Proposal − Minor Project

TITLE: Real Estate Price Predictor

Submitted by:

Dimple Saraogi(075BCT034)

Erosha Paudel(075BCT035)

Bishal Choudhary (075BCT027)

Beenaya Basnet(075BCT021)

Submitted To:

Department of Electronics and
Computer Engineering

ABSTRACT

Every year, house prices rise, necessitating the development of a method to forecast future house values. House price prediction can assist a developer in determining the selling price of a home and can also assist a client in determining the best time to buy a home. Physical conditions, concept, and location are the three aspects that impact the price of a home. The goal of this project is to use regression analysis to estimate housing prices with the least amount of error.

ACKNOWLEDGEMENT

Table of Contents:

INTRODUCTION

Real estate valuation is required for a variety of purposes, including financing, sales listing, investment research, property insurance, and taxation. However, the most common use of real estate valuation is establishing the asking the purchase price of real estate to a broker or an agent.

An accurate appraisal depends on the methodical collection of data. Specific data, covering details regarding the particular property, and general data, pertaining to the nation, region, city, and neighborhood wherein the property is located, are collected and analyzed to arrive at a value. Appraisals use three basic approaches to determine a property's value.

Because no two properties are precisely alike, adjustments will be made to the comparables' sales prices to account for differences in features and other factors that affect value, such as:

- Age and condition of buildings
- Terms and conditions of sale
- Location
- Physical features, including lot size, type and quality of construction, number and type of rooms, square feet of living space, kitchen upgrades etc.

Accurate real estate valuation is important to mortgage lenders, investors, insurers and buyers, and sellers of the property. While appraisals are generally performed by skilled professionals, anyone involved in a real transaction can benefit from gaining a basic understanding of the different methods of real estate valuation.

For the major part of this project we will be focusing on the model that will predict the estimate price of the house that is to be sold with the datasets. The project will be done by supervised learning process and regression analysis.

Problem Statement

The appraisal of real estate is traditionally conducted by a licensed professional, who would carry out a numerous surveys based on several factors such as location, surroundings, areas, and facilities of a real estate. Nonetheless, the appraisers' criteria and vested interests would almost certainly be included in a manual appraisal. This potential risk would likely cause a biased or subjective evaluation of a particular real estate, resulting in losses for investors or consumers.

Objectives

Any academic project is incomplete without holding a strong ground of achievable objectives. In the beginning phases of brainstorming the project, "Real Estate Price Predictor", we have assumed certain goals that we shall push to achieve at the end of this project. They are tabulated below:

1. To learn Machine Learning that solves real world problems.
2. To solve problems with supervised learning.
3. To improve the ability to work and cooperate in a team.

Scope and Applications

The main motive of this project is to learn how the concepts we have learned in class can solve real world problems. The code will be modular thus it will help for extension of this project in future as we will be able to add different features and go beyond houses and this model could check for new data time to time, incorporate them to expand the dataset and produce better and more accurate results.

Literature Review

Many studies on the study and forecast of housing values have been undertaken during the last few years. Ian D Wilson created an artificial neural network that assisted in forecasting future house price patterns in England. A regression model created by A.S. Mark and W.B. John was beneficial in assessing house price changes in a given area. Sampath Kumar and Santhi employed the multiple linear regression technique to estimate house prices in a given area, as well as the growth in land prices over a one-year period. In order to create their house price forecast model, they used economic parameters rather than traditional criteria. There is also a similar website called REALas.com which does the same thing.

Methodology

Advanced research methodologies such as machine learning and artificial intelligence have been widely implemented in many parts of modern property industry research. They are used not only to determine the price and worth, but also to determine potential future applications and potential obstacles. Machine learning and artificial intelligence have been widely adopted in the property business, transforming it from an experience-driven industry with significant arbitrage opportunities to an intelligent and data-driven enterprise.

Multiple linear regression is the most reliable method for predicting home prices, and it is used with the Spearman correlation coefficient to find the influential elements influencing house prices. For model creation, we used a data set of home prices to train and test the parameters of this multiple linear regression model. The most accurate way to examine and anticipate the house price is to use a multiple linear regression model.

When the distribution of the response variable Y (response variable) is continuous and approximately regular, the most commonly employed model is the linear regression model. The process for estimating the coefficients of a linear equation with at least one independent variable that best predicts the value of a dependent variable is known as linear regression. Our goal is to predict the outcome Y based on the given values of predictor variables.

Here are three common evaluation metric for regression problems:

Mean Absolute Error (MAE) is the mean of the absolute value of the errors and it calculates average error.

$$\frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

Mean Squared Error (MSE) is the mean of the squared errors. It is more popular than MAE because MSE "punishes" larger errors, which tends to be useful in the real world.

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

Root Mean Squared Error (RMSE) is the square root of the mean of the squared errors. RMSE is even more popular than MSE because RMSE is interpretable in the "y" units.

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

Importing Data and Checking out
First the data files will be read using pandas check needed rows of the data frame.

Training of Regression Model
Then the training of regression model will be done. First the data will be split into an X list that contains the features to train on, and a y list with the target variable.

Split Data into Train, Test
The training set will be going to use for training the model and testing set for testing the model. We are creating a split of 40% training data and 60% of the training set.

Creating and Training the Linear Regression Model
Linear Model will be imported and created and fit the training dataset in it. And then the prediction of test set will be seen on the basis of how it will perform.
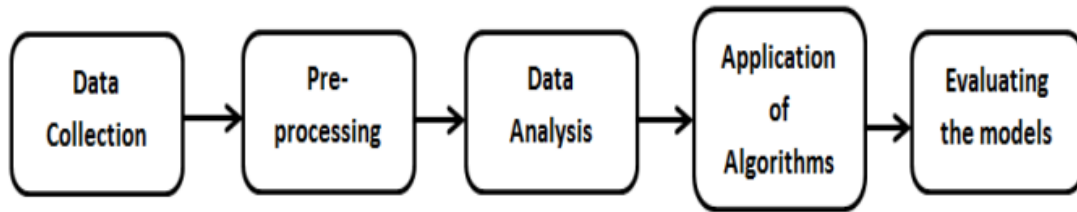
Block Diagram

Fig. System Block Diagram

Softwares Used

We will be using Python as our coding tool as it is fast, robust and compatible. We will also be using various platforms like Visual Studio Code, Jupyter Notebook and Github.

For the data manipulation and analysis we will be using Pandas library as it offers data structures and operations for manipulating numerical tables and time series.

Expected Outcomes

The price of various houses is predicted using a regression model. It falls under the category of supervised learning, which is one of the machine learning categories. The multiple linear regression model is shown to be appropriate for predicting property prices and this project is hoped to give the price prediction with minimum errors.

Project Schedule

The rough estimation of schedule is shown below based upon active working days

- Workflow Brainstorming: 4 days
- Learning Schedule: 10 days
- Dataset Collection and Manipulation : 7 days
- Coding Implementation: 14 days
- Revision: 7 days
- Conclusive Report/Finalizing: 3 days

The project assumes 45 active working days as per the Project Schedule. This is a rough estimation and might vary depending upon the complexity of obstacles we might face while taking on the project.

Bibliography and References

Wilson, I.D., Paris, S.D, Ware, J.A., & Jenkins, D.H. Residential Property Price Time Series Forecasting With Neural Networks. Journal of Knowledge-Based Systems; 2002, 15: 335-341

Mark, A.S., & John, W.B. Estimating Price Paths for Residential Real Estate. Journal of Real Estate Research; 2003: 25, 277–300

https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3833734

https://linuxhint.com/house-price-prediction-linear-regression/#:~:text=It%20is%20an%20algorithm%20of,tasks%20like%20house%20price%20prediction.