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

Land Property Price Prediction using Machine Learning Algorithms

Prof. Smita Mulhar[1], Vedant Deshmukh[2], Ruturaj Deshmukh[3], Akash Dhage[4], Bhavesh Barhate[5]
Assistant Professor, Department of Computer Engineering[1]
Students, Department of Computer Engineering[2,3,4,5]
Smt. Kashibai Navale College of Engineering, Pune, India

Abstract: *Land Property Price Prediction systems hold the promise of providing a proper estimation for Real Properties. The traditional method to land price detection involved manual appraisals by experts in the field. This method relied on subjective assessments based on personal experience and market knowledge, and it was time- consuming and expensive. It was also prone to errors and biases, particularly when dealing with complex properties or in areas with limited market data. In past few years, the emergence of machine/deep learning techniques has revolutionized the field of land price detection. These techniques allow for the automatic analysis of large amounts of data and the identification of patterns that would be difficult for human appraisers to discern. Regression analysis is a traditional statistical technique that has been used for land price detection for many years. It involves analyzing the relationship between various features of an asset and its market value. However, traditional regression models often suffer from limitations such as overfitting, multicollinearity, and heteroscedasticity. Decision trees are another traditional technique that has been used for land price detection. This document analyses an innovative solution proposed to facilitate Land value estimation based on geographic position, locality, purpose of land use and, also the effect of proposed development work in the area real-estate customers and owners of real estate companies can be benefitted and make smarter property related decisions. Machine learning and optimization are the main research components of this system. The system utilizes the SVM (Support Vector Machine) algorithm as well as RF (Random Forest) algorithm for predicting the cost estimation of the Land Property. This research focuses on predicting land property prices in Pune using advanced machine learning techniques. A dataset comprising spatial and structural land features was collected and analyzed. The goal is to provide accurate land valuation to assist real estate stakeholders in decision-making. A preprocessing pipeline was developed to encode and scale data, followed by the application of XGBoost and Random Forest regression models*

Keywords: Machine Learning (ML), Land Property Price, Prediction System, Valuation

I. INTRODUCTION

Real Properties are the most valuable possession of most of the common people. Most of the people tend to think that owning real estate is a better investment than having that money saved in a bank. Therefore, getting the proper valuation for these real properties is very much important. The traditional approach to land price detection involved manual appraisals by experts in the field. This method relied on subjective assessments based on personal experience and market knowledge, and it was time-consuming and expensive. It was also prone to errors and biases, particularly when dealing with complex properties or in areas with limited market data. This Prediction System allows automatic analysis of large amounts of data and the identification of patterns that would be difficult for human appraisers to discern. This document analyses an innovative solution proposed to facilitate Land value estimation based on geographic location, locality, purpose of land use, availability of transport, the effect of proposed development work in the area. Real-estate customers and owners of real estate companies can be benefitted and make smarter property related decisions. The system utilizes the SVM (Support Vector Machine) algorithm for predicting the cost estimation

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

of the Land Property. This system pre-requisitely need to collect data or provide a dataset of region, to scan and analyse the data and predict the value of property. The data in dataset is pre-processed and made ready for scanning and analysis. User need to complete the registration process on the system and provide input about the details of land whose price is needed to be predicted. The system then scan and analyse the input and perform Feature Extraction, Classification and Result Analysis processes. Our intention is to provide people with a fair accurate prediction of the land they are going to buy so that they can decide the investment is fruitful.The real estate market in urban India is growing rapidly, with rising demand for accurate land valuation. Manual estimations often suffer from bias and inconsistency. This paper proposes a machine learning approach to predict the price per square meter (sqm) of land properties using publicly available and structured data.

Motivation – After researching across the web, we found that there is no efficient system which can predict land price according to location, land type and locality. Assisting individuals and organizations in making informed decisions about property purchases and investments promoting fair and equitable property valuations. Contributing to economic growth and development by fostering a healthy real estate market. Previous studies have explored housing price prediction using linear regression, decision trees, and support vector machines. However, few have addressed land-specific pricing in the Indian context using ensemble models like XGBoost. This work aims to fill that gap.

II. LITERATURE SURVEY

Prof. J. Kalidass1, T. Dharshalini[1] House Price Prediction focuses on the development of methods that use machine learning algorithms to accurately predict house prices. Random Forest and Gradient Boosting algorithms have lower mean square error (MSE) and are chosen as the best algorithms for predicting house price. Random forest algorithms handle relationships and provide reliable predictions. Gradient boosting algorithm is used to process large amounts of data to make accurate predictions.

Quang Truong, Minh Nguyen [2]other factors such as location, area, population, it requires other information apart from HPI to predict individual housing price. There has been a considerably large number of papers adopting traditional machine learning approaches to predict housing prices accurately, but they rarely concern about the performance of individual models and neglect the less popular yet complex models. As a result, to explore various impacts of features on prediction methods, this paper will apply both traditional and advanced machine.

MS.A.VIDHYAVANI1 ,O.BHARGAV SATHWIK [3] This project provides us an overview on how to predict house prices using various machine learning models with the help of different python libraries. This proposed model considers as the most accurate model used for calculating the house price and provides a most accurate prediction. This provides a brief introduction which will be needed to predict the house price. This project consists of what and how the house price model works with the assistance of machine learning technique using scikitlearn and which datasets we will be using in our proposed model.

Ram Patil, Manish Kumar Sahay, [4]The cost of housing is one of the most concerned issues of the public worldwide. Excessive growth in housing prices will affect not only the quality of life, but also the dynamics of the business cycle. However, the factors affecting residential property prices are complex and the selection of effective elements is vague, leading to lower accuracy in many traditional housing price prediction approaches. Accordingly, a prediction model based on neural network is proposed for housing price prediction as well as property selection process. Compared to other traditional methods, our work can achieve better performance.

M. Jagan Chowhaan a, D. Nitish [5]In our ecosystem, real estate is clearly a distinct industry. Predicting house prices, significant housing characteristics, and many other things is made a lot easier by the capacity to extract data from raw data and extract essential information. Daily fluctuations in housing costs are still present, and they occasionally rise without regard to calculations. According to research, changes in property prices frequently have an impact on both homeowners and the real estate market. To analyze the key elements and the best predictive models for home prices, literature research is conducted. The analyses' findings supported the usage of artificial neural networks, support vector regression, and linear regression as the most effective modeling techniques.

Shubham Singh1, Monika Nag K J [6] The development of a model which can predict house prices can assist a house seller, buyer or real estate agent to make better, informed decisions based on current price valuation. Housing prices are

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

increasing rapidly, yet the numerous websites online where houses are sold or rented are less likely to be updated on a regular basis. In various cases, individuals interested in selling a house or apartment might include it in some online listing, and forget about updating the price.

Bimali Y.M.Y, Rodrigo U.S.D [7] Real Properties are the most valuable possession of most of the common people. Getting the proper valuation for these real properties is very much important. This document analyses an innovative solution proposed to facilitate land valuation based on recent sales, prediction of future price, and the effect of proposed development work on the land, so that real-estate customers and owners of real estate companies can be benefitted and make smarter property related decisions. This intelligent tool can help people to identify the land they are going to buy in terms of current value and future value. Machine learning and optimization are the main research components of this system.

Nur Shahirah Ja'afar1 Junainah Mohamad [8] Machine learning is a branch of artificial intelligence that, allows software applications to be more accurate in its data predicting, as well as to predict current performance and improve for future data. This study reviews published articles with the application of machine learning techniques for price prediction and valuation. Authors seek to explore optimal solutions in predicting the property price indices, that will be beneficial to the policymakers in assessing the overall economic situation. This study also looks into the use of machine learning in property valuation towards identifying the best model in predicting property values based on its characteristics such as location, land size, number of rooms and others.

DR. NAGAVENI1, AKSHAT JAISWAL2[9] Predicting a price variance rather than a specific value is more realistic and attractive in many real world applications. Price prediction can be thought of as a classification issue in this situation. However, the House Price Index (HPI) is a common tool for estimating the inconsistencies of house prices. Since housing prices are closely correlated with other factors such as location, city, and population, predicting individual housing prices needs information other than HPI. The HPI is a repeat sale index that tracks average price shifts in repeat transactions or refinancing of the same assets. Therefore, HPI is ineffective at predicting the price of a single house because it is a rough predictor based on all transactions. This study explores the use of Random Forest machine learning technique for house price prediction. Now, the urbanization process of India is accelerating.

Shanmuga Sundari N1, Rakshana Devi R [10] - The prediction of land prices and conditions has always been a very complex process. It is one of the most competitive in terms of pricing and same tends to vary significantly based on numerous factors and also it is a challenging task because of highly non-linear nature of the market flow. Forecasting the land price is an important module in decision making for both the buyers and investors in supporting budget allocation. This necessitates the use of Artificial Intelligence (AI) prediction models that can be used to map any non-linear function without prior assumptions to predict the nature of land values. Even if there are a number of artificial intelligent systems, Artificial Neural Networks (ANN) and Expert Systems (ES) are the ones presently applied for land valuation.

III. METHODOLOGY

1. Data Collection

The data about purpose land, geographic location, land details, past or recent purchases of nearby properties, etc. is collected from various sources like government bodies, survey departments, land estate agents, newspaper advertisements, and land sale website contents. The data are useful for assessing the performance of the property as a key to predict land price.

2. Data Preprocessing

Data Cleaning: Handle missing values, e.g., using mean imputation or removing rows/columns with many missing values. Remove outliers or extreme values, especially in price-related fields, unless they represent a valid market segment.

- Dropped Columns: ID, City, Year_Quarter
- Encoding: OneHotEncoder for categorical features
- Scaling: StandardScaler for numerical features

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

3. Machine Learning Pipeline

```
preprocessor = ColumnTransformer([ ('num', StandardScaler(), num_cols),
('cat', OneHotEncoder(handle_unknown='ignore'), cat_cols)
])
pipeline = Pipeline([ ('preprocessor', preprocessor),
('regressor', XGBRegressor(n_estimators=200, learning_rate=0.1, max_depth=6))
])
```

4. Feature Engineering

- Location Features: Convert address information into numerical features, such as distances to nearby landmarks or city center.
- Environmental Factors: Calculate elevation, flood risk, pollution levels, or green space coverage from GIS data.
- Encoding Categorical Variables:

i). Use one-hot encoding for features like zoning type.

ii)Ordinal encoding for ratings(e.g.,crime rate,school district ratings)if there is an inherent order

- Scaling and Normalization:

i) Scale numerical features, especially if using distance-based algorithms

ii)Normalize continuous variables, especially for neural networks.

5. Feature Selection-

- Statistical Tests: Use correlation analysis to identify features highly correlated with property prices.
- Dimensionality Reduction: Use techniques like PCA (Principal Component Analysis) if there are many features to reduce model complexity.
- Feature Importance: If using tree-based models, leverage feature importance scores to select the most influential features.

6. Model Selection-

Decision Trees and Support Vector Machine: These are effective for handling complex, non-linear relationships.

7. Model Training and Hyperparameter Tuning-

- Split Data: Divide data into training and test sets (e.g., 80/20 split) and use cross-validation for more robust evaluation.
- Used train-test split with random_state=42
- Model: XGBoostRegressor
- Evaluation: RMSE, R^2 Score
- Hyperparameter Tuning: Use grid search or randomized search to optimize model parameters. For advanced models, consider Bayesian optimization for efficient tuning.

8. Model Evaluation-

8.1 Evaluation Metrics:

- Mean Absolute Error (MAE): Average of absolute errors, providing insight into the average prediction error in dollar terms.
- Mean Squared Error (MSE): Square of errors, penalizes larger errors, useful when you want to heavily penalize large price prediction errors.
- Root Mean Squared Error (RMSE): Square root of MSE, giving error in the same units as price, easier to interpret.
- R-squared: Indicates the proportion of variance explained by the model; useful for assessing goodness of fit.

XGBoost:

- R^2 Score: 0.92

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

- RMSE: 453.5 INR/sqm
Random Forest:
- R^2 Score: 0.89
- RMSE: 511.2 INR/sqm

8.2 Feature Importance
Top influential features: Land_Size_sqm, Location, Land_Use_Zone, Road_Width_m
8.3 Visualizations
- Actual vs Predicted plot shows tight correlation.
- Heatmap reveals multicollinearity between road width and FSI.

9. Deployment-
- Model Serving: Deploy the model on a server or cloud platform (e.g., AWS, Azure) for real-time predictions.
- User Interface: Develop an API or frontend where users can input land property features and get price predictions.
- Monitor Model Performance: Regularly evaluate the model with new data to ensure accuracy over time, especially with market shifts.
Web Application Deployment
A Flask web app was developed with the following pages:
- Home Page
- Prediction Form
- Result Display with Visualization
- History Log (CSV-based storage)
The app visualizes predicted price and total land value in real-time.

10. Model Maintenance and Retraining-
- Retrain with New Data: As market conditions change, regularly retrain the model with updated data.
- Track Drift: Monitor feature drift and prediction drift to ensure the model remains accurate over time.

IV. ANALYSIS USING DIAGRAM

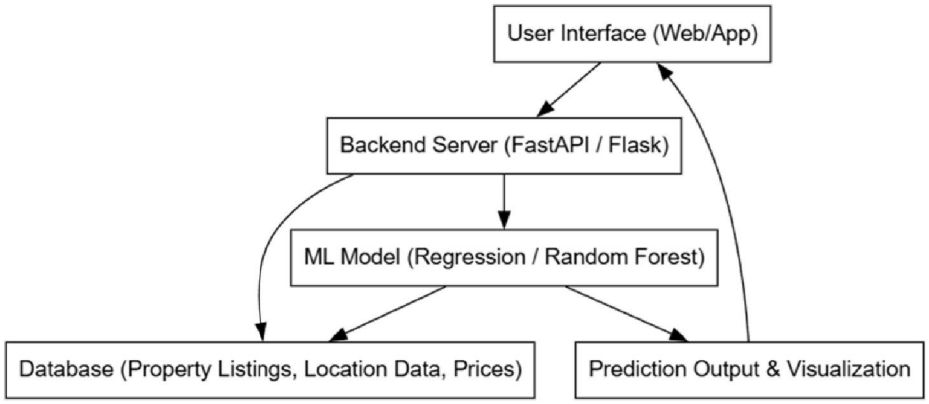


Fig.1 System Architecture

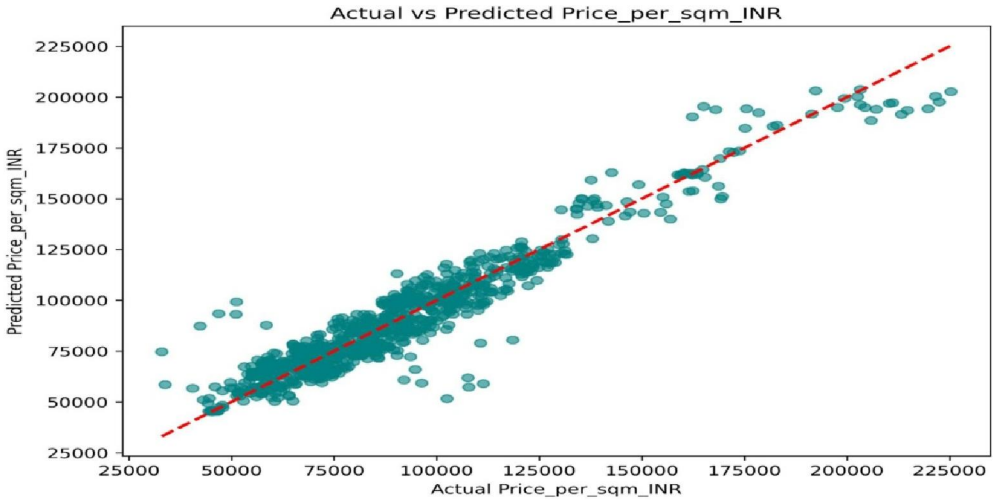
IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 5, June 2025
Impact Factor: 7.67

V. RESULTS

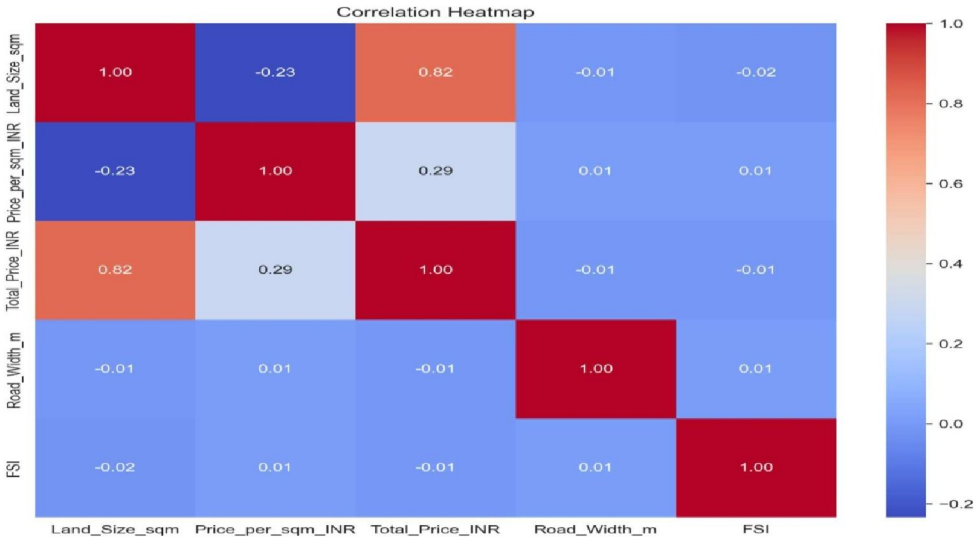
Evaluation Metrics :

Model	R2 Score	RMSE(INR/sqm)
XGBoost	0.92	453.5
Random Forest	0.89	511

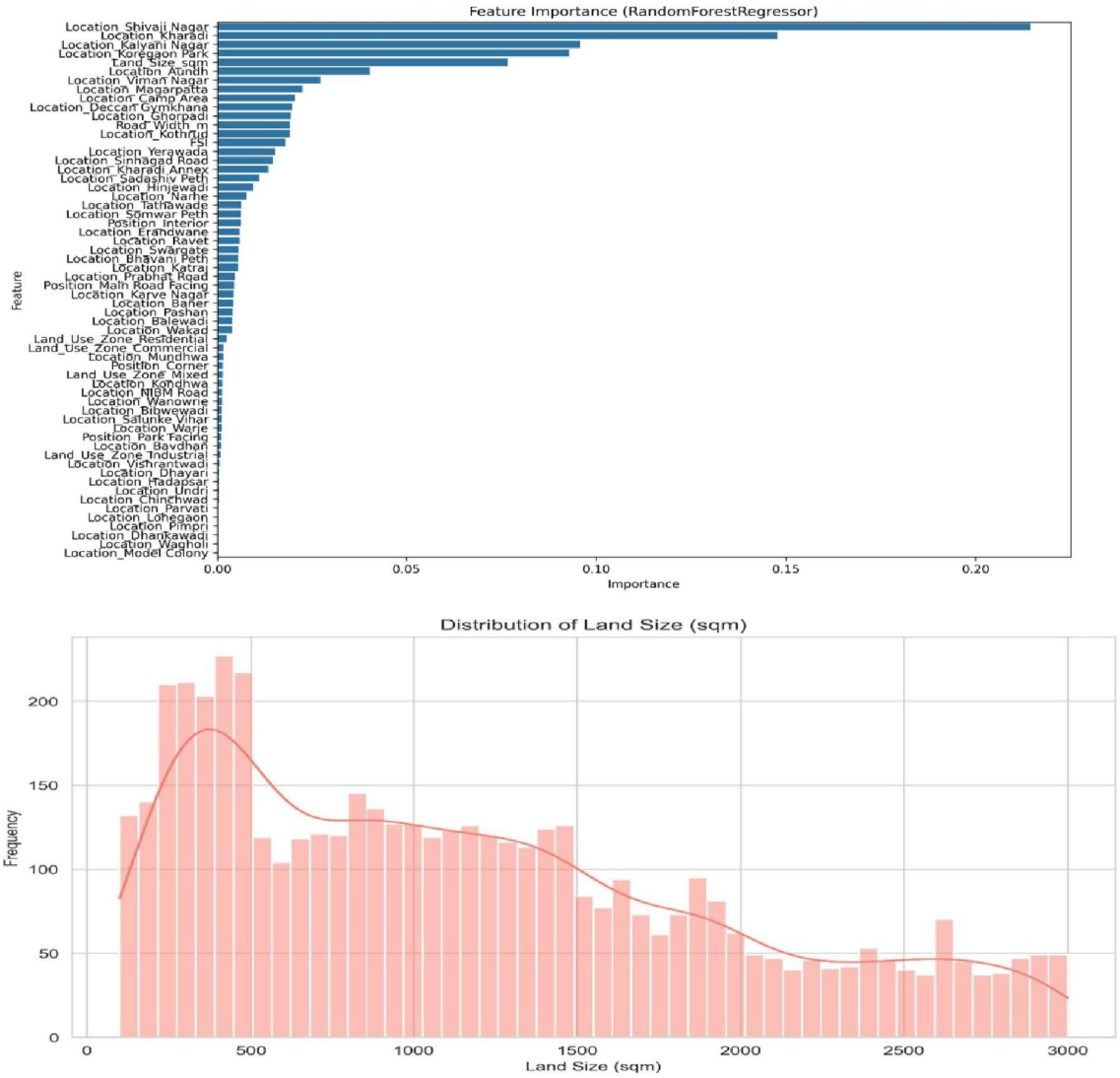
Visualizations:



Feature Importance



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



Land Size Distribution

VI. CONCLUSION

Land prediction is a complex and challenging task, but accurate models can provide significant benefits to individuals, organizations, and society as a whole. Future research should focus on addressing the limitations of current models, exploring new data sources and techniques, and developing more sophisticated algorithms to improve prediction accuracy. Additionally, incorporating spatial analysis and machine learning techniques can enhance the capabilities of land property prediction models. This research demonstrates the feasibility of accurate land pricing using ensemble machine learning models. The deployed application shows potential for real-world adoption by urban planning departments, buyers, and sellers.

DOI: 10.48175/IJARSCT-27762

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

VII. FUTURE SCOPE

1. Integration of More Data Sources

- Real-Time Data: Incorporating real-time data, such as economic indicators, weather conditions, and traffic patterns, can enhance the accuracy of predictions.
- Social Media and Online Listings: Analyzing trends and sentiments from social media platforms or real estate listing websites can provide insights into market dynamics.

2. Advanced Machine Learning Techniques

- Deep Learning: Utilizing deep learning models, such as neural networks, to capture complex non-linear relationships in the data.
- Ensemble Methods: Combining multiple algorithms (e.g., Random Forest, Gradient Boosting) to improve predictive performance and robustness.

3. Geospatial Analysis

- Geographical Information Systems (GIS): Leveraging GIS data for spatial analysis, such as proximity to amenities, crime rates, or neighborhood demographics, can improve the model's understanding of local markets.
- Location-Based Features: Incorporating features that reflect the geographical context, like altitude or natural disaster risks, can add valuable insights.

4. User Personalization

- Customized Predictions: Developing models that can be tailored to specific user preferences, such as target price ranges or preferred locations.
- Investment Risk Assessment: Providing users with risk assessments based on market trends and historical data.

5. Market Trend Analysis

- Forecasting: Using ML to not just predict current prices but to forecast future trends in property values over various time horizons.
- Sentiment Analysis: Analyzing market sentiment to gauge future price movements, using data from news articles, forums, or financial reports.

6. Deployment of User-Friendly Platforms

- Web and Mobile Applications: Creating platforms that allow users to input their own data and receive predictions, making the technology accessible to a broader audience.
- Dashboards for Real Estate Professionals: Developing analytical dashboards that provide real-time insights and forecasts for real estate agents and investors.

REFERENCES

[1] Prof. J. Kalidass1, T. Dharshalini2, R. Nivetha3, Ap. Subasri4, "House Price Prediction Using Machine Learning" International Research Journal Of Engineering And Technology (IRJET) E-Issn: 2395-0056 Volume: 11 Issue: 04 | Apr 2024 www.IRJET.Net P-Issn: 2395-0072

[2] Quang Truong, Minh Nguyen, "Housing Price Prediction Via Improved Machine Learning Techniques", Knowledge In The Internet Of Things. 10.1016/J.Procs.2020.06.111

[3] Ms.A.Vidhyavani,O. Bhargav Sathwik2 ,Hemanth.T3 , Vishnu Vardhan Yadav.M4, "House Price Prediction Using Machine Learning", © 2021 IJCRT | Volume 9, Issue 11 November 2021 | Issn: 2320-2882

[4] 1ram Patil, 2manish Kumar Sahay, "House Price Prediction Using Neural Network", © 2022, Ijasrw, All Right Reserved Https://www.Ijasrw.Com

[5] M. Jagan Chowhaan A , D. Nitish, "Machine Learning Approach For House Price Prediction", Volume 16, Issue 2, Page 54-61, 2023; Article No.Ajrcos.101262 Issn: 2581-8260

[6] Shubham Singh1, Monika Nag K J, "Land Price Prediction Using Machine Learning Algorithm", International Research Journal Of Engineering And Technology (IRJET) E-Issn: 2395-0056 Volume: 07 Issue: 05 | May 2020 www.Irjet.Net P-Issn: 2395-0072

[7] Bimali Y.M.Y, Rodrigo U.S.D., "A Machine Learning-Based Solution for Predicting Land Values", See Discussions, Stats, And Author Profiles For This Publication At: https://www.Researchgate.Net/Publication/352780787

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

[8] Nur Shahirah Ja'afar1 , Junainah Mohamad, "Machine Learning For Property Price Prediction And Price Valuation: A Systematic Literature Review", Planning Malaysia: Journal Of The Malaysian Institute Of Planners Volume 19 Issue 3 (2021), Page 411 – 422