

House Price Prediction by Machine Learning

Sangeetha Nachiyar.S, Subhaksha.V, Mohammed Sintha Haarish P, Aadhithya. T

UG Students, VELS Institute of Science, Technology and Advanced Studies(VISTAS), Pallavaram, Chennai

Abstract: *In recent years, the real estate industry has witnessed significant growth, resulting in an increased demand for accurate property valuation. House price prediction plays a vital role in assisting buyers, sellers, and investors in making informed decisions. Property prices are influenced by multiple factors such as location, size, number of rooms, and market conditions, making manual estimation difficult and often inaccurate.*

This paper presents a House Price Prediction system using Machine Learning techniques. The system processes housing data through various stages including data preprocessing, feature engineering, model training, and evaluation. Linear Regression is used as the primary algorithm to predict house prices based on input features.

The model is trained using historical data and evaluated using performance metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and R^2 score. The results indicate that the model can predict house prices with good accuracy. The proposed system provides a simple, efficient, and scalable solution for property price estimation and supports data-driven decision-making in the real estate domain.

Keywords: House Price Prediction, Machine Learning, Linear Regression, Data Analysis, Predictive Modelling, Real Estate

I. INTRODUCTION

House price prediction is an important application of machine learning in the real estate sector. Accurate estimation of property prices is essential for buyers, sellers, and investors to make informed financial decisions. The price of a house depends on various factors such as location, size, number of bedrooms, amenities, and economic conditions.

Traditional methods of estimating house prices rely on manual analysis and expert judgment, which are time-consuming and often inaccurate. With the advancement of technology, machine learning has emerged as a powerful tool for analyzing large datasets and predicting house prices based on historical data.

Machine learning models can identify patterns and relationships between input features and house prices, enabling accurate and efficient prediction. This paper aims to develop a machine learning-based system that can predict house prices using regression techniques.

II. LITERATURE SURVEY

Several studies have been conducted on house price prediction using machine learning techniques. Early approaches used statistical methods such as Linear Regression to model relationships between variables. These methods were simple but limited in handling complex data.

Recent research has explored advanced algorithms such as Decision Trees, Random Forest, and Gradient Boosting, which provide better accuracy by capturing non-linear relationships. Deep learning techniques have also been applied to large datasets for improved performance.

However, these methods often require high computational resources and large datasets. Additionally, challenges such as data quality, feature selection, and overfitting remain significant issues.

Therefore, this study focuses on using Linear Regression as a simple and effective approach for house price prediction.



III. PROBLEM DEFINITION

In the real estate market, accurately predicting house prices is a challenging task due to the influence of multiple factors. Traditional methods rely on manual estimation, which is time-consuming and prone to errors.

Existing systems often lack automation and fail to process large datasets efficiently. They may also ignore important features, leading to inaccurate predictions.

Furthermore, real estate data is dynamic and constantly changing, making it difficult to maintain accurate predictions over time. There is a need for an automated system that can analyze housing data and provide reliable predictions.

- Generate actionable insights through AI-based decision support
- Enable easy interaction through chatbot-based interfaces

The proposed AI-Driven Business Analytics System aims to address these challenges by integrating data preprocessing, analytics, machine learning, visualization, and AI-based modules into a single platform. This system is designed to enhance decision-making, improve business performance, and provide a scalable solution for modern data-driven organizations.

IV. PROPOSED SYSTEM

The proposed system is a machine learning-based House Price Prediction system that estimates property prices based on input features.

The system follows a structured workflow:

- Data collection from housing datasets
- Data preprocessing to clean and prepare data
- Feature engineering to select important features
- Model training using Linear Regression
- Model evaluation using performance metrics
- Prediction of house prices based on user input

The system also includes a user-friendly interface that allows users to enter property details and obtain predicted prices instantly.

The proposed system provides an efficient and scalable solution for house price prediction.

V. CONCLUSION

The model was trained using historical housing data and evaluated using standard performance metrics. The results indicate that the Linear Regression model performs well in predicting house prices.

The evaluation metrics showed:

- Low Mean Absolute Error (MAE)
- Low Mean Squared Error (MSE)
- High R^2 Score

The predicted values were close to actual values, indicating good model performance. However, some deviations were observed for extreme values due to outliers in the dataset.

Overall, the results demonstrate that the proposed system is effective for house price prediction.

This paper presents a House Price Prediction system using machine learning techniques. The system successfully predicts house prices based on input features and demonstrates good accuracy.

The use of Linear Regression provides a simple and efficient approach for prediction. The system helps users make informed decisions and reduces reliance on manual estimation.

Future work may include the use of advanced machine learning models, real-time data integration, and inclusion of additional features to improve prediction accuracy.



REFERENCES

- [1]. Tan, P. N., Steinbach, M., & Kumar, V. (2019). *Introduction to Data Mining* (2nd ed.). Pearson Education.
- [2]. Géron, A. (2022). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow* (3rd ed.). O'Reilly Media.
- [3]. McKinney, W. (2018). *Python for Data Analysis* (2nd ed.). O'Reilly Media.
- [4]. Pedregosa, F., et al. (2011). *Scikit-learn: Machine Learning in Python*. Journal of Machine Learning Research, 12, 2825–2830.
- [5]. Hunter, J. D. (2007). *Matplotlib: A 2D Graphics Environment*. Computing in Science & Engineering, 9(3), 90–95.
- [6]. Plotly Technologies Inc. (2023). *Plotly: Interactive Graphing Library for Python*. Available at: <https://plotly.com>
- [7]. Streamlit Inc. (2023). *Streamlit Documentation*. Available at: <https://docs.streamlit.io>
- [8]. Witten, I. H., Frank, E., & Hall, M. A. (2016). *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann.
- [9]. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An Introduction to Statistical Learning* (2nd ed.). Springer.
- [10]. Provost, F., & Fawcett, T. (2013). *Data Science for Business*. O'Reilly Media.
- [11]. Chollet, F. (2021). *Deep Learning with Python* (2nd ed.). Manning Publications.

