Location Based Real Estate Price Prediction

Kushagra Singh
Department Of Computer
Science & Engineering
PES University
Bangalore, India
skushagra11@gmail.com

Nikhil Raju Mohite
Department Of Computer
Science & Engineering
PES University
Bangalore, India
nikhilmohitelhs@gmail.com

Jeffrey S Varghese
Department Of Computer
Science & Engineering
PES University
Bangalore, India
jeff071604@gmail.com

Guide: Dr. Jayashree R
Department Of Computer
Science & Engineering
PES University
Bangalore, India
jayashree@pes.edu

***Abstract*- Real estate is the least transparent industry in our ecosystem. Housing prices keep changing day in and day out and sometimes are hyped rather than being based on valuation. Predicting housing prices which are realistic and depend on real factors is the main crux of our research project. We aim to make our evaluations based on parameters on which the price of a property is dependent on. Various Machine Learning Algorithm's such as linear regression, Lasso, Decision Tree Regressor, Random Forest Regressor and XGBoost Regressor are used and the model with the highest accuracy is used to predict the price of different Metropolitan Cities of India.**

Keywords—Regression, Machine Learning, Price Prediction

I. INTRODUCTION

In today's every prospering world, investing money in the right places is very crucial. As the stock market rises, Real Estate is and will always be one of the best choices to invest money in. Housing prices keep fluctuating throughout the year and there is a requirement to obtain realistic prices for different properties.
This project is designed keeping in mind that the people who look for new properties and it reduces the time and effort they need to put in to look for a property and decide upon whether to invest in it or not. People new to metropolitan cities of India or people who have insufficient time to look up the right property can use this model to narrow down the property based of on their requirements and budget. This project eliminates the middle men as they reduce the broker and commission hassle encountered by the buyer.

Project uses different regression techniques to predict the property prices which are dependent on so many different independent parameters.
Language used for implementing is Python and the regression techniques used are- - linear regression, Lasso, Decision Tree Regressor, Random Forest Regressor and XGBoost Regressor. Grid search cv is used to obtain the model with the best accuracy out of linear regression, lasso and decision tree regressor.

II. RELEVANT WORKS

Property prices of metropolitan cities are dependent on various features like, area of plot, no. of bedrooms, geographical location , resale value etc. A lot of research has been done to find the best
properties at the best price for the customers. Numerous technologies and techniques have been used for the same.

Mohammadzaman Zamani and Hansen Andrew Schwartz, [1] have shown that language in twitter is predictive of foreclosure rates and price fluctuations are observed. They used a residualised control approach to combine language features with traditional variables which lead to more accurate models.

Manjula, Shubham Jain, Sharad Srivastava and Pranav Rajiv Kher [2] presents various important features to use when forecasting property prices with good precision using a regression model.

Alisha Kuvalekar and Shivani Manchewar andSidhika Mahadik [3] have taken into

consideration features like air quality and crime rate into their dataset which can influence people's decisions while purchasing a property. The Decision tree machine learning algorithm is used to construct a prediction model to predict potential selling prices for any real estate property.

Nihar Bhagat and Ankit Mohokar and Shreyash Mane [4] This paper analyses previous market trends and price ranges, to predict future prices using linear regression algorithims, in an efficient manner

Varma [5] created a model which uses real-time data to get precise real-world predictions using Google maps.

III. System Design.

Data Set/Data Collection

Data was collected for different metropolitan cities namely- Bangalore, Chennai, Delhi, Hyderabad, Kolkata and Mumbai. The data sets has features like Carpet area, No. of Bedrooms, Location (Latitude & Longitude), resale value, Lift availability etc. The price of a property is dependent on these features. All ML models require data collection and this is the first step to building an efficient and accurate model.

Data Preprocessing

Data is cleaned in this process. Categorical data is changed to encoded numerical data. Outliers and missing data is removed. Important features which affect the price is used and the other features are dropped.

Training the model

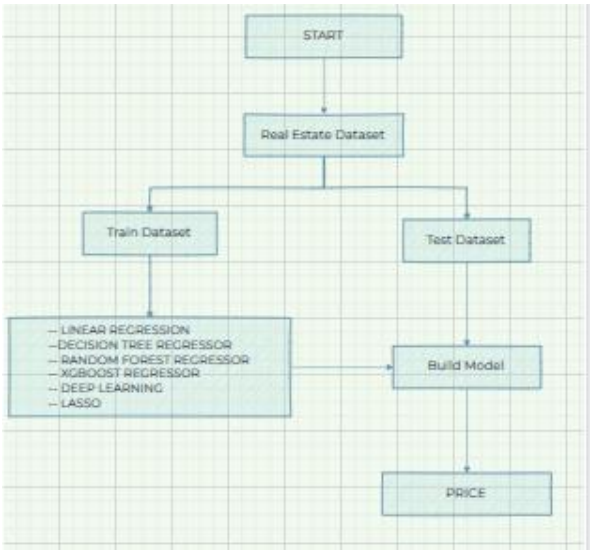
The data set is divided into training and testing data. Training data is used to train the model using different regression models namely- linear regression, Lasso, Decision Tree Regressor, Random Forest Regressor and XGBoost Regressor. Accuracy of each model is noted and the best model is chosen for the price prediction.

Testing/Evaluating

The best model is chosen and tested using the testing data from the data set, to predict the prices of the properties.

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper; use the scroll down window on the left of the MS Word Formatting toolbar.

IV. Flowchart



V. Algorithms Used

Linear Regression

It is an ML model based on supervised learning. It predicts the value of price (y) which is the dependent variable based on the independent variables(x) which are the different features from the data set. It finds a linear relationship b/w the prices and the features

Lasso

Decision Tree Regressor

This model is used to create a model which predicts the value of the target variable which is the price of the property by learning different decision rules obtained from the training data set. Data is trained in the form of a Tree.

Random Forest Regressor

it is a powerful technique which uses ensemble learning method for regression. It predicts results using multiple algorithms (more than 1) by merging multiple decision trees to ensure high efficiency and accuracy

XGBoost Regressor

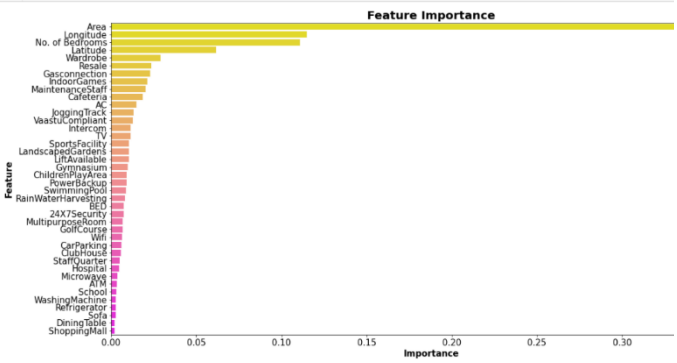
Model is trained in an iterative manner. New trees are added to remove the error of the previous trees and minimizes the loss. It basically learns from the previous trees and improves upon the accuracy.

VI. Implementation and Results

Data was preprocessed by removing outliers. Rows with missing data has been removed. Categorical data -location was encoded using one hot encoding method.

Latitude and Longitude features were added to improve the accuracy of the predictions.

Data visualization is done to obtain the features which have a higher impact on the prices. Other features have been dropped from the data set.



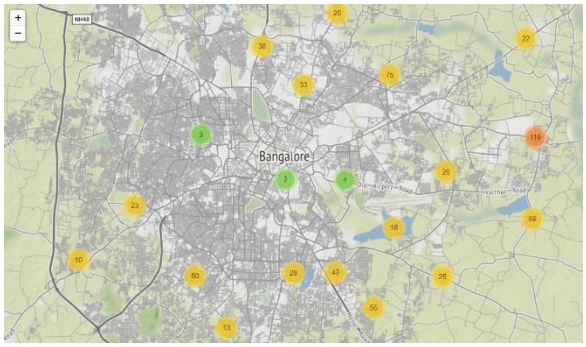
The above graph shows the feature importance of the merged data set.

The different models mentioned above are trained and the accuracies are obtained.
The results are as follows:

	model	best_score	best_params
0	linear_regression	0.817875	{'normalize': True}
1	lasso	0.817256	{'alpha': 2, 'selection': 'random'}
2	decision_tree	0.892889	{'criterion': 'friedman_mse', 'splitter': 'best'}

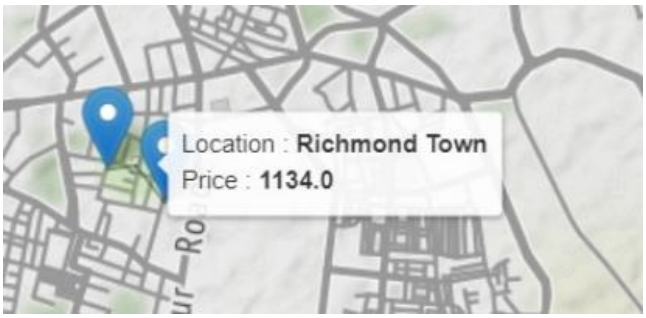
```
Algorithm - Random Forest Regressor
MAE - 11.740278512966535 MSE - 1071.7395999650169
Accuracy - 96.0%

Algorithm - XGBoost Regressor
MAE - 9.728143989127737 MSE - 721.5650573890657
Accuracy - 97.0%
```

Predicting prices:



Map feature has been included to show the no. of properties in a particular location.

And it also shows the predicted prices of properties of the pinpointed location.



Predicted prices in Richmond Town

Predicted prices of a plot in Nayandahalli

The test data set(20% of the data set) is used to test the model for its accuracy, and the predicted prices are compared to the actual prices.

```
1  predict_price(3340,4,0,1,12.265594,76.646540)
```

```
{'linear regression': 366.2125003298742,
 'decisiontree': 270.2222218055556,
 'Random forest': 299.66,
 'XGBregressor': 300.0}
```

Different Models predicting the price of a particular property in JP Nagar.

VII. Future Scope

In the Future we would like to deploy this project on a website. We have made a website titled “Enigma Estate” and we will integrate the ML model with the website which was built using a react app, html and java script. This will create a user friendly interface for the end user to interact with our model so that they can find a home of their choice is the easiest way possible.

We also plan to improve the model by implementing deep learning using Artificial Neural Networks.

VIII. Conclusion

Various ML Algorithms were used to predict the price of the property based on different features. Grid Search CV is used to obtain the best model with highest accuracy among linear regression, Lasso and Decision tree algorithms. Out of the three, decision tree had the

highest accuracy. Linear regression model gave an accuracy of 81.79%, Lasso gave an accuracy of 81.72% and Decision tree gave an accuracy of 89.28%. In addition Random forest regression and XGBoost regression techniques were also applied which gave accuracies 96% and 97% respectively. Latitude and Longitude of the places are included in the data set to improve its efficiency and accuracy.

IX. REFERENCES

[1] Mohammadzaman Zamani and Hansen Andrew Schwartz, "Using Twitter Language to Predict the Real Estate Market" Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics Valencia, Spain, April 3-7, 2017.

[2] R Manjula, Shubham Jain, Sharad Srivastava and Pranav Rajiv Kher, "Real estate value prediction using multivariate regression models" 14th ICSET-2017 IOP Publishing IOP Conf. Series: Materials Science and Engineering 263 (2017)

[3] Alisha Kuvalekar and Shivani Manchewar and Sidhika Mahadik, "HOUSE PRICE FORECASTING USING MACHINE LEARNING," Proceedings of the 3rd International Conference on Advances in Science & Technology (ICAST) 2020

[4] Nihar Bhagat and Ankit Mohokar and Shreyash Mane, "House Price Forecasting using Data Mining," International Journal of Computer Applications (0975 – 8887) Volume 152 – No.2, October 2016.

[5] 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.