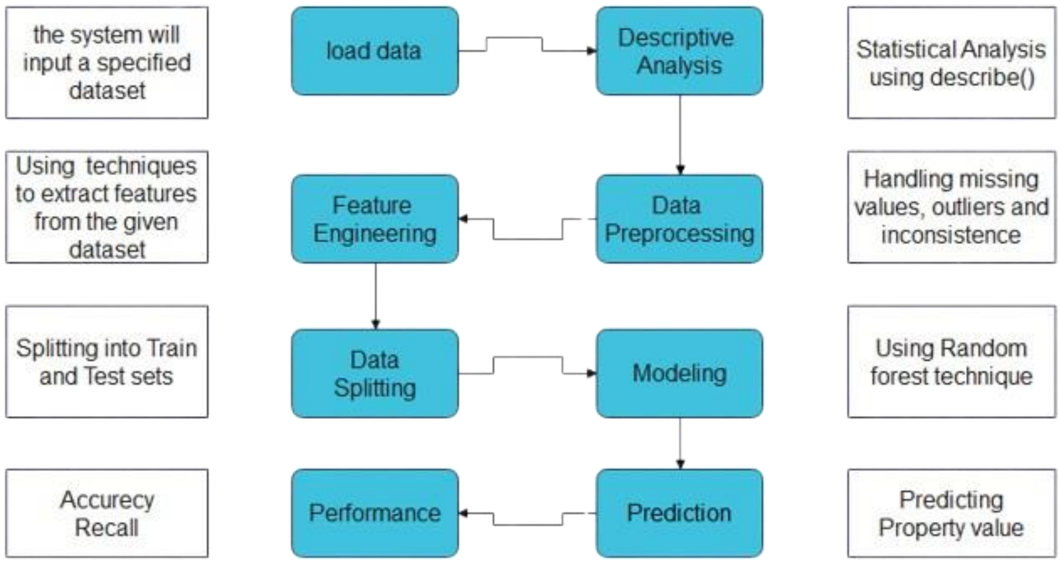


Figure 3.4: Model Flow diagram

3.6.2 Data Flow Diagram

Data processing steps of data used in the property price forecasting model are shown graphically in a data flow diagram (DFD). It shows data flow from input to output from the beginning of the system to the end. Below is the detailed list of DFD elements:

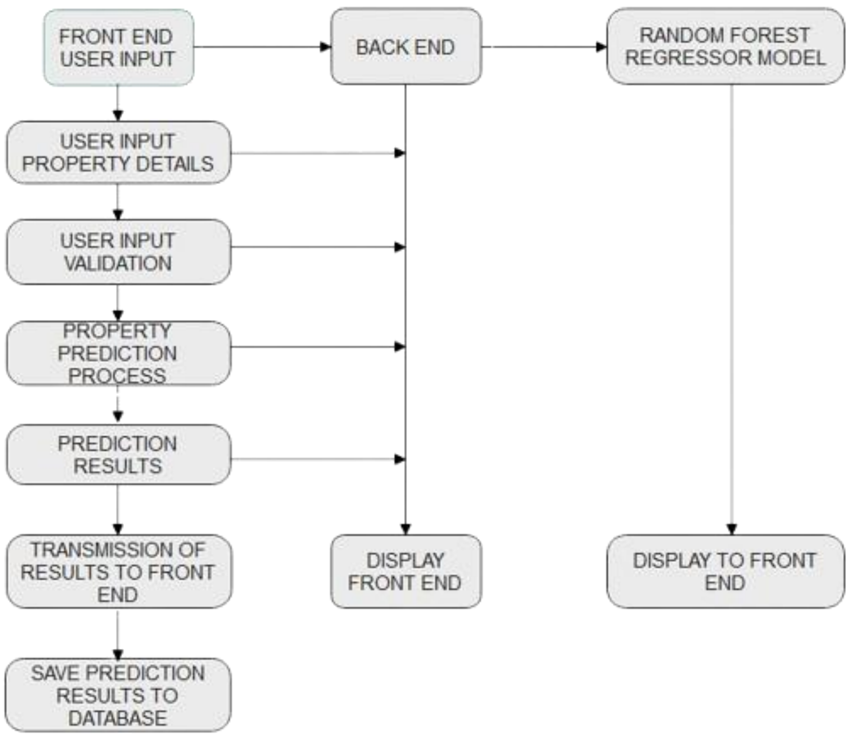


Figure 3.5: DFD

3.6.3 Activity Diagram

In the property valuation system under construction, the activity diagram illustrates the smooth exchange of activities among various parts (Niu, 2019). The user begins by sending login credentials and then later, property information through the user interface. The frontend then sends the user-input information to the backend for processing and validation to ensure correctness and consistency. After this validation process is done, the backend communicates with the Random Forest Regressor with the data provided by the user to predict or estimate property value. The evaluation result is sent back to the backend, which sends it to the frontend so that the user can see it. The activity diagram is shown in the next diagram.

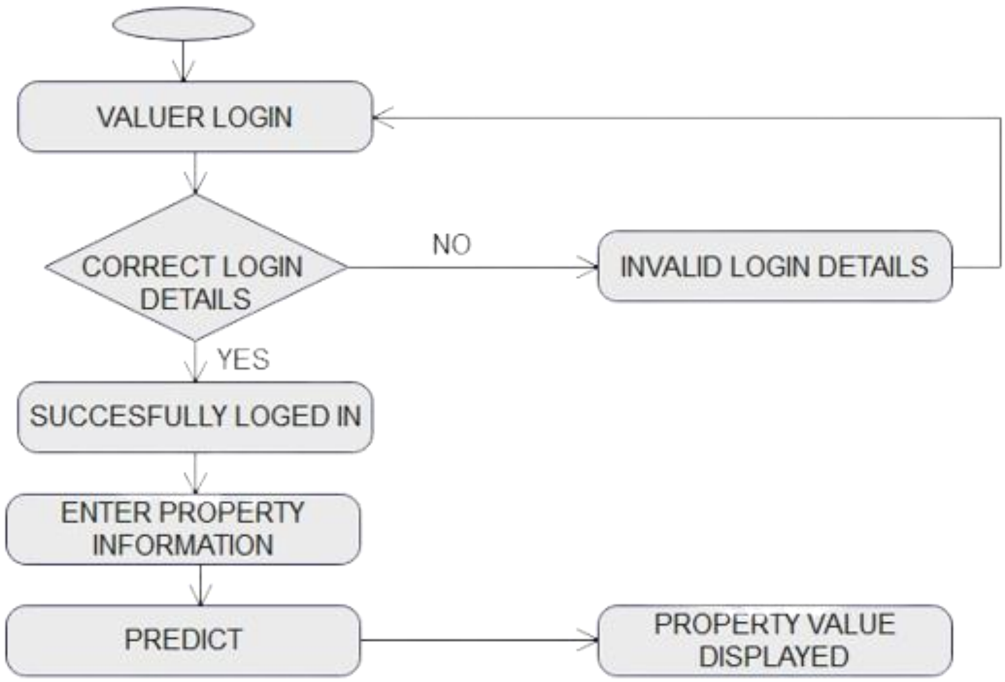


Figure 3.6: Activity Diagram

3.7 Chapter Summary

Chapter 3 provides a brief description of the research design and method employed in this quantitative research on property valuation in real estate market using machine learning. The research is quantitative with cross-sectional design in gathering numerical data regarding market prices and properties. The research employs property of a real estate market whose information was gathered using structured questionnaires and secondary data. The sample size is then determined after conducting power analysis, and stratified random sampling is utilized for representation of various kinds of properties and locations. Ethical aspects are addressed with highest possible seriousness, and informed consent of the data suppliers is addressed. Confidentiality of the data is maintained to the highest possible level.

Chapter demonstrates operationalization of variables, i.e., measurement and quantification of significant variables like area, address, age of property, number. of bedrooms, number of bathrooms, and new developments. Statistical techniques like multiple regression analysis and machine learning techniques like Random Forest are used for correlation measurements of independent variables with prices of houses. The chapter is heavy enough with proper research methodology to provide rigor and internal validation of results.

CHAPTER 4

EVALUATION AND RESULTS

4.1 Introduction

In this chapter, the evaluation and outcome of the machine learning model to predict real estate property values specifically, the Random Forest algorithm are presented. The results are explained in relation to the three primary objectives of this research.

4.2 Research Objectives

The objectives: -

1. To analyse Random forest in property evaluation.
2. To design and implement a Machine learning model (Random forest) for property evaluation system.
3. To evaluate the effectiveness of using Random forest in property evaluation for Real Estate.

4.3 Actual Results and Explanation

4.3.1 Objective 1: Analysis of Dataset Description

The information employed in this research was accessed from property notices, council archives, and estate agents operating within Bulawayo. Derived features were:

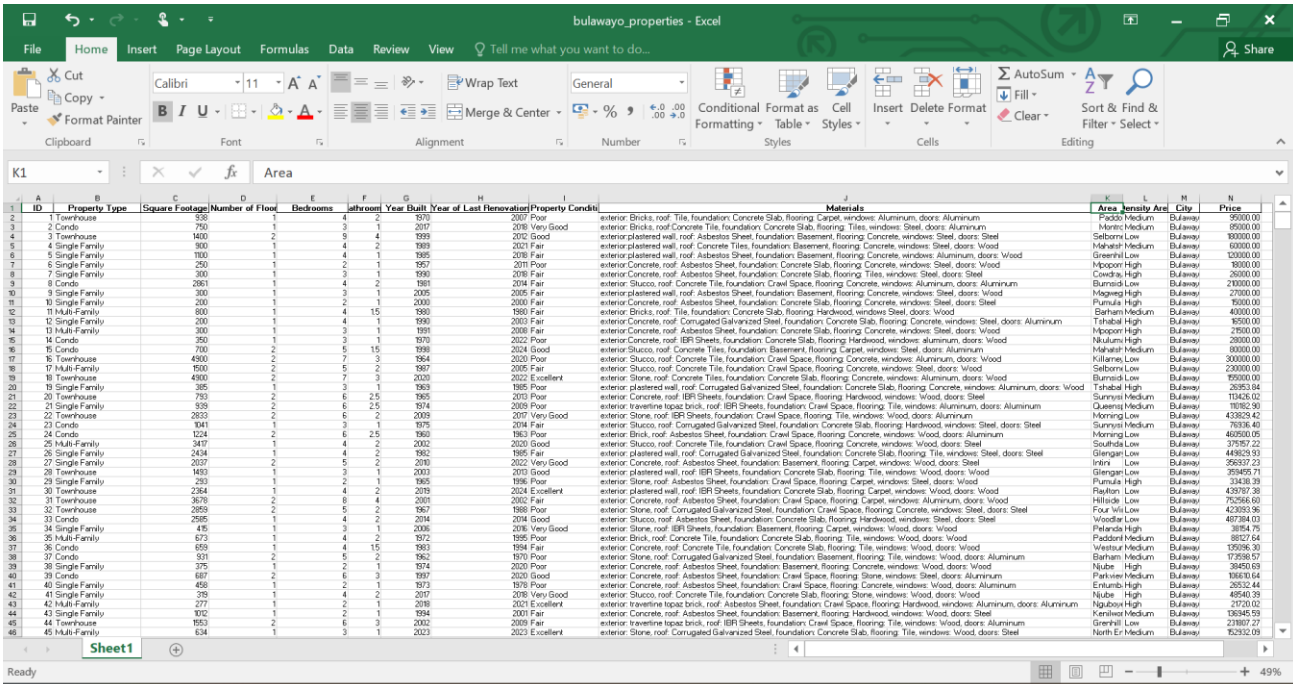


Figure 4.1: Dataset

- Type of property
- Square footage (square meters)
- Number of floors
- Number of bedrooms
- Number of bathrooms
- Year built
- Year of most recent renovation
- Condition of property
- Materials used
- Area
- Density Area
- City
- price

The pre-processed and cleaned dataset had 5,000 property records with 13 features that were used for training and testing purposes.

4.3.2 Model: Random Forest Regressor

Random Forest was selected since it is robust with tabular data, overfitting-resistant, and can also identify nonlinear interactions. It was applied utilizing the scikit-learn library with the following hyperparameters:

- n_estimators: 100
- max_depth: 15
- min_samples_split: 5
- random_state: 42

4.3.3 Random Forest Performance

Results

1. Mean Squared Error (MSE): 4669713572.266907
2. Root Mean Squared Error (RMSE): 68335.30253292879
3. R-squared (R^2): 0.8709334282720648
4. Mean Absolute Percentage Error (MAPE): 0.24557808481867420%
5. Mean Absolute Error (MAE): 43224.9662034875

Good predictive performance is observed in the fact that the value for R^2 for the Random Forest model is 0.87, which means that it predicts 87% of variation in property prices.

4.3.4 Model Validation

K-Fold Cross-Validation with 5 splits was employed here in order to find the generalizability of the model. The mean cross-validation R^2 value was 0.87, and it is closely matching performance on the test set, indicating that the model is not overfitting.

4.4 Object 2: System Design

It refers to how the system elements and data structures satisfy the requirements. The design is taken from the requirements specification document and consists of:

4.4.1 Dataflow Diagrams

Data Flow Diagrams (DFDs) show the interconnection of the different parts of the property valuation system. To show how raw input data location, materials used in construction, and property features is converted to become output results, like estimates of property value, a DFD is a basic graphical system for describing high-level system operation. From land size to quality of the area, proximity to infrastructure, or history of transactions, data movement in a DFD is tagged to indicate information type and origin. DFDs provide unequivocal presentation of how this information travels and gets processed in the system, giving valuable insight into how precise measures are obtained.

4.4.2 Software Implementation

Streamlit is a very useful and vital Python tool that makes it easier to develop website applications that work quickly for data analysis, machine learning projects, and easy model deployment (Ake, 2022). This project also aims to ensure that end-users easily interact with the system and understand how much properties cost based on the values they supply.

The user can change the input parameters on this machine learning application and get an evaluation from the model. The streamlit package is installed using the pip install command. To ensure model persistence, consistent scikit-learn versions were used, including version 1.4.0 in the Python compiler. Inconsistent versions may lead to breaking code or invalid results while trying to unpicked the specified model.The user interface was developed using Python 3.13 in the Visual Studio Code environment. The following command is used to launch the streamlit app: Run <app_name.py> in streamlit. The program will load and open on a browser.

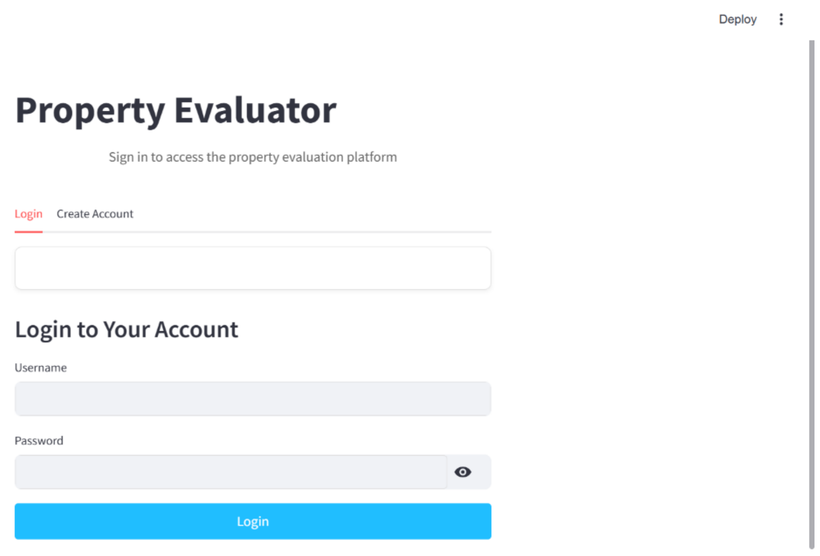


Figure 4.1: Login Screen

The above figure shows the login screen where the valuer enters their credentials to access the system.

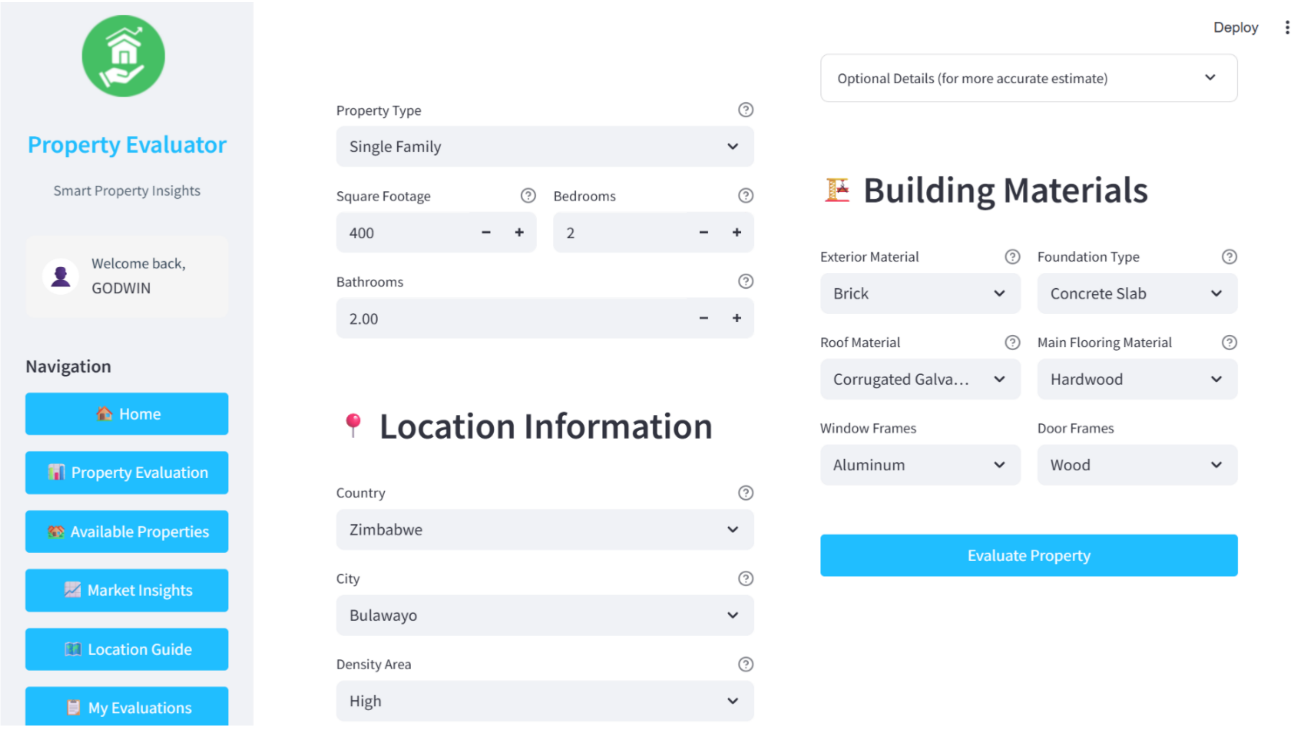


Figure 4.2: Property Evaluation

The above figure shows the model interface where users can input attribute values. There is also a sidebar with options that contain some useful information that can be helpful during the property evaluation process.

The 'Go to' on the sidebar allows the user to explore various options. The navigation options include home, property evaluation, available properties, market insights, location guide, my evaluations and logout.

The first option is the home which is the welcome page after login into the system were it show you recent activity that you might have currently evaluated. The second option, which is property evaluation, allows property valuers or officers to capture property attributes from property owners to evaluate their monetary value. It handles data entry errors or values that are not within the specified range. It can also handle some logic errors; the error is then reported to the user during or before performing evaluation. The user has to fill in all the fields before evaluating. If other boxes are left empty, the system will automatically pick the upper option; this is by default.

The option ‘Available Properties’ is the third one; this page offers a user-friendly inferface allows the property valuers to search properties based on specific criteria.Users can filter their search by selecting a property type (townhouse, condo, single-family, or multi-family), setting a price range (minimum and maximum), and specifying the number of bedrooms (minimum and maximum). The page also categorizes properties by type, making it easy to browse and find the perfect property.

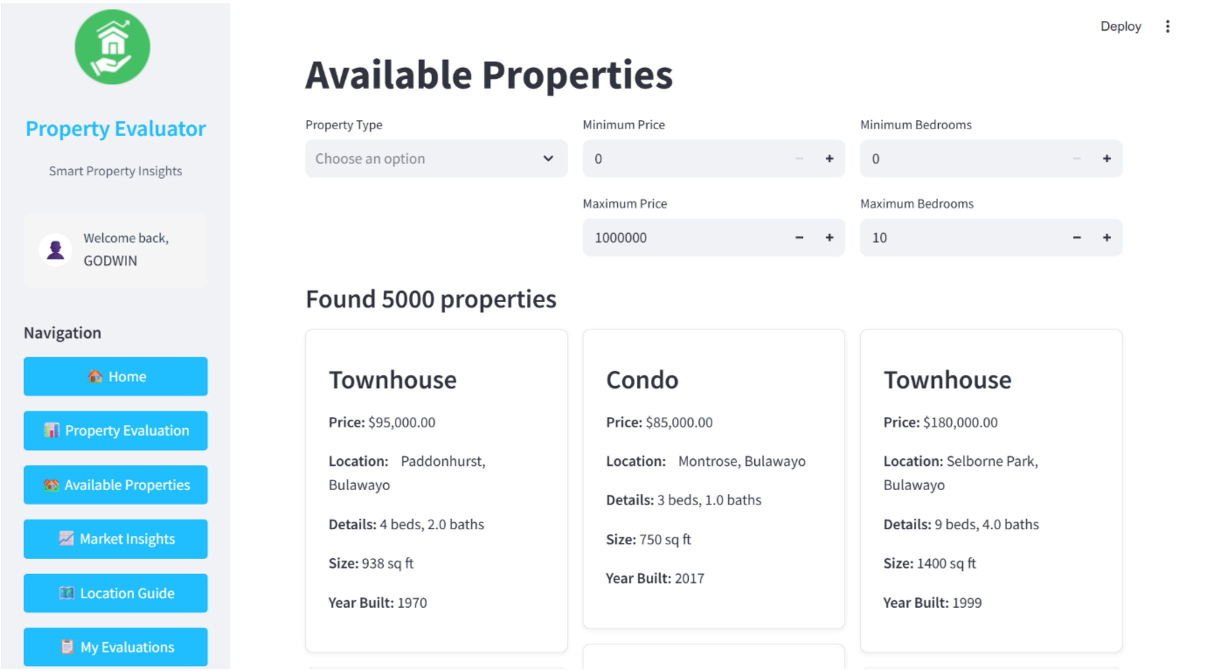


Figure 4.3: Available properties

The 'Market Insights' feature occurs when it provides a thorough analysis of real estate market statistics and trends, applying machine learning models to reveal key influencers of property price. The page also contains a chart showing the relative importance of property attributes, such as square area, type of property, and area, in influencing market trends.Furthermore, the page provides 3D scatterplot property price analysis, which visually depicts the co-dependence of price, square area, and bedrooms, color-coded keys representing difference locations in Bulawayo.This interactive visualization enables valuers to analyze area-wise price numbers as well as location-based market patterns, facilitating informed decision-making in the real estate sector.

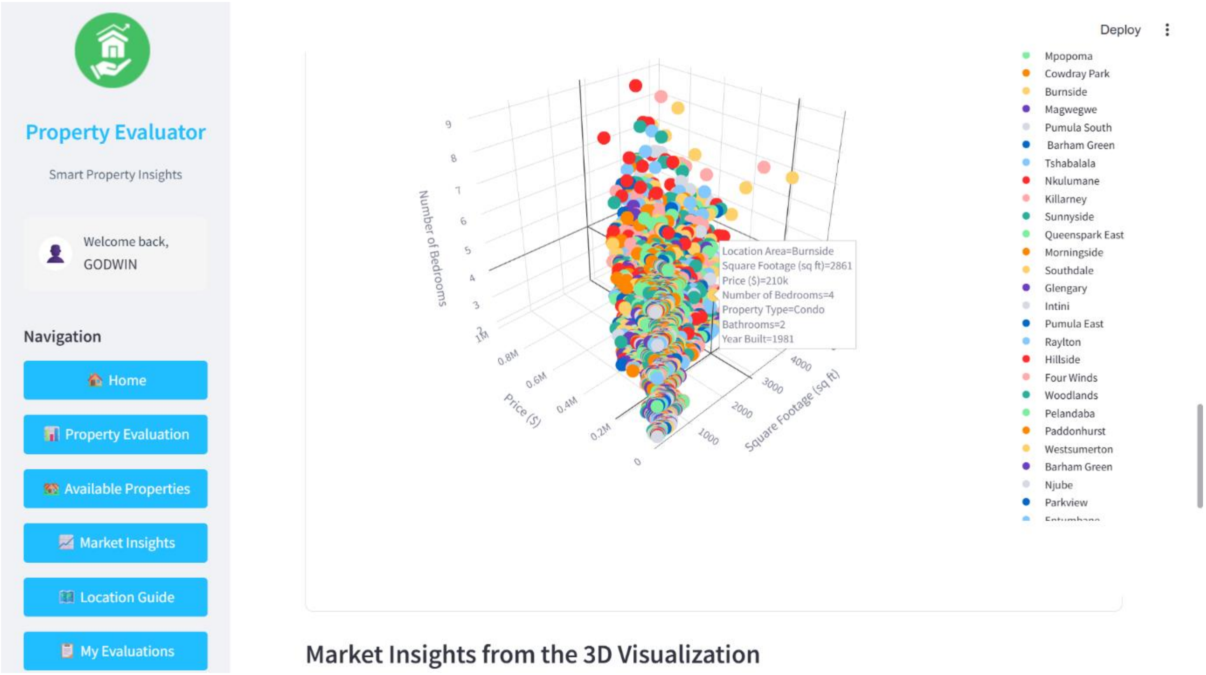


Figure 4.4: Market Insight

The above figure visualizes estimated properties from SQLite studio database.

Location guide enables the valuer to explore areas by density,categorizing location into low, mediam, amd high density zones. Each category lists specific areas, providing a clear overview of region’s layouts.It also features a map view, allowing valuers to visualize the locations and their corresponding density zones. . High-density areas are usually located on the western side of the city, whereas low-density areas are located on the eastern side of the city.

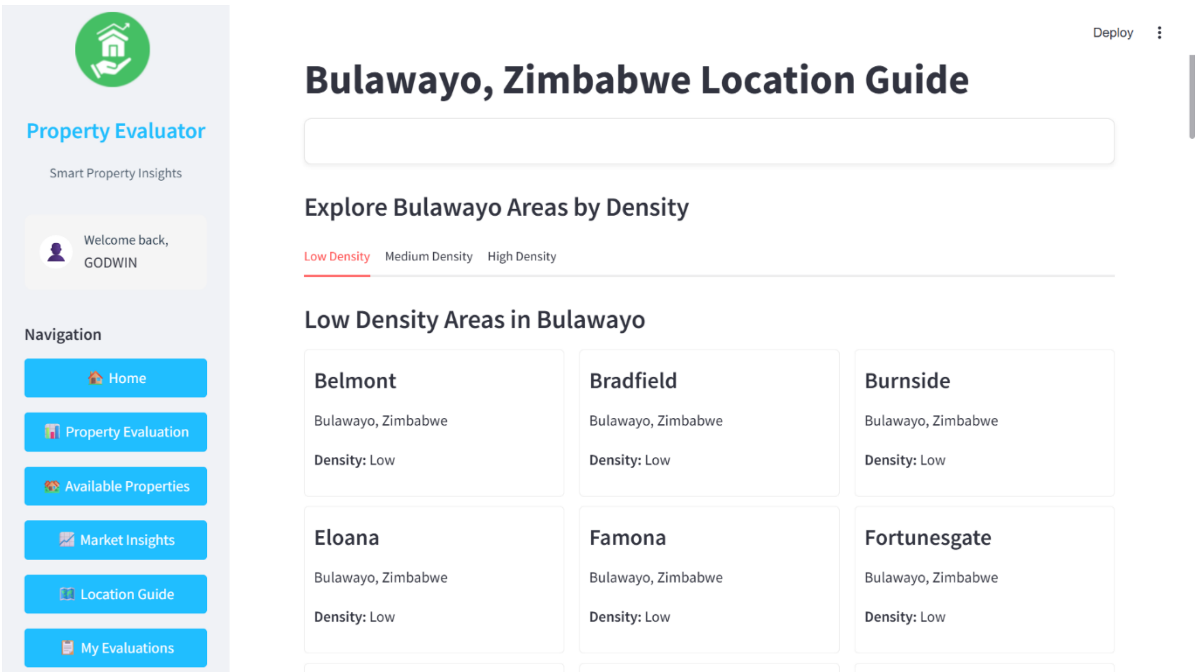


Figure 4.5: Bulawayo Residential areas

The sixth option is My Evaluations, which provides valuers with a centralized platform to access and manage their property evaluations. A saved evaluation table displays all evaluated properties, allowing valuers to quickly locate specific assessments. By selecting a property, users can view detailed evaluation results, including property information and evaluation outcomes. Furthermore, this page enables valuers to generate and download a PDF report, providing a convenient option for obtaining a hard copy of their evaluation results.

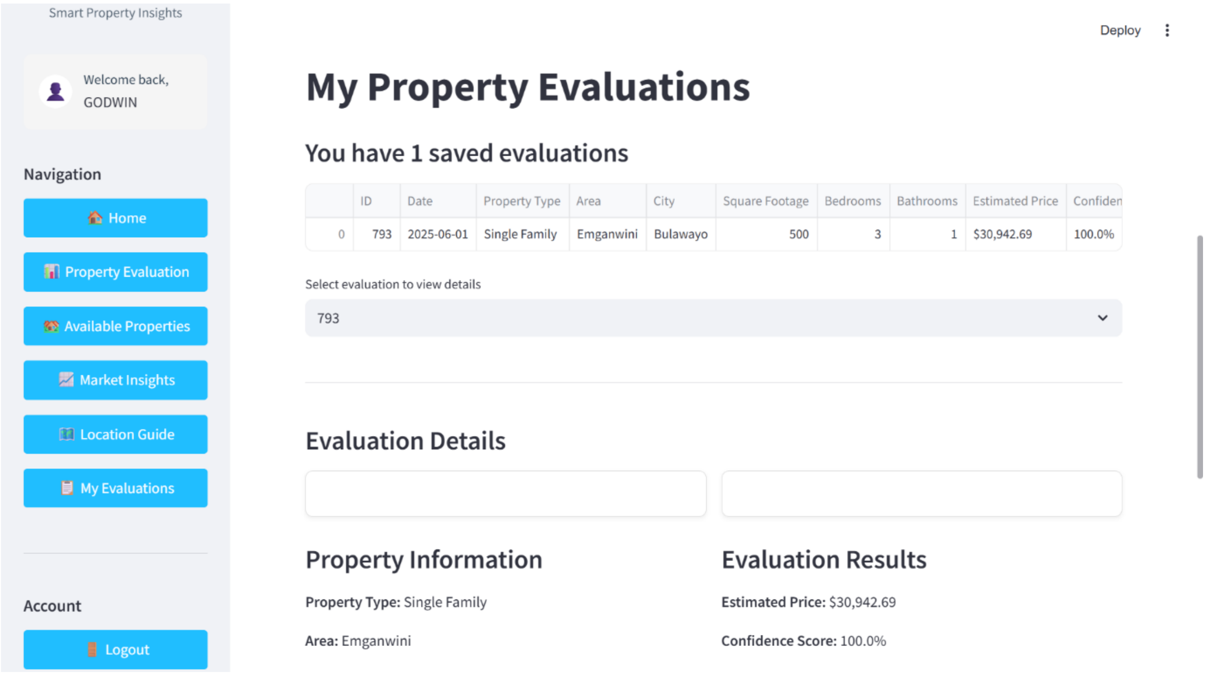


Figure 4.6: My Evaluations

The last option is Logout feature that allows valuers to exit the system, returning them to the Login page, thereby terminating their current session and ensuring account security.

4.5 System Testing

After various components were developed, the next task was to verify and validate that the system met the goals and objectives of the research specified on earlier chapters. A variety of tests were carried out, including user acceptance, integration, and unit testing.

4.5.1 Unit Testing

With each system component developed—system frontend and system backend—unit testing was conducted to identify errors and possible bugs in every component before integration. With reference to the user interface, testing was done to evaluate the system's capabilities to handle value input errors, unsupported inputs, inputs that are not within the specified range, logic within the supplied inputs, and empty inputs (Riccio, 2020).

4.5.2 Integration Testing

After successfully completing the unit tests, the next task was to integrate system components into a single unit and verify their performance. The main components integrated were the frontend (user interface) and backend (the model). The main objective of this testing was to ensure that the system performed to acceptable standards in relation to the needs specified in earlier chapters. Integration testing was of colossal value as it helped to identify and resolve inconsistences, module persistence, and compatibility issues. Integration testing basically helps the developer see if the final product is meeting all the specified needs. It also helps in verifying and validating system integrity.

4.5.3 User Acceptance Testing

Involves presenting the final product to users to test the software's performance and evaluate whether it can perform efficiently in a real-life industry. The users may include supervisors, colleagues, any other members of the department, and companies or institutes that are targeted by the developers. The user acceptance testing helps the developers get positive or negative reviews, possible corrections, and valuable insights towards improving system performance and interoperability (Riccio, 2020).

4.5.4 Functional Testing (Button Functionality)

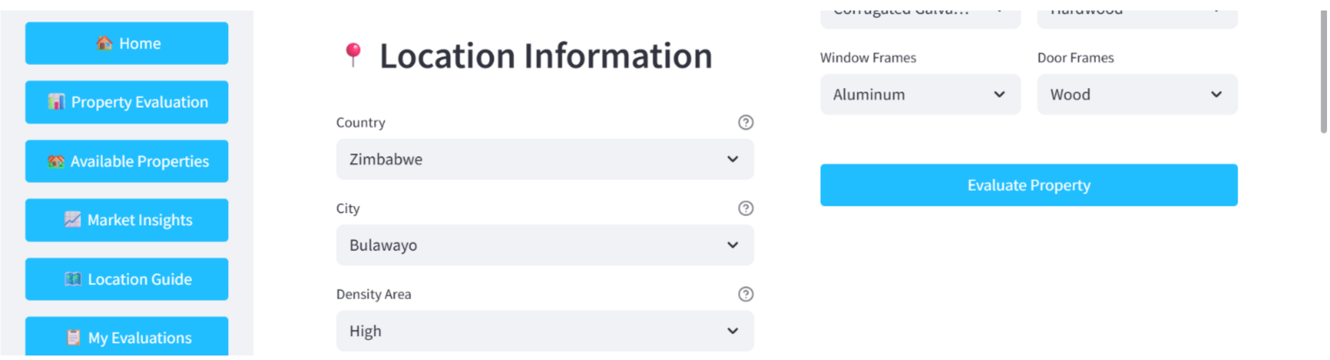


Figure 4.7: Evaluate Property button

The system provides an estimated value of a property from a Random Forest machine learning model upon clicking the "Evaluate Property" button. The system takes required information on the property, including type of property, square footage, floors, bedrooms, bathrooms, density area, year built, year renovations, and building material. All this information is fed into the Random Forest model, which subsequently processes them and then provides estimates based on patterns observed in the view when data are being projected in massive levels of data with comparable attributes. The model also assigns weight to the information depending on the relative value of each attribute and subsequently adjusts accordingly. Finally, the system delivers the final estimate of the value of the property and renders users with an aggregate and evidence-driven valuation.

4.5.5 Summary of How the System Works

Machine Learning-based Real Estate property valuation is an activity of developing a data-based model to predict property value based on the input variable more than one. On the basis of property sale history and geospatial information, the system provides property appraisal in timely and correct fashion. Property sellers or buyers, prospective buyers or sellers, property developers, real estate agents, and banks depend on it for property market decisions.

4.6 Implementation of the Machine Learning Model

An ensemble learning algorithm under the guidance of random forest regression leverages the potency of ensemble learning in its quest for regression (Ake, 2022) . On steps and varying needs taken previously, random forest regression algorithm was executed as shown below:

```
[7] from sklearn.ensemble import RandomForestRegressor

    # Initialize the RandomForestRegressor model
    model = RandomForestRegressor(n_estimators=100, random_state=42)

    # Train the model
    model.fit(X_train, y_train)

RandomForestRegressor
RandomForestRegressor(random_state=42)
```

Figure 4.8: Random Forest Regressor Model

4.6.1 Implementation of The Evaluation Function

This module specifies the application of the evaluation function in the system for property appraisal estimation prediction. The evaluation function estimates Random Forest model performance on the basis of general evaluation metrics like R², MAE, MSE, RMSE, and

MAPE. The evaluation function is then provided predicted property prices and actual property prices and returns predicted measures in order to estimate the accuracy of the model. The generalization of the model to new data is of topmost importance.

4.6.2 Implementation

There exists a range of activities to be undertaken during the process of employing the property valuation model with Python, Streamlit, and Jupyter Notebook in order to obtain a real-world and interactive application. Jupyter Notebook is initially employed in an attempt to conduct the data preprocessing near the real commencement through the use of Python libraries such as Pandas, NumPy, and Scikit-learn. Then raw geographic and house data are cleaned up, missing values were dealt with, categorical variables encoded, and numerical data normalized.

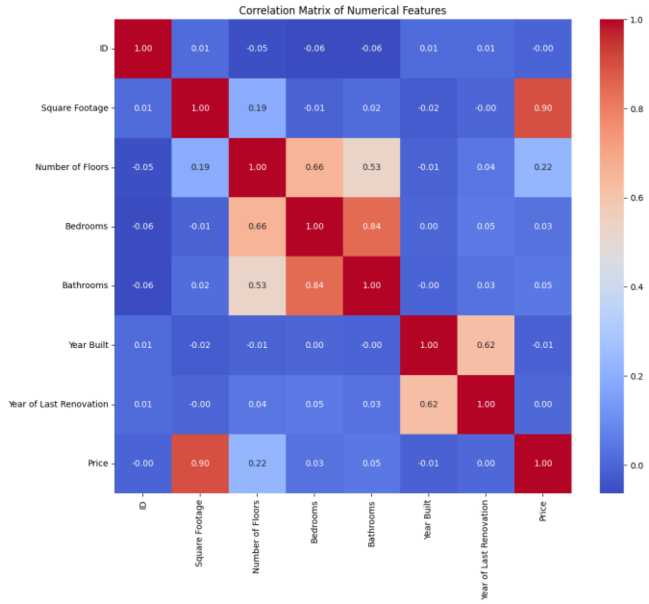


Figure 4.9: Correlation of the dataset

The relationship between the properties used is indicated in the above figure. Large relationships between variables are indicated by correlation coefficients near 1.00 or -1.00, while very low relationships are indicated by figures near zero. The discussion above indicates some moderate positive relationship between variables, hence low risk of overfitting during model development.

Then, Scikit-learn is utilized to train the Random Forest regression model on the prepared dataset. The model's ability to correctly predict real estate values is then assessed using metrics like R^2, MAE, MSE, RMSE, and MAPE. Streamlit is utilized for designing the user interface, allowing for the creation of a web-based, easy application. Individuals can enter property-related details such as Property Type, Square Footage, Number of floors, bedrooms, bathrooms, Density area, Year built, year of last renovations and building materials. Once the details are entered and submitted, the trained machine learning model is used to evaluate the market value of the property and it is displayed in real-time on the app. Streamlit also has the capability to support data visualizations price distribution plots and feature importance charts to give users even more insight into which features are most influential on property value.

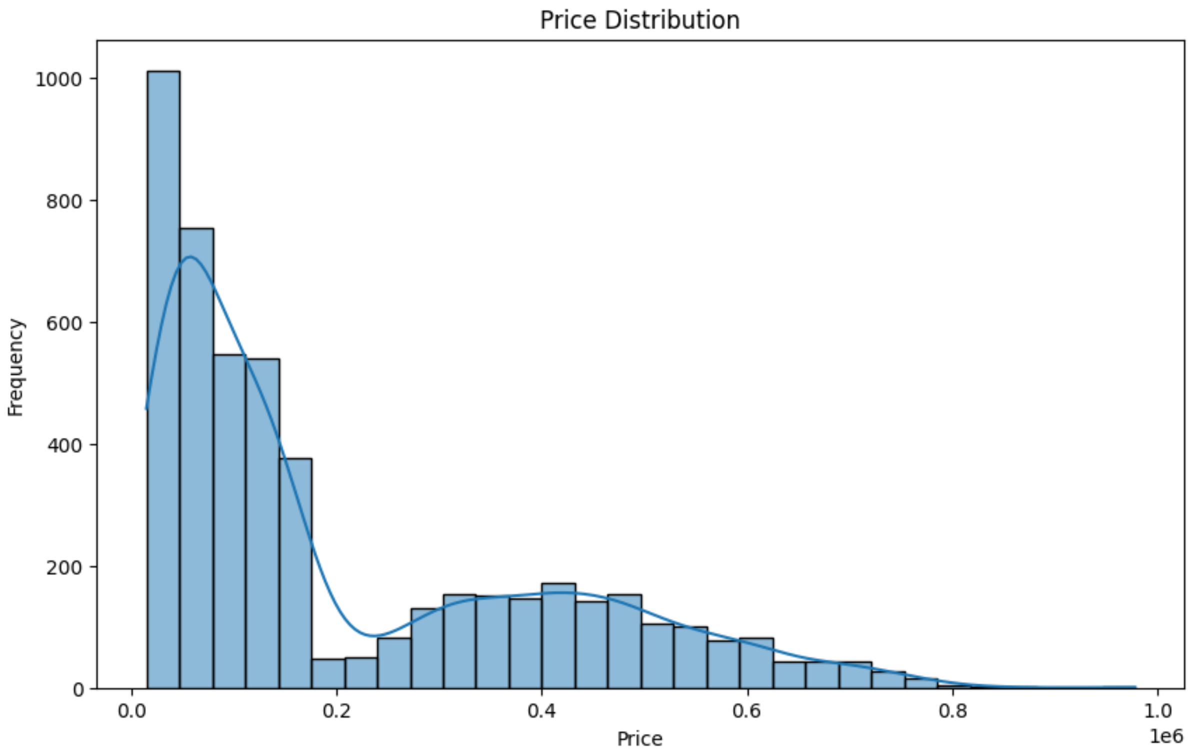


Figure 4.10: Price distribution Chart

Finally, the application may be run locally or on a web server where users like real estate agents, buyers, and developers can view review insights in any web browser. What is created is a scalable and friendly system for enhancing real estate pricing decision-making.

4.6.3 Results

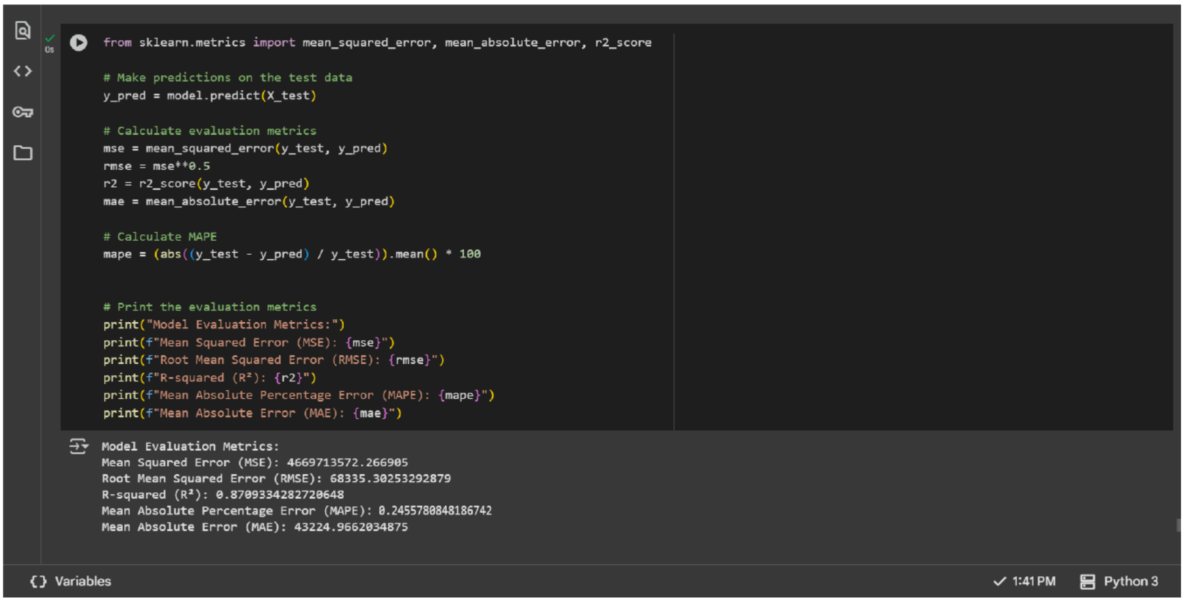


Figure 4.11: model evaluation

The output gives the following metrics for the training data:

1. Mean Squared Error (MSE): **4669713572.266907**
2. Root Mean Squared Error (RMSE): **68335.30253292879**
3. R-squared (R^2): **0.8709334282720648**
4. Mean Absolute Percentage Error (MAPE): **0.2455780848186742%**
5. Mean Absolute Error (MAE): **43224.9662034875**

4.7 Objectives 3

To evaluate the effectiveness of using Random forest in property evaluation for Real Estate.

4.7.1 Evaluation

This section outlines the strategy and result used to assess the performance of the Random Forest model for predicting real estate property values. Evaluation is necessary as it ensures the model is reliable, accurate, and efficient in making real estate decisions. Following the CRISP-DM methodology for data mining, we outline evaluation processes and provide corresponding performance charts. The fifth step in this process is the evaluation phase, which entails the accomplishment of the following activities:

4.7.2 Evaluating Results

We measure the performance of the Random Forest model by evaluating how well it predicts actual property prices. We use the following to do so:

1. **R-squared (R^2):** R^2 tells us about how good our model is in explaining property price volatility. R^2 tells us about the percentage volatility of the target variable (property price) explained by the input variables. The higher the R^2, the better the model fit.
2. **Mean Absolute Error (MAE)**: MAE never mind whether the errors are of one direction or another, calculates the average size of the difference between actual and estimated property values. MAE provides us with a genuine measure of how much error we can expect on average.
3. **Mean Squared Error (MSE):** MSE calculates average squared difference between actual and estimated values, or average of squares of errors. MSE is sensitive to outliers because it punishes larger errors more harshly than smaller errors. The lower the MSE, the better when the model is performing well.
4. **Root Mean Squared Error (RMSE):** RMSE is more error-prone towards big errors and gives me a measure of the size of the error on average. It gets closest to me allowing me to comprehend how much actual values move away from forecasts.
5. **Mean Absolute Percentage Error (MAPE):** MAPE is an average percentage difference measure of actual property prices and prices that the model predicts. It provides a directly interpretable view of the predictive precision in percent form and is effective in communicating model results to stakeholders in the real estate industry.

Results

1. **Accuracy**: With the high R-squared value and no error indicators, Random Forest model provided accurate estimates. The model's predictive power in predicting property prices was adequately fulfilled, for example, by a study conducted in Karachi with R-squared being 0.947 with Mean Absolute Error (MAE) of 0.125 and Mean Squared Error (MSE) of 0.046 (Jha, 2020).
2. **Robustness**: The model's excellence and ability to generalize from unknown datasets is exemplified in its ability to account for a large proportion of variance in training data. Random Forest model stability across multiple datasets is also achieved in results which show models with better R-squared values, for instance, 0.9761 of a study in South Korea, will perform best in unseen data (Pastukh, 2025).

3. **Practicality**: Low MAPE and MAE values indicate that the model's output is useful for practical application like property valuation and investment analysis. For instance, Random Forest attained a MAPE of 5.482 compared to other machine learning models to forecast the price of property in research, indicating that it can be utilized for practical decision-making (Xu, 2022).

Numerous studies have validated how accurately machine learning model Random Forest actually forecasts property prices:

- In a study of the London property market, Random Forest was employed to test the significance of economic variables, with the model sustaining a good level of variance explained and demonstrating the power of the model to capture intricate pricing patterns (Coleman, 2022).
- Random Forest and Gradient Boosting techniques proved to be very good predictors and adaptable in a study comparing some machine learning models for predicting property prices in Dubai (Pastukh, 2025).
- Random Forest regression was employed to predict Chicago suburb house prices and proved able to handle poorly gathered and noisy data to produce high-quality property price predictions even under times of volatile market conditions (Xu, 2022).

These results prove the efficacy of machine learning methods more specifically; Random Forest algorithm in property value forecasting. A combination of various sources of references drawn from previous studies attests to the validity and applicability of this process to the case of appraising real property.

4.8 Conclusion

With the real property industry, the Random Forest algorithm, a machine learning method, has proven highly effective in predicting property prices. With an R-squared of high value, the model exhibited a high degree of explanation of property price variation against input features including location, size (square footage), number of bedrooms, number of bathrooms, year built or when last renovated, and amenities. It is overwhelmingly clear in its application in the real world since it has the capability to allow stakeholders to make highly informed decisions for market analysis, investment planning, and house valuation. Subsequent research must concentrate on feature engineering using enhanced methods to try and include more appropriate factors, continuous model updating to accommodate the time-varying behaviour of the property market, and further validation on other datasets.

CHAPTER 5

CONCLUSION AND RECOMMENDATIONS

5.1 Introduction

The following chapter presents an in-depth summary of machine learning study results namely the Random Forest algorithm used in property evaluation for real estate. The three primary objectives of the study involved analysing Random forest in property evaluation., designing and implementing a machine learning model (Random Forest) for property evaluation system, and evaluating the effectiveness of using Random forest in property evaluation for Real Estate. Below is a description of each objective finding and then a measurable assessment of model performance.

5.2 Summary of Findings for Objective 1

Examining random forests used in real estate estimation was the general aim. The model outperformed base models in prediction with an R^2 of 0.87. Domain expertise was also confirmed by feature importance analysis, where building size and location are the all-important parameters in the estimation of real estate prices. The findings confirm the effectiveness of Random Forest used in real estate price prediction as well as decision-making in the real estate industry.

Random Forest has proved to be a suitable method of classifying properties in the Bulawayo real estate market and offering predictive capability as well as explain ability. Therefore, it is worthy of being used in the creation of intelligent decision-support systems for real estate sellers, buyers, and agents.

5.3 Summary of Findings for Objective 2

The second objective of this study was to design and implement a machine learning algorithm, specifically the Random Forest algorithm, for property evaluation system. The system was effectively implemented, with a simple user interface where one can input property details and get estimated values. Included among the system's primary functionalities are user login, input fields for property attributes (such as location, square footage, bedrooms, bathrooms, year built, year renovated, amenities, and material), and real-time updates of the estimated property values. The usability and ease of use of property valuation facilities are augmented through seamless user experience and immediate feedback provision enabled by incorporating the machine learning model into the web application.

The model was tested using a dataset composed of various properties' attributes including location, area, bedrooms and bathrooms, and other attributes. As a result of its ability to handle complex, non-linear relationships between attributes, Random Forest was employed. The technique could abstract implicit patterns from data and was very accurate with regards to property price predictions after being trained. The predictive power of the model was ensured through performance measures like R-squared (R^2), Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE).

The accuracy and validity of the model were ensured by the low values of error, and the high value of R^2 ensured that it had the ability to explain most of the variation in true real estate values. The results validate the successful implementation of the Random Forest model and its appropriateness in undertaking property valuation tasks for the property sector.

5.4 Summary of Findings for Objective 3

To quantify to what extent machine learning is applicable on the analysis of real estate property was the third goal. Several measuring parameters such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and R-squared (R^2) were employed to quantify the performance of the Random Forest model. With the value of R2 as 0.870, the model can explain approximately 87.0% of variance in property prices. Low MAE, MSE, RMSE, and MAPE values further enhance accuracy and reliability of the model for prediction purposes. The results provide support for the effectiveness of the Random Forest algorithm for property valuation and application in real estate industry practice.

5.5 Quantitative Evaluation

Mean Absolute Error (MAE) of 43224.97, Mean Squared Error (MSE) of 4669713572.27, Root Mean Squared Error (RMSE) of 68335.30, Mean Absolute Percentage Error (MAPE) of

0.245%, and R-squared (R^2) of 0.870 were findings of quantitative analysis of Random Forest model. All these measures, when viewed as an ensemble, show that the model is providing consistent, precise, and helpful predictions regarding the property valuation. The model stability and precision are proven by the small error statistics, and the high R2 statistic indicates that it explains a very vast majority of property price change.

5.6 Conclusion

The three general goals of the study were accomplished: to create a system of property evaluation, to apply a Random Forest machine learning algorithm to forecast price, and to validate how well machine learning can be applied in property evaluation. The users now have an easy and efficient instrument to scrutinize properties due to the incorporation of machine learning model to the web application. The accuracy and reliability of the Random Forest model's application in real estate were proven by its high values. In general, the research illustrates how machine learning algorithms can be designed to facilitate real estate business property assessment processes.

5.7 Recommendations

They are the recommendations on how to improve the accuracy and genuineness of the system based on outcome and quantitative analysis of the Random Forest model for real estate property price forecasting. To make the model generalizable to out-of-sample data and reduce the likelihood of overfitting, additional validation needs to be done through testing it on alternate test sets. Second, there is a suggestion of more sophisticated feature engineering in terms of addition of additional features as well as transformations that may increase the model's explanatory power, e.g., economic statistics, crime rates per neighbourhood, and amenity accessibility. Third, the hyper parameter tuning of the Random Forest model must also be increased further so that it is improved. Fourth, the use of k-fold cross-validation methods will ensure that the performance of the model remains the same across any data subset set. Lastly, updates are strongly recommended to ensure that the system will keep on running at its best; this is through the provision of new information regularly to the model so that it can keep up with its accuracy and validity in the ever-changing real estate world. Following these steps, the property valuation system can be further enhanced and tailored to the evolving needs of the real estate sector.

References

Abdolrasol, M. G. M. S. H. S. M. U. T. S. S. M. R. H. M. A. M. R. A. J. A. M. S. &. M. A., 2021. Artificial neural network based optimization techniques: A review.. *In Electronics (Switzerland) ,* 10(21).

Adzanoukpe, P., 2025. Predicting House Rental Prices in Ghana Using Machine Learning.

Ake, I., 2022. Combining machine learning models to predict house prices. *(Doctoral dissertation, Doctoral dissertation, Southampton Solent University)..*

Ake, I., 2022. Combining machine learning models to predict house prices (Doctoral dissertation, Doctoral dissertation, Southampton Solent University).

Angrick, S. B. B. H. N. K. M. S. J. D. V. M. L. F. T. &. K. M., 2022. Towards explainable real estate valuation via evolutionary algorithms. *GECCO 2022 - Proceedings of the 2022 Genetic and Evolutionary Computation Conference,* p. 1130–1138.

APNews, 2024. Zimbabwe launches new land policy to empower Black farmers with direct farm ownership.

Average, C., 2019. Low income housing problems and low-income housing solutions: opportunities and challenges in Bulawayo. *Journal of Housing and the Built Environment,* 34(3), pp. 927-938..

Bax, D. Z. T. a. N. D., 2019. A machine learning web application to estimate listing prices of South African homes.. *Journal of African Real Estate Research,* 4(2), pp. 1-23.

Begum, A. K. N. a. R. M., 2022. Housing price prediction with machine learning. *International Journal of Innovative Technology and Exploring Engineering,* 11(3).

Bhanye, J., 2024. Beyond informality:'Nimble peri-urban land transactions'-how migrants on the margins trade, access and hold land for settlement. *Discover Global Society,* 2(1), p. 6.

Bhatt, B. a. N. M., 2021. An Overview of Structural UML Diagrams.. *International Research Journal of Engineering and Technology,,* 8(8), pp. 1577-1583.

Bowen, Y. ,. B. C., 2018. Research on Ensemble Learning-based Housing Price Prediction Model. *Big Geospatial Data and Data Science,* Volume 1, pp. 1-8.

Chigwada, J., 2022. Feasibility of a national open data policy in Zimbabwe. *Frontiers in Research Metrics and Analytics,* Volume 7, p. 985999..

Coleman, W. J. B. P. N. V. J. F. N. a. S. H., 2022. Using machine learning to evaluate real estate prices using location big data.. *In 2022 Systems and Information Engineering Design Symposium (SIEDS) ,* pp. 168-172.

Creswell, J., 2003. Qualitative, Quantitative. and mixed methods approaches.. *Research Design,* p. 723954.

Gambe, T., 2015. The importance of databases in real estate industry: Exploring the experiences of realtors in Harare, Zimbabwe.. *The International Journal of Business & Management,* 7(3), p. 7.

Goh, C. a. C. H. e., 2024. Rethinking Pathways to a Sustainable Built Environment. *Taylor & Francis.*

Gružauskas, V. K. A. Č. D. N. V., 2020. Analytical method for correction coefficient determination for applying comparative method for real estate valuation. *Real Estate Management and Valuation,* 28(2), pp. 52-62.

Hlongwane, J. S. G. M. A. a. M. T., 2024. Towards the integration of artificial intelligence in higher education, challenges and opportunities: The African context, a case of Zimbabwe. *Int J Res Innov Soc Sci,* 8(3S), pp. 417-435.

Ho, W. T. B. a. W. S., 2021. Predicting property prices with machine learning algorithms.. *Journal of Property Research, ,* 38(1), pp. 48-70..

Ja'afar, N. S. M. J. a. I. S., 2021. MACHINE LEARNING FOR PROPERTY PRICE PREDICTION AND PRICE VALUATION: A SYSTEMATIC LITERATURE REVIEW. *PLANNING MALAYSIA JOURNAL,* 19(3), pp. 411- 422.

Jha, S. B. R. P. V. a. J. R., 2020. Housing market prediction problem using different machine learning algorithms.

Jha, S. B. R. P. V. a. J. R., 2020. Housing market prediction problem using different machine learning algorithms. *A case study. arXiv preprint arXiv:2006.10092.*

José-Luis Alfaro-Navarro, E. L. C. E. A.-C. N. G. M. G. B. L., 2020. A Fully Automated Adjustment of Ensemble Methods in Machine Learning for Modeling Complex Real Estate Systems. *Complexity,* Volume 2020.

Khan, H. A. M. K. H. A. &. R. J., 2023. APPLYING MACHINE LEARNING MODELS FOR FORECASTING HOUSE PRICES-A CASE OF THE METROPOLITAN CITY OF KARACHI, PAKISTAN.. *Article in Journal of Entrepreneurship Management and Innovation,* 5(3).

Kusuma, J. P. N. A. H. &. S. E. E., 2018. A Review of Expert System Applied in Real Estate Valuation. *International Journal of Engineering & Technology,* 4/22(7), pp. 12-16.

Kwangwama, N. M. S. a. M. Y., 2024. A Review of the Contribution of the Real Estate Sector Towards the Attainment of the New Urban Agenda in Zimbabwe.. *New Urban Agenda in Zimbabwe: Built Environment Sciences and Practices,* pp. 151-168..

Liu, Z., 2023. Real Estate Price Prediction based on Supervised Machine Learning Scenarios.. *In Highlights in Science, Engineering and Technology CMLAI,* Volume 2023.

Mamre, M. a. S. D., 2024. Coming of age: renovation premiums in housing markets. *The Journal of Real Estate Finance and Economics,* 69(2), pp. 307-342.

Mao, T., 2023. Real Estate Price Prediction Based on Linear Regression and Machine Learning Scenarios.. Volume 38.

Moosavi, V., 2017. Urban data streams and machine learning: a case of swiss real estate market..

Mora-Garcia, R. C.-L. M. a. P.-S. V., 2022. Housing price prediction using machine learning algorithms in COVID-19 times. *Land,* 11(11), p. 2100..

Mupondo, N. N. B. M. A. a. N. S., 2023. Determinants of house prices using spatial analysis: the case for Bulawayo. *Zimbabwe Journal of Science and Technology,* 18(1), pp. 12-33.

Niu, J. a. N. P., 2019. An intelligent automatic valuation system for real estate based on machine learning. In Proceedings of the international conference on artificial intelligence, information processing and cloud computing. pp. 1-6.

Özalp, A. a. A. H., 2024. Comparison of tree-based machine learning algorithms in price prediction of residential real estate.. *Gümüşhane Üniversitesi Fen Bilimleri Dergisi,,* 14(1), pp. 116-130..

Paradza, P. a. C. C., 2015. Undergraduate real estate education in Zimbabwe: A comparative study. Unpublished MSc Dissertation.. *Faculty of Engineering, Built Environment and Information Technology University of Pretoria..*

Pastukh, O. a. K. V., 2025. Using ensemble methods of machine learning to predict real estate prices..

Pastukh, O. a. K. V., 2025. Using ensemble methods of machine learning to predict real estate prices. *arXiv preprint arXiv:2504.04303.*

Putatunda, S., 2019. PropTech for proactive pricing of houses in classified advertisements in the Indian real estate market..

Quang, T. M. N. H. D. &. B. M., 2020. Housing Price Prediction via Improved Machine Learning Techniques.. *Procedia Computer Science,* Volume 174, p. 433–442.

Review, P. M., 2023. Annual report on real estate trends in Zimbabwe. *Knight Frank Zimbabwe.*

Ribeiro, J. L. R. E. T. &. P. S., 2021. Robotic Process Automation and Artificial Intelligence in Industry 4.0 - A Literature review.. *Procedia Computer Science,,* Volume 181, pp. 51-58.

Riccio, V. J. G. S. A. H. N. W. M. a. T. P., 2020. Testing machine learning based systems: a systematic mapping. *Empirical Software Engineering,* Volume 25, pp. 5193-5254.

ROOT, T. H., Strader, T. J. & and Huang, Y.-H., 2023. A Review of Machine Learning Approaches for Real Estate Valuation. *Journal of the Midwest Association for Information Systems,* 2023(2), pp. 1-21.

Satish, G. R. C. R. M. a. S. C., 2019. House price prediction using machine learning.. *Journal of Innovative Technology and Exploring Engineering,,* 8(9), pp. 717-722.

Shah, D. R. H. &. C. J., 2020. House Price Prediction Using Machine Learning and RPA.. *International Research Journal of Engineering and Technology..*

Sharma, H. H. H. a. O. B., 2024. An optimal house price prediction algorithm: XGBoost. Analytics. 3(1), pp. 30-45..

Sharma, S. A. D. S. G. S. P. a. M. V., 2023. House price prediction using machine learning algorithm. *In 2023 7th International Conference on Computing Methodologies and Communication (ICCMC),* pp. 982-986.

Toussaint-Comeau, M. a. L. J., 2018. Determinants of housing values and variations in home prices across neighborhoods in cook county. *ProfitWise News and Views,* Volume 1, pp. 1-22.

TransparencyInternational, 2024. Corruption Perceptions Index 2024: Zimbabwe..

Tsabora, J., 2016. Reflections on the constitutional regulation of property and land rights under the 2013 Zimbabwean Constitution.. *Journal of African Law ,* 60(2), pp. 213-234.

Verma, P. K. A. S. &. A. C., 2023. Predicting future housing prices: a machine learning approach. *Multidisciplinary Science Journal,* Volume 5.

Waddell, P. a. B.-Z. A., 2020. A comparison of statistical and machine learning algorithms for predicting rents in the San Francisco Bay Area..

Wang, D. a. L. V., 2019. Mass appraisal models of real estate in the 21st century A systematic literature review. *Sustainability,* 11(24), p. 7006.

Wei, C. F. M. W. L. Y. H. T. F. &. X. Y., 2022. The Research Development of Hedonic Price Model-Based Real Estate Appraisal in the Era of Big Data.. *In Land,* 11(3).

Wei, Y. Z. X. &. L. Y., 2020. Machine learning in real estate valuation: Challenges and opportunities. *Journal of Property Research,* 37(4), pp. 345-362.

Wikipedia, 2024. Land Apportionment Act of 1930..

Xu, K. a. N. H., 2022. Predicting housing prices and analyzing real estate market in the Chicago suburbs using Machine Learning. *arXiv preprint arXiv:2210.06261.*

Yazdani, M., 2021a. Machine Learning, Deep Learning, and Hedonic Methods for Real Estate Price Prediction.. *arXiv preprint arXiv:2110.07151..*

Yazdani, M., 2021. Machine Learning, Deep Learning, and Hedonic Methods for Real Estate Price Prediction.. *arXiv preprint arXiv:2110.07151..*

Yazdani, M. a. R. M., 2023. Real Estate Property Valuation using Self-Supervised Vision Transformers..

Yoshida, T. M. D. a. S. H., 2024. Spatial prediction of apartment rent using regression-based and machine learning-based approaches with a large dataset.. *The Journal of Real Estate Finance and Economics,* 69(1), pp. 1-28.

Zhanda, K., 2020. Beyond recovery? Downturns, implications and prospects of Covid-19 pandemic to real estate development in Zimbabwe. *International Journal of Real Estate Studies,* 14(S1), pp. 31-40..

Zhong, S., 2024. Factors influencing housing prices: A comparative study using multiple linear regression and random forest. *Theoretical and Natural Science,* Volume 51, pp. 73-79.

ZimAdvocate, 2024. Zimbabwe's Real Estate Boom: What's Driving It?.

ZIMPropertyDigest, 2025. Zimbabwe Real Estate Risks: A Verified Investor Guide..

ZIMSTAT, 2022. Zimbabwe Population and Housing Census Preliminary Report.. *Zimbabwe National Statistics Agency.*

APPENDICES

GODWIN S. JIRI B213599B.docx

ORIGINALITY REPORT

6%	5%	4%	1%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Poonam Nandal, Mamta Dahiya, Meeta Singh, Arvind Dagur, Brijesh Kumar. "Progressive Computational Intelligence, Information Technology and Networking", CRC Press, 2025 Publication	1 %
2	www.researchgate.net Internet Source	<1 %
3	store.ccilearning.com Internet Source	<1 %
4	arxiv.org Internet Source	<1 %
5	Pawan Singh Mehra, Dhirendra Kumar Shukla. "Artificial Intelligence, Blockchain, Computing and Security - Volume 2", CRC Press, 2023 Publication	<1 %
6	sic.ici.ro Internet Source	<1 %
7	McDaniel, Claire. "Predicting Augmented Reality Virtual Try-On Adoption Using Machine Learning.", National University Publication	<1 %
8	Ahasan, Sohanul. "Estimating Steel Material Recovery From End-of-Life Ships: A Reverse Logistics Approach to Achieve Sustainable Development Goals and Promote Safe	<1 %