IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

"House Price Prediction using Machine Learning"

Prof. Amisha Naik [1], Omkar Jadhav[2]

[1]Faculty Computer Engineering Vidya Prasarini Sabha's College of Engineering and Technology, Lonavala

[2]Student Computer Engineering Vidya Prasarini Sabha's College of Engineering and Technology, Lonavala

ABSTRACT: House price prediction is an important application of Machine Learning that helps estimate property prices based on various features such as location, area, number of rooms, amenities, and market conditions. Accurate prediction models are useful for buyers, sellers, real estate companies, and investors in making better financial decisions. Traditional methods of property valuation often depend on manual analysis and experience, which can be time-consuming and less accurate. Machine Learning techniques provide a data-driven approach that improves prediction accuracy and efficiency.

This project focuses on developing a house price prediction system using Machine Learning algorithms. The dataset used for the model contains multiple attributes related to houses, including size, number of bedrooms, bathrooms, location, parking facilities, and other relevant factors. Data preprocessing techniques such as handling missing values, feature selection, encoding categorical variables, and normalization are applied to improve the quality of the dataset.

Different regression algorithms such as Linear Regression, Decision Tree Regression, Random Forest Regression, and Support Vector Regression are implemented and compared to identify the most accurate model. The performance of the models is evaluated using metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared score. Among the tested algorithms, ensemble methods like Random Forest often provide better accuracy due to their ability to handle complex relationships within the data.

1.INTRODUCTION

The real estate industry plays a vital role in the economy, and property pricing is one of the most important factors for buyers, sellers, investors, and real estate agencies. Determining the correct price of a house is a complex task because it depends on multiple factors such as location, size, number of rooms, infrastructure, nearby facilities, and market trends. Traditionally, house prices are estimated through manual analysis and expert judgment, which may not always provide accurate results and can be time-consuming.

With the advancement of technology, Machine Learning has emerged as an effective solution for predictive analysis in various domains, including real estate. Machine Learning enables computers to learn patterns from historical data and make accurate predictions without being explicitly programmed. By analyzing large amounts of housing data, Machine Learning models can identify relationships between property features and prices, leading to faster and more reliable estimations.

House Price Prediction using Machine Learning involves collecting housing data, preprocessing it, and applying regression algorithms to predict the expected price of a property. Various algorithms such as Linear Regression, Decision Tree Regression, Random Forest Regression, and Support Vector Regression are commonly used for this purpose. These models help improve prediction accuracy and assist users in making informed decisions.

The main objective of this project is to design and develop a predictive system that estimates house prices based on important housing features. The proposed system aims to reduce human effort, improve accuracy, and provide quick price predictions. This project demonstrates the practical application of Machine Learning in solving real-world problems and highlights its importance in the modern real estate market.

2. PROPOSED SYSTEM

The architecture is designed to develop an intelligent and efficient stock market prediction system using machine learning techniques. The system focuses on analyzing historical stock data to identify patterns and forecast future price movements. It integrates data preprocessing, exploratory data analysis, feature selection, and machine learning models to improve prediction accuracy. The computation is performed using Python-based tools, where data processing and model training are carried out efficiently to generate reliable predictions. The implementation is evaluated based on model accuracy, error metrics, and overall performance.

2.1 Data Processing Module:

Handles data collection, preprocessing, and preparation for analysis. Historical stock data is collected from sources such as Yahoo Finance and Kaggle. The module ensures that the data is clean and structured by handling missing values, removing duplicates, converting data types, and normalizing the dataset for better model performance.

2.2 Exploratory Data Analysis (EDA) Module:

This module is responsible for analyzing and visualizing the dataset to understand patterns and trends. It uses graphs such as line charts, moving averages, and heatmaps to identify relationships between variables like stock prices and trading volume. This helps in gaining insights into market behavior.

2.3 Feature Selection Module:

Selects the most relevant features that influence stock price prediction. Important attributes such as open price, close price, high price, low price, and volume are selected. This improves model efficiency and reduces unnecessary complexity.

2.4 Machine Learning Model Module:

Implements various machine learning algorithms to predict stock prices. In this system:

- Linear Regression is used as a baseline model
- Decision Tree is used to handle non-linear data
- Random Forest is used to improve accuracy through ensemble learning The

models are trained on historical data to learn patterns and relationships.

2.5 Prediction Module:

After training, the model is used to predict future stock prices. The system takes new input data and generates predicted values based on learned patterns. These predictions help in understanding future market trends.

2.6 Evaluation Module:

Evaluates the performance of the machine learning models using metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and R^2 score. This ensures that the best-performing model is selected.

3. IMPLEMENTATIONS AND RESULTS:

The proposed stock market prediction system was implemented using Python and various data science and machine learning libraries. The implementation was carried out in an interactive development environment such as Jupyter Notebook/Google Colab, which allows efficient execution of code, visualization, and analysis.

The system begins with data collection, where historical stock market data is obtained from reliable sources such as Yahoo Finance and Kaggle. The dataset includes attributes such as opening price, closing price, highest price, lowest price, trading volume, and date. This data is stored in CSV format and loaded into the system using the Pandas library.

After data collection, preprocessing is performed to clean and prepare the dataset. Missing values are handled using appropriate techniques such as forward filling or mean substitution. Duplicate records are removed to ensure data consistency. Data types are converted into suitable formats, especially date and numerical fields. Normalization techniques are applied to scale the data, improving the performance of machine learning models.

Exploratory Data Analysis (EDA) is then conducted to understand the dataset. Visualization libraries such as Matplotlib and Seaborn are used to plot graphs like stock price trends, moving averages, and correlation heatmaps. These visualizations help in identifying patterns, trends, and relationships among different variables.

Feature selection is performed to choose the most relevant attributes affecting stock prices. Features such as open, close, high, low, and volume are selected as input variables for the machine learning models. This step reduces unnecessary complexity and improves model efficiency.

The core implementation involves training machine learning models using Scikit-learn. Three algorithms are implemented: Linear Regression, Decision Tree, and Random Forest. The dataset is divided into training and testing sets using the train-test split method, typically in an 80:20 ratio. The models are trained on the training data and tested on unseen data to evaluate their performance.

Once trained, the models are used to predict stock prices. The prediction results are compared with actual values to determine accuracy. Performance evaluation is carried out using metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and R^2 score.

The performance of the models was evaluated using metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and R^2 score.

Among all models, Random Forest provided the best results with higher accuracy and lower error values. Decision Tree performed moderately well, while Linear Regression showed lower accuracy for complex data.

The results show that machine learning models can effectively predict stock trends, with Random Forest being the most reliable. However, predictions are not completely accurate due to the unpredictable nature of the stock market.

4. CONCLUSION:

This paper presented a machine learning-based approach for stock market prediction using historical data. Various algorithms such as Linear Regression, Decision Tree, and Random Forest were implemented to analyze stock price trends and forecast future values. The system involved key steps including data collection, preprocessing, exploratory data analysis, feature selection, model training, and evaluation.

The results demonstrated that machine learning techniques can effectively identify patterns in stock market data and provide useful insights for prediction. Among the implemented models, Random Forest showed better performance in terms of accuracy and error reduction compared to other models.

However, stock market prediction remains a challenging task due to its highly volatile and unpredictable nature, influenced by external factors such as economic conditions, political events, and market sentiment. Therefore, while the proposed system improves prediction capability, it cannot guarantee completely accurate results.

Overall, this study highlights the potential of machine learning in financial analysis and decision-making, and it can serve as a foundation for developing more advanced and intelligent stock prediction systems in the future.

5. REFERENCES:

[1] Kaggle.

Stock Market Dataset. Available at: https://www.kaggle.com

[2] Yahoo Finance.

Historical Stock Market Data. Available at: https://finance.yahoo.com

[3] Scikit-learn Documentation.

Machine Learning in Python. Available at: https://scikit-learn.org

[4] Pandas Documentation.

Data Analysis and Manipulation Tool. Available at: https://pandas.pydata.org

[5] NumPy Documentation.

Scientific Computing with Python. Available at: https://numpy.org

[6] Matplotlib Documentation.

Data Visualization Library. Available at: https://matplotlib.org

[7] Seaborn Documentation.

Statistical Data Visualization. Available at: https://seaborn.pydata.org