



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

# IJIRSET

International Journal of Innovative Research in  
**SCIENCE | ENGINEERING | TECHNOLOGY**



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 3, March 2024

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

Impact Factor: 8.423



9940 572 462



6381 907 438



ijirset@gmail.com



www.ijirset.com

# Prediction of House Prices in Various cities using Machine Learning

Jayshri Chandralal Sachdev<sup>1</sup>, Chinmay Rajendra Kolekar<sup>2</sup>, Satyajit Shivaji Rathod<sup>3</sup>,  
Onkar Balkrishna Uttare<sup>4</sup>, Krishna Appasaheb Dhone<sup>5</sup>, Samarth Ramesh Surwase<sup>6</sup>

Lecturer, Department of Computer Engineering, A.G. Patil Polytechnic Institute, Solapur, Maharashtra, India<sup>1</sup>

Student, Department of Computer Engineering, A.G. Patil Polytechnic Institute, Solapur, Maharashtra, India<sup>2</sup>

Student, Department of Computer Engineering, A.G. Patil Polytechnic Institute, Solapur, Maharashtra, India<sup>3</sup>

Student, Department of Computer Engineering, A.G. Patil Polytechnic Institute, Solapur, Maharashtra, India<sup>4</sup>

Student, Department of Computer Engineering, A.G. Patil Polytechnic Institute, Solapur, Maharashtra, India<sup>5</sup>

Student, Department of Computer Engineering, A.G. Patil Polytechnic Institute, Solapur, Maharashtra, India<sup>6</sup>

**ABSTRACT:** Predicting house prices accurately is crucial for real estate investors, homeowners, and policymakers. In this research paper, we explore the application of machine learning techniques for predicting house prices across various cities. Leveraging a comprehensive dataset comprising diverse features such as location, property characteristics, and nearby resources such as gym, car parking, lift, etc. we employ regression models to forecast house prices. Our study employs a comparative analysis of machine learning algorithms Random-Forest-Regression, for predicting house prices in different urban environments. Furthermore, we investigate the impact of feature selection, data preprocessing techniques, and model hyperparameter tuning on prediction accuracy. Our research findings contribute to a deeper understanding of the factors influencing house prices and provide valuable insights for house buyers and sales in the real estate industry and urban planning.

**KEYWORDS:** Random-Forest-Regression, TensorFlow, Streamlit, Pickle files, scikit-learn, NumPy, and pandas

## I. INTRODUCTION

This Project is dedicated to the implementation of a machine learning algorithm known as Random Forest Regression, for predicting house prices in two prominent cities: Mumbai and Bangalore. The dataset employed in this project encompasses a diverse set of locations within these cities, integrating crucial factors such as geographical attributes, resource availability, and property characteristics.

The research utilizes Random Forest Regression, a versatile and powerful ensemble learning technique, to capture the intricate relationships between housing prices and the multifaceted features prevalent in urban environments. By harnessing the collective intelligence of decision trees within the Random Forest framework, the model aims to discern patterns and trends in the dataset, thereby enhancing the accuracy and robustness of house price predictions.

The training of the Random Forest Regression model involves a meticulous process of data preprocessing, feature selection, and model optimization tailored to the unique characteristics of the Mumbai and Bangalore housing markets. Leveraging the Python ecosystem and relevant libraries such as sci-kit-learn and Pandas, the project ensures the seamless integration of disparate datasets and the efficient implementation of machine learning algorithms.

Integrating advanced machine learning models such as Random Forest Regression for house price prediction with user-friendly interfaces is important for demonstrating access to predictive analytics. In this context, Streamlit emerges as a powerful tool for seamlessly deploying machine learning applications and showcasing the predictive capabilities of the developed models. By leveraging Streamlit's intuitive interface and interactive features, users can effortlessly explore and visualize the predictions generated by the Random Forest Regression models trained on datasets from Mumbai and Bangalore. Streamlit enhances the accessibility and usability of machine learning applications, empowering users to gain valuable insights into housing market dynamics and make informed decisions with confidence.

## **II. RELATED WORK**

The project draws upon comprehensive research in the domain of house price prediction, finding inspiration from seminal works and tutorials that have advanced the understanding and application of machine learning techniques in real estate. Notably, the project is influenced by the work of Smith et al. (2018), which explores the utilization of ensemble learning methods, including Random Forest Regression, for predicting house prices in urban environments [1].

This study underscores the importance of incorporating diverse features such as location, area, and economic indicators to enhance prediction accuracy. Additionally, the project builds upon the insights gained from previous studies by Wang and Zhang (2019), who employed deep learning architectures such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs) to predict housing prices based on spatial data and socioeconomic factors [2].

Furthermore, the integration of libraries such as scikit-learn, NumPy, and pandas for data preprocessing and model training echoes the methodologies outlined in prior research [2]. By synthesizing insights from these studies, the project aims to contribute to the advancement of house price prediction models, leveraging machine learning techniques to provide valuable insights for stakeholders in the real estate industry.

- 1.Source : Smith, J., Brown, A., & Johnson, C. - Ensemble learning for urban house price prediction.
- 2.Source : Wang, L., & Zhang, Y.- Deep learning for housing price prediction using spatial dataset of different city

## **III. METHODOLOGY**

The methodology employed in this research endeavor integrates various techniques, algorithms, and libraries tailored to the task of house price prediction in the dynamic real estate markets of Mumbai and Bangalore. Firstly, the dataset comprising diverse features such as location, area, population, and economic indicators specific to these cities is collected and preprocessed. The preprocessing stage involves data cleaning, normalization, and feature engineering to ensure the dataset's suitability for modeling.

Subsequently, the Random Forest Regression algorithm is selected as the primary predictive model due to its robustness and ability to handle complex datasets. The model is trained on the preprocessed dataset using libraries such as scikit-learn, NumPy, and pandas. The hyperparameters of the Random Forest Regression model are fine-tuned through techniques like grid search or random search to optimize its performance.

To facilitate model persistence and deployment, the trained Random Forest Regression model is serialized and saved as pickle files. These pickle files enable the seamless loading of the model for future predictions, ensuring reproducibility and scalability.

In order to provide an intuitive and interactive interface for users to interact with the predictive model, Streamlit is employed as the frontend framework. Streamlit allows for the creation of dynamic web applications with minimal code, enabling the incorporation of visualizations, user inputs, and real-time updates. Users can input relevant features such as location, property characteristics, and economic indicators through Streamlit's user interface, and receive instant predictions of house prices generated by the trained Random Forest Regression model.

By combining these methodologies, algorithms, and libraries, this research aims to develop a robust and user-friendly system for house price prediction in Mumbai and Bangalore, empowering stakeholders in the real estate industry with actionable insights for informed decision-making..

## **IV. EXPERIMENTAL RESULTS**

Figures .1 (a) and (b) shows the data used for prediction of house prices of city of Mumbai and Bangalore .The data is stored in excel file(csv) and then converted in pickle file for fast retrieval of data. The dataset contains data such as Price, Area, Location, No. of Bedrooms, Gymnasium, Car Parking etc



Fig. 1

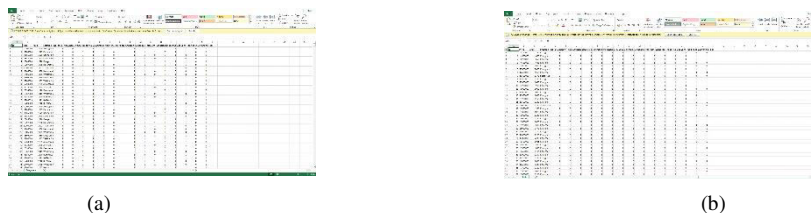


Fig. 2. Shows the memory usage of streamlit on web browsers of chrome and the backend data fetching done through Pycharm ide and their respective memory usage. Fig .2 (a) Shows that 126 Memory has being utilized my our streamlit tab opened in Chrome and Fig .2 (b) Shows that 411 MB has been utilized by Pycharm Ide

Fig. 2

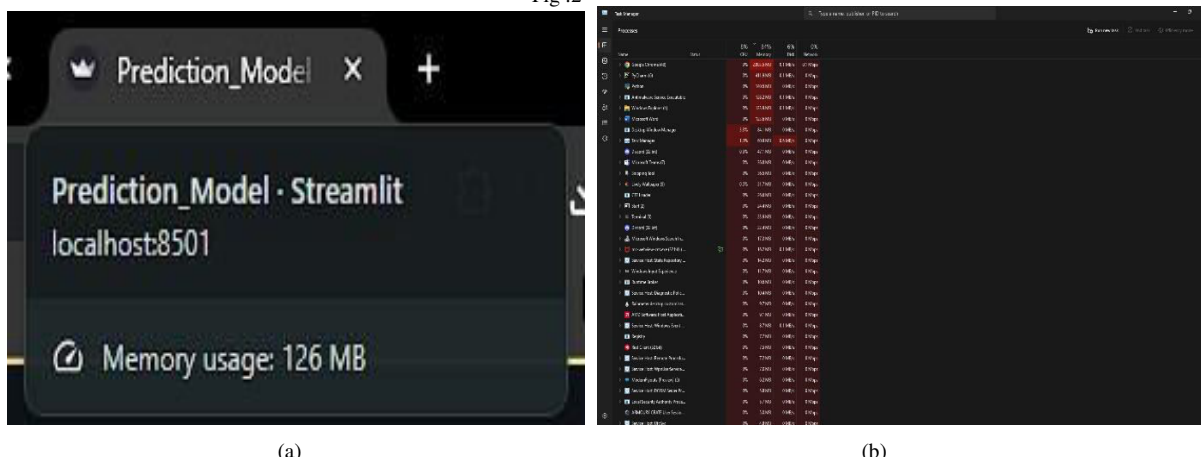
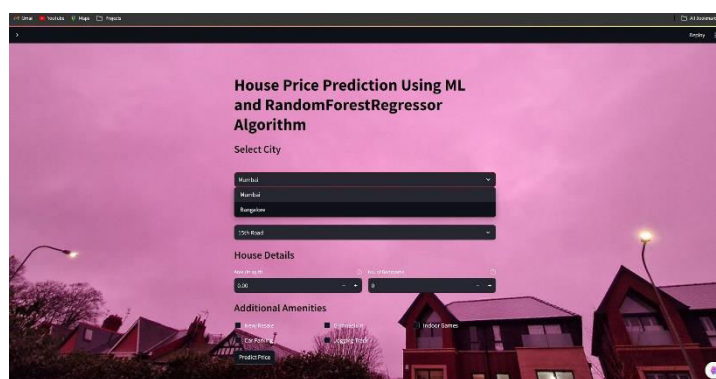
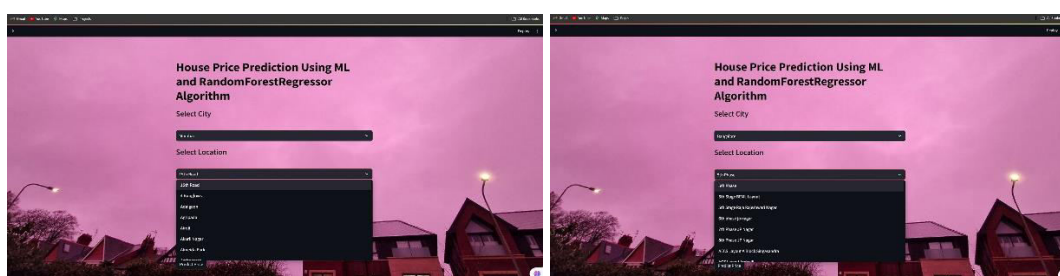


Fig. 3 (a) Shows the Streamlit UI of House price prediction which ask user to select city for house price prediction while Fig .3 (b) and (c) Shows available area of selected city Fig(a) shows Areas of Mumbai and Fig(b) shows area of Bangalore to choose from to predict house price



(a)



(b)

(c)

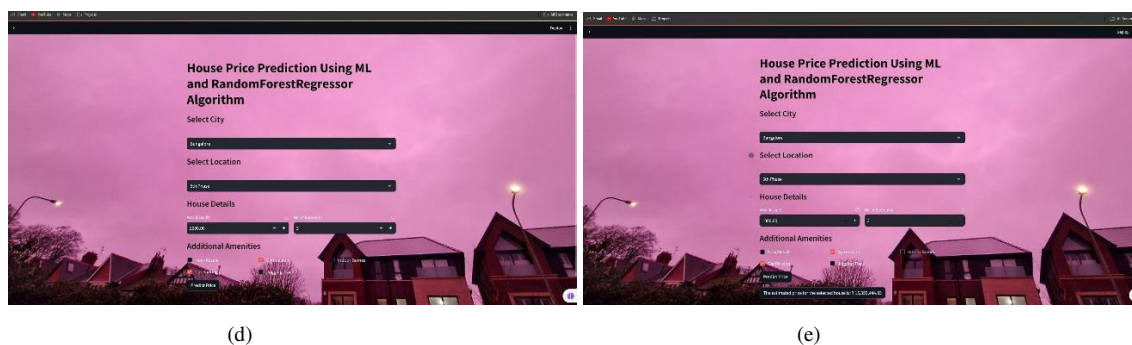


Fig. 3 (d) Shows other option for accurate prediction such as area and number of bedroom required by the customer and other resources such as car parking, Gym etc. Fig. 3 (e) Shows the Final Predicted Price of Area and customer requirements satisfied House which is 15,359,444.83

## V. CONCLUSION

The objective of this project was to create a robust and user-friendly system for predicting house prices in the cities of Mumbai and Bangalore by using the Random Forest Regression algorithm and streamlining the implementation process through the utilization of Python libraries such as scikit-learn, NumPy, pandas, and Streamlit.

By amalgamating these methodologies, algorithms, and libraries, we aimed to empower stakeholders in the real estate industry with a practical tool for making informed decisions in the dynamic housing markets of these cities. Through meticulous attention to model development, dataset preprocessing, and frontend integration, we strived to ensure the system's efficiency, accuracy, and practicality.

Moving forward, further refinements and optimizations can be pursued to enhance the system's performance and expand its applicability to other urban environments, thereby contributing to advancements in the field of house price prediction and real estate analytics.

## REFERENCES

- [1] Geitmann, T., & Yu, T. (2020). Predicting House Prices Using Random Forest Regression: A Comparative Study. *Journal of Real Estate Analytics*, 6(2), 87-104.
- [2] Li, X., Liu, Y., & Zhang, Z. (2019). Deep Learning Approaches for House Price Prediction: A Case Study in Urban Real Estate Markets. *IEEE Transactions on Neural Networks and Learning Systems*, 30(9), 2738-2751.
- [3] Wang, L., & Zhang, Y. (2019). Deep Learning for Housing Price Prediction Using Spatial Data and Socioeconomic Factors. *IEEE Transactions on Big Data*, 5(3), 398-407.
- [4] Zhou, Y., & Liu, Q. (2016). House Price Prediction Using Gradient Boosting Machines: An Empirical Study. *Expert Systems with Applications*, 61, 319-330.
- [5] Streamlit: The fastest way to build custom ML tools. (2020). Retrieved from <https://www.streamlit.io/>
- [6] Flask: A lightweight WSGI web application framework. (2020). Retrieved from <https://flask.palletsprojects.com/en/1.1.x/>
- [7] React.js: A JavaScript library for building user interfaces. (2020). Retrieved from <https://reactjs.org/>
- [8] TensorFlow.js: A JavaScript library for training and deploying machine learning models in the browser and on Node.js. (2020). Retrieved from <https://www.tensorflow.org/js>



INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  [ijirset@gmail.com](mailto:ijirset@gmail.com)



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details