IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Chatbot Enhanced House Price Prediction Using Ensemble Technique

Sarang Kadakia, Department of Information Technology, Dwarkadas J. Sanghvi College Of Engineering, Mumbai,

Ishan Desai, Department of Information Technology, Dwarkadas J. Sanghvi College Of Engineering, Mumbai,

Harshit Jain, Department of Electronics & Telecommunication, Dwarkadas J. Sanghvi College Of Engineering,Mumbai,

Prof. Neha Katre, Department of Information Technology, Dwarkadas J. Sanghvi College Of Engineering, Mumbai,

ABSTRACT

The rapid culmination of the real estate market has made accurate house price prediction an essential tool for homebuyers, sellers, and investors. The comprehensive house price prediction software uses cutting- edge machine learning algorithms to display property prices. To produce accurate price projections, the platform combines a various number of models to maximize the accuracy of the software. People get real-time price predictions which are calculated by formulating around 20 parameters.

Keywords: *Machine Learning, ADAboost, SVM, Random Forest, Chatbot*

1.INTRODUCTION

Homebuyers, sellers, and investors have been facing difficulty estimating a property's value due to the volatility and unpredictability of the real estate market. However, developments in data science and machine learning have made it possible to create house price prediction softwares, which could fundamentally alter how to asses and anticipate the value of real estate. The idea of developing a house price prediction software is presented in this introduction, which also emphasizes its potential advantages and the significance it bearsfor various real estate industry players [1].

Software predicts house prices by making use of complex algorithms and a dataset of specific region to deliver precise estimations of property values [2]. This software produces insightful analysis of the property that support decision-making by a client looking at various elements including crime rate (per capita) by town, average number of rooms per residence, pupil-teacher ratio by town, etc. This software can be an important tool for anyone wishing to make an informed purchase allowing investors and buyers trying to find lucrative prospects for a specific region. The software utilizes various machine learning techniques like ADABoost and Regression models with Stacking as final ensemble technique to find patterns and relationships in the data [3].

By practising on previous sales data and mixing diverse factors, the programme can learn to provide forecasts that are in accordance with market trends and accurately reflect property prices. Furthermore, the programme can constantly update its models with the most recent data, guaranteeing that the predictions are relevant and up to current.

One of the primary advantages of house price prediction software is its ability to aid decision- making for people involved in real estate transactions. Homebuyers can use the software to appraise properties, determine their value, and make informed offers. The software's insights can assist investors in identifying successful investment possibilities and forecasting future market trends [4**].**

2.LITERATURE REVIEW

"House Price Prediction" has been one of the most popular topic of study in the domains of machine learning and data science. Many machine learning techniques and methodologies have been developed to estimate property prices with high accuracy [5]. This literature review underscores on the current research of housing price prediction by implementing applications of various machine learning techniques.

2.1 Regression Techniques:

Regression models are required to forecast house prices. The most common techniques includes "Linear Regression", "Multiple Regression" and "Polynomial Regression". To capture complicated associations and boost prediction accuracy, advanced regression methods including Support Vector Regression (SVR) and Random Forest Regression have also been used [2].

2.2 Ensemble Methods:

To improve prediction accuracy, ensemble approaches bring together a variety of machine learning models [6]. Common ensemble approaches include bagging, boosting, and stacking. According to studies, ensembling models like AdaBoost, Gradient Boosting, and XGBoost can significantly increase the generalisation and resilience of housing price prediction models.

2.3 Model Evaluation and Comparison:

To determine which machine learning techniques and algorithms provide the most reliable and accurate prediction models, comparative evaluations of these methods are crucial. House price prediction model performance is frequently assessed and compared using metrics like mean squared error (MSE), root mean squared error (RMSE), standard deviation (SD), mean(scores) [5].

3.DATA SET

The dataset incorporated in the system is taken from the StatLib library being maintained at Carnegie Mellon University. It has statistical information for house prices and features from 506 unique locations of Boston. It consists of 7084 records with 14 parameters that have the possibility of affecting the property prices. However, out of these 14 parameters, only 13 were chosen (per capita crime rate, proportion of residential landzoned for lots over 25,000 sq.ft., proportion of non-retail business acres, Charles River dummyvariable, nitric oxides concentration, rooms per dwelling, owner-occupied units built prior to 1940, distances to five Boston employment centers, accessibility to radial highways, property-tax rate per $10,000, pupil- teacher ratioby town, 1000(Bk - 0.63)^2 where Bk is the proportion of blacks by town, % lower status of the population,) along with 1 added parameters (Owner-occupied homes in $1000's) which are bound to have a major effect on housing prices. The area is the total built-up area in square feet. Charles River dummy variable is a binary variable (= 1 if tract bounds river; 0 otherwise).

3.1 Data Pre-processing

Data Preprocessing is a major step in transforming the dataset into an efficient format. This includes replacing the missing values with the existing median values from the dataset. The model combines different attributes in order to enhance its overall efficiency (eg: TAX and RM attributes are combined together to form a new TAXRM attribute).

3.2 Data Analysis

Every single parameter in the dataset is analyzed with every other parameter to check the dependence and correlation using a heatmap. The correlation is measured in the range of -1 to +1 where a higher absolute score shows better correlation and the lower absolute score worse the relation. Below Fig.1 depicts the correlations among the 14 parameters which affect the house prices.

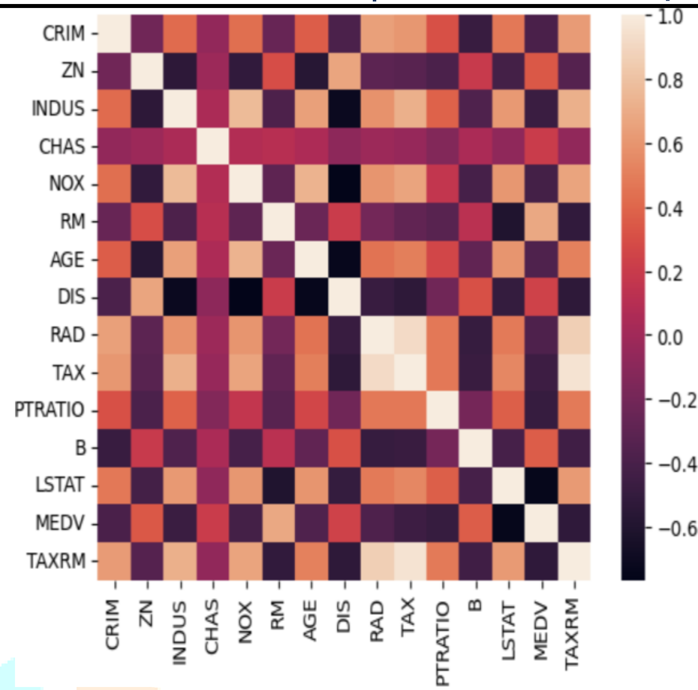

Fig.1. Correlation Matrix Heatmap

4.METHODLOGY

The next phase is training the model to assist us in forecasting house prices after data analysis and visualization [1]. It has been found out that mixture of algorithms or models performed better than a single algorithm on the data, i.e., delivered lower error values, throughout the process of constructing the model for training the data. Testing was done on a number of regression techniques including Linear Regression, Decision Tree, Random Forest, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Bagging, and Boosting, where Bagging, Boosting, and Random Forest were ensemble models in and of themselves [2]. Finally, the initial output of every individual algorithm is combined and used in ensemble learning technique to get the final output. The properties of the dataset serve as the sole determinant of the methods for merging models. In order to enhance the performance of models for tasks like classification, prediction, function approximation, etc., ensemble learning is primarily used. Stacking is an advanced version of ensemble learning where all sub-models equally contribute to create a new model on the basis of individual performances [7].

To create the train and test sets, we divided the data 80/20. We discovered three algorithms— AdaBoost Regressor, SVR, and RandomForest Regressor—to be the best fit after combining several models for ensemble learning. In comparison to other combinations, a stacking of the predictions from these three algorithms had the narrowest error range [3].

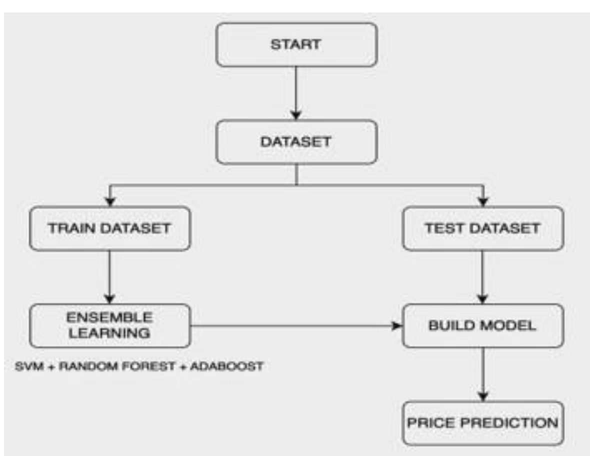

Fig.2. Workflow

4.1 AdaBoost Regressor

The Adaptive Boosting algorithm (AdaBoost) is one of the Boosting techniques used as Ensemble method in Machine Learning. Adaptive Boosting involves re-assigning the weights to each instance and assigning higher weights to wrongly classified instances. In supervised learning, Adaptive Boosting plays an important role in reducing the bias and variance. It works mainly on the principle that learners grow sequentially. Except for the first learner, every subsequent learner grows from the previously grown learner. In other words, the weak learner is converted into the strong learner. Fig.3 below shows the correlation between the actual and the predicted value using the AdaBoost Regressor model.

Model	Mean (Score)	SD (Score)	RMSE
AdaBoost	3.67627547	0.72154954	3.57157255

Table No.1

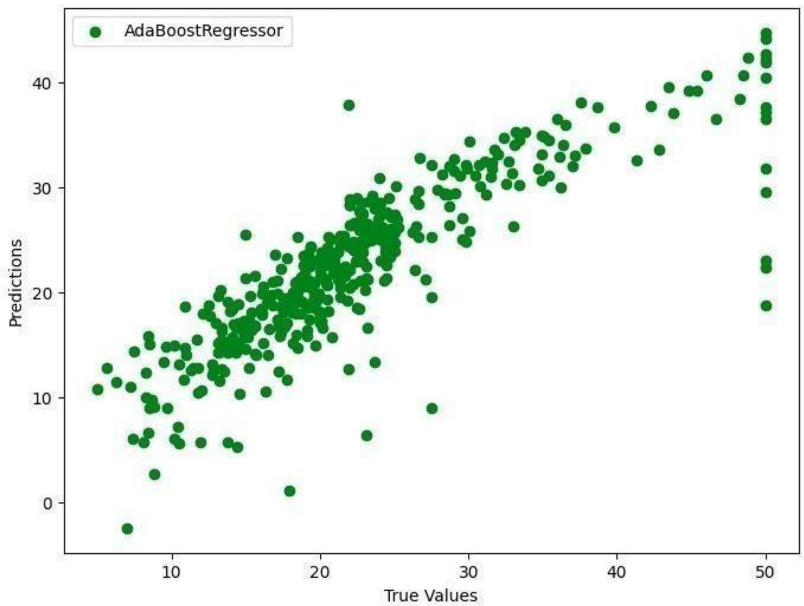

Fig.3. AdaBoost Scatter Plot

4.2 SVR (Support Vector Regression)

SVR stands for Support Vector Regression. Support vector regression is a machine learning algorithm that is mainly used in regression analysis. The main objective of support vector regression is to reduce the prediction error to a function that is close to the relationship of the input variables to the continuous target variable. The goal of the Support Vector Regressor is to find the hyperplane which is most suitable for the data point in the continuous space. The input variables are mapped to a high dimensional feature space and the hyperplane is found which maximizes the distance (distance) from the hyperplane to the nearest data point. This helps in minimizing the prediction error. As shown in Fig.4 below, the scatter-plot plot shows the relationship between the real and the predicted values using Support Vector Regressor Model.

Model	Mean (Score)	SD (Score)	RMSE
SVR	5.60671288	1.70961043	4.45040326

Table No.2

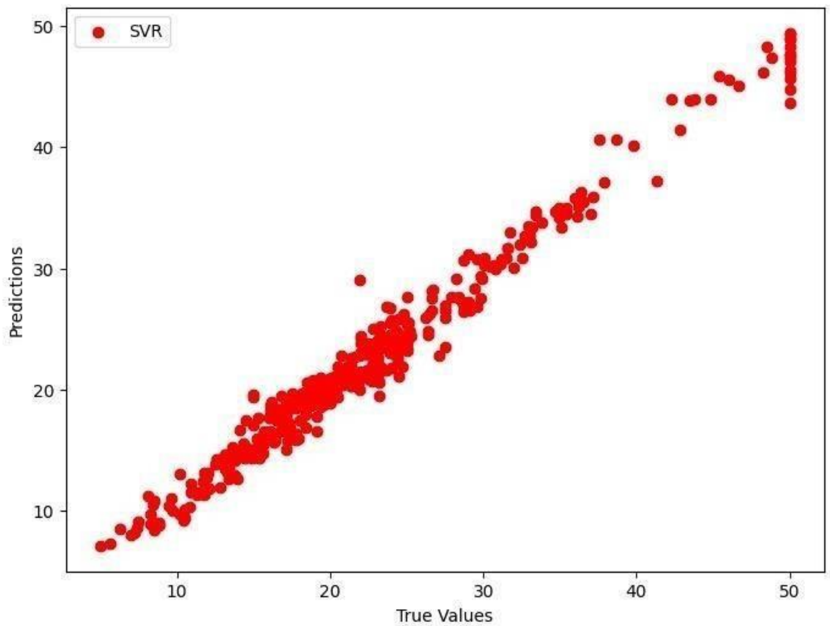

Fig 4. SVR Scatter Plot

4.3 Random Forest Regressor

Random forest regressor (RFR) is a bootstrap algorithm. It is a combination of ensemble learning methods and decision tree framework that creates multiple randomly generated decision trees from data. The results are then averaged to get a new output, often resulting in robust predictions/classes. In simple terms, it builds a set of decision trees that are each trained on different subsets of data. Finally, the predictions are combined to form a more reliable prediction. Fig.5 below is a scatter- graph that shows the relationship between the actual values and the predicted values using Random Forest regressor Model.

Model	**Mean (Score)**	**SD (Score)**	**RMSE**
Random Forest	3.32470054	0.6982874	2.98980323

Table No.2

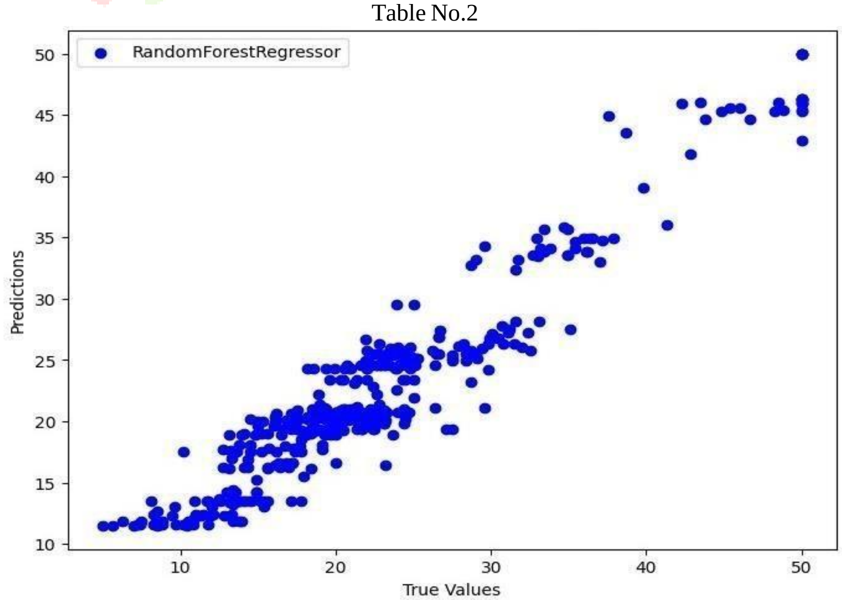

Fig 5. Random Forest Scatter Plot Scatter Plot

4.4 Chatbot

The model makes use of chatbot that predicts house prices using machine learning, speech recognition, and text generation. It begins by importing necessary libraries and modules, captures user voice input through a microphone, generates text responses via the OpenAI API, and converts text to speech for user interaction. In the main loop, the chatbot listens for house price- related queries, collects numeric input for specific house features, and predicts house prices using a trained machine learning model. The predicted prices are then communicated back to the user. Note that certain code segments related to non- numeric word replacement and value extraction through regular expressions are commented out and may be intended for future enhancements.

5.RESULT

The stacking of the three models gives a much-improved RMSE value than the individual models. On performing ensemble technique (stacking) along with hyperparameter tuning, error-rate is reduced to a considerable extent. AdaBoost, RandomForest & SVR models gave better accuracy with stacking rather than the normally used weighted average method for ensemble learning with better house price prediction value.

Model	Mean (Score)	SD (Score)	RMSE
Ensemble (Stacking)	5.07243524	1.21528614	2.92306593

Table No.4

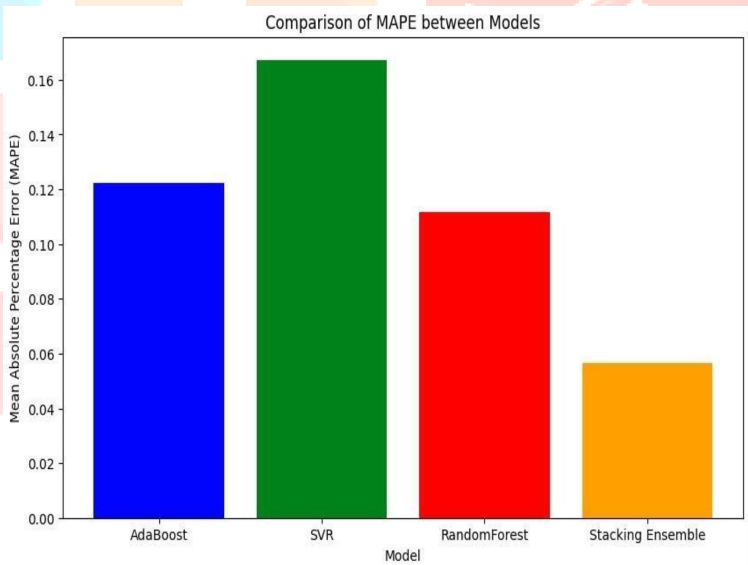

Fig.6. MAPE Comparison

```
A warm welcome to your personalized house price predictor. ChatMe AI will help you predict house prices with precision.
Listening
Recognizing...
You said: what is house price
Sure, please provide the necessary details for house price prediction.
What is the crime rate?
0.00632
What is the proportion of residential land zoned for lots over 25,000 sq.ft.?
18.0
What is the proportion of non-retail business acres per town?
2.31
Is the property near the Charles River? (1 if it is, 0 otherwise)
0.0
What is the nitric oxides concentration (parts per 10 million)?
0.538
What is the average number of rooms per dwelling?
6.575
What is the proportion of owner-occupied units built prior to 1940?
65.2
What is the weighted distances to five Boston employment centers?
4.09
What is the index of accessibility to radial highways?
1.0
What is the full-value property-tax rate per $10,000?
296.0
...
The predicted house price is 25.91267098985947 thousand dollars.
```

6.FUTURE SCOPE

Enhancing user experience, incorporating real-time data, and enhancing natural language comprehension in order to acquire more precise prediction. It is also crucial to expand to new geographic areas, integrate with mobile applications, and adhere to real estate laws. User engagement and happiness can be further increased via personalization, speech recognition advancements, and feedback mechanisms. In order to be relevant and useful in the changing real estate market, the chatbot can be trained to investigate multimodal inputs, AI explainability, and scalability. Enhanced UI for user-interaction and convenience can be created for better functionality with input as a non-numeric value with output as predicted price as required.

7.CONCLUSION

There are numerous elements that affect how much an apartment unit costs. The price estimation technique must therefore take into account a variety of intrinsic and indicative elements. Typically, the algorithms are tested on specific data sets to determine their efficacy. Thedata sets' sample spaces have an impact on the prediction accuracy as well. There is research on a variety of prediction approaches in the literature.The majority of them adopt the concept of contrasting many algorithms and choose the best performing algorithms as the prediction model. Inplace of employing individual algorithms, this research proposes a revolutionary approach tohousing price prediction based on ensemble learning.

8.ACKNOWLEDGEMENT

A sincere thanks to Dwarkadas J. Sanghvi College of Engineering for providing a platform and necessary resources to develop this project. We are also deeply thankful to Prof. Neha Katre (Dept. Of Information Technology), for her assistance, dedication and collaboration in the successful completion of this project. We sincerely appreciate the efforts of my colleagues and fellow researchers for this stimulating conversations, which played a significant role in refining my ideas. This work would not have been possible without the contributions of all those mentioned above.

9.REFERENCES

[1] Sushant Kulkarni, Shefin Shajit, Akshay Mohite, Dr. Swati Sinha, "House Price Prediction Using Ensemble Learning," *International Journal of Creative Research Thought,* 2021.

[2] Quang Truong, Minh Nguyen, Hy Dang, Bo Mei, "House Price Prediction via Improved Machine Learning Techniques," *International Conference on Identification, Information and Knowledge in the Internet of Things,* 2019.

[3] CH. Raga Madhuri, G. Anuradha, M. Vani Pujitha, "House Price Prediction Using Regression Techniques: A Comparative Study," *International Conference on Smart Structures and Systems,* 2019.

[4] Anand Rawool, Dattatray V. Rogue, Sainath G. Rane, Dr. Vinayk A. Bhardi, "House Price Prediction Using Machine Learning," *Iconic Research and Engineering Journals,* vol. 4, no. 11, 2021.

[5] Alisha Kuvalker, Shivani Manchewar, Sidhika Mahadik, "House Price Forecasting Using Machine Learning," *Proceedings of the 3rd International Conference on Advances in Science & Technology,* 2020.

[6] Bindu Sivasankar, Arun P. Ashok, Gouri Madhu, Fousiya S., "House Price Prediction," *International Journal of Computer Sciences and Engineering,* vol. 8, no. 7, 2020.

[7] Ayush Kumar Tiwari, Aman Goyal, Akshita Sharma, Pragya Tiwari, "House Price Prediction Using Machine Learning," *International Journal of Mechanical Engineering* , vol. 7, no. 4, 2022.

[8] Neelam Shinde, Kiran Gawande, "Valuation of House Prices using Predictive Techniques", *International Journal of Advances in Electronics and Computer Science*, 2018.