Republic of the Philippines
Western Mindanao State University
College of Computing Studies
DEPARTMENT OF COMPUTER SCIENCE
Zamboanga City

Web-Based House & Lot Price Prediction Using Gradient Boosting in Zamboanga City

A Thesis presented to the faculty of
Department of Computer Science
College of Computing Studies

In partial fulfillment of the requirements for the degree of
Bachelor of Science in Computer Science

James Kheil S. Francisco
Christle Joyce U. Santuyo
Researchers

Engr. Edwin I. Arip, MEnggEd - ICT
Adviser

April 8, 2022

Western Mindanao State University
College of Computing Studies
DEPARTMENT OF COMPUTER SCIENCE
Zamboanga City

Approval Sheet

The Thesis attached hereto, entitled **"Web-Based House and Lot Price Prediction using Gradient Boosting in Zamboanga City"**, prepared and submitted by James Kheil S. Francisco and Christle Joyce U. Santuyo, in partial fulfillment of the requirements for the degree of Bachelor of Science in Computer Science, is hereby **recommended for Oral Examination**.

ENGR. EDWIN I. ARIP, MEnggEd-ICT
Adviser

APPROVED by the Oral Examination Committee on **January 10, 2022** with a rating of **PASSED**.

MS. LUCY F. SADIWA, MSCS
Chairperson

MR. JAYDEE C. BALLAHO
Member

ENGR. MARJORIE A. ROJAS
Member

ACCEPTED in partial fulfillment of the requirements for the degree of **Bachelor of Science in Computer Science**

ENGR. ODON A. MARAVILLAS JR, MSCS
Head, Department of Computer Science

RODERICK P. GO, Ph.D.
Director, College of Computing Studies

Acknowledgement

First and foremost, praises and thanks our Almighty God for His blessings and spiritual guidance throughout this study and for giving us the strength and knowledge to finish this thesis.

This study will not have been completed and successful without the guidance of our former adviser **Engr. Marvic A. Lines** for patiently guiding us on how to start our thesis study.

To our thesis adviser **Engr. Edwin I. Arip** and with the help of **Mr. Jaydee C. Ballaho**, for the immense patience, effort, understanding, and vast knowledge they have given and shared with us. Thank you so much!

To **Mr. Arnel V. Bayot** of the City Assessor's Office, without you we would not have the data that we need for us to complete our study. Thank you for your patience and understanding, sir. You will always have our gratitude!

And lastly, to our beloved family, the **Francisco Family** and **Santuyo Family,** who served as our inspiration and strength. Thank you for your love, financial support, emotional support, and understanding. Without you, we would not be able to finish our study. Thank you so much!

To the people who have been part of this research journey, classmates, teachers, friends, colleagues, and loved ones, thank you so much for the love and support!

To God be the Highest Glory!

Abstract

People are cautious when looking for a new home and regarding budgeting and market techniques. Machine-based forecasting technologies are not still known and utilized to estimate houses or other real estate properties in the Philippine Real Estate market. Thus, the researchers developed a web-based house and lot price prediction and embedded it with a machine-learning algorithm to forecast house and lot property prices in the Philippines, particularly in Zamboanga City. This study focuses on developing a system that predicts the calculated cost of a house and lot property. The information on the property listing was taken from historical assessed data of house and lot from the City Assessor's Office of Zamboanga City. The study uses three (3) different machine learning algorithms and chooses Gradient Boosting Regression to predict the property's price. The Gradient Boosting Regression technique has a positive relationship between the dependent and independent variables. Compared to all other algorithms for predicting property prices, the gradient boosting approach has a high accuracy value. The measure can be improved even more by applying some preprocessing before fitting the data. The goal of this study is to help the seller sell a property at a reasonable price and help the buyer know the factors that affect the selling price of the property. Some of the related factors that impact the cost were also considered, such as location, physical conditions, and materials used.

Keywords: *Machine Learning, Gradient Boosting, House Price Prediction*

Table of Contents

List of Figures

List of Tables

List of Appendices

CHAPTER I
INTRODUCTION

Background of the Study

Machine learning is a branch of Artificial Intelligence (AI) that uses algorithms and technologies to extract valuable facts from large data. Machine learning approaches are suited for big data since manually processing massive amounts of data would be difficult without the assistance of machines. Machine learning tries to solve problems algorithmically or step by step rather than mathematically in computer science. [1] The center of developing algorithms is to allow the machine to learn. On the other hand, machine learning has two categories: supervised and unsupervised. It is a Supervised algorithm if datasets are trained on a predetermined set of data to predict and present new data. The program tries to uncover the hidden pattern and relationship between the data when run unsupervised. Many regression methods rely on an unknown number of variables to produce the predicted value. The performance is tested by forecasting house prices. Specifications can determine the cost of a home. Houses have a variety of characteristics that may or may not have the exact cost depending on their location. For example, a large house in a desired wealthy neighborhood may command a higher price than one in a poor community.

A variety of factors determines the values of properties. In real estate, there is an explanation regarding a rise in the market by increasing the local population's income. As a result of the COVID-19 pandemic, the Philippine economy fell by 9.5 percent in 2020 compared to the previous year, the most significant drop since the Philippine Statistics Authority (PSA) began collecting statistics in 1946. From 2010 to 2019, the economy increased at an annual rate of 6.4 percent. [8] However, rigorous study reveals that these characteristics, such as demand-oriented variables and others, can only momentarily suggest rising real estate prices. As a result, it can deduce that the variables are subject to change throughout time. The average household income determines house prices in the area, the availability of housing stock, and the payment mechanism (whether it accepts installment payments or requires cash payment).

Real estate corporations in Europe and other advanced countries are battling to build algorithms that can more precisely estimate real estate property prices. Researchers rely on well-known housing datasets, such as Boston and King City in the United States. The lack of a comprehensive housing dataset is one of the gaps in the Philippines, particularly in Zamboanga City. Some Philippine real estate property sites provide a good approximation of the housing market; however, they do not currently use house price predicting methods. Competitions on Kaggle are organized by websites like Zillow, a US real estate marketplace, to motivate researchers to develop accurate house price forecasting algorithms. Because such issues are uncommon in the Philippines, the only sources of housing data are those acquired in the Assessor's Office of a particular area in the Philippines, making it extremely difficult for researchers to develop such forecasting algorithms for the Philippine real estate housing data.

Machine-based forecasting technologies are not yet used to estimate houses or other real estate properties in the Philippine real estate market. Thus, the researchers developed a house and lot price prediction with the application of a machine-learning algorithm to forecast the price of a house and lot property that includes a Land property price prediction because some other related study about price prediction is just prediction of a house property. The researchers add this feature because the Philippines is also a hotspot for local and foreign property investors in building their businesses as well as building their own homes. [16] Because of the low property rate of the country and great opportunities for investments, many foreign investors are willing to own a property such as a house and lot. The outcome of the study will be helpful for the people who want to know their property's value and for interested buyers who want to know the value of their desired property. This will also answer how the price anticipated by the system is close enough to the actual cost of the house and lot utilizing a machine learning algorithm.

Statement of the Problem

The lack of a comprehensive housing dataset is one of the gaps in the Philippines, particularly in Zamboanga City. Some Philippine Real Estate property sites approximate the housing market; however, they do not currently use house price-predicting methods. There is still an absence of machine-based forecasting technologies to estimate houses or other real estate properties in the Philippine real estate market.

Thus, the initialization machine learning project predicts the house and lot property price with the application of machine learning algorithms such as Gradient Boosting. It will determine whether an independent variable positively affects the dependent variable.

Objectives

The general objective of the study is to predict the price of a house and lot property with the use of Gradient Boosting. Gradient boosting is a technique for building prediction models, and it is used in improving cost functions by selecting weak hypotheses or a function with a negative gradient iteratively.

Specifically, the study will:

- Gather datasets of sold properties needed for the prediction model to work.
- Create a predicting model using Gradient Boosting with at least 80% predictive accuracy.
- Compare the predicted price on the actual selling price of a property to have fair standard pricing of a property.
- Give an idea to the buyers, investors, and sellers about the selling price of a house and lot property.
- Prove that the use of Gradient boosting will have a positive relationship between the attributes of the property and the selling price.

Scope and Limitations

Scope

The scope of the study is to predict the estimated house and lot price depending on factors such as the location of the property, lot area, floor area, and other related attributes that can contribute to the price of the house and lot in Zamboanga City.

The data used in the study would be the properties that have been assessed by the Assessor's Office that contain different attributes specifically such as the house area, lot area, the location of the property, type of the house, other amenities and other related attributes that can contribute to the price of the house and lot.

Limitations

The study is limited in assessing the house properties that are located within the seven (7) kilometer radius and limited in evaluating the land properties within and outside of the seven (7) kilometer radius of Zamboanga City. This study constrains the system's capability of evaluating the materials used in building the house, appliances, and furniture of the house, if there is any. Value-Added Tax (VAT) is also excluded from the calculation.

Significance of the Study

The result of this study will be significant to the following:

Buyers. Through this study, they will have an idea of how much is the value of their desired house and lot. This system will give them a planned preparation to save or prepare the money to purchase their desired property.

Investors. For them to know if the selling price of the house and lot is close enough to its actual price. This will give them an insight into the property's attributes that contributes to the price of the house and lot.

Sellers. They can sell the house and lot property at a fixed and calculated price. This system will help them to be credible. This study constraint can prove that the price is accurate or close enough to the real price.

Real Estate Company. For them to have a tool that helps them calculate the price of the house and lot in a fair market and improves their technical skills. The system will help them determine the selling price of the property.

Future Researchers. To guide, assist, and give them an idea of the possible improvements and developments they can include in this study.

Definition of Terms

Term	Definition
1. Accuracy	– the measurement's proximity to a specified value.
2. Datasets	– a collection of standardized data to be used in the system.
3. Gradient Boosting	– a machine learning technique used in regression and classification tasks, among others. It gives a prediction model in the form of an ensemble of weak prediction models, which are typically decision trees.
4. House & Lot	– a type of property than can be sold individually or in a package deal.
5. Machine Learning	– allows an algorithm to learn from the data and to improve its experience with less human interaction.
6. Prediction	– the act of estimating something (result or value) that will happen in the future or will be the result of the present.
7. Price	– the amount of money that is desired, demanded or offered in exchange for something, the value of a certain thing.
8. Property	– the possession or possessions of a particular owner or individual.

9. Testing data	– a set of data that is used to evaluate the model or system's accuracy.
10. Training data	– a set of data that is used to train a machine learning application to identify patterns and perform the specification.
11. Data cleaning	– is the process of detecting and correcting (or removing) corrupt or inaccurate records from a record set, table, or database and refers to identifying incomplete.
12. Data processing	– is simply converting raw data to meaningful information through a process.
13. Linear regression	– attempts to model the relationship between two variables by fitting a linear equation to observed data.
14. Ridge regression	– is a method of estimating the coefficients of multiple-regression models in scenarios where linearly independent variables are highly correlated. *(Great Learning Blog)*
15. Appraisal	– a valuation of the property by the estimate of an authorized person *(Merriam-Webster)*

Table 1 Definition of Terms

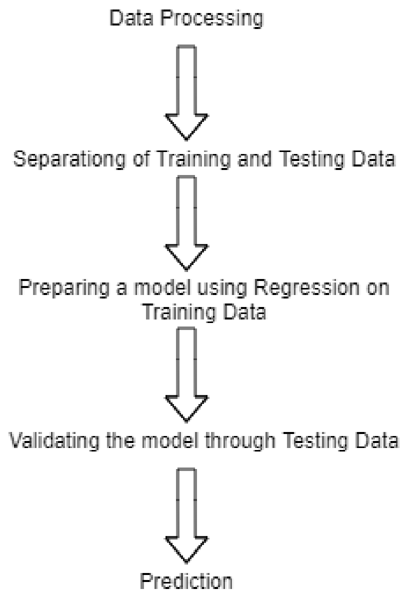
CHAPTER II
REVIEW OF RELATED LITERATURE

Related Studies

In the study of Kumar et al. [4], numerous studies have proved the impact of a property's location and surrounding neighborhood in determining its price. However, some neighborhood characteristics may be unobservable, and the optimum way to acquire location data remains an unanswered subject for the price forecasting model. The goal of their study is to predict the predicted price of a house property using a gradient boost regressor. The initial stage is to gather raw data from multiple sources, and the dataset might be any historical data. The attributes are the number of stories, the number of bedrooms, the number of bathrooms, the availability of a garage, a swimming pool, a fireplace, the year the house was built, the square footage, and the sale price. The researchers used these features to train the machine on Gradient boost regression and forecast the house price. They also calculated the accuracy by comparing the test set predictions to the actual values. The suggested system calculates the anticipated price. They used a decision tree classifier, by which 70% of the data were used for training, and 30% was used for testing the data. The decision tree algorithm generates the categorization rules. The data can be tested using the training data. Using this model made it possible to get output or an accurate predicted stock price.

Revend [13] describes how they preprocessed the data and the steps they made before training the model. Their study focuses on predicting the price of a house located in the countryside of Sweden using boosted decision trees. They first find the geographical coordinates, latitude, and longitude to improve the positional accuracy of the areas in Sweden. Then they sort out the residences that do not belong to the countryside. They only selected the villas type of residence, and they filtered out the other type of residence. Then they use the KT filter to exclude the residences that are under or overvalued. Since their data was sold properties, so they need to adjust the residence price. They also treat the missing data by removing the redundant data and replacing some variables with a large percentage of missing values with -1, and replacing some not a number (nan) deals with the mean. They used the method of feature scaling to normalize the variables. A scaling algorithm, RobustScaler is used to robust the outliers. They also used 70% of their data set for training data, and 30% was used for testing data. They did hyper-parameter to know the number of iterations, step size of the gradient boosting algorithm, and the maximum tree depth.

Aalam & Gaurav [7] uses a path that they used for preparing the model for house price prediction purpose.



Path used for preparing a model for prediction (Aalam & Gaurav, 2018)

The above diagram shows the steps or way of preparing a model for predicting house price. The first step is the preprocessing of data. It plays a vital role because the data needs to be investigated and preprocessed first before usage to remove some redundancy in data if any, and null values. After preprocessing the data, it will now be the separation of the training data and the testing. The preliminary data collection used by computers to learn how to apply methods such as linear regression to generate correct outcomes is known as training data. On the contrary, testing data are used if the building model is complete. Testing data will determine how the built model performs when a new set of data is inputted. After separating the sets of data to be used, it is now time to prepare the model with the Regression approach to the training data. Here, they used multiple regression because they used several features.

In the study of Ho et al. [9] they used different machine learning algorithms to know what best performs in predicting prices, and each algorithm has other techniques. They employed a regression method called Support Vector Machine (SVM) to split objects into distinct groups while maximizing the distance between data points. Support Vector Regression (SVR), a regression approach developed by SVM, is used to train the model using a supervised learning method. The outcome of their dataset has been divided into five (5) folds as a result of their SVM estimation, and these folds are the subsets that will be used in the iteration for training and testing the dataset. In other words, the resampling procedure or cross-validation is designed for evaluating machine learning models. Gradient Boosting was utilized to turn a weak feature into a strong learner. Gradient Boosting uses the error rate to compute the gradient of the error function, then uses the gradient to figure out how to tweak the model parameters to lower the error with each iteration. According to their study, advanced machine learning algorithms such as SVM, RF, and GBM have been demonstrated to be potential tools for property researchers to utilize in house price predictions. The use of machine learning in property research provides methodological and empirical contributions to property evaluation, as well as giving an alternative approach to the valuation of housing values.

In a study conducted by Uzut [17], the goal of a study was to forecast the price of a real estate property. It examined multiple techniques (linear regression, random forest regressor, and gradient boosting) and discovered that gradient boosting outperformed the others. This model is used to make predictions. Based on the holdout method, three distinct sizes of testing sets were chosen from the dataset: 10%, 20%, and 30%. Then, taking these sizes into account, prediction models were created, and accuracy (highest) and error rates were determined. They utilized 80% of their dataset for the training set and the remaining 20% for the testing set. According to their results, the gradient boosting method delivered the best performance for real estate price prediction. A new dataset with extra parameters can be used to expand this research.

Monika et al. [12] conducted research on *House Price Forecasting using Machine Learning methods* that focus on the idea of predicting the price of a house property. They utilized machine learning algorithms and discovered a reasonable forestalling price with a low fault rate. They used various machine learning algorithms such as Extreme Gradient Boosting, Gradient Boosting, Light Gradient Boosting, Support Vector Machine, and Random Forest Regression, and these algorithms were compared according to their performance in predicting prices. The evaluation of the algorithms used was done by calculating the error value. The greater the error value, the higher the accuracy rate of the regression model. The performance was measured using the Mean Square Error (MSE) and the Root Mean Square Error (RMSE).

Imran et al. [10] conducted a research study about House Price prediction using various machine learning algorithms to determine which algorithms performed better in predicting prices. They used Gradient Boosting, Linear Regression, Bayesian Regression, Support Vector Regressor, Stochastic Gradient descent, ElasticNet Regression, Passive Aggressive Regression and Theil-Sen Regression. To evaluate the performance matrices of each algorithm, they used Mean Absolute Percentage Error (MAPE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE). The outcome of their study resulted in being successful and shows that the Support Vector Regression performs best than the other algorithms they used. Their study shows which of the algorithms they had used has a better performance in predicting prices.

Abdulal and Aghi [1] propose a study that focuses on the performance comparison between machine learning regression and Artificial Neural Network (ANN). Multiple linear, Least Absolute Selection Operator (Lasso), Ridge, and Random Forest are the regression techniques employed in this work. Furthermore, this research seeks to find the most relevant elements that influence property prices in Malmö, Sweden, by analyzing the correlation between variables. In this investigation, two datasets labeled public and local were employed. They show home values in Ames, Iowa, and Malmö, Sweden, respectively. The training model's root square and root mean square error scores are used to assess the prediction's accuracy. After applying the necessary pre-processing methods and separating the data into two halves, the test is carried out. However, one component will be used during training, and the other will be used for testing. They also introduced a binning approach that increased the model's accuracy.

Kokasih and Paramita [11] created software that can predict the price for property rental. Listing feature, neighborhood, review, date, and property owner information are the variables they considered in their study. The user's dataset is used to generate a prediction model, which is then processed using the Extreme Gradient Boosting algorithm and saved in the system. The findings of their study are likely to be used to develop property rent price prediction models for property owners and tourists to consider when renting a property. The result of their study shows that the Extreme Gradient Boosting algorithm can estimate property rental prices with an average RMSE of 10.86 percent (or 13.30 percent).

Singh and J [15] developed a research study of price prediction for a Land property with the use of a Machine Learning Algorithm. In their study, they constructed a machine learning application that considers numerous elements in the real estate market in real-time, such as houses that are advertised at a lower price than the market price, as well as demographics and other factors. They combined two alternative machine learning architectures based on Random Forest (RF) and Linear Regression to predict property prices (LR). As they combined these two machine learning algorithms, the result shows that the enriched Random Forest performs well with numeric features. Since they predict the prices from online advertisements, it requires the insight of data combined with powerful machine learning algorithms.

A research study made by Sahithi et al. [14] *Crop Price Prediction System using Machine learning Algorithms* focuses on identifying appropriate data models that aid in achieving high price forecast accuracy and generality. On various data sets, various data mining approaches were assessed. It describes a system that employs data analytics approaches to forecast crop prices. The suggested system will use machine learning algorithms to forecast crop prices based on several criteria such as harvested area, planted area, and so on. This gives a farmer an idea of what the future price of the crop he'll be harvesting will be. The study creates a system by combining data from a variety of sources. Data analytics and prediction analysis can assist farmers in predicting the crop's target price and boost their profit margins in the long run. After conducting extensive research, they have determined that XGBoost is the best-fit machine learning for their project.

Synthesis

Comparison of the Related Systems

Existing System	Graphical Presentation	Price Prediction	Web-Based	Land Property	Machine Learning
House Price Calculator – Nationwide	□	✓	✓	□	□
Acadata	✓	✓	✓	□	✓
Property Solvers	□	✓	✓	□	□
Springbok Properties	□	✓	✓	□	□
Web-based House and Lot Price Prediction using Gradient Boosting Regression	✓	✓	✓	✓	✓

Table 2 Comparison of the Related Systems

The table shows the different related existing systems that have price prediction.

Property Solvers let the user input the postcode of a certain address, and the system will find the location and display all listings properties within that area.

House Price Calculator and **Acadata** have a feature house price calculator where users can calculate their property's current value. Users must fill in the fields such as property type, purchase price, purchase year, and postcode for them to get the current value of their property.

Springbok Properties is offering a fast transaction of selling a property. Users can search the location of the property they want by filtering in depending on their minimum or maximum price of the property, type of the property, and minimum or the maximum number of bedrooms. This gives the user an idea of what location has the property they are planning to buy.

Most of the predicting systems do not have a prediction for Land properties. The researchers opted to include other functionalities in the system to be suitable in Philippine settings.

The reason why the researchersinclude Land Property is that in the Philippines, it is common to have an empty land property for the people who wants to have an investment. They will acquire a property, and if they will not use or build something on it, they will sell it. Some people just want to know the estimated price of their property. This system will be a tool for them to calculate the price of their property without going outside of their hous

Conceptual Framework

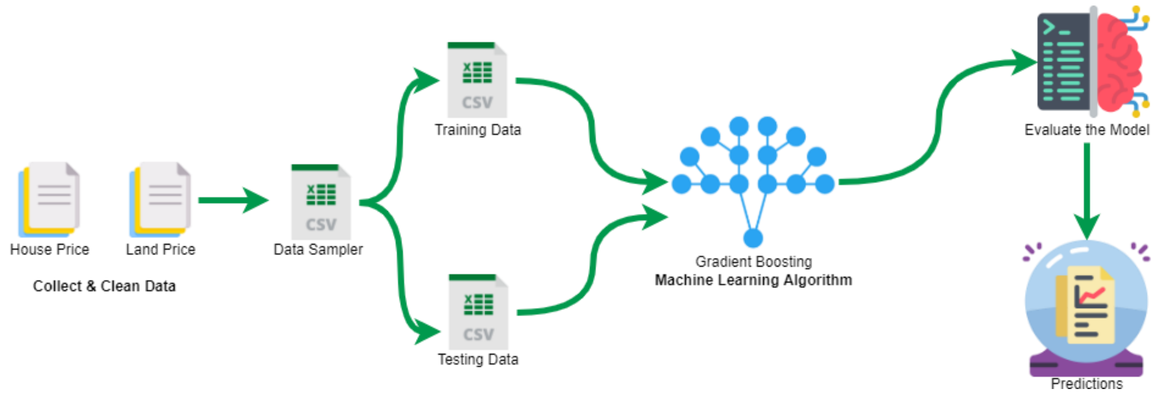


Figure 2 Conceptual Framework

The above diagram shows the steps or path in preparing a model for predicting house prices. The first step is the preprocessing of data. It plays a vital role because the data needs to be investigated and preprocessed first before usage to remove some redundancy in data if any, and null values. After preprocessing the data, it will now be the separation of the training data and the testing. The preliminary data collection used by computers to learn how to apply methods such as gradient boosting to generate correct outcomes is known as training data. Then, the testing data are used if the building model is complete. Testing data will determine how the built model performs when a new set of data is inputted. After separating the sets of data to be used, it is now time to prepare the model with the Regression approach on the training data. The researchers used linear regression in predicting the final value of the house and lot based on the inputted features.

CHAPTER III

METHODOLOGY

Research Design

The purpose of this study is to apply Gradient Boosting in predicting the price of a house and lot property.

This study used an experimental research design to solve a specific problem and provide relatively new solutions to the issues or concerns, or factors that affect society. Empirical research best fits this study. It could be helpful because it includes a scientific approach and uses statistical tools to analyze data. The researchers used this kind of research design to attempt to predict the price of a house and lot property with a machine learning algorithm.

Respondents

The target users of this study are the following: The Buyers will provide them with a strategy for saving or preparing funds to purchase their desired property. The Investors will help them know if the selling price of the house and lot is close enough to the actual worth. The Sellers can sell the house and lot property at a fixed and calculated price. The Real Estate Companies will have a tool that helps them calculate the cost of the house and lot in a fair market and improves their technical skills. Future researchers will make this study a guide, assistance, and giver of the idea of the possible improvements a house and lot can make.

Data Gathering Instruments, Techniques, and Procedures

This section describes how data is gathered, processed, trained, and tested. The data are historical properties assessed and collected at the Assessor's Office of Zamboanga City. The researchers prepared a request letter to gather the said data on the said office. Due to the pandemic, transactions and offices in Zamboanga City, including the Assessor's Office, had closed; therefore, there is a limitation in needed data.

Location, total floor area, and lot area are mostly to consider in predicting the property's price. These would be the independent variables, and the expected price would be the dependent variable. The data collected are from the historical assessed property records. Researchers standardized the data by turning categorical values into numerical values so that the language recognizes them and has equal and no null and redundancy within the data; data are clean and pre-processed. Once done, the data are categorized into two (2) training and test data. Training data is training data set with the machine learning algorithm.

In contrast, in Testing data, a data set will verify if the machine learning has already learned from the fed or trained data. The machine learning algorithm used in training will be gradient boosting, appropriate for predicting continuous values, and used to indicate the house and lot price. Before applying linear regression, determine whether there is a relationship between the variables. To determine that, a scatter plot will be helpful. It is a graphical tool for visualizing the relationship between variables. The machine learning model evaluates and analyzes the data after separating the data processing to predict the house and lot price.

Statistical Tools

Regression Analysis

In the Statistical model, regression analysis collects statistical procedures for calculating the relationships among variables. It includes some techniques for modeling and examining various variables when emphasizing the relationship between the dependent variable and more individual variables or predictors. More specifically, regression analysis helps this study see how the standard measure of this dependent variable (or criterion quantity) shifts when any of these individual variables are different. In contrast, it does fix other independent variables.

Median

In the collected raw data, there is an empty value for the feature land area, floor area, and price; this can contribute to an error in developing the model for houses and lot. The statistical median is the middle value or number in a given sequence of numbers. For the data cleaning, the researchers have used the statistical median to fill up the empty values for the land area, floor area, and the market price. These features are the type of double or floating values, and the researchers did not include other features that should fill with a whole number. It may cause an error for the data because a median could result in decimal numbers and should not be for the other features.

Mode

Categorical data may have an empty value after converting the collected raw data. The non-numerical data incorporates categorical data such as barangay, ceiling, flooring, roofing, roof framing, and other data features. The researchers used mode statistics to fill in the blanks to find the most frequently observed value in a data set. It must be related to the statistical median used for the whole number. The researchers did not use the median for categorical data because there is a tendency to get a decimal result, which is inappropriate for categorical data

Pearson Correlation Coefficient

The researchers used the Pearson correlation coefficient as a test statistic for evaluation in this study. The statistical relationship, or association, between two continuous variables known as the best method for quantifying the relationship between variables of interest because they relate it to covariance. It provides information on the magnitude and direction of the relationship's link or correlation.

Scatter Plot

Before applying linear regression, determine whether there is a relationship between the variables. To determine that, a scatter plot will be helpful. It is a graphical tool for visualizing the relationship between variables.

R-Squared (R^2)

The researchers used R^2 metrics to identify if the data fit or was close to its accuracy. A higher r-squared indicates a better fit for the model. The r-squared value of 0.3 and less than 0.5 is generally considered a weak or low effect size and weak accuracy scores. If the r-squared value of 0.5 and less than 0.7, the value is regarded as a moderate effect, and if the r-squared value is more significant than 0.8,

it is considered a substantial effect size.

Analytical Tools

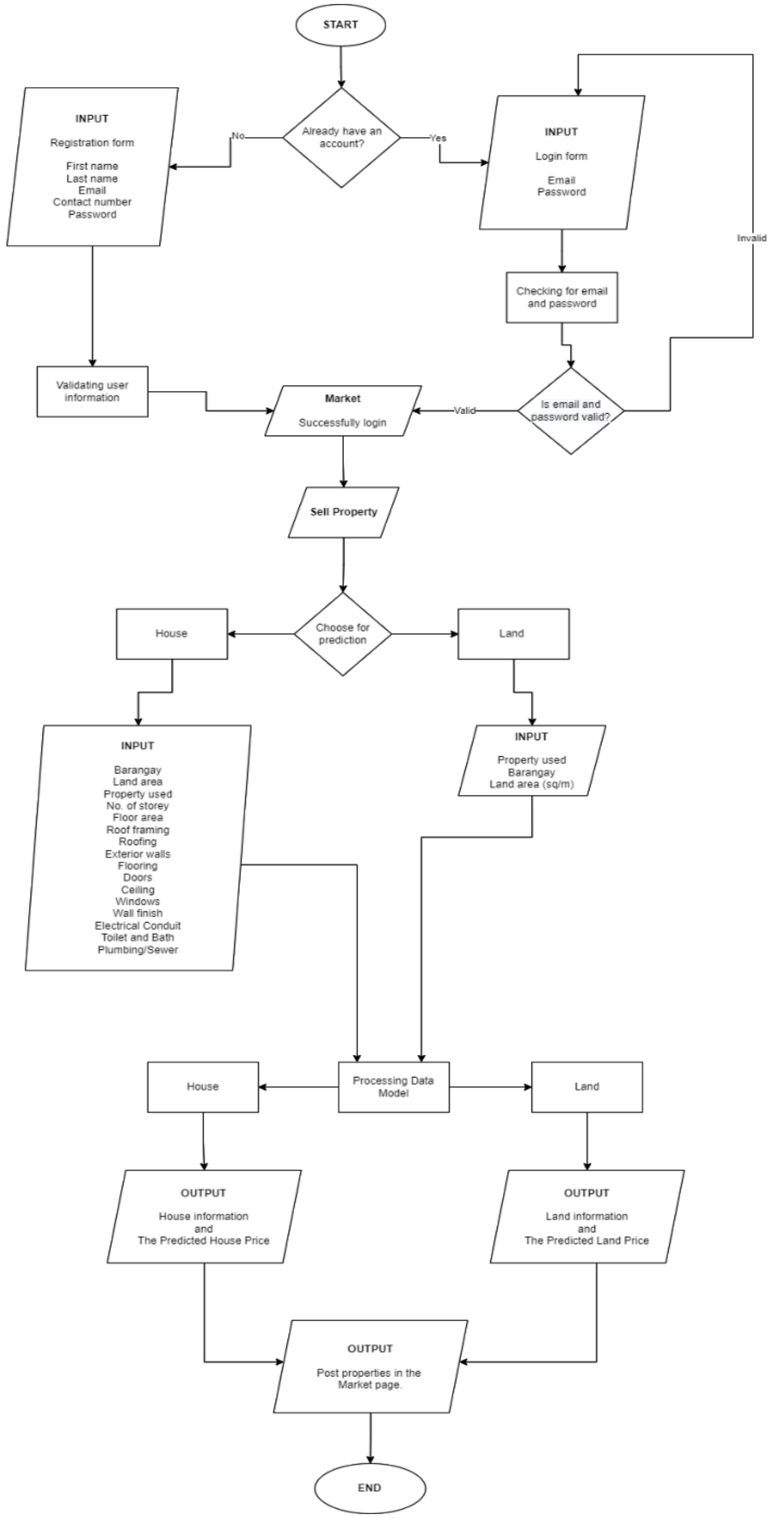


Figure 3 Systems flowchart

For this flowchart, there will be a discussion of the following stages.

1. Start
2. Users must first create an account.
3. To log in, users must give an email address and a password.
4. It will check if the email and password exist or have gone under validation after entering all of the required information for login.
5. The system will display the market page or dashboard if an email and password are legitimate.
6. Select the property where users want to estimate/predict by clicking Sell property.
7. Fill in the required information for each feature shown on the evaluation form, then click "Continue."
8. It will show the market property value after it processes the data model.
9. It will take you to the posting page. You will get another form to fill out to list your property information and the system's estimated price.
10. Once you click the "Post" button, It will now be visible to your property market after you click the "Post" button.
11. End

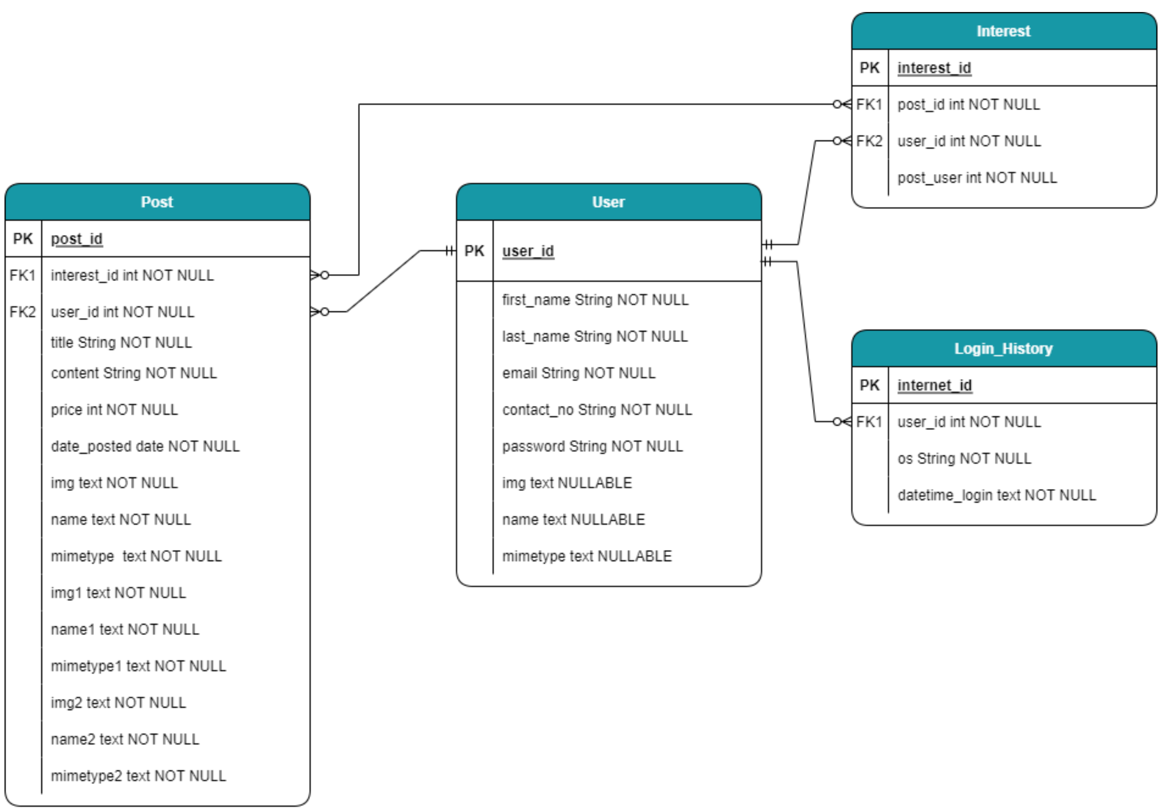


Figure 4 Entity-Relationship Diagram (ERD)

The system allows the users to fill up a data form that will serve as their account information. It contains the first name, last name, email address, contact number, password, and image and its name and mime type for the User class.

The system allows the user to fill out the posting form, which contains the following: title, content, date posted, the requisite three photos, and the estimated price, after completing the relevant data for the house and lot forms.

The system collects minor device data for the Login History class, such as the type of operating system used by the user device and the Date Time when the user logged in.

The system displayed the class of interest as a wish list, allowing users, buyers, and even sellers to add their interest properties to the wish list. The goal is to add interest without an immediate intent to buy.

In the Entity-Relationship Diagram (ERD) process, the user class can have zero or more Posts, and Posts can have one-to-one relationships with the user class. The same logic for the type of interest. A user can have zero or more interests, and the User to Login History class, where a user can have zero or more login records.

Technical Tools

Technical Tools

This section discusses the different tools or components used to develop and complete the system that contains technical terms and the packages. The following are the technical components used in developing the system.

Integrated Development Environment (IDE)

- Visual Studio Code is a multilingual programming IDE that supports a variety of plugins, packages, and languages such as HTML, CSS, and JavaScript suited for the developer.

Data Management

- Orange is an open-source data visualization and analysis tool.
- Microsoft Excel is a spreadsheet used for calculation, computation, and graphing tools.

Front-End Frameworks

● Bootstrap is an open-source front-end development framework for creating websites and web apps. It is built on HTML, CSS, and JavaScript to facilitate the development of responsive, mobile-first sites and apps.

Backend Frameworks

● Jinja is a fast, expressive, extensible templating engine. It allows writing code like Python syntax, and the template will pass data to render the final document.

● Flask is a lightweight WSGI web application framework designed to make getting started quick and easy, with the ability to scale up to complex applications. It has become one of the most popular Python web application frameworks.

● JavaScript is high-level, often just-in-time compiled, and multiparadigm. It has curly-bracket syntax, dynamic typing, prototype-based object orientation, and first-class functions.

● Jupyter is a fast, small, and feature-rich JavaScript library. It makes things like HTML document traversal and manipulation, event handling, animation, and Ajax much simpler with an easy-to-use API that works across a multitude of browsers.

● NodeJS is an open-source, cross-platform, backend JavaScript runtime environment that runs on the V8 engine and executes JavaScript code outside a web browser.

● Jupyter Notebook is an open-source web application that allows us to share documents, equations, and visualizations. The use of jupyter also includes data cleaning and transformation and statistical modeling, which the system used.

● Python is an object-oriented, high-level programming language used in coding and designing the system's backend.

Software Process Model

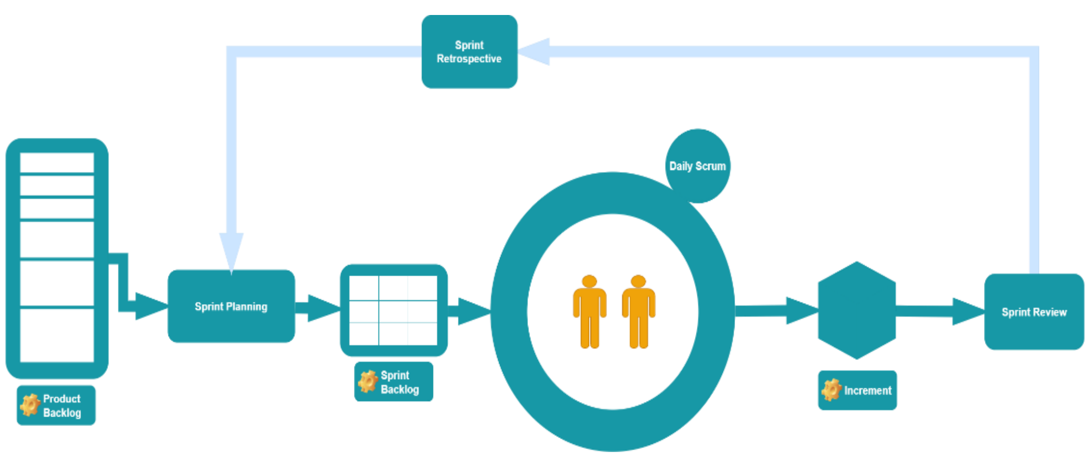


Figure 5 Scrum Methodology

The software development methodology used in this study is the Scrum methodology. Scrum is the iterative and progressive model for managing development. In this system, many essential requirements need to do again to achieve the system's goal. For the process model, the researchers must get all the vital requirements first and then start working on the requirements gathered. In Scrum Methodology, usually, it goes with Sprint. In Sprint, this project might take one year to finish, and it takes time for the client to wait for the entire year to deliver the product. The client will not wait long for the researchers to deliver the product so Sprint will do the process every 2-4 weeks. The client is updated on the product progress so that the sprints will give an initial or potentially shippable product increment. For this visualization, the flows are explained below.

A small rectangle or box represents an essential requirement in the product backlog. It has to break it down for the current Sprint or the current week and its sub-parts. It can define later or every 2-4 weeks depending on the due of the requirements.

The sprint backlog is the process that can call the current sprint backlog, but before that, the researchers must plan the Sprint (Sprint planning) of what will take and what the requirements have. The group got a product backlog, then a plan (sprint planning), and then got a sprint backlog. Once the group has the sprint backlog after 2-4 weeks of working on the system, the next is to build the product. The responsibility for it is the scrum team. The team consists of only two members to do this system.

For the Daily Scrum, the group makes some meetings on the google meet to discuss what to do next for the product. The Sprint Review showed the feeds about the progress and the updates of the product for the first increment. The group deployed more updates about the product in the first two weeks, the first increment. But, in the first Sprint, the group did not complete all the sprint backlog, so they sent the remaining backlog in the next Sprint.

In the previous Sprint, the group will now complete the last Sprint and take some extra knowledge from the product backlog. Still, before going back to the sprint backlog, the group has to do the Sprint Retrospective. The sprint backlog that the researcher has learned will be used in the Sprint Retrospective [1].

In conclusion, Scrum emphasizes teamwork in project management. It stresses accountability and is iterative progress towards a well-defined goal. Scrum is part of agile software development, and teams practice agile. The name comes from rugby, where Scrum is a formation where everyone plays a specific role. However, everyone is working towards a quick adoption of strategies.

System Architecture

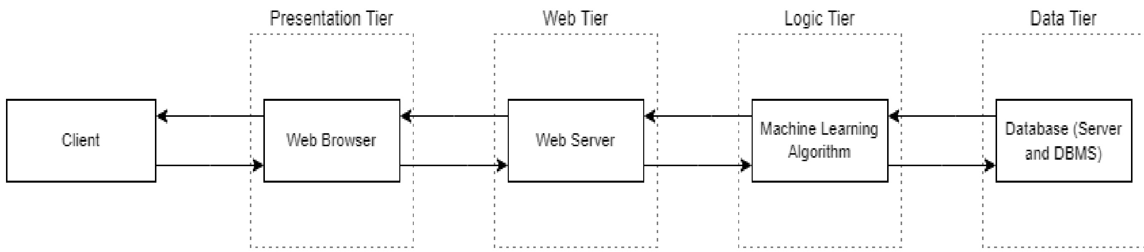


Figure 6 System Architecture

Presentation Tier

The system architecture used in this study is a 4-tier system architecture since the system is a web application or web-based. The buyer, investors, and real estate company will be the client, and they will be able to manipulate the data, such as the barangay, land area (sq/m), floor area (sq/m), classification, storey, roof framing, roofing, exterior walls, flooring, doors, ceiling, windows, wall finish, electrical conduit, toilet, plumbing, that's all for the house. For the Land, there are only three features, the barangay, land area (sq/m), and classification, whether it is commercial or residential.

Web Tier

In this system architecture, the web tier allows the webserver to communicate between the users' web browsers to the logic tier.

Logic Tier

The web application handles the commands and the logic of the data decision, the data analysis and evaluation of the data from the fields from the presentation tier to the machine learning model, and the data management.

Data Tier

In the data tier, all data provided by users in the presentation tier will be collected and kept; it stores all necessary data to be used in the system. Users' data are collected, including their first and last names, email addresses, phone numbers, profile images, and passwords, as well as the sort of operating system they used to log in. The title, content of the property, price of the property, date of posting, and three photos of the property are all included in the data for posting the property.

Implementation Plan

This section will discuss the implementation plan of the system for the client or target users. The target users are buyers, sellers, real estate agents, and future researchers. To gain an understanding of implementation strategy first is to determine the strategic idea. A strategic idea is a process of determining the strategy by which the researchers achieve a specific goal and made decisions. The organizations create strategic plans to take the organizational position, one specific sector's campaign, or any program or enterprise.

For the client to know the system, it will be advertised in a commercial advertisement. Commercial ads often seek to generate increased consumption of the developed system or the services through branding, which associates the system's name and logo with certain qualities in the minds of consumers. On the other hand, ads that intend to elicit an immediate sale are known as direct response advertising.

Include software, hardware, network, and database requirements

- **Hardware Requirements**

  The following are the system requirements that are needed by the client. They are required to have internet connectivity to access it in the browser. The following are the standard minimum browser requirements:

  - CPU: Pentium 4 and above
  - Memory: 512 MB
  - Storage: 150 MB

The system requirements used for development:

- CPU: Intel(R) Core or Newer
- Memory: 8GB
- Storage: 136GB

- **Software Requirements**

	Desktop	Mobile
Operating System	Windows 7 and above, Mac OS X 10.2	Android 5 and above, iOS or later
Browser	Chrome, Firefox, Microsoft Edge, Opera mini, Brave, UC Min. Disk Space: 150 MB Rec. Disk Space: 70 MB Rec. Disk Space 64-bit: 120 MB	Chrome, Safari, Opera mini, Firefox, UC, Brave

Table 3 Software Requirements

The packages used for the development and deployment of the system by the command of *pipreqs* includes the following:

- Flask==2.0.1
- Flask_Login==0.5.0
- Flask_SQLAlchemy==2.5.1
- numpy==1.21.3
- pytz==2021.3
- SQLAlchemy==1.4.23
- Werkzeug==2.0.1

The system software requirements used for development:

- **Network Requirements**

Internet Connection

A stable, and high-speed internet connection is required for accessing the web-application. The response time for each assessment depends on the reliability and speed of the network.

Bandwidth

Network Min. Bandwidth: 500 kbps

The part of the Network architecture:

- Client
- Internet
- Router
- Web Server
- Internet Service Provider (ISP)
- Domain Name System (DNS)

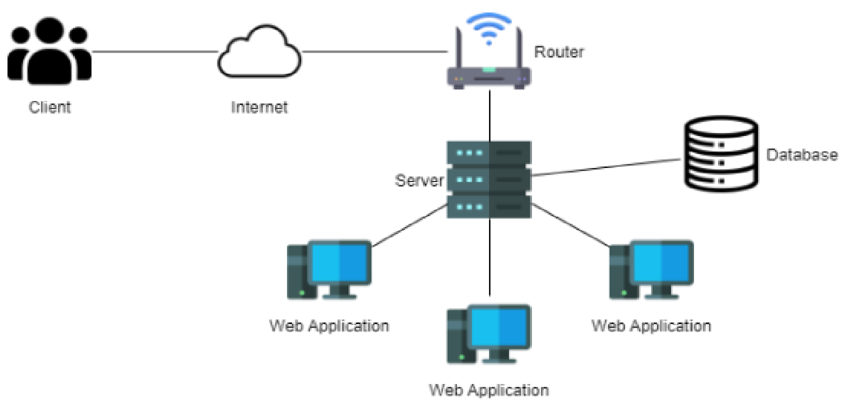


Figure 7 Network Architecture

- **Database Requirements**

A relational database was used for this project. The relational database management system (RDMS) is a program that allows users to create, update, and administer a relational database. It also allows users to store numerous data, related tables, and data in rows and columns inside the tables. The researchersneed to design a database schema for the House and Lot price prediction system based on the following.

- **USERS** The system allows the users to fill up a data form that will serve as their account information, which contains the first name, last name, email address, contact number, password, and image and its name and mime type for the User class.
- **POST** The system allows the user to fill out the posting form, which contains the following: title, content, date posted, the requisite three photos, and the estimated price, after completing the relevant data for the house and lot forms.
- **LOGIN HISTORY** The system collects minor device data for the Login History class, such as the type of operating system used by the user device and the DateTime, when the user logged in. 28 The system displayed the class of Interest as a wish list, allowing users, buyers, and even sellers to add their interest properties to the wish list.
- The goal is to add interest without an immediate intent to buy. In the Entity-Relationship Diagram (ERD) process, the class of User can have zero or more Posts, and Posts can have one-to-one relationships with the class of User. The same logic for the class of Interest, where a user can have zero or more interests, and the class of User to Login History, where a user can have zero or more login records.
- Database type: Relational Database, SQL Server

CHAPTER IV
RESULTS AND DISCUSSION

Introduction

This chapter presents the results and discussion gathered from the collected data, data analysis, and the results of the three-machine learning algorithm, the linear regression, ridge, and the gradient boosting algorithm, it will also present the process of cleaning the data, the formula used when there is a null, empty fields, the data processing after cleaning, evaluation of the model, and the confusion matrix.

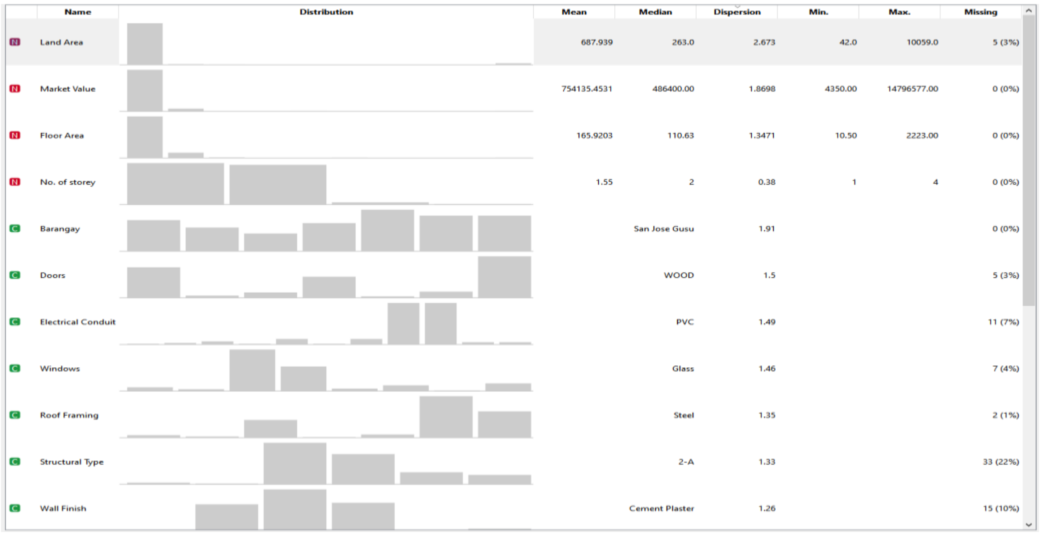
Initial Data Analysis

The researchers promptly encode the hard copy of the request data to excel after receiving it from the City Assessor's Office. Due to the pandemic, transactions and offices, including the Assessor's Office in Zamboanga City, were closed and has a limited employee to process the request and resulting to have a limited dataset acquired.

The real-property field appraisal sheet for Building or House contains 145 instances and 18 features, while the real-property field evaluation sheet for Land contains 158 instances and 3 features. The researchers found 106 samples that were suitable for the house and 114 samples that were suitable for the land. Using Slovin's formula, the study calculated the appropriate amount depending on the desired confidence level of 95%. After calculating, the analysis requires 106 samples for the home and 114 samples for the land using Slovin's formula. All eighteen (18) feature of the house is related to the house price impact, whereas three (3) features of the land are related to the land price impact.

After receiving the housing dataset, it was discovered that there is a possibility that the limited number of datasets that the researchers have can affect the gap in the predicted price. It was considered as errors in the datasets such as abbreviation or shortcut value of each feature that resulted to some confusion in analysis and missing values in each feature. The following are the statistics of the house dataset to see the missing values in each feature.

Real-Estate Property Field Appraisal Sheet – Building | House



	Name	Distribution	Mean	Median	Dispersion	Min.	Max.	Missing
N	Land Area		687.939	263.0	2.673	42.0	10059.0	5 (3%)
N	Market Value		754135.4531	486400.00	1.8698	4350.00	14796577.00	0 (0%)
N	Floor Area		165.9203	110.63	1.3471	10.50	2223.00	0 (0%)
N	No. of storey		1.55	2	0.38	1	4	0 (0%)
C	Barangay			San Jose Gusu	1.91			0 (0%)
C	Doors			WOOD	1.5			5 (3%)
C	Electrical Conduit			PVC	1.49			11 (7%)
C	Windows			Glass	1.46			7 (4%)
C	Roof Framing			Steel	1.35			2 (1%)
C	Structural Type			2-A	1.33			33 (22%)
C	Wall Finish			Cement Plaster	1.26			15 (10%)

Figure 8 Feature Statistics (Checking for null or missing values)

The first step in the data analysis is to see if there are any missing values if the data distribution is unbalanced. Finding the median of each feature is the way to replace the missing values. This is where to get the missing values from the dataset. As shown in Fig. 11 Barangay Distribution, the means are calculated, and the distribution of graphs is displayed. It was decided to drop these features to avoid low accuracy in our prediction because the Structural type has the most missing values in the dataset.

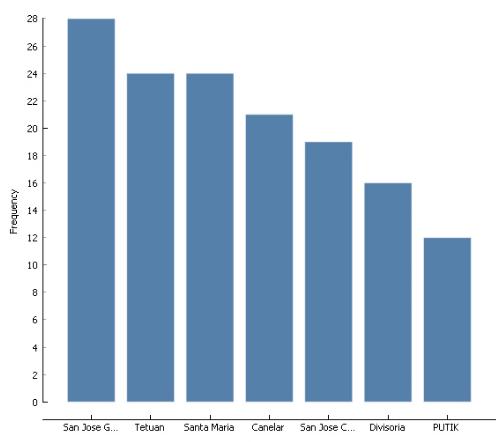


Figure 9 Distribution of Barangay

After determining the missing values, the researchers look for the distribution of each feature, as shown in figure 12. The San Jose Gusu has the most data samples in the distribution visualization for the feature of barangay, and barangay Putik is the least. In this scenario, it can say that there will be bias in predicting the house price because the distribution is unbalanced.

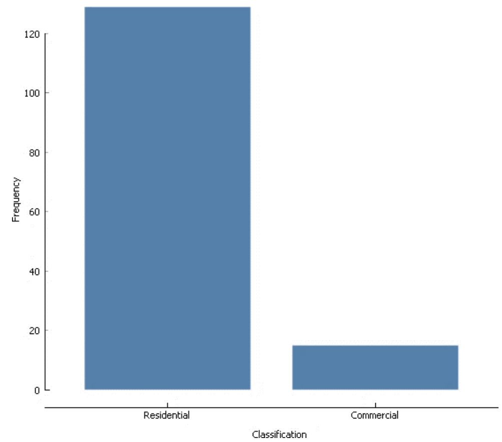


Figure 10 Distribution of Classification

In this feature of Classification, the distribution of sample data is imbalanced because there are 129 samples for residential, which estimates for 89.58 percent of all the data collected, and 15 samples for commercial, which accounts for only 10.42 percent of the total data, the housing price estimate could be erroneous. Since the data cannot be dropped or removed to have an equal distribution so, to achieve the 80% accuracy rate of the model, the researchers used a different algorithm that fits the number of data has.

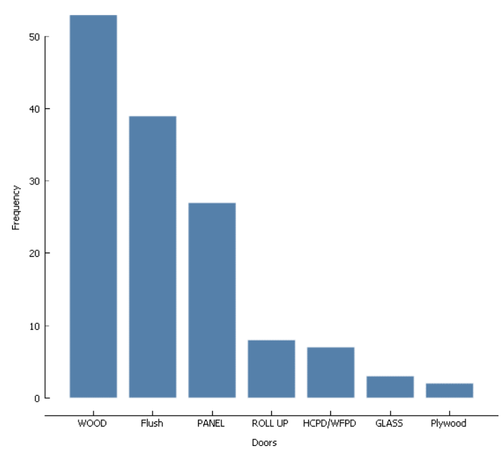


Figure 11 Distribution for Doors

As shown in figure 11 above, wood is the most common with 53 (38.13 percent) data samples, followed by a flush with 39 (28.06 percent), panel door with 27 (19.42 percent), roll with 8 (5.76 percent), HCPD or Customs Panel Doors with 7 (5.04 percent), glass with only 3 (2.16 percent), and plywood with 2 (1.44 percent). Because of the unbalanced data for the last four data occurrences, the prediction in this feature may be biased.

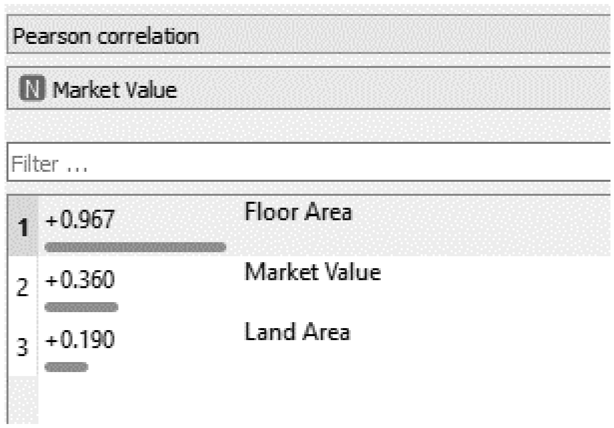


Figure 12 Pearson Correlation

It can observe from this correlation that the Floor area is the most significant or relevant feature in predicting market value.

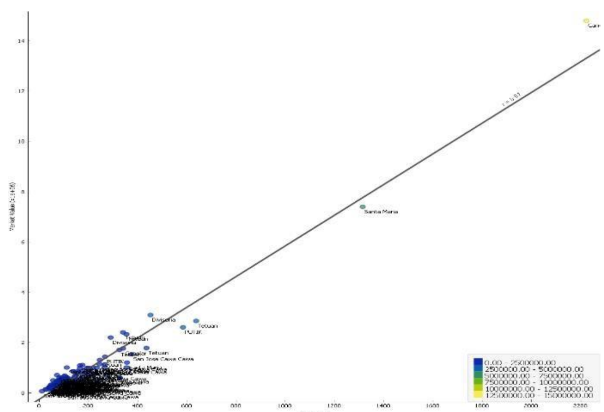


Figure 13 Scatter Plot (Floor Area)

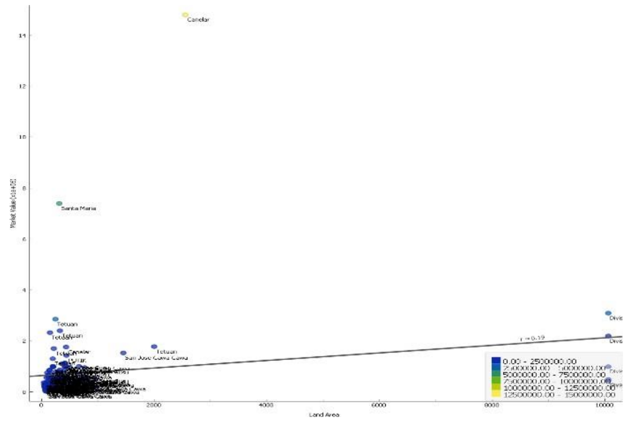
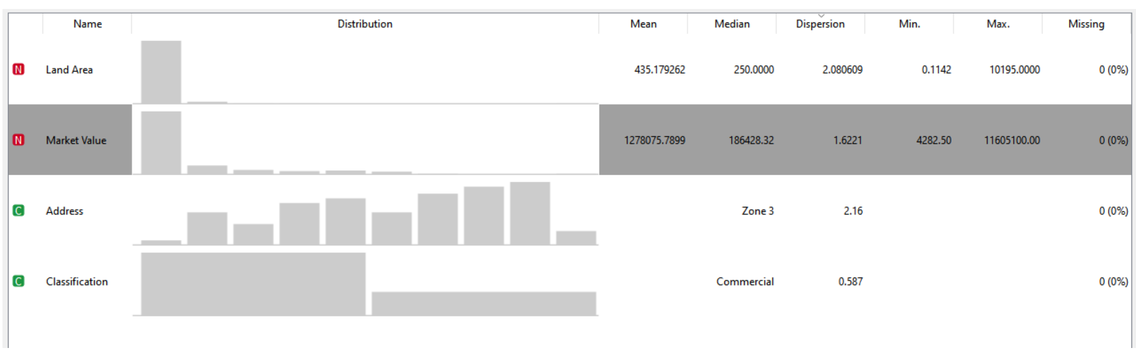


Figure 14 Scatter Plot (Land Area)

As shown in figure 15, the floor area and the land area is the most relevant feature in predicting the market value of a house property. It shows a positive correlation in predicting the data model.

Real-Estate Property Field Appraisal Sheet – Land



	Name	Distribution	Mean	Median	Dispersion	Min.	Max.	Missing
N	Land Area		435.179262	250.0000	2.080609	0.1142	10195.0000	0 (0%)
N	Market Value		1278075.7899	186428.32	1.6221	4282.50	11605100.00	0 (0%)
C	Address			Zone 3	2.16			0 (0%)
C	Classification			Commercial	0.587			0 (0%)

Figure 15 Feature Statistics (Checking for null or missing values)

The model was split into two parts: one for the house pricing model and another for the land model. Figure 17 shows the data's feature statistics for the land dataset, including the mean, median, dispersion, min, max, and missing values. Since there are no missing values in the land dataset, it was skipped on that part using the median to fill the gaps.

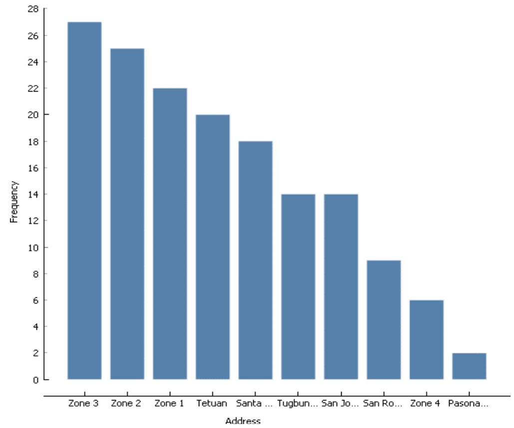


Figure 16 Scatter Plot (Barangay)

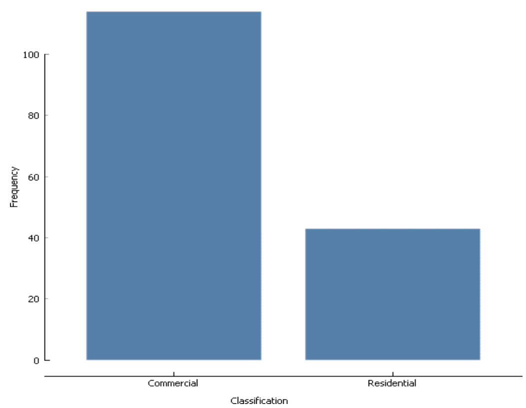


Figure 17 *Scatter Plot (Classification)*

In data analysis for the distribution of data, the barangay, and the classification has unbalanced the contribution of data. In Figure 18, the Distribution of Barangay shows that the data is unbalanced. Zone 3 has 27 samples, equivalent to 17.20% of the total data, Zone 2 got 25 samples (15.92%), Zone 1 has 22 samples (14.01%), Tetuan has 20 samples (12.74%), Santa Maria has 18 samples (11.46%), Tugbungan has 14 samples (8.92%) same with San Jose Gusu, San Roque with 9 samples (5.73%), Zone 4 with 6 samples and only 2 (1.27%) samples for Pasonanca. For the Classification, the commercial has 114 samples, 72.61% of the total data, and 27.39% for Residential with 43 samples.

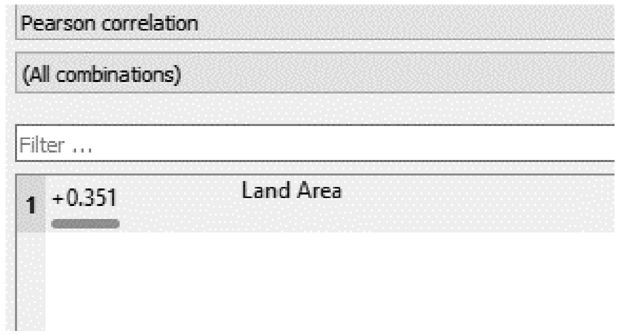


Figure 18 Pearson Correlation

Since the Pearson Correlation was used to determine what is the most relevant feature in predicting the Land price, it resulted in only one, which is the Land Area.

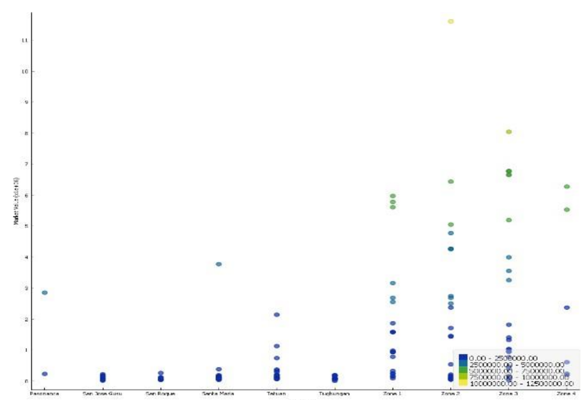


Figure 19 Scatter Plot-Categorial (Barangay)

Figure 20 Scatter Plot-Categorial (Classification)

In the figures above, the scatter plot for the categorical feature, the barangay, and classification reveals that Zone 3 has the highest market price, as indicated by yellow marks. The commercial data has the highest market worth, which is between 10,000,000 and 12,500,000.00 Philippine pesos, according to the data.

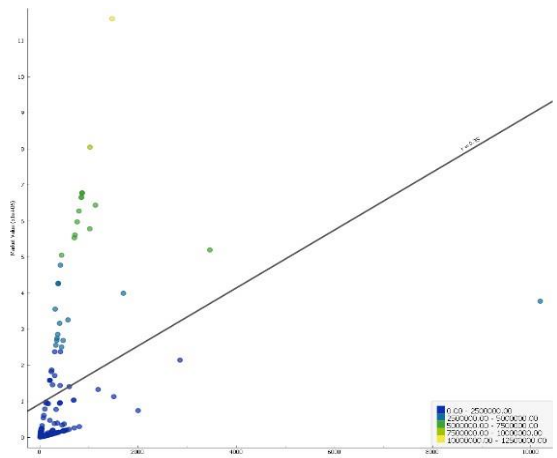


Figure 21 Scatter Plot (Land Area)

As shown in Figure 21, the land area is the most relevant feature in predicting the market value of land property. It shows a positive correlation in predicting the data model.

Initial Strategy, Regression

Figure 22 Means and Median

The initial strategy that has used to find the prediction of houses and lots is the use of a linear regression algorithm. The researchers had faced difficulty in using it because the accuracy was so low that they could not even reach half of the target accuracy rate. The outliers are too far from the trendline. Since there are many missing values, the highest missing value, which is the *Structural Type* and *Stairs was dropped*. After removing these, the strings or categorical data were converted into numerical data so they could be readable by the model. For other features, Median Statistics was used to get the missing values. See Figure 24. The median was used to get the most common value, and that value was used as anonymous in filling in those missing values. The researchers did not use the Mean statistics because there is a tendency that a whole number would be decimal, and if that happens, it will take time to convert it again to a whole number.

After converting all the categorical data and filling it up into the missing values with median statistics, next is to proceed to data processing.

For data processing, the following algorithms were used;

- **Linear Regression**

- **Ridge Regression**
- **Gradient Boosting**

Initial Strategy, Results

The results for the initial strategy were inconclusive at the least. The result for the processing has reached the 80% of accuracy for all the regression algorithms. The model scores are shown in Table 4.

- For regression model must be considered good, the following assumption must be met:
- The coefficient of determination or the R-Square(R^2) score of the model has reach the 0.8 or 80% and above for gradient boosting.
- The Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) are very low in all regression algorithm that researchers used, that will be considered as poor or not reliable metric in comparing the performance of the models.
- Also, the Mean Absolute Error (MAE) is not reliable metric because it has a low value.

Algorithms	MAE	MSE	RMSE	R^2
Linear Regression	466568.205	1382321751940.308	683.05	0.83
Ridge Regression	169393.23	36759305696.917	411.574	0.80
Gradient Boosting	120745.617	25211587219.354637	347.484	0.93

Table 4 Model Scores

The following metrics indicate that the model developed seems to be of poor quality. Despite of intensive testing and data processing. But the R^2 Model scores indicate that the gradient boosting model performed significantly better than the linear and ridge regression.

Model	Parameter
Ridge Regression	Alpha=0.5 Fit Intercept=True
Gradient Boosting	No. of estimators = 100 Max Depth = 3 Min Samples Split = 2 Learning Rate = 0.3

Table 5 Model Parameters

Table 5 shows the model parameters that were determined to achieve the accuracy of the model score for the algorithms.

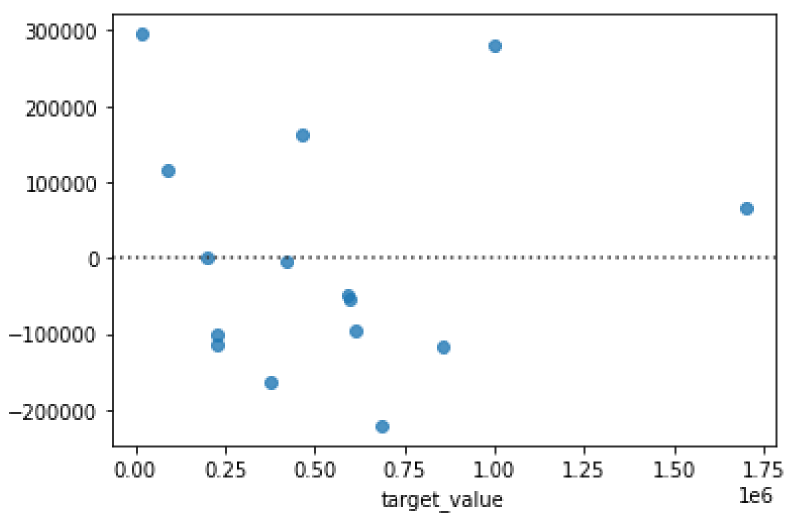


Figure 23 Residuals (Actual vs. Predicted prices)

Figure 23 shows the residuals plot that the errors follow in a distribution around the 0 value of the x-axis, and it does not form exactly in the specific near 0 point.

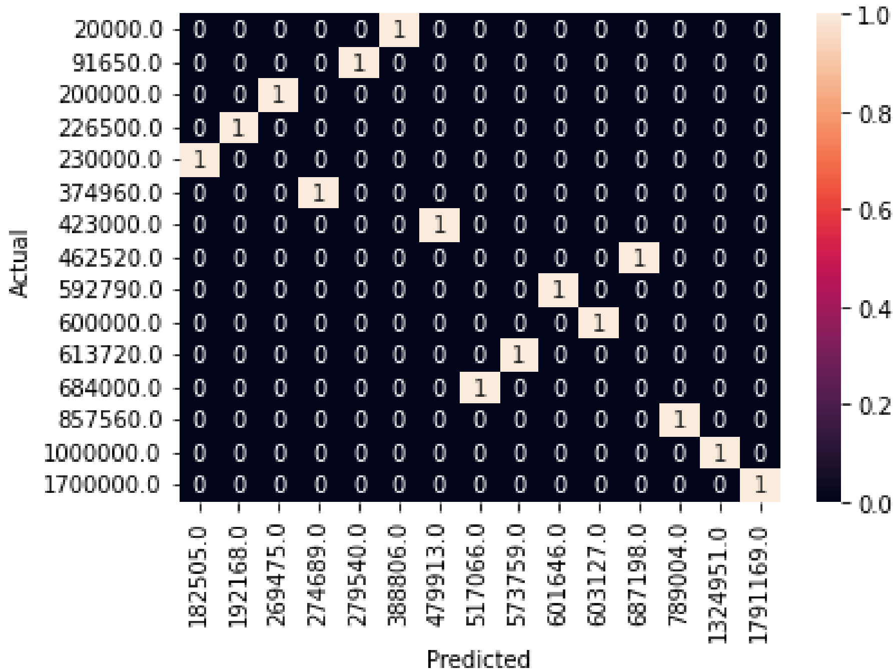


Figure 24 Gradient Boosting Confusion Matrix

Figure 24 shows the confusion matrix of the Gradient Boosting model. It is shown on the presentation that the actual price is on the y-axis and the predicted price on the x-axis. The 1 represents True, and the 0 represents False. This shows the interception of the x-axis (Actual price) and the y-axis (Predicted price) in the datapoint. The results are inconclusive; however, the accuracy rate reaches the target rate of 80%.

After visualizing and analyzing the errors of the different algorithms that have been used, It was reviewed, and the results on the use of the R^2 metrics of Gradient Boosting in developing the model have the highest accuracy among the other algorithms.

Gradient Boosting Implementation and Model Evaluation

This section describes all regression algorithms used and implemented in Jupyter notebook integrated development environment for the data cleaning, data processing, evaluation, and deployment of the model to the actual presentation system. The researchers used Orange Data Mining – a software application to visualize graphs, distribution of data, scatter plots, and confusion matrix.

Among the algorithms that have been used, the Gradient Boosting was the best choice in evaluating the model since its accuracy is consistent, and it has reached the target of 80% accuracy score. It also shows that the model that have developed has a positive result in predicting house prices.

For the deployment of the model into the system, the package of Pickle file was used. After creating the pickle file, the model pickle file was imported to the function form. The System already provided the form for the user to input the information of House and Lot, and after clicking the submit button, it will bring the form to the functions containing all the data needed to fit in the model. After fitting the data, now assign the variable prediction = model.predict(data[x])[0]. The model will predict and will pass the result or the predicted value to variable prediction, and it will submit it back to the System. Overall, the whole system successfully predicts the price of a house and lot despite the adjustments and errors that have been faced in developing the system.

Possible Reason of the Results

Although the scores of the house and land accuracy show an acceptable result of the study, there were still underlying problems that could have affected the results. Due to the pandemic, the approach or the way the researchers wanted to gather the data needed was limited. The movement of the transactions, the number of people, and the offices available were limited. Some of the data are confidential, which may also cause the lack, and enough data have for the model to work.

CHAPTER V
CONCLUSION AND RECOMMENDATIONS

Conclusion

The web-based price prediction system with the use of Gradient Boosting has shown acceptable scores and was proven to be effective in predicting the price of a house and lot property. Due to the pandemic, the collected data were limited. Thus, the created model was fed with a limited amount of data. Although the results showed a positive effect on the independent variable and the dependent variable, due to the lack and a limited number of data needed, the results can be further improved and can give a higher accuracy rate. As the researchers have performed the prediction, there are some changes made during the methodology stage. The first algorithm that is used is the Linear Regression algorithm. Since the data gathered is limited and due to many missing values, it used other algorithms to find out which gives an acceptable accuracy rate despite the lack of data. The Ridge Regression and Gradient Boosting were used as an alternative algorithm, and the result shows that Gradient Boosting gives the accepted accuracy rate of 80% despite the lack of data. Since the main purpose of this is to predict the price of a property at 80%, it was decided to use Gradient Boosting as the main algorithm in the prediction model. Upon using this algorithm, the Floor Area and the Lot Area are factors to have a high impact on the price of a property. These findings may be differed from other related studies because of some factors affecting the result such as limited data, but this can also prove that the floor area and lot area of a house and land property are considered to be a great factor in pricing a property.

In summary, although the unexpected pandemic caused alterations and adjustments to the data gathering procedure, the prediction system was still able to show acceptable results despite some problems encountered during the process. This study's outcome could be enhanced further, resulting in greater and better performance in terms of accuracy.

Recommendations

The prediction system may have reached the minimum requirements and accomplished the objectives of the study, but further improvements can help improve the effectiveness of the prediction system.

It is recommended that future researchers consider the following:

1. Adding more data for both house and land datasets.
2. To survey a larger population for better improvements to the system.
3. Use other algorithms to compare the prediction model using the same dataset you have gathered.
4. Expand the scope of the study and have a wide range of locations.
5. Try to consider increasing or decreasing the number of datasets utilized in each method or algorithm to see if the result is the same as expected.

References

[1] Abdulal, A., & Aghi, N. (2020). *House Price Prediction*. diva-portal.org. Retrieved April 7, 2022, from https://diva-portal.org/smash/get/diva2:1456610/FULLTEXT01.pdf

[2] Ashfaque, J. M. (2020, July). *What is Machine Learning?* researchgate.net. https://www.researchgate.net/publication/343231033_What_is_Machine_Learning.pdf

[3] Avanija, J., Sunitha, G., Madhavi, K. R., Kora, P., & Sai Vittal, R. H. (2021, April 5). *Prediction of house price using XGBoost regression algorithm*. Researchgate. Retrieved March 26, 2022, from https://www.researchgate.net/profile/Padmavathi-Kora-2/publication/350810698_Prediction_of_House_Price_Using_XGBoost_Regression_Algorithm/links/6074173b92851c8a7bbef01b/Prediction-of-House-Price-Using-XGBoost-Regression-Algorithm.pdf

[4] B.Vijay Kumar, B.Ashritha, Ch.Teja, M.Vineeth, "HOUSE PRICE PREDICTION USING GRADIENT BOOST REGRESSION MODEL", IJRAR - International Journal of Research and Analytical Reviews (IJRAR), E-ISSN 2348-1269, P- ISSN 2349-5138, Volume.7, Issue 1, Page No pp.52-55, March 2020, Available at : http://www.ijrar.org/IJRAR2001569.pdf

[5] Baldominos, A., Blanco, I., Moreno, A., Iturrarte, R., Bernárdez, Ó., & Afonso, C. (2018). Identifying Real Estate Opportunities Using Machine Learning. *Applied Sciences*, *8*(11), 2321. https://doi.org/10.3390/app8112321

[6] Dann, C., Grohmann, J., Ho, A., Khoo, S., Kim, A., & Lal, S. (2020). The Hedonic House Price Model in a Development Context – Evidence from the National Capital Region of the Philippines, 2000 - 2010. *Rationale*, *1*, 23–58.

[7] Gaurav, & Aalam, Z. (2018). *Prediction of House Price Using Machine Learning Approach*, *5*(11, 2018), 67–69. http://www.ijsrd.com/articles/IJSRDV5I110013.pdf

[8] Global Property. (2020). *Philippines' housing market is now very weak amidst an ailing economy*. Philippines' housing market is now very weak amidst an ailing economy. Retrieved April 8, 2022, from https://www.globalpropertyguide.com/news-philippines-housing-market-is-now-very-weak-amidst-an-ailing-economy-4247

[9] Ho, W. K. O., Wong, S. W., & Tang, B.-S. (2020, October). *Predicting property prices with machine learning algorithms*. Retrieved March 26, 2022, from https://www.researchgate.net/publication/346308101_Predicting_property_prices_with_machine_learning_algorithms

[10] Imran, Zaman, U., Waqar, M., & Zaman, A. (2021). Using Machine Learning Algorithms for Housing Price Prediction: The Case of Islamabad Housing Data. koreascience.or.kr. Retrieved April 7, 2022, from https://www.koreascience.or.kr/article/JAKO202122260767075.pdf

[11] Kokasih, M. F., & Para, A. S. (2020, September 30). *Property Rental Price Prediction Using the Extreme Gradient Boosting Algorithm*. ijiis.org. Retrieved April 7, 2022, from http://ijiis.org/index.php/IJIIS/article/download/65/27

[12] Monika, R., Nithyasree, J., Valarmathi, V., Hemalakshmi, G. R., & Prakash, N. B. (2021, May 10). *House price forecasting using machine learning methods*. Turkish Journal of Computer and Mathematics Education (TURCOMAT). Retrieved April 8, 2022, from https://turcomat.org/index.php/turkbilmat/article/view/6435

[13] Revend, W. (2020, September 8). *Predicting House prices on the countryside using boosted decision trees*. DIVA. Retrieved April 1, 2022, from https://kth.diva-portal.org/smash/record.jsf?pid=diva2%3A1464970

[14] Sahithi, B., Saheli, T., Ramanika, D., & Anil Kumar, N. (2020). *Crop Price Prediction System using Machine learning Algorithms*. www.questjournals.org. Retrieved April 7, 2022, from http://questjournals.org/jses/papers/Vol6-issue-1/B06011420.pdf

[15] Singh, S., & Nag K J, M. (2020, September 30). *Land price prediction using machine learning algorithm*. IRJET. Retrieved April 8, 2022, from https://www.academia.edu/44199114/IRJET_Land_Price_Prediction_using_Machine_Learning_Algorithm Retrieved April 8, 2022, from https://www.academia.edu/44199114/IRJET_Land_Price_Prediction_using_Machine_Learning_Algorithm

[16] Sohlberg, M. (2021, March 5). *Philippines Real Estate Market Outlook in 2021: A Complete Overview*. Asia Property HQ. https://www.asiapropertyhq.com/philippines-real-estate-market/

[17] Uzut, G., & Buyrukoglu, S. (2020, May). *PREDICTION OF REAL ESTATE PRICES WITH DATA MINING ALGORITHMS*. researchgate.net. Retrieved April 6, 2022, from https://www.researchgate.net/publication/342782159_Prediction_of_real_estate_prices_with_data_mining_algorithms

[18] Ziweritin, S., Ukegbu, C., Taiwo Adisa Oyeniran, T. A., & Ulu, I. O. (2021, February). *A Recommendation Engine to Estimate Housing Values in Real Estate Property Market*. researchgate.net. Retrieved April 6, 2022, from

https://www.researchgate.net/publication/349668577_A_Recommendation_Engine_to_Estimate_Housing_Values_in_Real_Estate_Property_Market

Appendix A

Gantt Chart

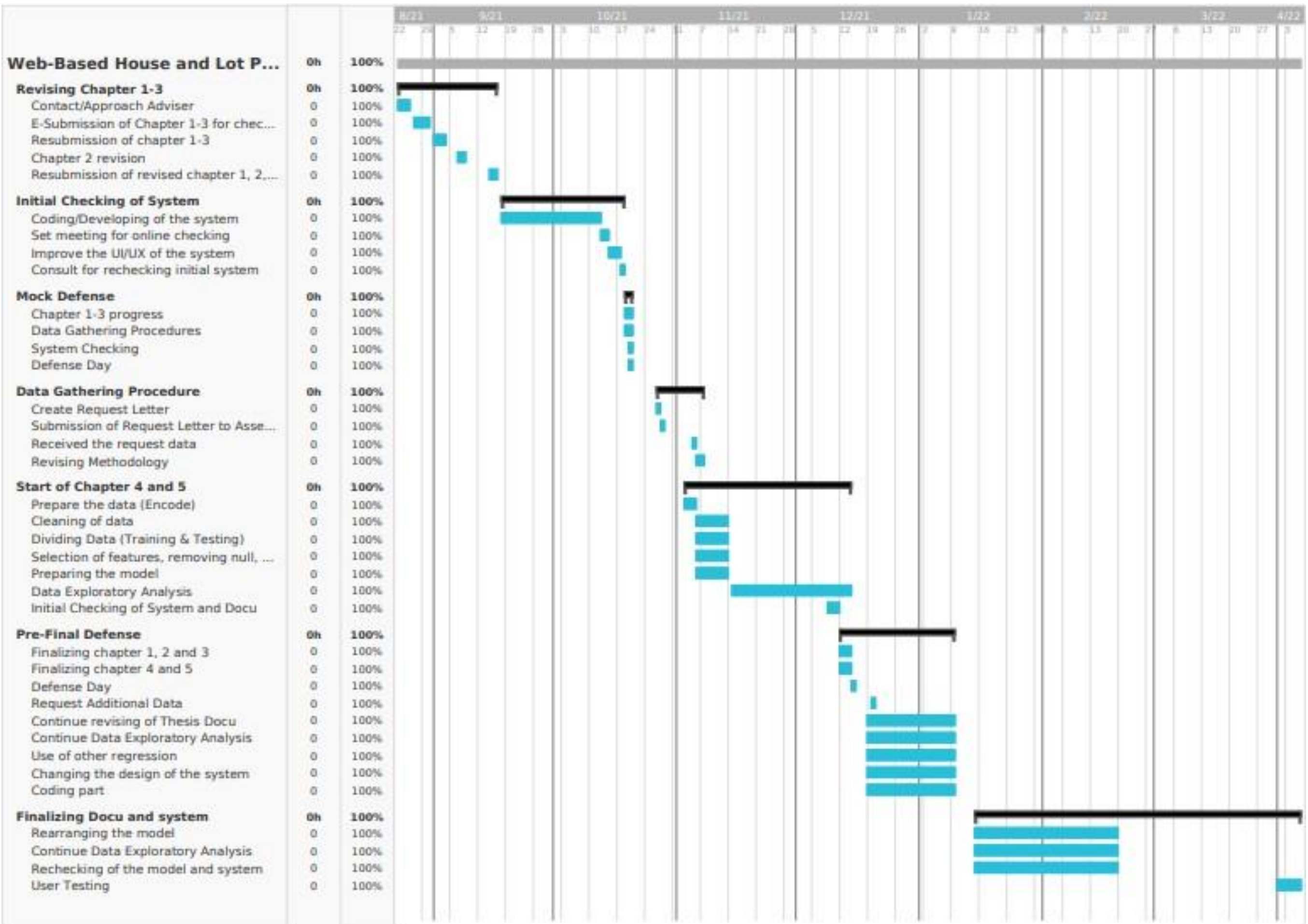


Figure 25 Gantt Chart

Appendix B

Evaluation Tool

Gradient Boosting

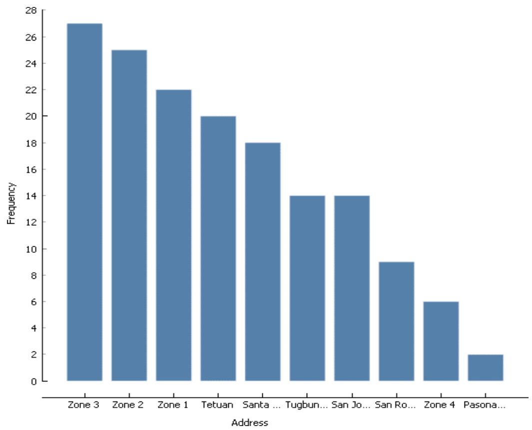


Fig. 16 Distribution (Barangay)

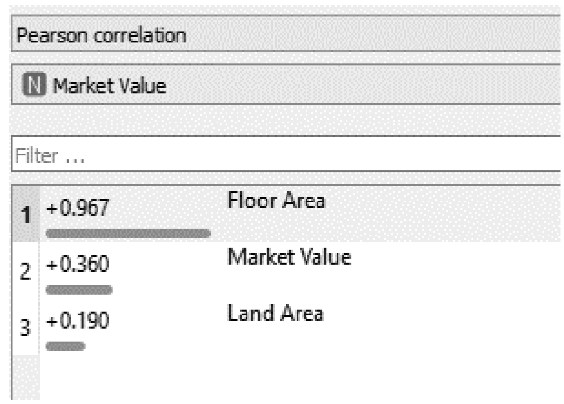


Fig. 17 Pearson Correlation

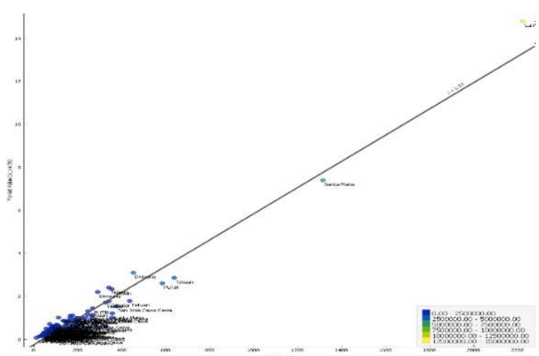


Fig. 16 Scatter Plot

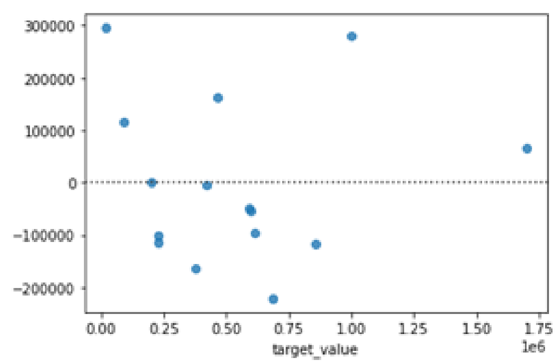


Fig. 17 Residuals (Actual vs. Predicted prices

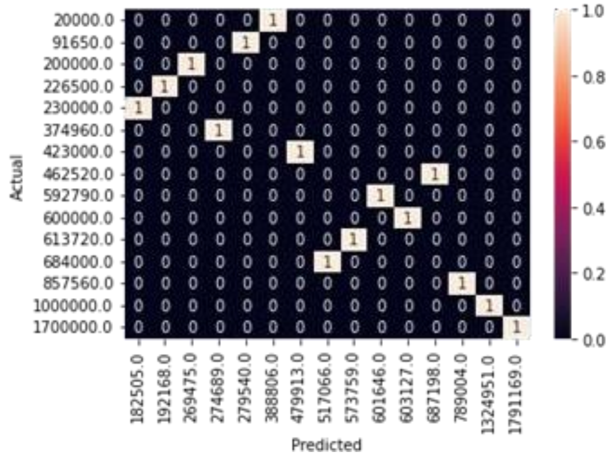


Fig. 18 Confusion Matrix

Appendix C

Photo Documentation of Testing

Photo Documentation of Testing

1	Property	Barangay	Land Area	Floor Area	Classifica	No. of sto	Structural	Roof Fran	Roofing	Exterior W	Flooring	Doors	Ceiling	Windows	Stairs	Wall Fini	Electrical	Toilet & B	Plumbing	Market Val
2	Property 1	Canelar	665	354.08	Residenti	2	2-A	Steel	GI	Cement P	Tiles	WOOD	Plywood	Glass	CONC.	Cement P	PVC	Tiles	PVC	1000000
3	Property 2	Canelar	104	71.25	Residenti	1	3-A	Steel	Long Spar	CHB	Tiles	WOOD	Plywood	Alum. Cas	CONC.	Plain Cem	Polyflex	Tiles	PVC	413000
4	Property 3	Canelar	406	210.3	Residenti	1		Steel	GI	CHB	Tiles	WOOD	Plywood	Glass	CONC.	Cement P	PVC	Tiles	PVC	727910
5	Property 4	Canelar	240	81	Residenti	1	2-A	Steel	Long Spar	CHB	Tiles	WOOD	Plywood	Alum. Cas	CONC.	Cement P	Polyflex	Tiles	PVC	610000
6	Property 5	Santa Maria	314	1314	Commerci	4	2-A	Steel	GI	CHB	Tiles	Flush		Alum. Cas	CONC.	Cement P	PVC	Tiles	GI	7393688
7	Property 6	Santa Maria	300	98.83	Residenti	2	2-A	Steel	Long Spar	CHB	Tiles	WOOD	Plywood	Glass		CHB	PVC	Tiles	PVC	560650
8	Property 7	Santa Maria	206	111.26	Residenti	2		Steel	GI	CHB	Tiles	WOOD	Wood	Glass	CONC.	Cement P	PVC	Tiles	PVC	1000000
9	Property 8	Santa Maria	300	81	Residenti	2	2-A	Wooden	GI	CHB	Tiles	PANEL	Plywood	Jalousie		Plain Cem	Conduit P	Tiles	PVC	400000
10	Property 9	Santa Maria	237	160	Residenti	1	2-B	Wooden	GI	CHB	Concrete	ROLL UP	Plywood	Jalousie	WOODEN	CHB	Polyflex	CHB	PVC	440320
11	Property 1	Santa Maria	300	106.4	Residenti	2	3-A	Wooden	GI	Cement P	Concrete	Flush	Plywood	Wood	WOODEN	CHB	Open	CHB	PVC	505400
12	Property 1	Santa Maria	300	25.6	Residenti	2	3-A	Wooden	GI	Cement P	Concrete	Flush	Plywood	Wood	WOODEN	CHB	Open	CHB	PVC	132860
13	Property 1	Santa Maria	300	189.75	Residenti	2	3-A	Wooden	GI	Cement P	Tiles	WOOD	Plywood	Glass	WOODEN	CHB	PVC	Tiles	PVC	788630
14	Property 1	Santa Maria	137	40	Residenti	1	2-B	Steel	GI	CHB	Concrete	WOOD	Plywood	Glass		Cement P	Polyflex	Tiles	PVC	230000
15	Property 1	Santa Maria	688	97.7	Residenti	1	3-A	Wooden	GI	CHB	Concrete	WOOD	Plywood	Glass	WOODEN	CHB	PVC	Tiles	PVC	423000
16	Property 1	Santa Maria	851	98	Residenti	1	2-B	Lumber	GI	CHB	Tiles	Flush	Plywood	Jalousie		Plain Cem	Polyflex	Tiles	PVC	400000
17	Property 1	Santa Maria	138	99.12	Residenti	2	2-A	Wooden	Long Spar	Cement P	Tiles	WOOD	Plywood	Glass	RC	Cement P	PVC	Tiles	PVC	638080
18	Property 1	Santa Maria	150	48.8	Residenti	1	2-A	Steel	GI	CHB	Tiles	WOOD	Wood	Glass		Cement P	PVC	Tiles	PVC	320616
19	Property 1	Santa Maria		164.63	Residenti	2	2-A	Steel	Long Spar	Cement P	Tiles	PANEL	Plywood	Glass	RC	Cement P	PVC	Tiles	PVC	1081620
20	Property 1	Santa Maria	322	79	Residenti	2	2-A	Wooden	GI	CHB	Concrete	PANEL	Plywood	Jalousie	WOODEN	Cement P	PVC	Tiles	GI	300000
21	Property 2	Santa Maria	600	110	Residenti	2	3-B	Wooden	GI	Wood	Concrete	WOOD	Wood	Wood	WOODEN	Plain Cem	PVC	CHB		528000
22	Property 2	Santa Maria	800	140	Residenti	1	2-B	RC	Long Spar	CHB	Tiles	Plywood	Plywood	Jalousie		Plain Cem	PVC	Tiles	PVC	534896.3
23	Property 2	Santa Maria	200	63	Residenti	1	3-A	Wooden	GI	Cement P	Concrete	WOOD	Wood	Jalousie		Cement P	PVC	Tiles		200000
24	Property 2	Santa Maria	500	164	Residenti	1	2-B	Steel	Long Spar	CHB	Tiles	WOOD	Plywood	Glass		CHB	PVC	Tiles	PVC	970000
25	Property 2	Santa Maria	150	71.5	Residenti	1	2-A	Steel	GI	CHB	Tiles	PANEL	Plywood	Glass		CHB	Polyflex	Tiles	PVC	700000

Figure 26 Data samples (Testing)

< Back House & Lot

Real Property Field Appraisal & Assessment Sheet - House | Building

Make your property price get predict by filling up the data.

Note: Choices are fixed.

Note: The price prediction may not work, if you've chose [others], No prediciton available.

Land Reference

Barangay :
Santa Maria

Land Area (sq/m)
688

Building Classificatio

Property Used :
Residential

y :

Structural Characterisıic

Floor Area (sq/m)
97.7

Roof Framing :
Wooden

Roofing :
Galvanized Iron

Exterior Walls :
Concrete Hallow Block (CHB)

Flooring :
Concrete

Doors :
Wood

Ceiling :
Plywood

Windows :
Glass

Wall Finish :
Concrete Hallow Block (CHB)

Electrical Con. :
Polyvinyl Chloride (PVC)

Toilet & Bath :
Tiles (Ceramic,etc)

Plumbing/Sewer :
Polyvinyl Chloride (PVC)

-- Data is ready for prediction --

Continue

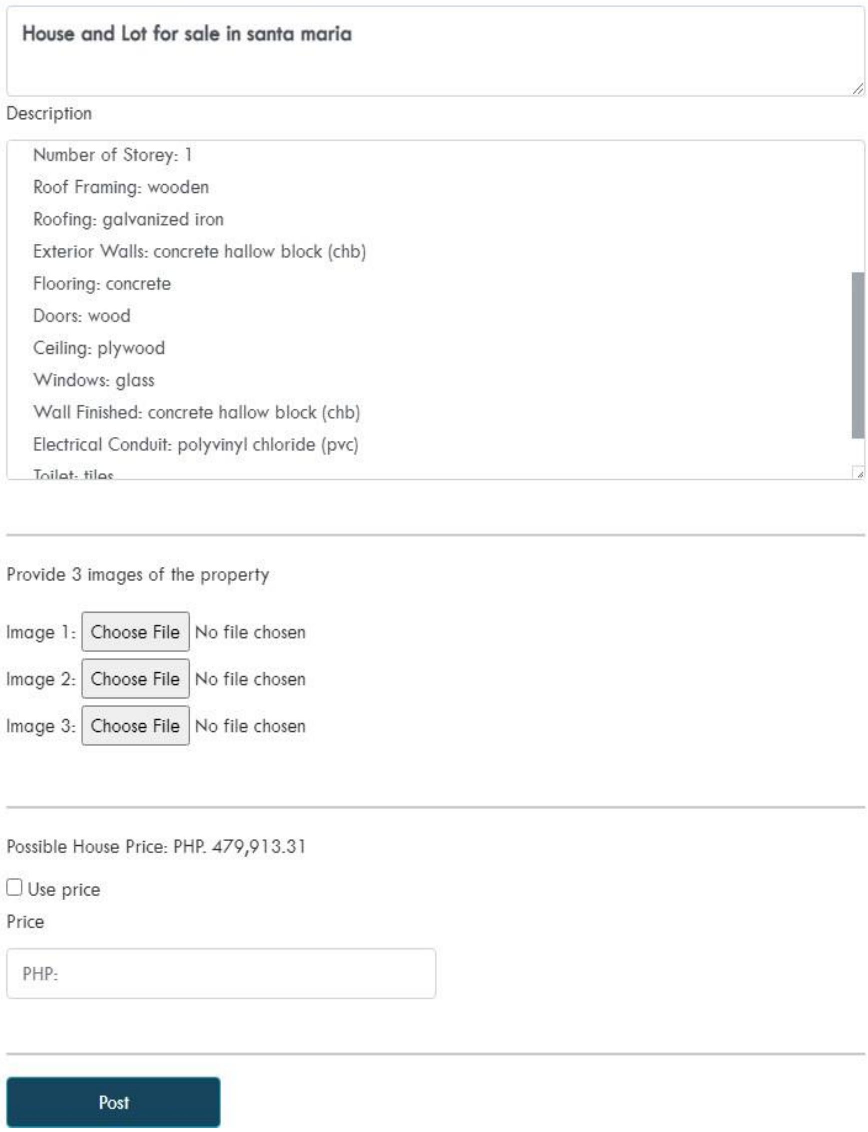


Figure 27 Screenshot of system Testing

Appendix D

Algorithms

Regression

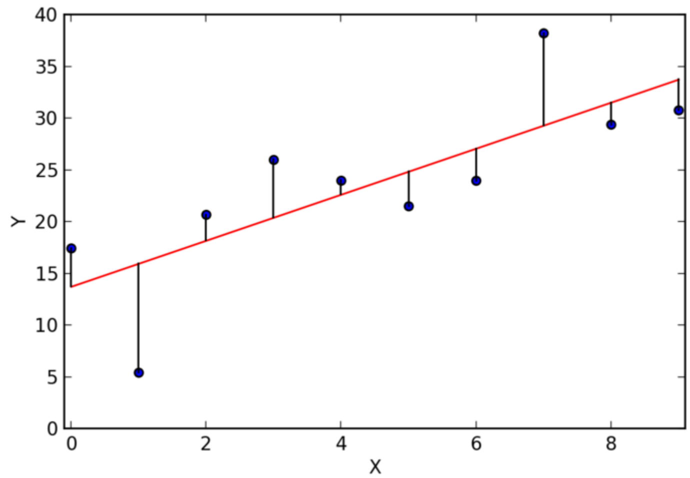


Figure 28 Linear Regression Best Fit line – residuals in black (Thomas, 2013)

Ridge Regression

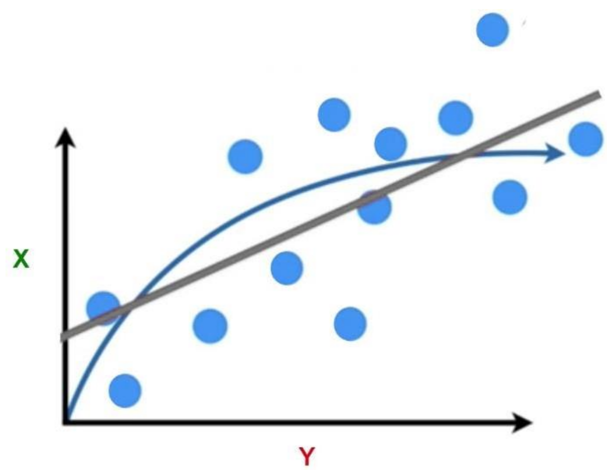


Figure 29 Ridge Regression

Gradient Boosting

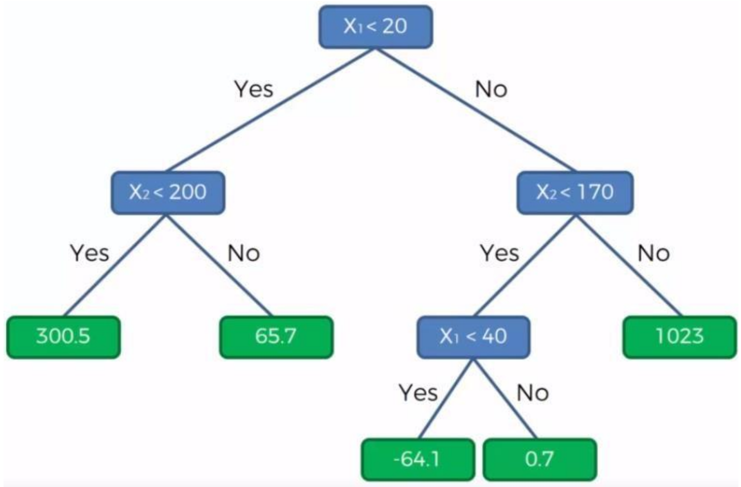


Figure 30 Decision Tree Regression (Girgin. 2019)

Appendix E

Relevant Source Code

Relevant Source Code:

https://youtube.com/playlist?list=PLeo1K3hjS3uu7clOTtwsp94PcHbzqpAdg

https://youtu.be/yNwpn9Bh4DA

Web Application Source Code:

https://github.com/FranciscoJamesKheil/broker.git

Machine Learning Model Source Code:

https://github.com/FranciscoJamesKheil/gradientBoosting_model_broker.git

Data Model Code Snippet 1 (Fitting)

```
from sklearn import ensemble
from sklearn.inspection import permutation_importance

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.1, random_state=20)

params = {
    "n_estimators": 100,
    "max_depth": 3,
    "min_samples_split": 2,
    "learning_rate": 0.3,
    "loss": "squared_error",
}
reg = ensemble.GradientBoostingRegressor(**params)
reg.fit(X_train, y_train)
```

Data Model Code Snippet 2 (Web application Integration)

```
# Implementing model to system

import pickle
import numpy as np

__locations = None
__data_columns = None

modelhouse = pickle.load(open('./broker/model/model.pkl','rb'))
@views.route('/predict_house_price',  methods=['GET','POST'])
def predict_house_price():
    brgy = request.form['brgy']
    land_area = request.form['land_area']
    floor_area = request.form['floor_area']
    classification = request.form['classification']
    storey = request.form['storey']
    roof_framing = request.form['roof_framing']
    roofing = request.form['roofing']
    ext_walls = request.form['ext_walls']
    flooring = request.form['flooring']
    doors = request.form['doors']
    ceiling = request.form['ceiling']
    windows = request.form['windows']
    wall_finish = request.form['wall_finish']
    elect_conduit = request.form['elect_conduit']
    toilet = request.form['toilet']
    plumbing = request.form['plumbing']

    brgy1 = request.form['brgy1']
    classification1 = request.form['classification1']
    roof_framing1 = request.form['roof_framing1']
    roofing1 = request.form['roofing1']
    ext_walls1 = request.form['ext_walls1']
    flooring1 = request.form['flooring1']
    doors1 = request.form['doors1']
    ceiling1 = request.form['ceiling1']
    windows1 = request.form['windows1']
    wall_finish1 = request.form['wall_finish1']
    elect_conduit1 = request.form['elect_conduit1']
    toilet1 = request.form['toilet1']
    plumbing1 = request.form['plumbing1']

    list_of_others =
[brgy,land_area,floor_area,classification,storey,roof_framing,roofing,ext_walls,flooring,
        doors,ceiling,windows,wall_finish,elect_conduit,toilet,plumbing
    ]
    list_of_sub =
[brgy1,land_area,floor_area,classification1,storey,roof_framing1,roofing1,ext_walls1,flooring1,
```

```
        doors1,ceiling1,windows1,wall_finish1,elect_conduit1,toilet1,plumbing1
    ]
    final_list = []

    others = 0
    push = 0 # push data on the list

    for x in list_of_others:
        if x == 'others':
            final_list.append(list_of_sub[push])
            others = others + 1
        else:
            final_list.append(x)

        push = push + 1

    if others > 0:
        prediction = 0.00
    else:
        # convert final list to text
        brgylist = ['canelar','divisoria','putik','san jose cawa-cawa', 'san jose gusu', 'santa maria', 'tetuan']
        classlist = ['commercial', 'residential']
        framinglist = ['c-purlins','galvanized iron','lumber','open deck','reinforced concrete',
'steel','wooden']
        roofinglist = ['corr. / galvanized iron','concrete','galvanized iron','long span', 'open deck',
'plywood', 'reinforced concrete', 's-tile']
        ext_wallslist = ['adobe type', 'cement plaster','concrete hallow block (chb)', 'painted', 'wood']
        flooringlist = ['concrete','tiles', 'wood']
        doorslist = ['flush', 'glass', 'hcpd/wfpd','panel','plywood','roll up','wood']
        ceilinglist = ['plywood','reinforced concrete','unfinished', 'wood']
        windowslist = ['aluminum casement','french window', 'glass','jalousie','scfw','steel
casement','wfpd','wood']
        wallfinishlist = ['adobe','cement plaster','concrete hallow block (chb)','plain cement paint','sand
blast','wood']
        electconlist = ['breakers','concealed','conduit pipe','mouldflex','open','painted','pdx electrical
wire','polyflex','polyvinyl chloride (pvc)','tube','unplasticized polyvinyl chloride (upvc)']
        toiletlist = ['concrete hallow block','tiles','wood']
        plumbinglist = ['galvanized iron','open','polyvinyl chloride (pvc)']

        brgy = brgylist.index(brgy)
        classification = classlist.index(classification)
        roof_framing = framinglist.index(roof_framing)
        roofing = roofinglist.index(roofing)
        ext_walls = ext_wallslist.index(ext_walls)
        flooring = flooringlist.index(flooring)
        doors = doorslist.index(doors)
        ceiling = ceilinglist.index(ceiling)
        windows = windowslist.index(windows)
        wall_finish = wallfinishlist.index(wall_finish)
```

```
    elect_conduit = electconlist.index(elect_conduit)
    toilet = toiletlist.index(toilet)
    plumbing = plumbinglist.index(plumbing)

    global_data_columns
    global_locations

    with open("./broker/model/columns.json", "r") as f:
      __data_columns = json.load(f)['data_columns']
        __locations =_data_columns[0:]

    x = np.zeros(len(_data_columns))
    x[0] = brgy
    x[1] = land_area
    x[2] = floor_area
    x[3] = classification
    x[4] = storey
    x[5] = roof_framing
    x[6] = roofing
    x[7] = ext_walls
    x[8] = flooring
    x[9] = doors
    x[10] = ceiling
    x[11] = windows
    x[12] = wall_finish
    x[13] = elect_conduit
    x[14] = toilet
    x[15] = plumbing


    prediction = modelhouse.predict([x])[0]


  tablea = final_list[0]
  tableb = final_list[1]
  tablec = final_list[2]
  tabled = final_list[3]
  tablee = final_list[4]
  tablef = final_list[5]
  tableg = final_list[6]
  tableh = final_list[7]
  tablei = final_list[8]
  tablej = final_list[9]
  tablek = final_list[10]
  tablel = final_list[11]
  tablem = final_list[12]
  tablen = final_list[13]
  tableo = final_list[14]
  tablep = final_list[15]
```

```
    return render_template("total.html",
a=tablea,b=tableb,c=tablec,d=tabled,e=tablee,f=tablef,g=tableg,h=tableh,i=tablei,j=tablej,k=tablek,l=t
ablel,m=tablem,n=tablen,o=tableo,p=tablep, prediction_alert="{:,.2f}".format(prediction),
user=current_user, fname=current_user.first_name, lname=current_user.last_name)
```

Appendix F

Screenshot/Picture of the System

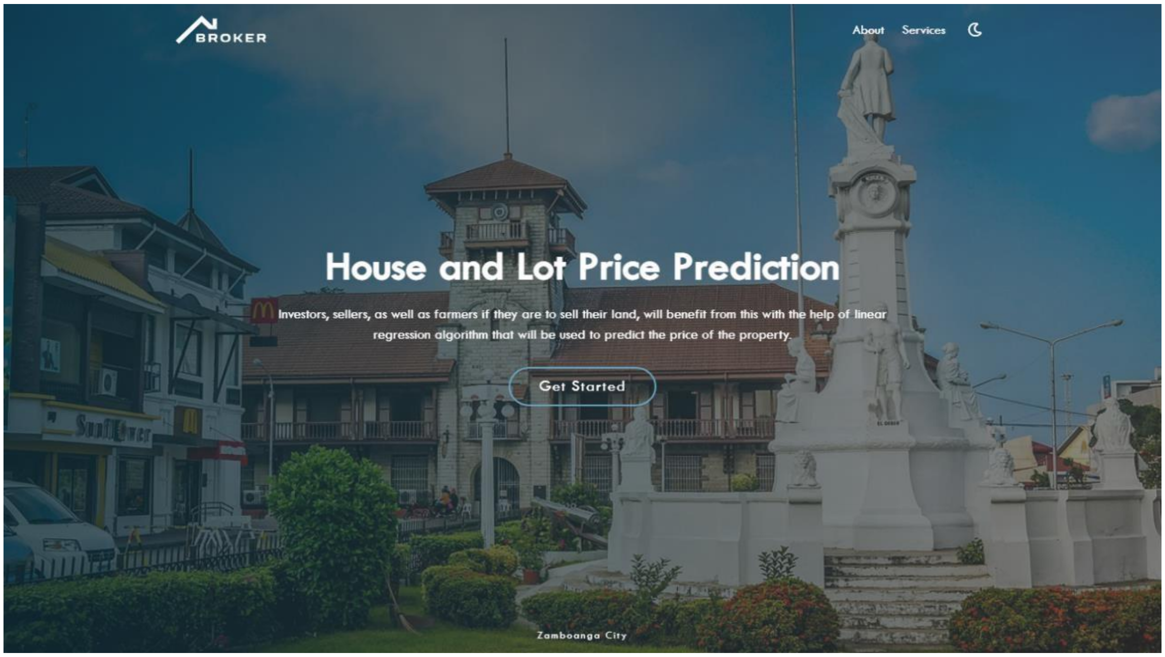


Figure 31 Landing Page of the System

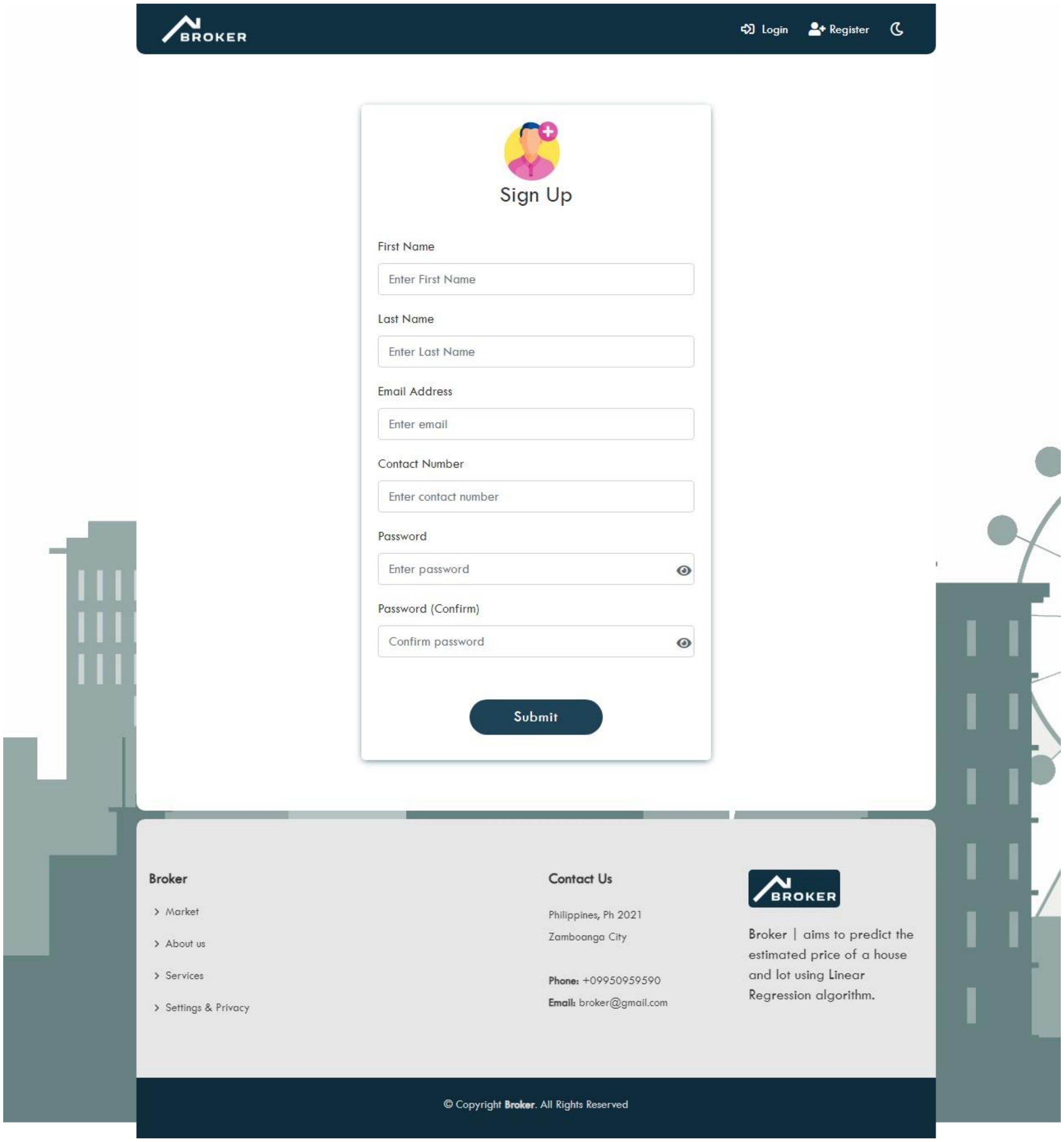


Figure 32 Registration Page of the System

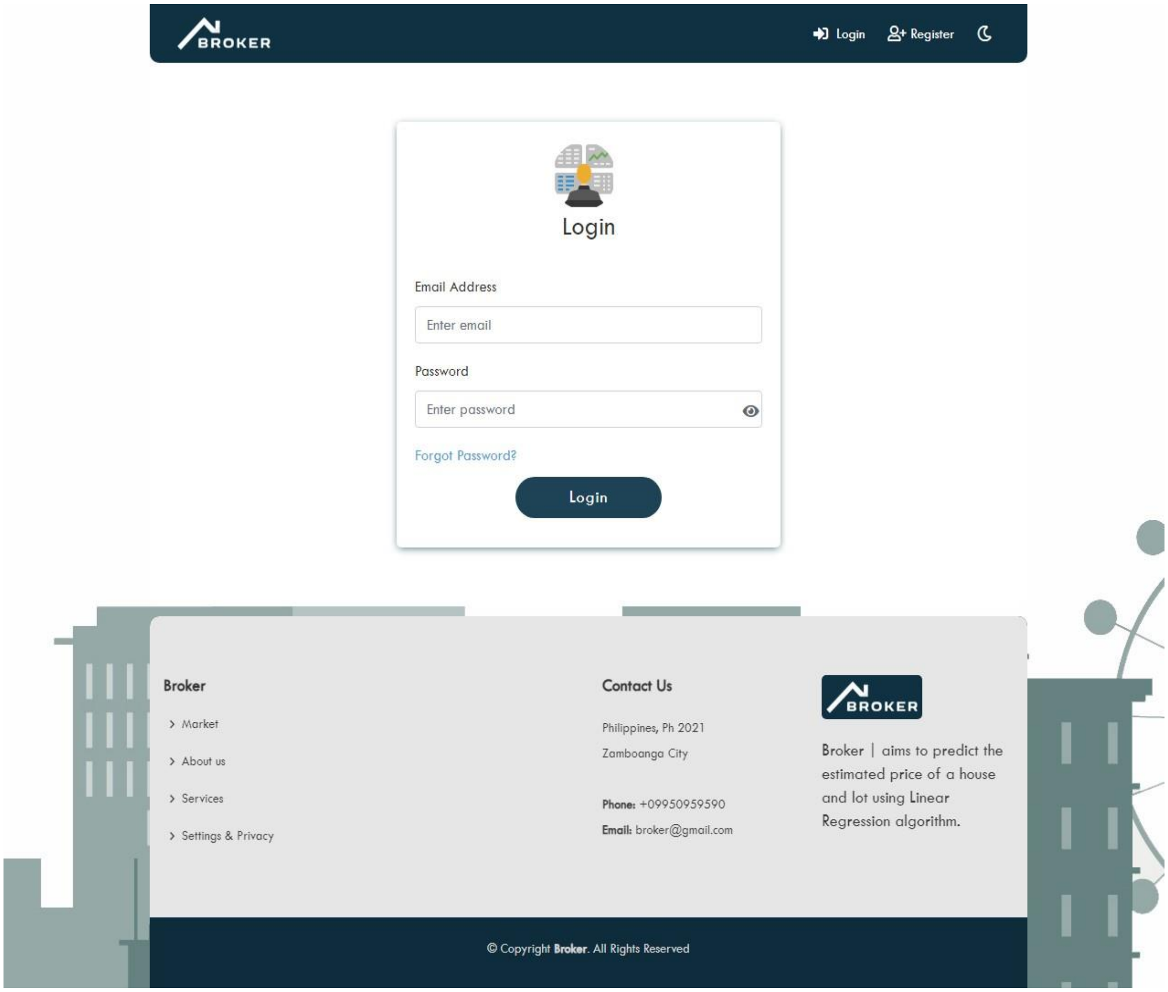


Figure 33 Login Page of the System

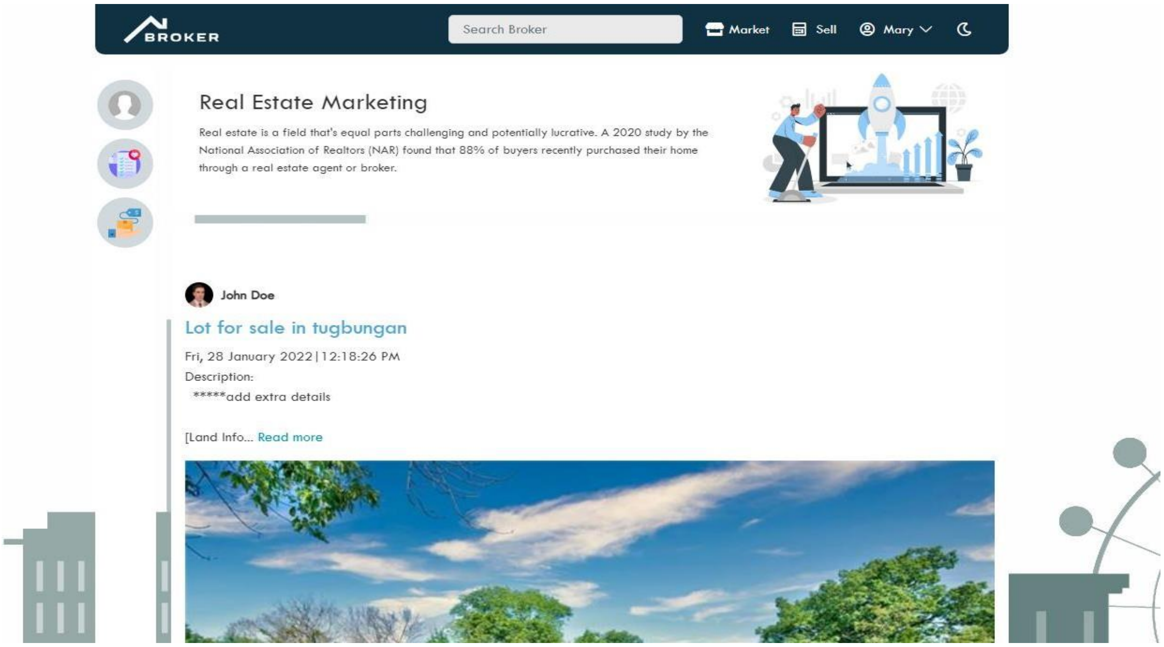


Figure 34 Market page of the system

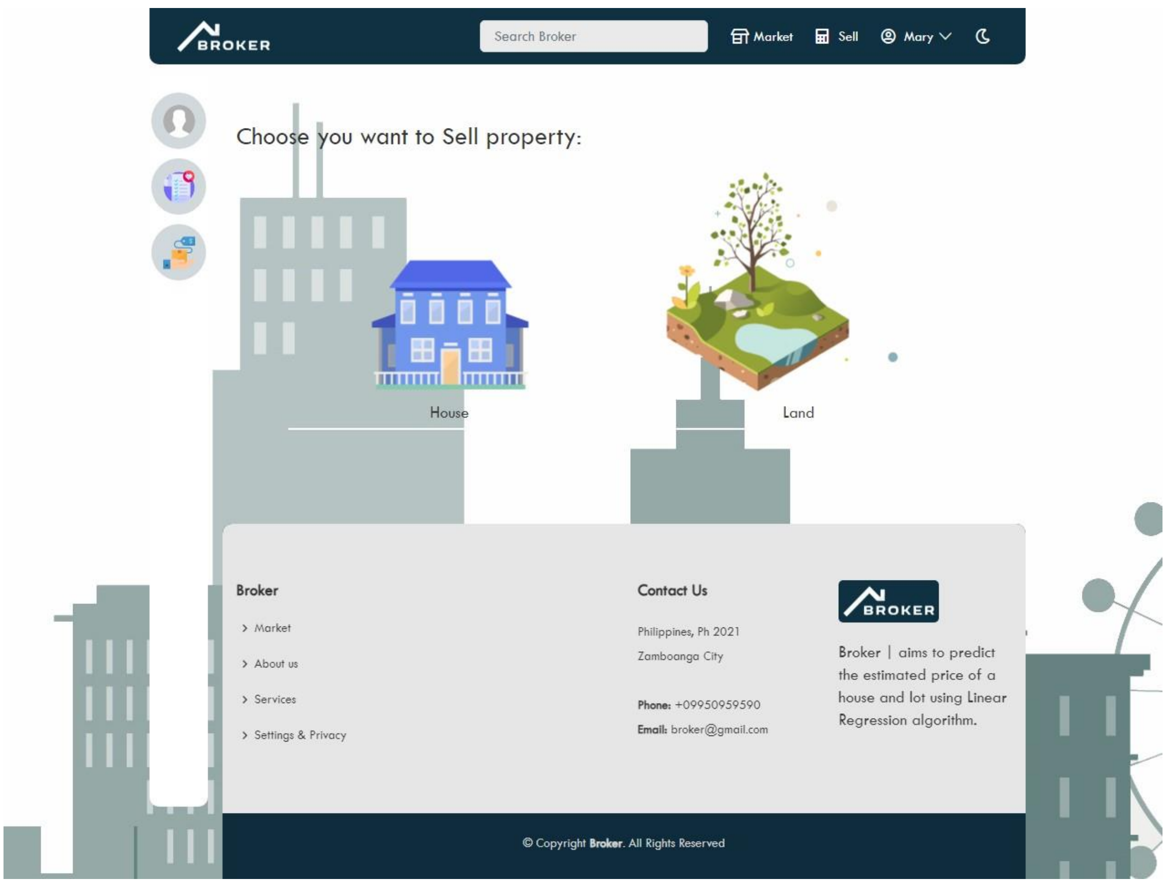


Figure 35 Sell property (core of the study)

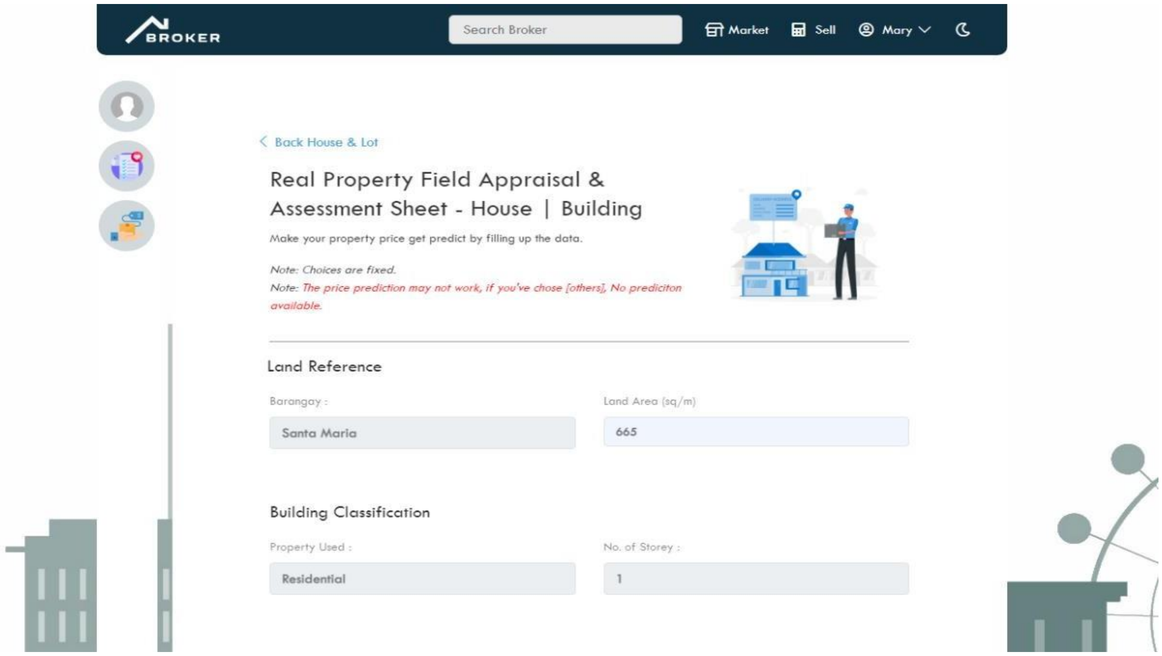


Figure 36 House prediction form

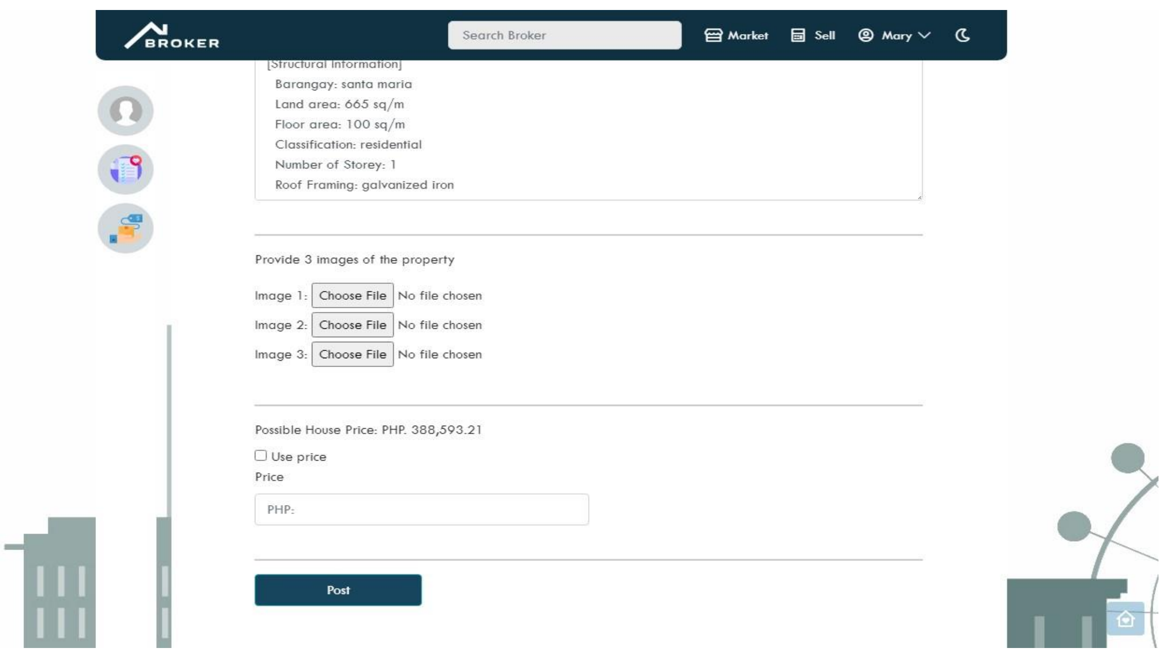


Figure 37 House estimated price

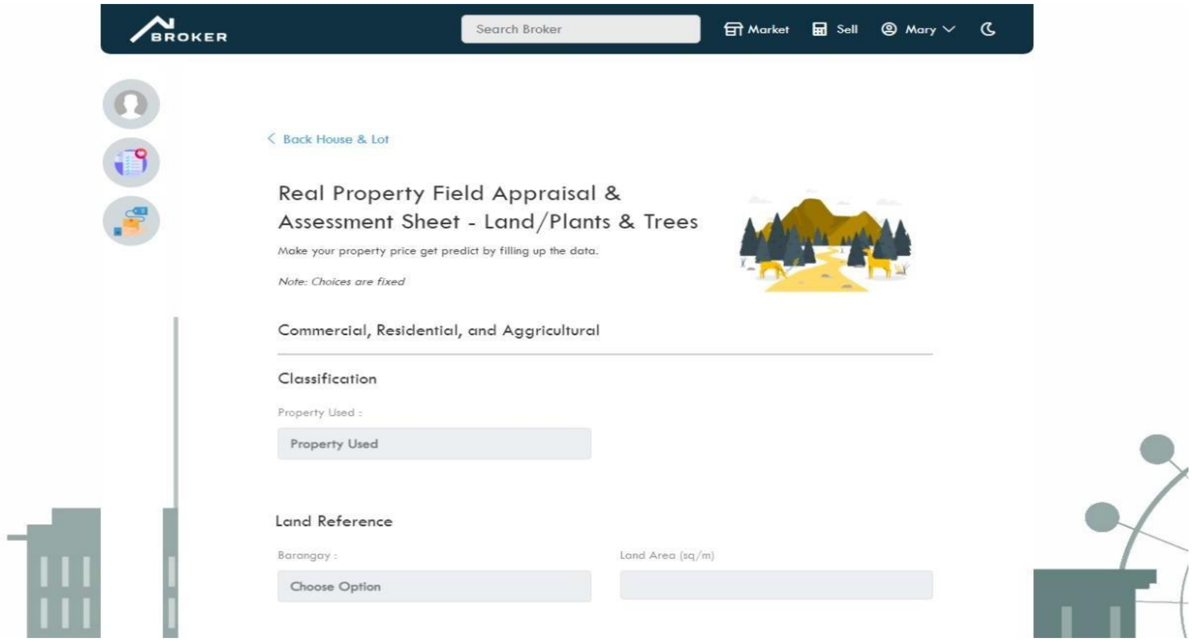


Figure 38 Land prediction form

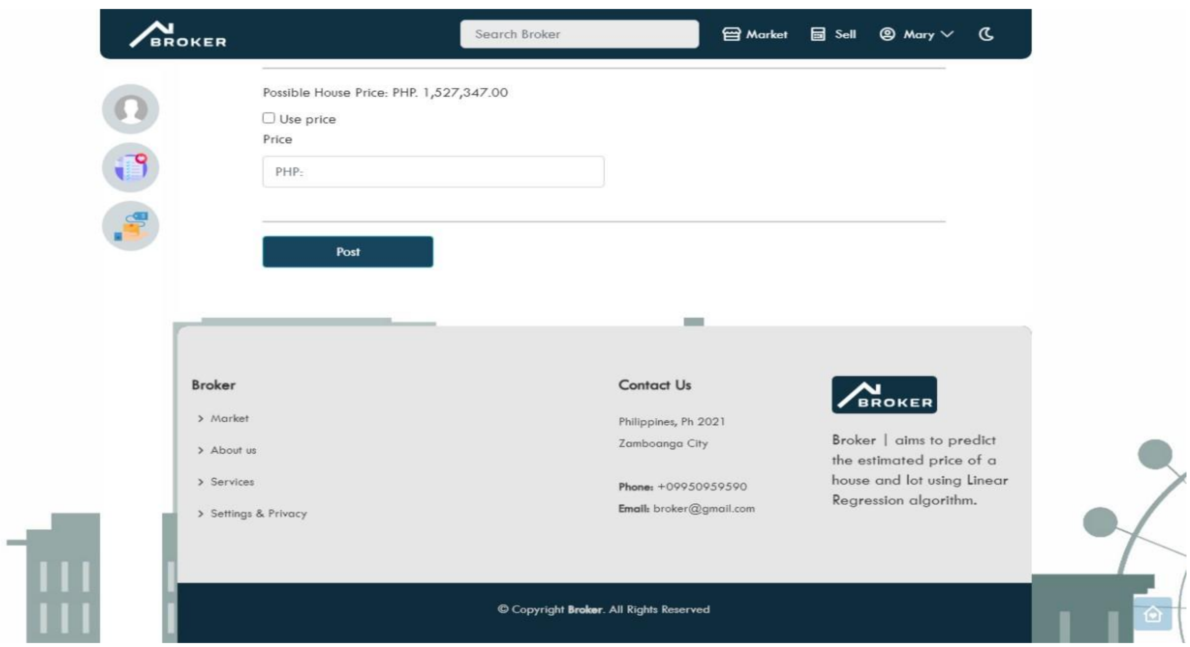


Figure 39 Land estimated price

Appendix G

Test Cases and Results

Test Cases

Project Name:	House and Lot price prediction using Gradient Boosting in Zamboanga City
Module Name:	Login
Reference Document:	House and Lot Test Plan
Date of creation:	October 12, 2021
Date of review:	April 1, 2022

Test Case ID	Test Scenario	Test Case	Pre-Condition	Test Steps	Test Data	Expected Result	Post Condition	Actual Result	Status (Pass/Failed)
Login_User_01	Verify the user's account	Enter valid email and valid password	Requires a valid user’s account to login	1. Enter email address	<valid email address>	Logged in successfully!	House and Lot Market Page displayed	House and Lot Market Page displayed	PASS
Login_User_01	Verify the user's account	Enter valid email and invalid password	Require a valid user’s account to login	1. Enter email address 2. Enter password 3. Click “Login” button	<valid email address> <invalid password>	Incorrect password, try again.	Error message prompt	Error message prompt	PASS
Login_User_01	Verify the user's account	Enter invalid email and	Require a valid user’s accou	1. Enter email address	<invalid email address>	Email does not exist.	Error message prompt	Error message prompt	PASS

		valid password	nt to login	2. Enter password	<valid password>				
				3. Click “Login” button					
Login_User_01	Verify the user’s account	Enter invalid email and invalid password	Require a valid user’s account to login	1. Enter email address	<invalid email address>	Email does not exist.	Error message prompt	Error message prompt	PASS
				2. Enter password	<invalid password>				
				3. Click “Login” button					

Table 6 Test Cases Report

Appendix H

Bug Report

Test Result Summary

Build/Version	Total Test Case	Passed Test Case	Failed Test Case	Remarks
1.0	4	4	0	

Table 7 Test Result Summary

Appendix I

Survey Questionnaire

SURVEY FORM

To Respondents:

The researchers are students of Western Mindanao State University taking BS Computer Science and is currently working on a thesis entitled " Web-Based House and Lot Price Prediction using Gradient Boosting in Zamboanga City".

The aim of the study is to determine what factors or attributes may influence the overall price of the house and lot property and to provide a system with high predictive results as possible. The outcome of the study will answer the question what is the probability that the price predicted by the system is close enough to the actual price of the house and lot using a machine learning algorithm.

Rest assured that your answers to this survey will be treated with the utmost confidentiality. The survey result will be analyzed and interpreted for academic purposes only.

Your cooperation is highly appreciated as it will contribute to the conduct of the study and will help the researchers analyze what are the factors that affects in calculating the cost of a house and lot property. Thank you, and stay safe!

Fullname: (Optional)
Occupation:
Status:

- o Single
- o Married
- o Widowed
- o Other:

1. Type of Respondent:

- o Buyer
- o Seller
- o Investor
- o Real Estate Agent
- o Researcher
- o Other:

2. Do you desire to have your own house and lot?

- o Yes
- o No
- o Maybe

* If your answer in item no. 1 is “Yes”

3. In what location do you want to have your house?

- o Residential Area
- o Urban Area
- o Rural Area
- o Other:

4. What do you consider the most in having your own house and lot property?

*Select all that applies.

- Backyard
- Fence and gate
- Spacious grounds
- Balcony
- Bedrooms
- Garage
- Spacious living room
- Bathroom
- Parking
- Dining and kitchen area
- Other:

5. How many bedrooms do you want to have?
    - 1-2 bedrooms
    - 3-5 bedrooms
    - 5 and more
6. How many bathrooms do you want to have?
    - 1 only
    - 2 only
    - More than 2
7. How much budget do you have to buy your desired house and lot?
    - ₱ 300,000 - ₱ 500,000
    - ₱ 500,000 - ₱ 1.5M
    - ₱ 1.5M - ₱ 2M
    - ₱ 2M and above
8. Do you want to have an application that can help you predict the price of your dream house?
    - Yes
    - No
    - Maybe

Appendix J

Curriculum Vitae

James Kheil Francisco

web developer

Contact

jameskheilsalon@gmail.com

(+63) 936-177-8011

Profile

I am a highly creative and multi-talented front-end web developer, hardworking, dedicated, and versatile person with moderate analytical and critical thinking skills. I am always excited to learn new things.

Objectives

To expand my learnings, knowledge, and skills through self-learning and experience.

References

Salimar B. Tahil
College of Computing Studies
(+63) 917-177-1654,
tahil.salimar@wmsu.edu.ph

Jaydee Ballaho
College of Computing Studies
jaydee.ballaho@wmsu.edu.ph

Education

Bachelor of Science in Computer Science — 2018 - 2022
Western Mindanao State University
Award: Academic Award
GPA: 1.9702

Information and Communications Technology — 2016 - 2018
System Technology Institute, Zamboanga
Senior High School

Work Experience

Front-end Developer — 2017 - 2018
Gavilan Team is a local private company IT group of developers who are specifically developed a software application and game development.

Achievements

Java Programming Competition, 3rd placer — 2017
Java Festival
STI College Zamboanga City

Final Project, Best Project — 2018
Web Development Project
STI College Zamboanga City

Skills

Graphic Designing

Programming

Communication

Document Processing

PC Troubleshooting

Team Work

Christle Joyce Santuyo

web developer

Contact

christlejsan@gmail.com

(+63) 997-445-3367

Profile

Dynamic and determined team player with well-developed written and verbal communication abilities.

Objectives

To have an experience and to expand my learnings, knowledge, and skills in areas relate to science and technology.

References

Salimar B. Tahil
College of Computing Studies
(+63) 917-177-1654,
tahil.salimar@wmsu.edu.ph

Jaydee Ballaho
College of Computing Studies
jaydee.ballaho@wmsu.edu.ph

Education

Bachelor of Science in Computer Science — 2018 - 2022
Western Mindanao State University
Award: Academic Award
GPA: 1.8096

Science, Technology, Engineering and Mathematics — 2016 - 2018
Western Mindanao State University
Senior High School

Special Science Class — 2012 - 2016
Zamboanga City High School Main

Work Experience

Software Engineer Project - ICT Faculty Club Management System — 2020 - 2021
Software engineering is a required course in Computer Science degree wherein students are task to develop a working system and deliver it for their respective clients organization/business

Achievements

University Quiz Bowl 2019 2nd runner-up
College of Engineering AVR, Western Mindanao State University
November 7, 2019

Skills

- Microsoft Office
- Web Development
- Document Processing
- Communication
- PC Troubleshooting
- Team Work