House Price Prediction Using Machine Learning

### A Project Work Synopsis

*Submitted in the partial fulfillment for the award of the degree of*

# BACHELOR OF ENGINEERING

### *Specialization in*

### BIG DATA ANALYTICS

### Submitted by:

**EALURI YESHWANTH**

**IQBAL SINGH**

**SHIV MURAT VERMA**

**ANKIT SINGH**

**SUDIKSHA**

### University Roll Number:

**18BCS3765**

**18BCS3766**

**18BCS3768**

**18BCS3771**

**18BCS3774**

### Under the Supervision of:

### SUPERVISOR’S NAME

### Ms. Seema



### 

### CHANDIGARH UNIVERSITY, GHARUAN, MOHALI - 140413,

**PUNJAB**

#### MONTH & YEAR

#### SEPTEMBER & 2021

**TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| **Sl. No.** | **Title** | **Page No.** |
| **1** | **Project Title and Brief Description** | **1** |
| **2** | **Abstract** | **1** |
| **3** | **Introduction** | **2** |
| **4** | **Aim and Purpose** | **2** |
| **5** | **Literature Survey** | **3** |
| **6** | **Methodology** | **4** |
| **7** | **References** | **5** |

**PROJECT TITLE AND BRIEF DESCRIPTION**

HOUSE PRICE PREDICTION USING MACHINE LEARNING

Thousands of houses are sold everyday. There are some questions every buyer asks himself like: What is the actual price that this house deserves? Am I paying a fair price? In this paper, a machine learning model is proposed to predict a house price based on data related to the house (its size, the year it was built in, etc.). During the development and evaluation of our model, we will show the code used for each step followed by its output. This will facilitate the reproducibility of our work. In this study, Python programming language with a number of Python packages will be used..

In this Synopsis, we’ll discuss the project where we worked on predicting the price of house based on different parameters using Machine Learning algorithms. The algorithms included Linear Regression, XGBoost, etc. The dataset has been taken from Kaggle.

ABSTRACT

Machine learning plays a major role from past years in image detection, spam reorganization, normal speech command, product recommendation and medical diagnosis. Present machine learning algorithm helps us in enhancing security alerts, ensuring public safety and improve medical enhancements. Machine learning system also provides better customer service and safer automobile systems. In the present paper we discuss about the prediction of future housing prices that is generated by machine learning algorithm. For the selection of prediction methods we compare and explore various prediction methods. We utilize linear regression as our model

because of its adaptable and probabilistic methodology on model selection. Our result exhibit that our approach of the issue need to be successful, and has the ability to process predictions that would be comparative with other house cost prediction models.

More over on other hand housing value indices, the advancement of a housing cost prediction that tend to the advancement of real estate policies schemes. This study utilizes machine learning algorithms as a research method that develops housing price prediction models. We create a housing cost prediction model In view of machine learning algorithm models for example, XGBoost, lasso regression and neural system on look at their order precision execution. We in that point recommend a housing cost prediction model to support a house vender or a real estate agent for better information based on the valuation of house. Those examinations exhibit that linear regression algorithm, in view of accuracy, reliably outperforms alternate models in the execution of housing cost prediction..

INTRODUCTION

Machine learning is a subfield of Artificial Intelligence (AI) that works with algorithms and

technologies to extract useful information from data. Machine learning methods are appropriate

in big data since attempting to manually process vast volumes of data would be impossible

without the support of machines. Machine learning in computer science attempts to solve

problems algorithmically rather than purely mathematically. Therefore, it is based on creating

algorithms that permit the machine to learn. However, there are two general groups in machine

learning which are supervised and unsupervised. Supervised is where the program gets trained

on pre-determined set to be able to predict when a new data is given. Unsupervised is where

the program tries to find the relationship and the hidden pattern between the data.

Several Machine Learning algorithms are used to solve problems in the real world today.

However, some of them give better performance in certain circumstances.

The performance will be measured upon predicting house prices since the prediction in many

regression algorithms relies not only on a specific feature but on an unknown number of

attributes that result in the value to be predicted. House prices depend on an individual house

specification. Houses have a variant number of features that may not have the same cost due to

its location. For instance, a big house may have a higher price if it is located in desirable rich

area than being placed in a poor neighbourhood.

The data used in the experiment will be handled by using a combination of pre-processing

methods to improve the prediction accuracy. In addition, some factors will be added to the local

dataset in order to study the relationship between these factors and the sale price in Bangalore.

AIM AND PURPOSE

The No Free Lunch Theorem state that algorithms perform differently when they are used

under the same circumstances. This study aims to analyse the accuracy of predicting house

prices when using Multiple linear, Lasso, Ridge, Random Forest regression algorithms and linear regression. Thus, the purpose of this study is to deepen the knowledge

in regression methods in machine learning.

In addition, the given datasets should be processed to enhance performance, which is

accomplished by identifying the necessary features by applying one of the selection methods

to eliminate the unwanted variables since each house has its unique features that help to

estimate its price. These features may or may not be shared with all houses, which means they

do not have the same influence on the house pricing resulting in inaccurate output.

The main objectives of this study are as follows:

* To apply data preprocessing and preparation techniques in order to obtain clean data
* To build machine learning models able to predict house price based on house features
* To analyze and compare models performance in order to choose the best model

LITERATURE SURVEY

EXISTING SYSTEM

The value of a particular property depends on the infrastructure, amenities, surrounding the property​. Recently, a few writers’ scopes for finding the best properties for the customers came along with various technologies. Raghunandhan mentioned the basic data mining concepts of how it works and supporting algorithms for the purpose of prediction. The most important part is which machine learning algorithm is best suited for predicting the house price. Often the location's environmental conditions decide what kind of price we can expect for different types of houses, Manjula presents various important features to use when forecasting property prices with good precision using a regression model. A. Varma designed a system that used real-time neighborhood data to get precise real-world valuations using Google maps.

Researchers also showed that there exist relationships between the visual appearance and non-visual attributes

such as crime statistics, housing prices, population density, etc. of a city. For instance, “City Forensics: Using Visual Elements to Predict Non-Visual City Attributes” [4], uses visual attributes to predict the sale price of the property. Hujia Yu, Jiafu Wu (2014) [5] used classification and regression algorithms. According to analysis, living area square feet, roof content, and neighborhood have the greatest statistical importance in estimating the selling price for a home. And prediction analysis can be improved by the PCA technique. Li Li and kai-Hsuan Chu (2017) studied various algorithms such as Backpropagation neural network (BPN) and Radial basis functional (RBF) neural networks. The use of RBF and BPN models is introduced to identify the difference between the house price index such as Cathy and sinny price index and complicated correlation function to detect the macroeconomic analysis.

Nihar Bhagat, Ankit Mohokar, Shreyash Mane (2016) studied linear regression algorithms for prediction of the houses. The goal of the paper is to predict the efficient price of real estate for customers with respect to their budgets and priorities. Analysis of past market trends and price ranges will predict future house pricing.

PROPOSED SYSTEM:

*Phase 1: Collection of data*

Data processing techniques and processes are numerous. We collected data for Mumbai’s real estate properties from various real estate websites. The data would be having attributes such as Location, carpet area, built-up area, age of the property, zip code, etc. We must collect the quantitative data which is structured and categorized. ​Data collection is needed before any kind of machine learning research is carried out. Dataset validity is a must otherwise there is no point in analyzing the data.

*Phase 2: Data preprocessing*

Data preprocessing is the process of cleaning our data set. There might be missing values or outliers in the dataset. These can be handled by data cleaning. If there are many missing values in a variable we will drop those values or substitute it with the average value.

*Phase 3: Training the model*

Since the data is broken down into two modules: a Training set and Test set, we must initially train the model. The training set includes the target variable. The decision tree regressor algorithm is applied to the training data set. The Decision tree builds a regression model in the form of a tree structure.

*Phase 4: Testing and Integrating with UI*

The trained model is applied to test dataset and house prices are predicted. The trained model is then integrated with the front end using Flask in python.

M​ETHODOLOGY

1. STUDIED ALGORITHMS:

In the process of developing this model, various regression algorithms were studied. SVM, Random Forest, Linear regression, Multiple linear regression, Decision Tree Regressor, KNN, all were tested upon the training dataset. However, the decision tree regressor provided the highest accuracy in terms of predicting the house prices. The decision to choose the algorithm highly depends upon the dimensions and type of data in the data used. The decision tree algorithm suited best for our dataset.

1. DECISION TREE REGRESSOR:

The decision tree regressor observes features of an attribute and trains a model in the form of a tree to predict data in the future to produce meaningful output. Decision tree regressor learns from the max depth, min depth of a graph and according to system analyzes the data. Grid Search CV is a way to deal with parameter tuning that will efficiently manufacture and assess a model for every mix of calculation parameters indicated in a grid. Grid Search CV in this algorithm is used to assess the best value for max-depth, using which the decision tree is constructed.

1. FLASK INTEGRATION

After building the model and successfully giving the result, the next step is to do the integration with the UI, for this purpose flask is used. Flask is a web framework. This means flask provides you with tools, libraries, and technologies that allow you to build a web application. Flask is easy to put away routes together and this framework is mainly used for integrating python models.

REFERENCES

[1] ​Lakshmi, B. N., and G. H. Raghunandhan. "A conceptual overview of data mining." 2011 National Conference on Innovations in Emerging Technology. IEEE, 2011.

[2] ​Manjula, R., et al. "Real estate value prediction using multivariate regression models." Materials Science and Engineering Conference Series​. Vol. 263. No. 4. 2017.

[3] ​A. Varma et al., “House Price Prediction Using Machine Learning And Neural Networks,” 2018 Second International Conference on Inventive Communication and Computational Technologies, pp. 1936–1939, 1936.

[4] ​Arietta, Sean M., et al. "City forensics: Using visual elements to predict non-visual city attributes." ​IEEE transactions on visualization and computer graphics 20.12 (2014)

[5] ​Yu, H., and J. Wu. "Real estate price prediction with regression and classification CS 229 Autumn 2016 Project Final Report 1–5." (2016).

[6] Li, Li, and Kai-Hsuan Chu. "Prediction of real estate price variation based on economic parameters." 2017 International Conference on Applied System Innovation (ICASI). IEEE, 2017.

[7] Nihar Bhagat, Ankit Mohokar, Shreyash Mane "House Price Forecasting using Data Mining" International Journal of Computer Applications,2016.

[8] N. N. Ghosalkar and S. N. Dhage, "Real Estate Value Prediction Using Linear Regression," 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), Pune, India, 2018.

[9] ​Pow, Nissan, Emil Janulewicz, and Liu Dave Liu. "Applied Machine Learning Project 4 Prediction of real estate property prices in Montréal." ​Course project, COMP-598, Fall/2014, McGill University​ (2014).

[10] Sampathkumar, V., Santhi, M. H., & Vanjinathan, J. (2015). Forecasting the land price using statistical and neural network software. ​Procedia Computer Science​, ​57​, 112-121.

[11] Banerjee, Debanjan, and Suchibrota Dutta. "Predicting the housing price direction using machine learning techniques." 2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI)​. IEEE, 2017.