



PROJECT REPORT

Submitted By

Aswini Arun Patil

NAME OF THE PROJECT

Housing Price Prediction

Acknowledgement:

- Primarily I would like to thank God to being able to complete this project with success. Then I would like to express my special thanks of gratitude to my SME,
- And I am thankful I am part of flip rob technology of employee, who given me the golden opportunity to do this wonderful project on the given topic which is also help me in doing a lot of research and I came to know about so many new things, I am really thankful to flip robo.

DATE:24/06/2022

ASWINI A. PATIL
Data Science course
Institute: Data trained education
Internship: Flip Robo technology
@Bangalore

Abstract:

House price forecasting is an important topic of real estate. The literature attempts to derive useful knowledge from historical data of property markets. Machine learning techniques are applied to analyze historical property transactions in India to discover useful models for house buyers and sellers. Revealed is the high discrepancy between house prices in the most expensive and most affordable suburbs in the city of Goa

Introduction:

Thousands of houses are sold every day. 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.

AIM and IMPORTANCE:

These are the Parameters on which we will evaluate ourselves-

- Create an effective price prediction model
- Validate the model's prediction accuracy
- Identify the important home price attributes which feed the model's predictive power

Problems facing during buying a house:

- 1) Buying a house is a stressful thing.
- 2) Buyers are generally not aware of factors that influence the house prices.
- 3) Many problems are faced during buying a house.
- 4) Hence real estate agents are trusted with the communication between buyers and sellers as well as laying down a legal contract for the transfer. This just creates a middle man and increases the cost of houses.

Goals of the Study

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

Paper Organization

This paper is organized as follows: in the next section, section 2, we examine studies related to our work from scientific journals.

In section 3, we go through data preparation including data cleaning, outlier removal, and feature engineering.

Next in section 4, we discuss the type of our problem and the type of machine-learning prediction that should be applied; we also list the prediction techniques that will be used.

In section 5, we choose algorithms to implement the techniques in section 4; we build models based on these algorithms; we also train and test each model.

In section 6, we analyze and compare the results we got from section 5 and conclude the paper.

How It Works:

- Collecting Data: First step was to collect data we collected data from different source and merge them together to from our training data set.
- Then trained the model using machine learning algorithm which in this case is multiple linear regression.
- Based on the generated graphs we predict the cost of the house

Features of a House:

- Year Built
- Total Basement in sqr.Ft.
- Lot Area
- Floor Area
- Overall Condition
- Lot Frontage
- Garage details
- Detail about fireplace

We will have two type of data:

- Training Data – This data will contain the information related to the year sold and sale price of house.
- Test Data – It will contain all the information about a house. And, based on all the given information, Logistic Regression Algorithm will predict the selling price of a house

Importing Data & EDA:

As I said before, we are going to work with the house price dataset that contains various features and information about the house and its sale price. Using the '*reads*' function provided by the Pandas package, we can import the data into our python environment. After importing the data, we can use the 'head' function to get a glimpse of our dataset.

Data Visualization:

In this process, we are going to produce three different types of charts including heatmap, scatter plot, and a distribution plot.

Heatmap:

Heatmaps are very useful to find relations between two variables in a dataset. Heatmap can be easily produced using the '*heatmap*' function provided by the seaborn package in python.

Literature Review:

In this section, we look at five recent studies that are related to our topic and see how models were built and what results were achieved in these studies.

Stock Market Prediction Using Bayesian-Regularized Neural Networks:

The model proposed in the study is built using Bayesian regularized neural network. The weights of this type of networks are given a probabilistic nature. This allows the network to penalize very complex models (with many hidden layers) in an automatic manner. This in turn will reduce the overfitting of the model.

The model consists of a feedforward neural network which has three layers: an input layer, one hidden layer, and an output

layer. The author chose the number of neurons in the hidden layer based on experimental methods. The input data of the model is normalized to be between -1 and 1, and this operation is reversed for the output so the predicted price appears in the appropriate scale.

Stock Market Prediction Using A

Machine Learning Model:

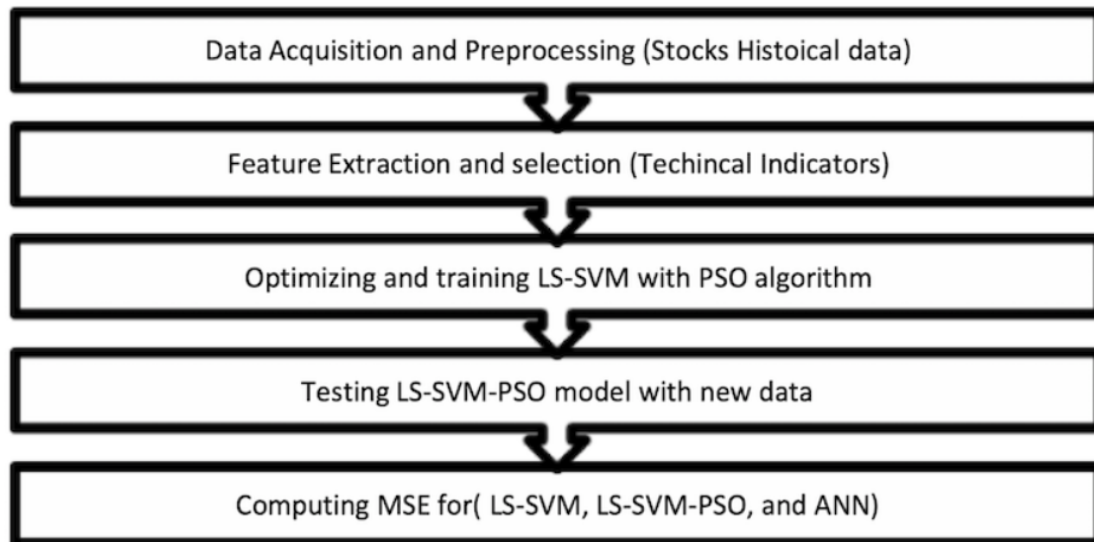
a system was proposed to predict daily stock market prices. The system combines particle swarm optimization (PSO) and least square support vector machine (LS-SVM), where PSO was used to optimize LV-SVM.

The authors claim that in most cases, artificial neural networks (ANNs) are subject to the overfitting problem. They state that support vector machines algorithm (SVM) was developed as an alternative that doesn't suffer from overfitting. They attribute this advantage to SVMs being based on the solid foundations of VC-theory. They further elaborate that LS-SVM method was reformulation of traditional SVM method that uses a regularized least squares function with equality constraints to obtain a linear system that satisfies Karush-Kuhn-Tucker conditions for getting an optimal solution.

The authors describe PSO as a popular evolutionary optimization method that was inspired by organism social behavior like bird flocking. They used it to find the optimal parameters for LS-SVM. These parameters are the cost penalty CC , kernel parameter γ , and insensitive loss function ϵ .

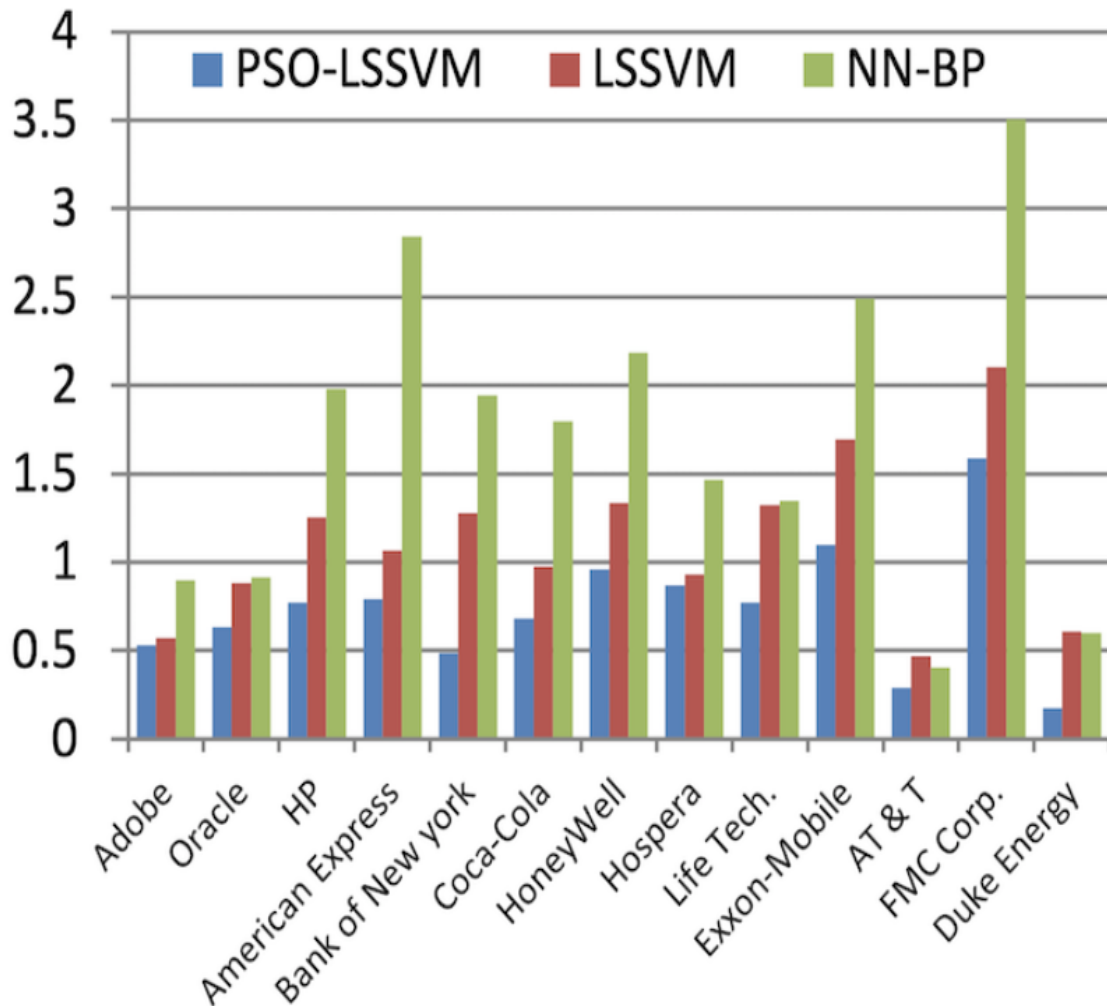
The model proposed in the study was based on the analysis of historical data and technical financial indicators and using LS-SVM optimized by PSO to predict future daily stock prices. The

model input was six vectors representing the historical data and the technical financial indicators. The model output was the future price. The model used is represented in Figure 2.



Regarding the technical financial indicators, five were derived from the raw data: relative strength index (RSI), money flow index (MFI), exponential moving average (EMA), stochastic oscillator (SO), and moving average convergence/divergence (MACD). These indicators are known in the domain of stock market.

The results obtained in the study showed that LS-SVM-PSO model had the best performance. Figure 3 shows a comparison between the mean square error (MSE) of the three models for the stocks of many companies.



House Price Prediction Using Multilevel Model and Neural Networks:

The multilevel model integrates the micro-level that specifies the relationships between houses within a given neighborhoods, and the macro-level equation which specifies the relationships between neighborhood's. The hedonic price model is a model that estimates house prices using some attributes such as the number of bedrooms in the house, the size of the house, etc.

Imputing Null Values:

With data this large, it is not surprising that there are a lot of missing values in the cells. In order to effectively train our model we build, we must first deal with the missing values. There are missing values for both numerical and categorical data. We will see how to deal with both.

For numerical imputing, we would typically fill the missing values with a measure like median, mean, or mode. For categorical imputing, I chose to fill the missing values with the most common term that appeared from the entire column.

Places Where NaN Means Something:

If you look at the data description file provided, you will see that for some categories, NaN actually means something. This means that if a value is NaN, the house might not have that certain attribute, which will affect the price of the house. Therefore, it is better to not drop, but fill in the null cell with a value called "None" which serves as its own category.

Data Preparation:

Ok, now that we have dealt with all the missing values, it looks like it's time for some feature engineering, the second part of our data preprocessing. We need to create feature vectors in order to get the data ready to be fed into our model as training data. This requires us to convert the categorical values into representative numbers.

It appears that the target, Sale Price, is very skewed and a transformation like a logarithm would make it more normally distributed. Machine Learning models tend to work much better

with normally distributed targets, rather than greatly skewed targets. By transforming the prices, we can boost model performance.

Model Building:

Now that we've preprocessed and explored our data, we have a much better understanding of the type of data that we're dealing with. Now, we can begin to build and test different models for regression to predict the Sale Price of each house.

Usages of housing price prediction:

Predicting house prices can help to determine the selling price of a house of a particular region and can help people to find the right time to buy a home.

Understanding the Client and their Problem:

A benefit to this study is that we can have two clients at the same time! (Think of being a divorce lawyer for both interested parties) However, in this case, we can have both clients with no conflict of interest!

Client House buyer: This client wants to find their next dream home with a reasonable price tag. They have their locations of interest ready. Now, they want to know if the house price matches the house value. With this study, they can understand which features (ex. Number of bathrooms, location, etc.) influence the final price of the house. If all matches, they can ensure that they are getting a fair price.

Client House seller: Think of the average house-flipper. This client wants to take advantage of the features that influence a house price the most. They typically want to buy a house at a low price and invest on the features that will give the highest return. For example, buying a house at a good location but small square footage. The client will invest on making rooms at a small cost to get a large return.

Conclusion:

- So, our Aim is achieved as we have successfully ticked all our parameters as mentioned in our Aim Column. It is seen that circle rate is the most effective attribute in predicting the house price.
- As a recommendation, we advise to use this model (or a version of it trained with more recent data) by people who want to buy a house in the area covered by the dataset to have an idea about the actual price. The model can be used also with datasets that cover different cities and areas provided that they contain the same features. We also suggest that people take into consideration the features that were deemed as most important as seen in the previous section; this might help them estimate the house price better.

