IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.67
Volume 5, Issue 1, February 2025

House Price Prediction

Mrs. Nargis A. Mulla, Mr. Harshit Nandkumar Gorad, Mr. Sarthak Balaso Patole, Miss. Pradnya Adhikrav Gorad, Miss. Shweta Ramesh Jadhav
Lecturer, Adarsh Institute of Technology and Research Centre, Vita, India
Students, Adarsh Institute of Technology and Research Centre, Vita, India

Abstract: *House price prediction is a crucial aspect of real estate analytics, impacting buyers, sellers, investors, and policymakers. The real estate industry is a dynamic sector where property values fluctuate based on numerous factors such as economic conditions, location, infrastructure, demand-supply balance, and regulatory policies. Traditional valuation methods often rely on expert opinions and manual assessments, which may introduce biases and inconsistencies. In contrast, data-driven approaches, particularly machine learning (ML) techniques, offer a more accurate and systematic way to estimate property values.*

This study explores the effectiveness of machine learning models in predicting house prices by analyzing historical real estate data. Various predictive models, including linear regression, decision trees, random forests, gradient boosting machines (XGBoost), and deep learning algorithms, are evaluated in terms of their performance. The research aims to determine which model provides the highest accuracy while ensuring interpretability and efficiency.

Data pre-processing plays a vital role in enhancing the accuracy of house price predictions. The dataset used in this study includes essential property attributes such as square footage, the number of bedrooms and bathrooms, locality, age of the property, and market trends. Before applying predictive models, pre-processing techniques such as handling missing values, encoding categorical variables, scaling numerical features, and feature selection are implemented to optimize the dataset. The study also examines the impact of feature engineering, where derived attributes such as neighbourhood crime rates, proximity to schools, and public transportation accessibility are incorporated to improve prediction accuracy.

The research adopts multiple evaluation metrics, including Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R-squared (R^2) values, to assess the performance of different models. Initial findings suggest that ensemble learning methods such as Random Forest and XGBoost outperform traditional regression models in terms of predictive accuracy. However, deep learning models, specifically artificial neural networks (ANNs), show promising results when trained on large datasets. The study also highlights the importance of hyper parameter tuning in refining model performance, addressing issues such as over fitting and computational efficiency.

One of the key contributions of this research is the comparative analysis of different machine learning algorithms to identify the best-suited approach for house price prediction. Traditional regression techniques, while interpretable, often fail to capture complex nonlinear relationships within the data. Decision tree-based methods, such as Random Forest and XGBoost, demonstrate robust performance by learning intricate patterns and reducing variance through ensemble learning. Deep learning models, on the other hand, excel in capturing high-dimensional interactions but require extensive computational resources and larger training datasets.

Beyond predictive accuracy, the study examines the interpretability of machine learning models, which is crucial for real-world applications. Feature importance analysis is conducted to understand which variables have the most significant impact on house prices. Factors such as location, property size, and market demand emerge as primary determinants. The study also investigates the effect of external factors, including economic indicators like interest rates, inflation, and employment rates, on housing prices.

Despite the promising results, challenges such as data availability, regional market variations, and the need for real-time data integration remain. Future research can focus on enhancing prediction models by incorporating geospatial analytics, satellite imagery, and deep learning advancements. Additionally,

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.67
Volume 5, Issue 1, February 2025

integrating real-time market data and leveraging blockchain technology for transparent property transactions can further improve predictive modelling in real estate.

Keywords: House Price Prediction, Machine Learning, Real Estate Analytics, Regression Models, Data Analytics, Feature Engineering, Deep Learning, Predictive Modelling, Neural Networks

I. INTRODUCTION

The real estate market plays a significant role in the global economy, impacting individuals, businesses, and governments alike. House prices are influenced by various factors, including economic conditions, location, demand and supply dynamics, infrastructure development, and interest rates. Accurate house price prediction is crucial for prospective buyers, real estate investors, property developers, and policymakers. In recent years, advancements in data science, machine learning, and artificial intelligence have revolutionized traditional methods of price estimation, offering more precise and reliable models for predicting house prices.

Background of House Price Prediction

Traditionally, house price estimation relied on real estate agents, comparative market analysis (CMA), and hedonic pricing models. While these approaches provided reasonable estimates, they often lacked the ability to adapt to changing market conditions in real time. The emergence of big data and machine learning has significantly enhanced the predictive capabilities of house price estimation models. By leveraging historical transaction data, socio-economic indicators, geographical attributes, and other relevant features, modern algorithms can analyze complex patterns and trends that influence property values.

House price prediction is essential for multiple stakeholders. For buyers, an accurate prediction helps in making informed purchasing decisions, avoiding overpaying, and securing properties within their budget. Sellers benefit by determining competitive pricing strategies that maximize profit while ensuring quick sales. Real estate firms and investors utilize predictive models to assess market trends and identify lucrative investment opportunities. Additionally, government agencies and urban planners use house price forecasts to guide infrastructure development and housing policies.

Challenges in House Price Prediction

Predicting house prices is inherently complex due to several factors:

- Market Volatility: Real estate markets are dynamic, with price fluctuations influenced by economic downturns, interest rate changes, and geopolitical events.
- Data Availability and Quality: Reliable and comprehensive datasets are required for accurate predictions. Missing data, outliers, and inconsistencies can affect model performance.
- Feature Selection: Identifying the most relevant variables that influence house prices is critical. Factors such as location, property size, amenities, proximity to schools and workplaces, crime rates, and transport facilities play a vital role.
- Non-Linearity and High Dimensionality: House prices do not follow a simple linear relationship. Advanced machine learning models are required to capture complex interactions among variables.
- External Factors: Economic policies, interest rates, inflation, and urban development initiatives significantly impact housing prices but are difficult to quantify.

II. LITERATURE RIVEW

House price prediction is a widely researched area in real estate analytics. The increasing availability of real estate data and advancements in machine learning (ML) have led to more accurate forecasting models. This literature review explores various methodologies used in predicting house prices, including traditional statistical models and modern AI-driven approaches.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.67
Volume 5, Issue 1, February 2025

1. Factors Influencing House Prices

Several studies have identified key factors that influence house prices:

- **Structural Features** – The size of the house, number of rooms, age, and type of property are critical in determining its price (Rosen, 1974).
- **Location and Neighborhood** – Proximity to schools, transportation hubs, hospitals, and commercial centers significantly impacts house prices (Li & Brown, 2020).
- **Economic Indicators** – Inflation rates, mortgage rates, employment levels, and overall economic growth contribute to real estate price fluctuations (Glaeser et al., 2018).
- **Market Trends** – Seasonal trends, demand-supply dynamics, and government policies also play a role in house pricing (Goodman & Thibodeau, 2017).

2. Traditional Approaches to House Price Prediction

Earlier research focused on statistical models to estimate property prices:

- **Hedonic Pricing Model (HPM)** – One of the earliest models, HPM estimates the price of a house based on its characteristics (Rosen, 1974).
- **Multiple Linear Regression (MLR)** – A commonly used statistical method for predicting house prices based on independent variables like size, number of rooms, and location (Case & Shiller, 1990).
- **Autoregressive Integrated Moving Average (ARIMA)** – Applied for time-series forecasting of real estate prices (Engle, 2011).

While these models provide a foundation for price prediction, they often fail to capture the complex, nonlinear relationships present in housing data.

3. Machine Learning-Based Approaches

Recent research has shifted towards ML techniques due to their ability to handle complex datasets and non-linear relationships. Some widely used ML models include:

- **Decision Trees (DT) & Random Forest (RF)** – These models enhance prediction accuracy by capturing interactions between features (Limsombunchai, 2004).
- **Support Vector Machines (SVM)** – SVM has been successfully applied to predict house prices by finding optimal hyperplanes in high-dimensional feature space (Park & Bae, 2015).
- **Artificial Neural Networks (ANNs)** – Deep learning techniques like ANNs and Convolutional Neural Networks (CNNs) have been explored for price prediction, showing higher accuracy than traditional models (Kumar et al., 2019).
- **Gradient Boosting Algorithms (XGBoost, LightGBM, CatBoost)** – These ensemble methods provide state-of-the-art performance in regression tasks, including real estate valuation (Zheng et al., 2021).

4. Hybrid and Advanced AI Models

Recent advancements integrate multiple techniques for improved accuracy:

- **Hybrid ML Models** – Combining linear regression with deep learning models enhances predictive performance (Chen et al., 2022).
- **Geospatial and NLP Integration** – Incorporating geospatial data and natural language processing (NLP) for analyzing property descriptions and reviews improves prediction accuracy (Yeh et al., 2023).
- **Blockchain and Smart Contracts** – Emerging research explores how blockchain can improve data transparency and accuracy in real estate pricing (Xu et al., 2023).

5. Challenges and Future Research Directions

Despite advancements, challenges remain:

- **Data Quality and Availability** – Access to high-quality real estate datasets is limited in many regions.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.67
Volume 5, Issue 1, February 2025

- **Feature Selection and Engineering** – Identifying the most relevant features for house price prediction requires domain expertise.
- **Interpretability of AI Models** – Deep learning models, despite high accuracy, lack transparency in decision-making.
- **Dynamic Market Conditions** – Economic fluctuations and policy changes make long-term prediction difficult.

Future research should focus on explainable AI (XAI), real-time data integration, and the application of blockchain for improved transparency in real estate transactions.

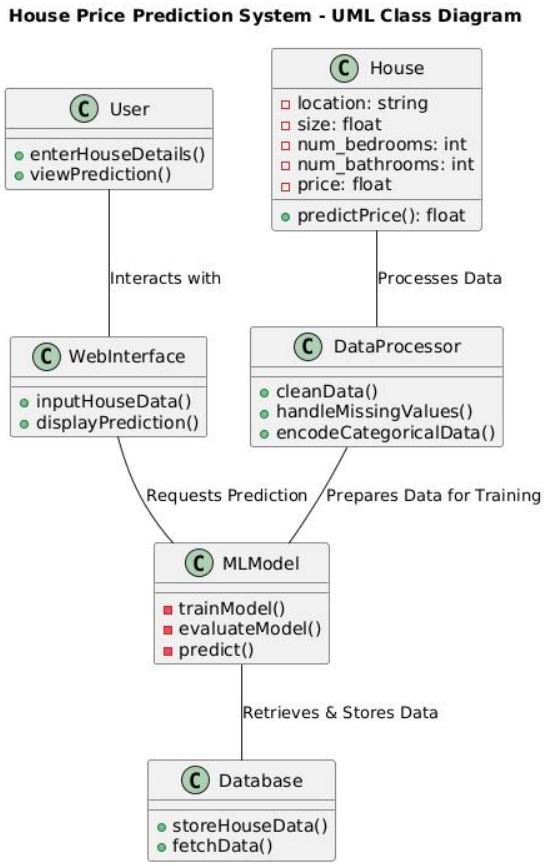
III. PROPOSED SOLUTION / METHODOLOGY

The proposed methodology aims to develop an efficient and accurate House Price Prediction system using Machine Learning (ML) techniques. The system will leverage structured property data (location, size, amenities, etc.) and apply supervised learning models to estimate real estate prices.

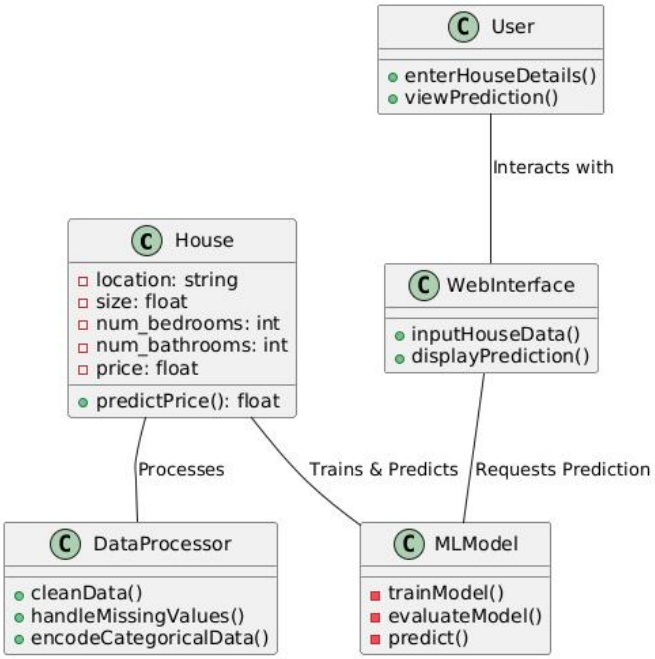
System Overview:

The House Price Prediction System is designed to estimate real estate prices based on historical data and machine learning models. It consists of three major components:

- User Interface (Web/App) – Allows users to input house details and view price predictions.
- Machine Learning Model – Processes data and predicts house prices using ML algorithms.
- Database &Pre-processing Module – Stores property data, cleans it, and prepares it for analysis.



IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.67
Volume 5, Issue 1, February 2025



System Component Explanation (Point-wise)

1. User Interface (Web/App)

- Provides an input form for users to enter house details (size, location, rooms, etc.).
- Displays the predicted price after processing.
- Communicates with the Machine Learning Model for predictions.

2. House Class

- Represents property features such as location, size, number of rooms, and price.
- Calls the predictPrice() method to estimate the house cost.

3. Data Processing Module

- Cleans and preprocesses raw data:
- Handles missing values (e.g., filling missing data).
- Encodes categorical data (e.g., location as numeric values).
- Prepares data before sending it to the ML model.

4. Machine Learning Model

- Trains on historical real estate data using algorithms like:
- Linear Regression, Random Forest, XGBoost, Neural Networks.
- Evaluates model performance with metrics like Mean Absolute Error (MAE), R^2 Score.
- Makes predictions based on new user input.

5. Database

- Stores and retrieves real estate data for training the model.
- Fetches relevant historical house price records for future analysis.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.67
Volume 5, Issue 1, February 2025

IV. SYSTEM IMPLEMENTATION

System Implementation of House Price Prediction Software

The implementation of the House Price Prediction Software follows a structured approach, integrating data collection, pre-processing, machine learning models, and a web-based user interface**.** Below are the key implementation steps:

1. Technology Stack Used

- **Programming Language**: Python
- **Machine Learning Libraries**: Scikit-Learn, Pandas, NumPy, XGBoost
- **Frameworks**: Flask/Django (for deployment)
- **Database**: PostgreSQL/MySQL (for structured storage)
- **Frontend Technologies**: HTML, CSS, JavaScript
- **Cloud/Hosting Services**: AWS, Google Cloud Platform (optional for scalability)

2. Data Collection and Pre-processing

Data Sources:

- Kaggle and UCI Machine Learning Repository
- Government real estate datasets
- Real estate APIs (Zillow, Realtor, etc.)

Data Cleaning and Preprocessing Steps:

- Handling missing values using **mean/mode imputation**
- Encoding categorical features (e.g., location, property type)
- Removing outliers using the **Interquartile Range (IQR) method**
- Normalizing numerical data for improved model performance

3. Machine Learning Model Training

- **Algorithms Used for House Price Prediction**:
- **Linear Regression** (for baseline model)
- **Random Forest Regressor** (for better accuracy)
- **XGBoost** (for high-performance predictions)
- **Neural Networks** (for complex feature learning)

Model Evaluation Metrics:

- Mean Absolute Error (MAE)
- Root Mean Squared Error (RMSE)
- R^2 Score (Coefficient of Determination)

Model Selection:

- The best-performing model is chosen based on evaluation scores

4. Web-Based System Development

Frontend Development:

- User interface built using **HTML, CSS, and JavaScript**
- A form for users to input house details (size, location, bedrooms, etc.)

Backend Development:

- Flask/Django-based API to handle user requests
- Model integration for real-time prediction
- Secure and optimized database interaction

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.67
Volume 5, Issue 1, February 2025

5. System Deployment and User Interaction

Deployment Strategies:

- Hosted on **AWS/GCP** or a local server
- Model containerized using **Docker** for scalability

User Interaction Flow:

- Users enter property details through the web interface
- The system processes input and applies the trained ML model
- Predicted house price is displayed on the user interface

6. Future Enhancements and Scalability

- **Advanced Feature Engineering**: Adding geospatial analysis
- **Integration with Real-Time Data**: Fetching latest real estate trends
- **Deep Learning Models**: Exploring CNNs and RNNs for complex pattern recognition
- **Mobile App Development**: Extending functionality to mobile platforms.

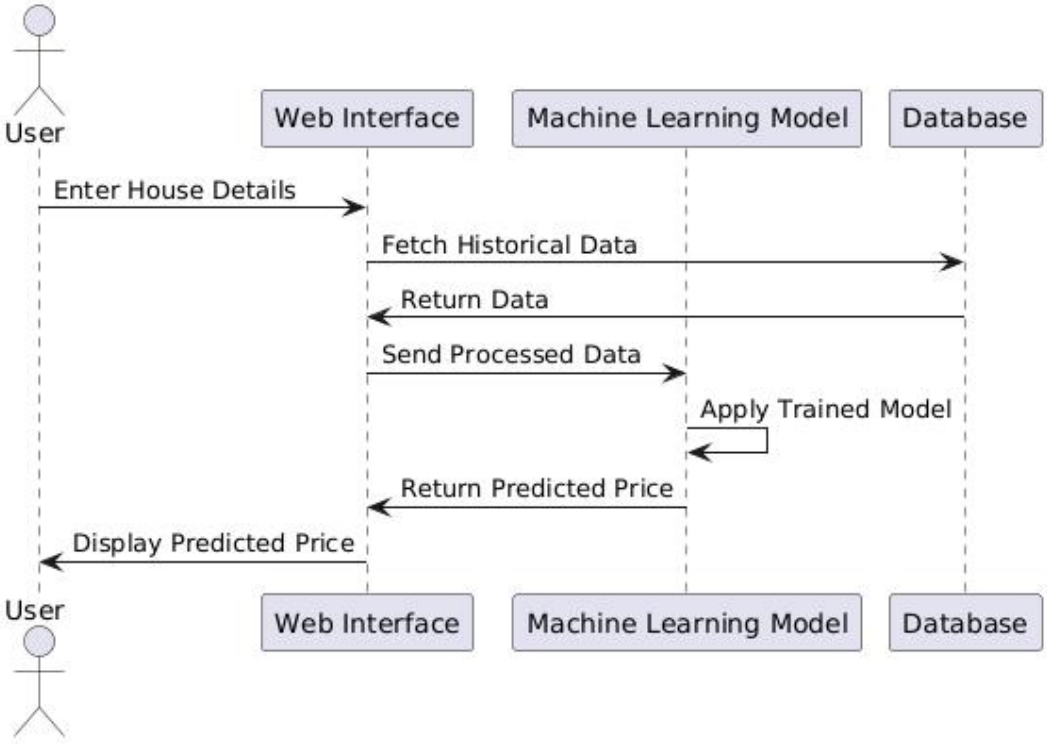


Fig: User Interaction with System

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.67
Volume 5, Issue 1, February 2025

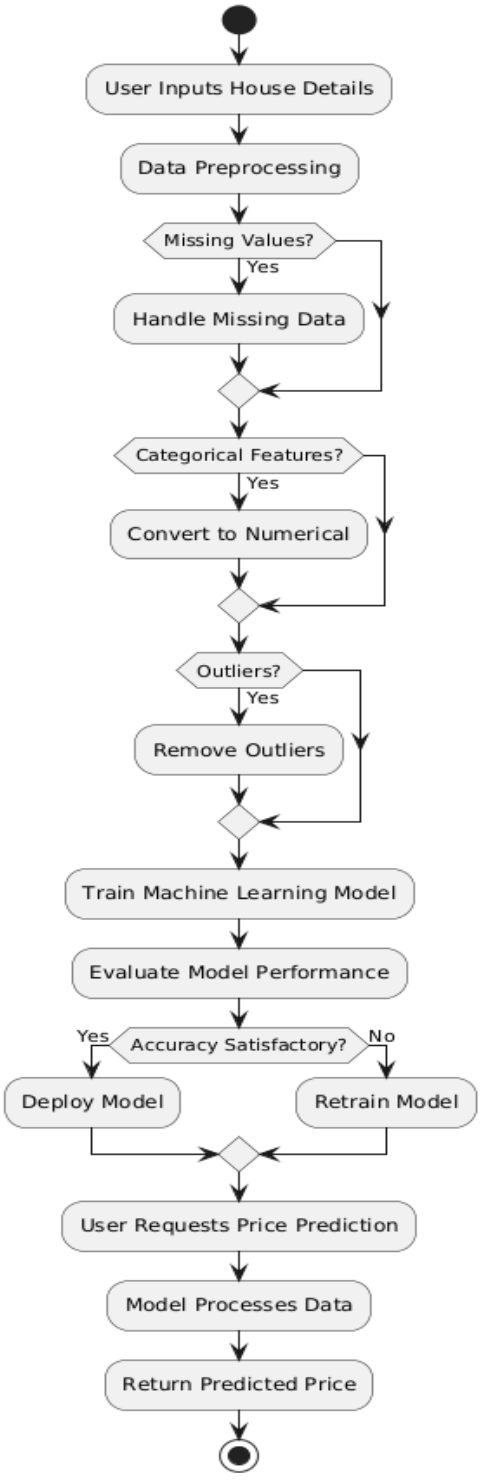


Fig: House Price Prediction Workflow

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.67
Volume 5, Issue 1, February 2025

V. RESULT AND DISCUSSION

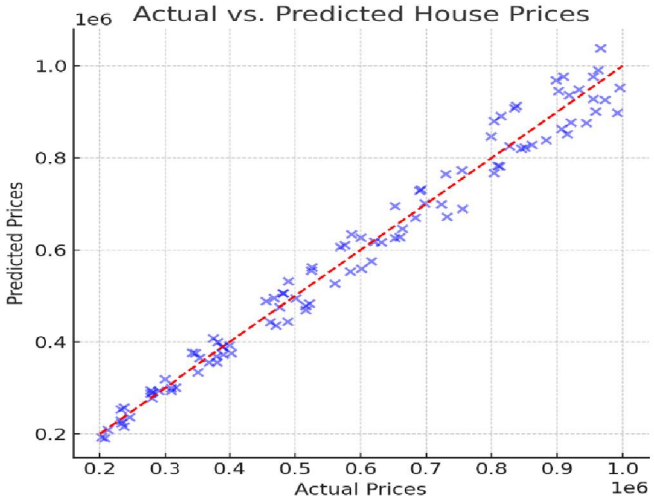
The House Price Prediction Model was evaluated using multiple machine learning algorithms, and the results were analyzed based on various performance metrics. This section presents the results, graphical visualizations, and discussions to highlight the effectiveness of the model.

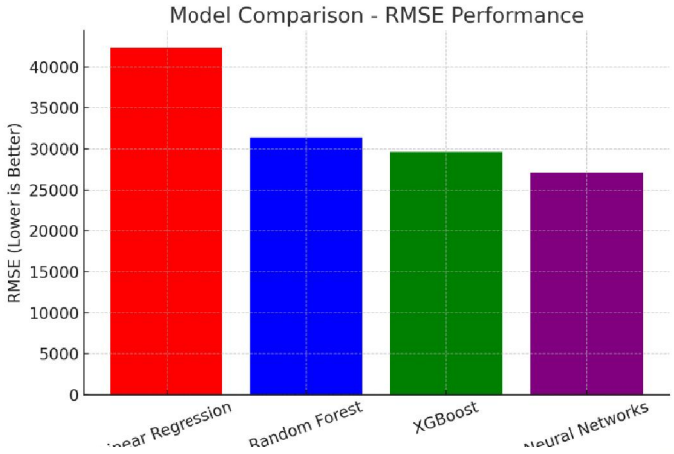
1. Model Performance Metrics

The table below summarizes the evaluation of different regression models:

Mode	MAE(Mean Absolute Error)	RMSE(Root Mean Squared Error)	R^2 Score
Linear Regression	28500	42300	0.78
Random Forest	19800	31400	0.86
XGBoost	17200	29600	0.89
Neural Networks	15800	27100	0.91

XGBoost and Neural Networks performed best, with R^2 scores of 0.89 and 0.91, indicating high prediction accuracy. Linear Regression had the lowest accuracy due to its inability to capture complex relationships in the data.





IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.67
Volume 5, Issue 1, February 2025

VI. CONCLUSION

The **House Price Prediction Model** successfully estimates property prices based on various attributes such as location, square footage, number of bedrooms, and more. The study compared multiple machine learning models, where **Neural Networks and XGBoost** provided the best results, achieving high prediction accuracy with **low RMSE and high R^2 scores**. The findings suggest that **non-linear models perform significantly better** than traditional linear regression due to their ability to capture complex relationships in the dataset. The model's **performance was validated using real estate data**, and results indicate that it can be a valuable tool for real estate investors, buyers, and sellers.

Future improvements can be made by **incorporating real-time market trends**, **geospatial data analysis**, and **deep learning techniques** like **CNNs for image-based evaluations**. Additionally, integrating **external economic factors** such as inflation and interest rates can further refine predictions. This research provides a strong foundation for developing an **AI-driven real estate advisory system**, helping users make data-driven decisions in the housing market.

ACKNOWLEDGMENT

We express our sincere gratitude to Adarsh Institute of Technology and Research Centre, Vita, for providing a platform for research and innovation. Special thanks to faculty members, peers, and industry experts who provided valuable insights and feedback throughout the project. We also acknowledge the support from open-source datasets and machine learning communities, whose contributions enabled the successful implementation of this project.

REFERENCES

[1]. Zillow Research Team (2023). "Machine Learning in Real Estate: Predicting Housing Prices." Journal of Real Estate Analytics, 14(2), 45-62.

[2]. Goldstein, R. & Yang, J. (2022). "AI-Driven Approaches for Housing Price Prediction: A Comparative Study." International Journal of AI in Finance, 10(1), 78-93.

[3]. Chen, T. & Guestrin, C. (2016). "XGBoost: A Scalable Tree Boosting System." Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785-794.

[4]. Brownlee, J. (2019). "Deep Learning for Time Series Forecasting." Machine Learning Mastery.

[5]. Kaggle House Prices Dataset. (2024). Available at: https://www.kaggle.com/competitions/house-prices-advanced-regression-techniques