

Group_ID : ITB-03

PROJECT TITLE: HOUSE PRICE PREDICTION

Student Names : Sushant Kulkarni

Shefin Shajit

Akshay Mohite

Guide: Prof. Swati Sinha

1. Abstract

With the ever-rising traffic to own a place, house or in general real estate, it is important to have the best tools at disposal which would guide the buyers about where to put their money into. As home-buying season is shifting into high gear, buyers prefer to make smart investments and purchases which in turn provide profits or prove to a steal for the price. Selling and purchasing a house is a section which invites lot of statistics and data. We will be using these numbers to make a prediction of the price of a house in accordance with the existence of amenities such as parking, pool, school, etc.; or not.

Our work is based on a set of data containing house prices of places in Mumbai along with the major parameters affecting the price such as area, location, swimming pool, etc obtained from open web source Kaggle Inc. and predict the price under the parameters. The model implemented incorporates ensemble learning i.e., a combination of machine learning algorithms instead of relying on a single algorithm for improved predictions. The ensemble model incorporated in our system (weighted average of Decision Tree, Linear Regression, and K-Nearest Neighbour) brings an added advantage over using solo algorithms in the process of obtaining minimum error and trying to get the predictions as accurate as possible.

2. Introduction

Housing being one of the basic needs of human, accounts to high percentages of national transactions per year. The real estate sector is a major sector influencing India's economy. In India, about 15 percent of the total jobs are generated by the real estate sector. Since property prices rarely decrease rapidly, it is a major contender for investment. The property prices depend on various intrinsic and extrinsic factors which directly or indirectly affect the long-term price values.

A common issue that normally emerges is the measurement of asset values for investment purposes. The heterogeneity exhibited by the real property values is given due weightage which derives a pattern of variation in the values of properties over a period. Careful attention must be given to the dynamics of various factors affecting the housing prices for full understanding on our research on predicting housing prices.

In 2021, the sector must adopt innovative ways of dealing with the requirements. While houses will continue to be sold, they will now be done with creative disruption. The reinvention will include technology playing a lead role in meeting altered norms being considered by home buyers. There are three factors that influence the price of a house which include physical conditions, concept and location. Physical condition are properties possessed by a house that can be observed by human senses, including the size of the house, the number of bedrooms, the availability of kitchen and garage, the availability of garden, the area of land and buildings, distance from school and stations, the age of the house. Location is an important factor in shaping the price of a house. This is because the location determines the ease of access to public facilities, such as schools, campus, hospitals and health centers as well as family recreation facilities such as malls, culinary tours, or even beautiful scenery.

We will be evaluating the data and statistics to provide the buyer with a price depending on various factors such as location, physical condition, house details, etc. This prediction will be implemented with the use of machine learning.

3. Problem Statement

To develop a system that would help the user to estimate the price for a particular house considering the real-life factors such as location of the house and the connectivity, leisure services, schools etc. that affect the price of the house along with the features of house and the amenities available.

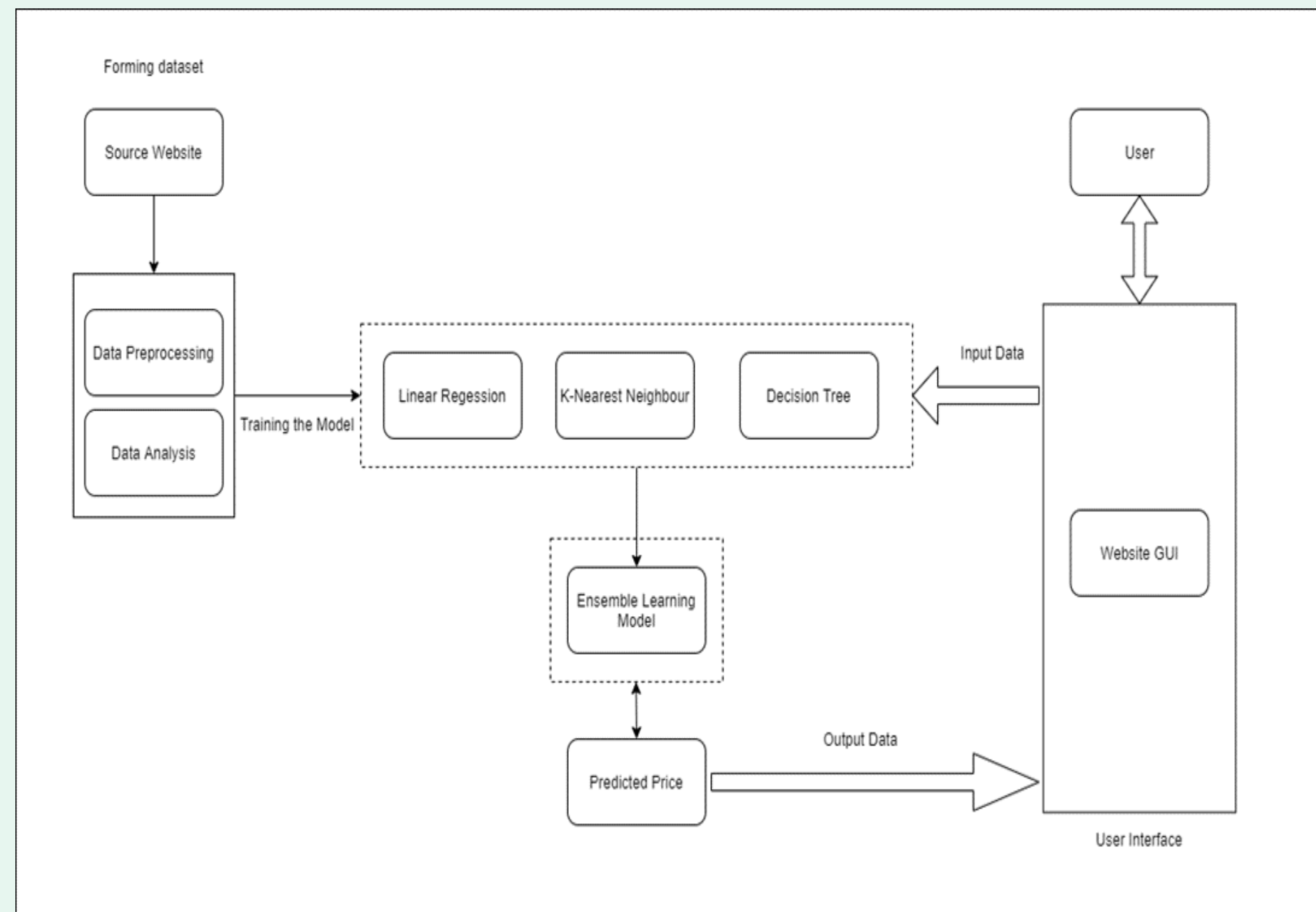
A fair share of India's economic condition affects property prices in the long run. This scenario calls for technology to bring out the best ways to guide the customer's investment decisions. It is important to have the best tools at disposal which would guide the buyers about where to put their money into. The main aim of the project is to predict the house pricing for buyers with respect to various factors and priorities affecting the prices. By analyzing current market trends and price ranges, prices will be predicted.

Objectives

- To study the current literature available for House price prediction.
- To identify the ML models suitable for accurately predicting House price.
- Design a website which accepts customers specifications and then combines the application with the trained model.
- To provide guiding information to the companies and customers in the sector and to contribute to the literature.

4. Proposed System

Flowchart



Methodology

Ensemble Learning:

Ensemble learning helps improve machine learning results by combining several models. This approach allows the production of better predictive performance compared to a single model. Ensemble methods are meta-algorithms that combine several machine learning techniques into one predictive model in order to **decrease variance** (bagging), **bias** (boosting), or **improve predictions** (stacking).

We reached a conclusion from our regression model results and literature survey that a hybrid model using ensemble learning needed to be deployed to gain superior predictive analysis and to minimize the error margin.

In our case we deployed various models, tested them and noted the error metrics for each model. These metrics helped us determine the 3 best algorithms matching our requirements and giving minimal error. These 3 algorithms/models are:

- Linear Regression
- K-Nearest Neighbors
- Decision Tree

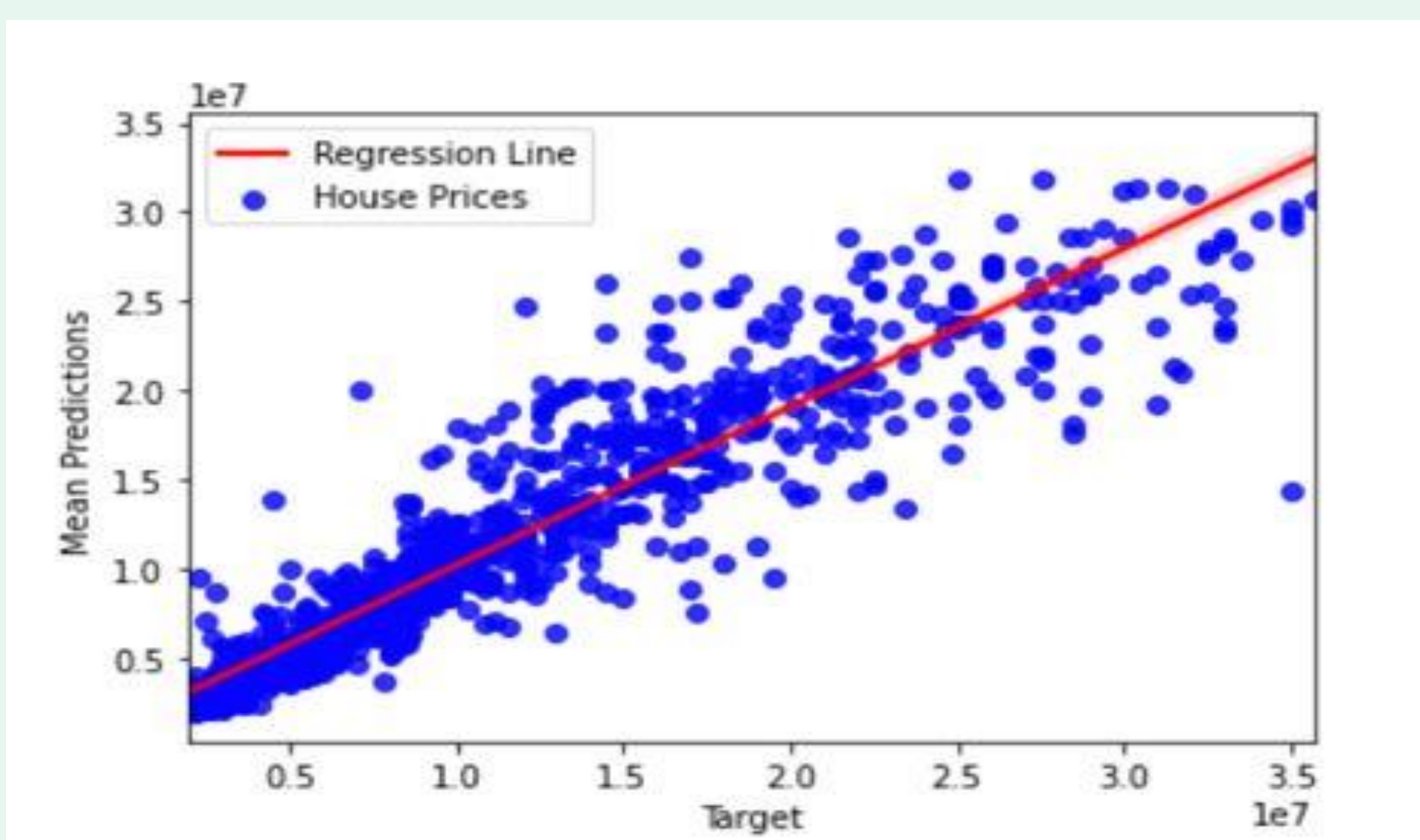
All the models and algorithms were implemented using the TensorFlow library.

Weighted averaging :

The ways of combining models depend exclusively on the features of the dataset. Simple averaging of results of ensemble members (models used in the combination) implies equal contribution to the final prediction. Weighted averaging is an extension to simple averaging which deals with the limitation of simple averaging when some models are known to perform better or much worse than the others. A weighted ensemble is a model where the weight contribution of each model participating (ensemble members) is computed according to the efficiency of predictions.

MODEL	MAE	MSE	RMSE	R2	W
Linear Reg.	2229371.20	9.2E+10	3040960.80	0.83	0.1
KNN	1948884.96	8.9E+10	2987929.06	0.84	0.3
Decision Tree	1851282.60	1.0E+09	3183553.80	0.81	0.6
Ensemble	1658701.38	7.3E+10	2694486.53	0.87	-

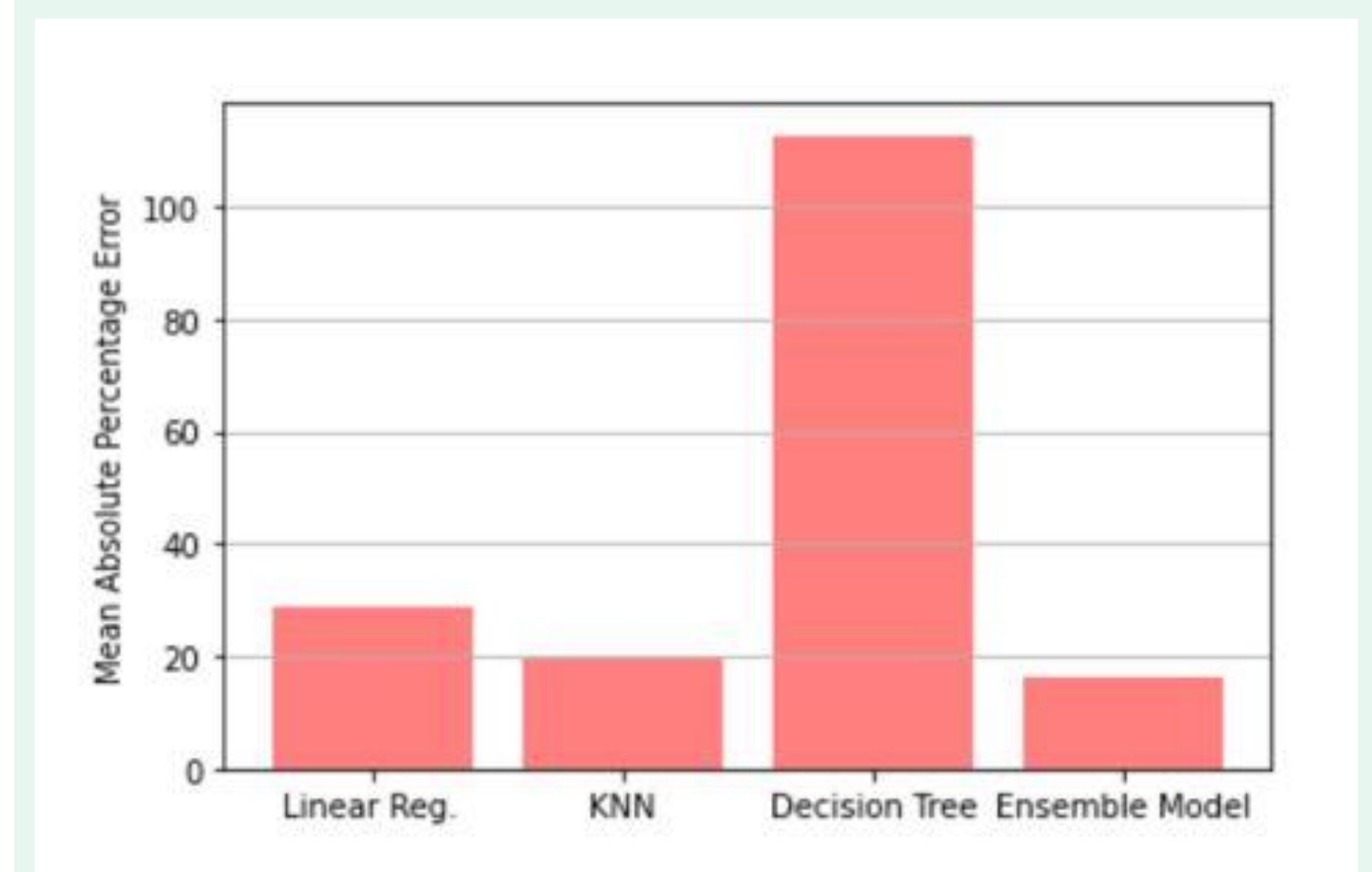
Error Metrics Comparison



Ensemble Model Scatter Plot

Experimental Results

The weighted average ensemble model performs substantially better than the individual models. The trained model attains an accuracy of 84%. The comparison of Mean Absolute Percentage Errors (MAPE) of the models used and the ensemble model is given below in the form of a bar chart



MAPE Comparison

On comparing the various models, we find that the ensemble model works best with the lowest Mean Absolute Percentage Error (MAPE).

5. References

1. Neelam Shinde, Kiran Gawande, "Valuation of House Prices using Predictive Techniques", International Journal of Advances in Electronics and Computer Science – 2018.
2. Joao Mendes Moreira, Alipio Mario Jorge, Carlos Soares, Jorge Freire de Sousa, "Ensemble Approaches for Regression: A Survey", ACM Computing Surveys – 2012.
3. Yashraj Garud, Hemanshu Vispute, Nayan Bisai, and Prof. Madhu Nashipudimath, "Housing Price Prediction using Machine Learning", International Research Journal of Engineering and Technology (IRJET) – 2020.
4. Prof. Pradnya Patil, Darshil Shah, Harshad Rajput, Jay Chheda, "House Price Prediction Using Machine Learning and RPA", International Research Journal of Engineering and Technology (IRJET) – 2020.K. Elissa, "Title of paper if known," unpublished.
5. Alisha Kuvalekar, Sidhika Mahadik, Shivani Manchewar, Shila Jawale, "House Price Forecasting using Machine Learning", 3rd International Conference on Advances in Science & Technology (ICAST) – 2020.
6. Zhongyuan Han, Jiaming Gao, Huilin Sun, Ruifeng Liu, Chengzhe Huang, Leilei Kong, Haoliang Qi, "An Ensemble Learning-based model for Classification of Insincere Question", FIRE – 2019.
7. Ayush Varma, Sagar Doshi, Abhijit Sarma, Rohini Nair, "House Price Prediction Using Machine Learning and Neural Networks ", Second International Conference on Inventive Communication and Computational Technologies (ICICCT) – 2018.
8. P. Durganjali; M. Vani Pujitha, "House Resale Price Prediction Using Classification Algorithms", International Conference on Smart Structures and Systems (ICSSS) – 2019.
9. CH. Raga Madhuri; G. Anuradha; M. Vani Pujitha, "House Price Prediction Using Regression Techniques: A Comparative Study", International Conference on Smart Structures and Systems (ICSSS) – 2019.
10. A. Adair, J. Berry, W. McGreal, "Hedonic modeling, housing submarkets and residential valuation", Journal of Property Research – 1996.
11. O. Bin," A prediction comparison of housing sales prices by parametric versus semi-parametric regressions", Journal of Housing Economics – 2004.
12. T. Kauko, P. Hooimeijer, J. Hakfoort, "Capturing housing market segmentation: An alternative approach based on neural network modeling", Housing Studies – 2002.
13. Li Li, Kai-Hsuan Chu, "Prediction of Real Estate Price Variation Based on Economic Parameters ", IEEE International Conference on Applied System Innovation – 2017.
14. G. Naga Satish, Ch. V. Raghavendran, M.D.Sugnana Rao, Ch. Srinivasulu, "House Price Prediction Using Machine Learning ", International Journal of Innovative Technology and Exploring Engineering (IJITEE) – 2019.
15. Adyan Nur Alfiyatin, Hilman Taufiq, Ruth Ema Febrita, Wayan Firdaus Mahmudy, "Modeling House Price Prediction using Regression Analysis and Particle Swarm Optimization", (IJACSA) International Journal of Advanced Computer Science and Applications.
16. Ayşe SOY TEMÜR, Melek AKGÜN2, Günay TEMÜR, "Predicting housing sales in Turkey using ARIMA, LSTM, and Hybrid models", Journal of Business Economics and Management ISSN – 2019.