Prediction of House Price using SVM in ML

Prof. Shikha Pachouly[1], Mr. Piyush Chordia[2], Mr. Pratik Konde[3], Ms. Suraya Jadhav[4], Hrithik Pandhare[5]

Department of Computer Engineering[1,2,3,4,5]

AISSMS College of Engineering, Pune, Maharashtra, India

Abstract: *House costs increment consistently, so there is a need for a framework to anticipate house costs later on. House cost expectation can assist the designer with deciding the selling cost of a house and can assist the client with orchestrating the perfect chance to buy a house. Three factors impact the cost of a house which include physical condition, idea, and area. House price forecasting is a crucial topic of land. Machine learning techniques are applied to research historical property transactions in World to get useful models for house buyers and sellers. In the current paper, we examine the expectation of future lodging costs that are produced by AI calculation. For the choice of forecast strategies we look at and investigate different expectation techniques. For the choice of prediction methods, we compare and explore various prediction methods. We utilize regression as our model due to its adaptable and probabilistic methodology for model selection. Our result exhibit that our approach of the difficulty got to achieve success, and has the power to process predictions that might be comparative with other house cost prediction models. This study utilizes machine learning algorithms as a search method that develops housing price-prediction models.*

Keywords: SVM, ML

I. INTRODUCTION

- Development of civilization is the foundation of the increase in demand for houses day by day. Accurate prediction of house prices has been always a fascination for buyers, sellers, and bankers also. Many researchers have already worked to unravel the mysteries of the prediction of house prices. Many theories have been given birth as a consequence of the research work contributed by various researchers all over the world. Some of these theories believe that the geographical location and culture of a particular area determine how the home prices will increase or decrease whereas other schools of thought emphasize the socio-economic conditions that largely play behind these house price rises.

- We all know that a house price is a number from some defined assortment, so obviously, the prediction of prices of houses is a regression task. To forecast house prices one person usually tries to locate similar properties in his or her neighborhood and based on collected data that person will try to predict the house price. All these indicate that house price prediction is an emerging research area of regression that requires the knowledge of machine learning. This has motivated me to work in this domain.

- Real estate appraisal is an integral part of the property buying process. Traditionally, the appraisal is performed by professional appraisers specially trained for real estate valuation. For the buyers of real estate properties, an automated price estimation system can be useful to estimate the prices of properties currently on the market. Such a system can be particularly helpful for novice buyers who are buying a property for the first time, with little to no experience.

II. LITERATURE SURVEY

Paper Name: Virtual Reality for Real Estate
Author: Bogdan, Alexandru Deaky , Luminita Parv.
This paper presents the results of the VR4RE (Virtual Reality for Real Estate) project, which aims at saving time and money for both real estate sellers and buyers by employing modern technologies. VR4RE is one of the innovative projects developed by Bluemind Software and it is in an advanced state. This paper also illustrates the history of in-house technological attempts at creating appropriate presentation tools for real estate properties with 3D and VR (Virtual Reality).

Paper Name: Developing Smart Commercial Real Estate

Author: Peter Ekman

In this paper, CNN-based detection and evaluation of infected patients. Tailored CNN models: A set of tailored models based on CNN have been designed to take three sets of image categories (e.g.; normal case, viral pneumonia case, and bacterial case). DenseNet169 architecture and RNN-based architectures are used.

To evaluate the potential of smart commercial real-estate (CRE) we studied a Swedish commercial real estate firm that has developed and deployed a technology-based self-service (TBSS) to help tenants reduce energy consumption.

Paper Name; An Optimized Support Vector Machine (SVM) based on Particle Swarm Optimization (PSO)

for Cryptocurrency Forecasting

Author :Nor Azizah Hitama, Amelia Ritahani Ismailb, Faisal Saeed

In this, we have studied various Support Vector Machine(SVM) and Particle Swarm Optimization (PSO) for forecasting cryptocurrency. Forecasting accurate future value is very important in the financial sector. An optimized Support Vector Machine (SVM) based on Particle Swarm Optimization (PSO) is introduced in forecasting the cryptocurrency's future price. It is part of Artificial Intelligence (AI) that uses previous experience to forecast the future price.

Paper Name: Comparison of Ensemble Methods for Real Estate Appraisal

Author::- Prathamesh Kumar, Ishan Madan, Ashutosh Kale

In this paper, four ensemble methods, namely Bagging, Random Forest, Gradient Boosting, and Extreme Gradient Boosting were analyzed and compared in terms of their efficiency in the appraisal of real estate in Mumbai. The property listings available on the real estate website 99acres were used as the data source for this study. The analysis showed that Extreme Gradient Boosting (XGBoost) model performed the best as compared to the rest of the ensemble models. The results confirm that ensemble models can be useful for estimating real estate prices.

Paper Name: Prediction of House Pricing Using Machine Learning with Python

Author: Mansi Jain, Himani Rajput, Neha Garg

This paper provides an overview of how to predict house costs utilizing different regression methods with the assistance of python libraries. The proposed technique considered the more refined aspects used for the calculation of house prices and provide a more accurate prediction. It also provides a brief about various graphical and numerical techniques which will be required to predict the price of a house. This paper contains what and how the house pricing model works with the help of machine learning and which dataset is used in our proposed model.

III. METHODOLOGY

3.1 Dataset

The dataset utilized during this work is obtained from land Agents within the US to gauge the prices of homes. This dataset was intended to help them find the foremost appropriate cost of the house within the respective locations within the US. For predicting the prices of the homes, the following attributes are identified and included, in conjunction with the acronym used for his or her representation within the dataset snippet as shown. Base Price (price), Size of Plot, House age, Number of Bedrooms, Number of rooms, building year, renovation year, Address, of the house. These attributes, or predictor variables, are the factors that are majorly considered during house purchase and thus influence the pricing of homes. The dataset utilized in this work is obtained from land Agents within the US to judge the costs of homes.

The data set we have used is a group of houses in the USA. The size of the dataset is 5,000 houses which are divided into training data and testing data. The dataset contains 7 columns and 5000 rows with CSV extension. The data contains the following columns:

a) 'Bedrooms' – No bedrooms in the house.

b) 'Sqft_living' – sq feet of the house.

c) 'floor' – The floor on which the house is.

d) 'Condition' –the condition of the house.

e) 'yr_built' – On which year the house is built.

f) 'Price' – Price that the house sold at.

g) 'yr_renovated' – on which year last time house was renovated.

h) 'Zipcode' – what is the zip code of the area.

3.2 Fetch the Data
Set with the help of the pandas in the python platform and analyze the data set.

3.3 Data Preprocessing
This is a process to convert raw data into meaningful data using different techniques. Data in the real world are incomplete, noisy, duplicated, or inconsistent, so we can convert this data into accuracy, completeness, and consistency through various data preprocessing techniques. Major steps in data preprocessing:

a) Data Cleaning

b) Data Integration

c) Data Reduction

d) Data Transformation

e) Data Discretization

3.4 Train Test Split Evaluation
It is a technique for evaluating the performance of a machine learning algorithm. We have segregated dependent and independent variables in two different vectors respectively x and y, we handled our data preprocessing. We have 500 records inside of our Data set and we split the data into a training set and a test set. Vector y is dependent or target variables. The training part of data is there that will pass to our machine learning model for the training purpose and one machine learning model is trained, the rest remaining data which is test data will use to validate the machine learning model. We split the data in percentages between 0 and 1 for either the train or test datasets. Here we take a test set with the size of 0.40 (40 %) means that the remaining percentage of 0.60 (60 %) is assigned to the training set.

A. Linear Regression
Linear regression is a statistical model that attempts to point out the connection between two variables with the equation. It is one of the easiest algorithms in machine learning. It is calculated by using the formula,

$$y = MX + b$$

y = Dependent variable, m = Coefficient rate and slope of the line, x= Independent variable, b = Where the line crosses the y-axis
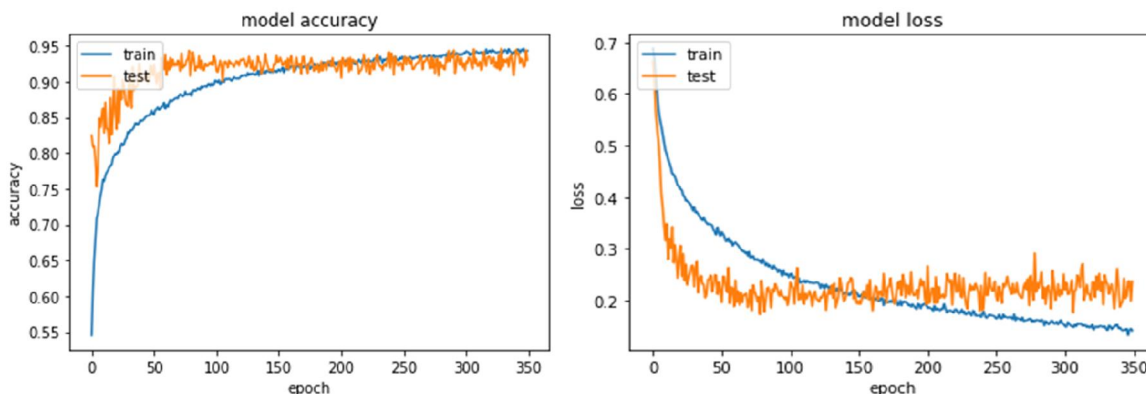
From the above formula, we try to find the value of x and y that every value of x has a corresponding value of y in it if it is continuous. The reason for this is linear regression is always continuous. The output of the linear regression is the value of the variable. The accuracy of the great fit is calculated by using the r squared method.

IV. OBJECTIVE
- To predict the house price according to the area.
- To calculate house price depending upon surrounding environment like railway station, hospital area, ATM, college, banks so, customers can purchase flats with full facilities.
- To provide a comparison of house pricing to customers.
- To suggest builder price prediction for the new constructions.
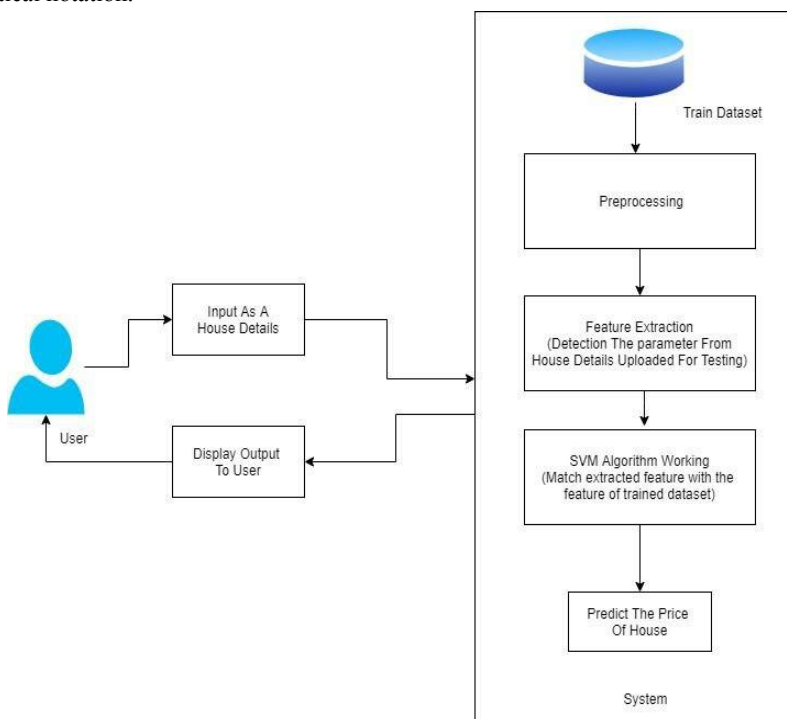
V. TRAINING & TESTING
- Train accuracy: 0.9371
- Test accuracy: 0.8894

Why we Select Linear Regression

The regression model predicts a continuous variable and the linear Regression predicted value is continuous it also removes the outliers there for our model will perform in a better way.

Linear regression is easily comprehensive and transparent. They can be understood very easily because it is represented by simple mathematical notation.

VI. CONCLUSION

The paper studies the SVM algorithm in machine learning for house price prediction. It takes data from the user and process it and classify it using pre-available data and uses various classification algorithm and classifies data and predict the accurate price of the property. It then confirms that accurate prediction result also depends on the population and quality of the training dataset. Results obtained earlier through SVM vs optimized SVM were then evaluated. we can use this theory and can implement it to find the predicted value of the house by considering various factors.

VII. ACKNOWLEDGMENT

REFERENCES

[1]. www.google.com

[2]. https://www.sciencedirect.com/science/article/abs/pii/S0030402613012515?via%3Dihub

[3]. https://www.turcomat.org/index.php/turkbilmat/article/download/6435/5333

[4]. https://towardsdatascience.com/a-data-science-web-app-to-predict-real-estate-price-d2366df2a4fd

[5]. An Optimized Support Vector Machine (SVM) based on Particle Swarm Optimization (PSO) for Cryptocurrency Forecasting nor Aziza Hiram, Amelia Rita Hani Ismail, Faisal Saeed

[6]. PropTech for Proactive Pricing of Houses in Classified Advertisements in the Indian Real Estate Market Sayan Putatunda Member, IEEE