AMERICAN JOURNAL OF AI CYBER COMPUTING MANAGEMENT
Peer Reviewed, Referred & Indexed Journal
E-ISSN:3069-0102
VOL.6, NO. 2(2026)

Predictive House Prices Using Machine Learning and Techniques

[1]Kuchipudi Krishna Vinay,[2]Garapati Jaya Lakshmi Akhil,[3]Mukku Lakshmi Naga Sasanka Varma,[4]Peethala Naga Venkata Vijay Kumar, [5]Mr.P. Krupavaram

[1,2,3,4]U. G Student, Department of CSE (Artificial Intelligence & Machine Learning),
D.N.R. COLLEGE OF ENGINEERING & TECHNOLOGY (AUTONOMOUS)
Balusumudi, Bhimavaram, West Godavari District, Andhra Pradesh -534202

[5]Assistant Professor, Department of CSE (Artificial Intelligence & Machine Learning),
D.N.R. COLLEGE OF ENGINEERING & TECHNOLOGY (AUTONOMOUS)
Balusumudi, Bhimavaram, West Godavari District, Andhra Pradesh -534202

ABSTRACT

House price prediction is a crucial task in the real estate sector for supporting buyers, sellers, and investors in making informed decisions. With the rapid growth of data, machine learning techniques have become effective tools for analyzing housing market trends. This project focuses on predicting house prices using advanced models such as Random Forest, Artificial Neural Networks (ANN) including LSTM, XGBoost, and Gradient Boosting by considering important features such as location, area, and number of rooms. Data preprocessing techniques, including normalization and feature selection, are applied to improve model performance and reliability. The models are trained using historical housing price data and compared with traditional approaches such as Linear Regression. Ensemble methods like Random Forest, XGBoost, and Gradient Boosting enhance prediction accuracy, while LSTM captures sequential patterns in data. Model performance is evaluated using metrics like Mean Squared Error. Experimental results show that the proposed models achieve higher accuracy compared to conventional methods. The system provides a reliable, scalable, and efficient solution for real estate price prediction and decision-making.

KEYWORDS: - Artificial Neural Networks (ANN) including LSTM, XGBoost, and Gradient Boosting, Machine Learning

INTRODUCTION

The real estate market plays a significant role in economic development and urban planning. Accurate house price prediction is essential for investors, home buyers, and policymakers to make informed financial decisions. Traditional prediction methods mainly rely on statistical and linear regression models, which often fail to capture complex and non-linear relationships among housing features. With the advancement of machine learning, advanced algorithms such as Random Forest, Gradient Boosting, and XGBoost have demonstrated improved predictive performance. Artificial Neural Networks (ANN), including Multilayer Perceptron and LSTM models, are capable of learning hidden patterns and handling both structured and sequential data effectively. These models can analyze large-scale datasets and uncover intricate relationships between features such as location, area, and amenities.

This study applies multiple machine learning techniques to enhance the accuracy of house price prediction. The proposed system compares the performance of ANN, LSTM, Random Forest, XGBoost, and Gradient Boosting with conventional regression approaches. Ensemble methods further improve prediction stability and reduce overfitting. By integrating intelligent learning models, the system aims to deliver accurate, efficient, and data-driven predictions. Overall, the approach provides a robust solution for real-world real estate analysis and decision-making.

RELATED WORK

Recent research in house price prediction has increasingly focused on the use of machine learning techniques to improve accuracy over traditional statistical methods. Early studies primarily used Linear Regression models, which were limited in handling non-linear relationships between housing features. To overcome this, researchers introduced algorithms such as Random Forest and Decision Trees, which provided better performance by capturing complex feature interactions. Gradient Boosting and XGBoost have been widely adopted due to their ability to enhance prediction accuracy through ensemble learning. Artificial Neural Networks (ANN) have also been applied to model intricate patterns in large datasets. More recent studies incorporate Long Short-Term Memory (LSTM) networks to capture temporal and sequential trends in real estate data. Comparative analyses show that ensemble methods and deep learning models outperform conventional approaches. Feature engineering techniques, including location-based and economic indicators, further improve model performance. Hybrid models combining multiple algorithms have also been explored for robust predictions. Overall, existing research demonstrates that advanced machine learning techniques significantly enhance the reliability and accuracy of house price prediction systems.

LITERATURE SURVEY

The Literature on house price prediction highlights the growing importance of machine learning techniques in improving prediction accuracy and efficiency. Early studies by David H. Wolpert and William G. Macready introduced the No Free Lunch Theorem, emphasizing that no single algorithm performs best for all problems, thereby supporting the use of multiple models for prediction. Research by Uyanık G.K. and Güler N. demonstrated the effectiveness of Multiple Linear Regression while highlighting its limitations in handling complex, non-linear relationships. Studies by Annina S and Mahima SD showed that machine learning techniques significantly improve prediction systems compared to traditional approaches. Real-world datasets such as Svensk Mäklarstatistik provide valuable insights into housing market trends and are widely used for model training and validation. Recent research focuses on advanced algorithms like Random Forest, Gradient Boosting, and XGBoost for better predictive performance. Artificial Neural Networks, including LSTM models, are applied to capture hidden and temporal patterns in housing data. Ensemble methods further enhance model robustness and accuracy. Feature engineering techniques, such as location and economic indicators, improve prediction quality. Overall, the literature confirms that combining multiple machine learning models leads to more accurate and reliable house price prediction systems.

EXISTING METHOD

The existing house price prediction system mainly depends on traditional statistical and machine learning techniques such as Linear Regression and basic regression models. These approaches assume a linear relationship between housing features and prices, which limits their ability to capture complex and non-linear patterns in the real estate market. Manual feature selection and data preprocessing are often required, making the process time-consuming and prone to human error. Existing systems are sensitive to noise and missing data, which reduces prediction accuracy. They also struggle to handle large and diverse datasets efficiently. Performance evaluation is usually limited to simple error metrics without robust optimization. Moreover, these systems operate mostly in offline environments and lack real-time prediction capabilities. As a result, the existing system provides less accurate, less flexible, and less reliable house price predictions.

PROPOSED METHOD

AMERICAN JOURNAL OF AI CYBER COMPUTING MANAGEMENT
Peer Reviewed, Referred & Indexed Journal
E-ISSN:3069-0102
VOL.6, NO. 2(2026)

The proposed system implements a machine learning-based framework for accurate house price prediction using multiple advanced algorithms. Initially, historical housing data is collected and preprocessed by handling missing values, normalization, and feature selection to improve data quality. Important features such as location, area, number of rooms, and amenities are extracted for model training. The system applies multiple models including Random Forest, Artificial Neural Networks (ANN) with LSTM, XGBoost, and Gradient Boosting for prediction. Each model is trained on the dataset to learn complex relationships between input features and house prices. Ensemble learning techniques are used to enhance prediction accuracy and reduce overfitting. LSTM models capture sequential and temporal patterns present in the data. The system compares model performance using evaluation metrics such as Mean Squared Error and accuracy. The best-performing model is selected for final prediction and deployment.

SYSTEM ARCHITECTURE

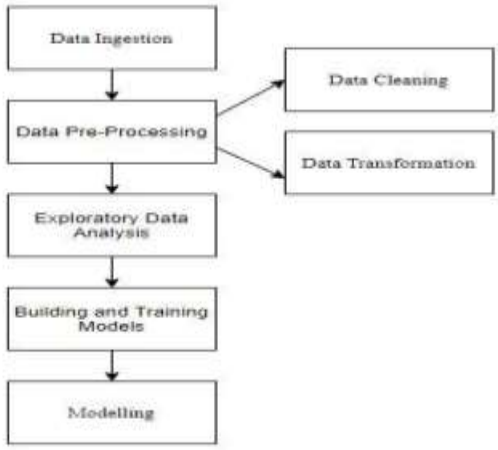

Figure 1: Architecture of the Project

METHODOLOGY DESCRIPTION

Data Collection: This stage involves gathering housing datasets from various sources such as CSV files, real estate websites, or public datasets. The collected data includes important features like location, area, number of rooms, and price. The data is stored in a structured format for further processing. Flask can be used to upload datasets through an HTML interface. This step ensures that sufficient and relevant data is available for building accurate prediction models.

Pre-processing and Visualization

In this stage, the collected data is cleaned and prepared for analysis. Missing values are handled, categorical features are encoded, and numerical data is normalized. Data visualization techniques such as graphs and plots are used to understand trends and relationships. Libraries like Matplotlib or Seaborn help in visual representation. This step improves data quality and provides insights before model training.

Data Splitting: The processed dataset is divided into training and testing sets to evaluate model performance. Typically, a percentage of data is used for training while the remaining is reserved for testing. This ensures that the model can generalize well to unseen data. The splitting process helps avoid overfitting and improves reliability. It is an essential step before applying machine learning algorithms.

Model Training and Saving: In this stage, multiple machine learning models such as Random Forest, ANN (LSTM), XGBoost, and Gradient Boosting are trained using the training dataset. Each model learns patterns and relationships between features and house prices. Performance is evaluated to identify the best model. After training, models are saved as .pkl files for future use. This enables faster predictions without retraining the models again.

Prediction and Result Display: This stage allows users to enter input details such as location, area, and number of rooms through an HTML form. The input data is preprocessed similarly to the training data. The saved model is loaded in Flask and used to predict house prices. The results are displayed dynamically on a web page. This module provides a user-friendly interface for real-time house price prediction.

RESULTS AND DISCUSSION

This project shows the details of profile how we can detect easily.

Figure 2.1: Flask Development Server Running in Anaconda Prompt

A terminal window running a Flask-based web application for predictive modeling using Python. It shows the development server status, debug mode, and local hosting on http://127.0.0.1:5000.

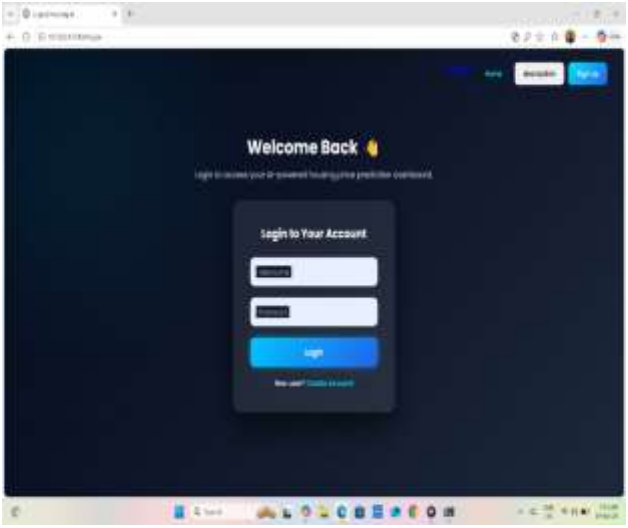

Figure 2.2: Login Page

An AI-powered web application login interface for accessing housing price prediction features built using Python. It allows users to securely enter credentials and navigate to the prediction dashboard.

AMERICAN JOURNAL OF AI CYBER COMPUTING MANAGEMENT
Peer Reviewed, Referred & Indexed Journal
E-ISSN:3069-0102
VOL.6, NO. 2(2026)

Figure 2.3: Dashboard Interface page

An AI-powered dashboard for predicting housing prices based on user inputs such as location, size, and amenities using Machine Learning. It provides prediction insights and allows users to enter property details for accurate estimation.

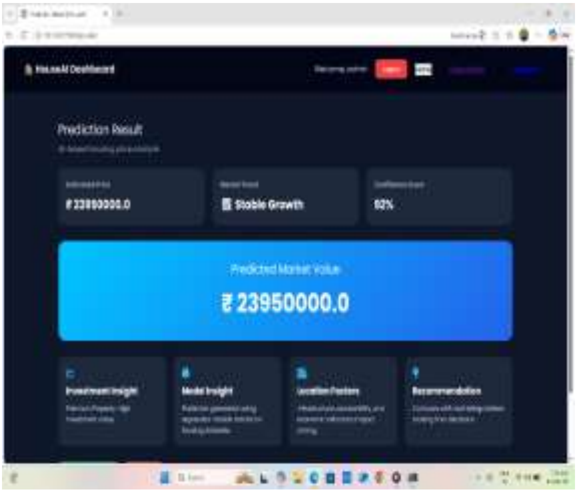

Figure 2.4: Prediction Result Page

An AI-based results interface displaying predicted housing prices, market trends, and confidence score using Machine Learning. It provides insights such as investment value, model analysis, and location-based factors.

CONCLSUION

The house price prediction system developed in this project demonstrates the effectiveness of machine learning techniques in real estate analysis. By using advanced models such as Random Forest, ANN (LSTM), XGBoost, and Gradient Boosting, the system achieves higher accuracy compared to traditional methods. Data preprocessing and feature selection significantly improve model performance and reliability. The integration of Flask and HTML provides a user-friendly interface for real-time predictions. Overall, the system offers a scalable and efficient solution for accurate house price estimation and decision-making.

FUTURE SCOPE

Future enhancements can focus on incorporating more advanced deep learning models and larger datasets to further improve prediction accuracy. Integration of real-time data sources such as market trends and economic indicators can make predictions more dynamic. The system can be extended to include location-based recommendations using GIS mapping. Deployment as a mobile or cloud-based application can increase accessibility for users. Additionally, incorporating explainable AI techniques can help users better understand the prediction results and improve trust in the system.

REFERENCES

1. Abiodun, O. I., et al. "State-of-the-art in artificial neural network applications: A survey." Heliyon, 2018.
2. Alzubaidi, L., et al. "Artificial Neural Networks: A Review of Applications in Real Estate Prediction." Journal of Real Estate Research, 2020.
3. Bishop, C. M. Pattern Recognition and Machine Learning. Springer, 2006.
4. Breiman, L. "Random Forests." Machine Learning, 2001.
5. Choudhury, S., et al. "Predicting Housing Prices Using Machine Learning Techniques." International Journal of Computer Applications, 2019.
6. Cortes, C., & Vapnik, V. "Support-Vector Networks." Machine Learning, 1995.
7. Friedman, J., et al. The Elements of Statistical Learning. Springer, 2001.
8. Gardner, M. W., & Dorling, S. R. "Artificial neural networks (the multilayer perceptron)—A review of applications in the atmospheric sciences." Atmospheric Environment, 1998.
9. Ghahramani, Z. "Probabilistic Machine Learning and Artificial Intelligence." Nature, 2015.
10. Goodfellow, I., et al. Deep Learning. MIT Press, 2016.
11. Gu, Q., et al. "House Price Prediction Based on Machine Learning Algorithms." IEEE Access, 2021.
12. Haykin, S. Neural Networks and Learning Machines. Pearson, 2009.
13. He, K., et al. "Deep Residual Learning for Image Recognition." CVPR, 2016.
14. Kotsiantis, S. B., et al. "Supervised Machine Learning: A Review of Classification Techniques." Emerging Artificial Intelligence Applications in Computer Engineering, 2007.
15. Li, W., et al. "A Hybrid Machine Learning Approach for House Price Prediction." Expert Systems with Applications, 2020.
16. Liu, H., et al. "Predicting Real Estate Prices Using Deep Learning Techniques." Journal of Big Data, 2021.
17. Mansour, M., et al. "Data Preprocessing in Machine Learning: Techniques and Methods." Journal of Artificial Intelligence Research, 2019.
18. Ng, A. Machine Learning Yearning. 2018.
19. Ozbayoglu, A. M., et al. "Deep Learning for Financial and Real Estate Forecasting." Applied Soft Computing, 2020.
20. Pan, S. J., & Yang, Q. "A Survey on Transfer Learning." IEEE Transactions on Knowledge and Data Engineering, 2010.
21. Raschka, S., & Mirjalili, V. Python

Machine Learning. Packt Publishing, 2019.

22. Ribeiro, M. T., et al. "'Why Should I Trust You?' Explaining the Predictions of Any Classifier." KDD, 2016.
23. Rumelhart, D. E., et al. "Learning representations by back-propagating errors." Nature, 1986.
24. Schmidhuber, J. "Deep Learning in Neural Networks: An Overview." Neural Networks, 2015.
25. Singh, A., & Kumar, P. "House Price Prediction Using Machine Learning Algorithms." International Journal of Engineering & Technology, 2020.
26. Vapnik, V. The Nature of Statistical Learning Theory. Springer, 1995.
27. Wang, H., et al. "Artificial Neural Networks for Real Estate Valuation: A Review." Sustainability, 2021.
28. Zhang, G., et al. "Forecasting with Artificial Neural Networks: The State of the Art." International Journal of Forecasting, 1998.
29. Zhao, L., & Song, H. "Machine Learning Techniques for Housing Price Prediction: A Comparative Study." Journal of Computational Science, 2020.
30. Zhou, Z.-H. Ensemble Methods: Foundations and Algorithms. CRC Press, 2012.