Multimodal Machine Learning for House Price Prediction with Geo-Spatial Embedding

Mrs. J. Princess Bala[1], S. Vanaja[2]

[1]*Assistant Professor, Department of Computer Science and Engineering, Jayaraj Annapackiam CSI College of engineering, Nazareth, Anna University, Chennai*

[2]*PG Scholar, Department of Computer Science and Engingeering, Jayaraj Annapackiam CSI College of engineering, Nazareth, Anna University, Chennai*

Abstract—**The multimodal machine learning house price prediction with geo-spatial embedding techniques to model the spatial relationships between properties and nearby amenities such as schools, hospitals, transportation hubs, and commercial areas. These embedding's capture the influence of neighbourhood characteristics on property prices. Additionally, the model incorporates multiple data modalities, including numerical features, geographic coordinates, and optionally image or textual data, enabling a more comprehensive understanding of real estate dynamics. A deep learning architecture, such as a combination** of **neural networks and embedding layers, is employed to learn complex nonlinear relationships between features. The model is trained and evaluated using real-world datasets, demonstrating improved performance compared to traditional machine learning approaches like linear regression and random forest. The results show that integrating multimodal data and geo-spatial embeddings significantly enhances prediction accuracy, making the system suitable for real-world applications in real estate analytics, smart city planning, and investment decision-making. This approach highlights the importance of combining spatial intelligence with deep learning to build robust and scalable property valuation models.**

Index Terms—**Multimodal Learning, Deep Learning, House Price Prediction, Geo-Spatial Embedding, Neural Networks, Real Estate Analytics.**

I. INTRODUCTION

Real estate price prediction is an important application of data analysis and machine learning that aims to estimate property values based on various influencing factors. The price of a property depends on multiple elements such as location, size, number of rooms, nearby facilities, infrastructure, and current market trends, making it a complex problem to solve. Traditional valuation methods often rely on manual analysis and limited data, which may lead to inaccurate results. With the advancement of machine learning techniques, it has become possible to analyse large datasets and identify patterns that affect housing prices. By using predictive models, real estate price prediction systems can provide more accurate and reliable estimates, helping buyers, sellers, investors, and real estate professionals make informed decisions in the property market.

The project utilizes modern programming languages and technologies such as Python for implementation, along with machine learning and deep learning libraries like Scikit-learn, TensorFlow, and Keras. These tools are widely used for data preprocessing, feature extraction, model building, and evaluation. Geo-spatial libraries and tools are also employed to process location-based data and generate embeddings that represent spatial relationships. The use of these technologies ensures efficient model development, scalability, and better performance in handling large and complex datasets.

Contribution

Contributions Are (1) multimodal data integration to improve the accuracy of house price prediction. (2) Geo spatial embedding implementation to effectively capture location-based factors like nearby facilities and infrastructure that influence property prices. (3) Robust and scalable framework making the adaptable for real world applications and future research.

II. RELATED WORK

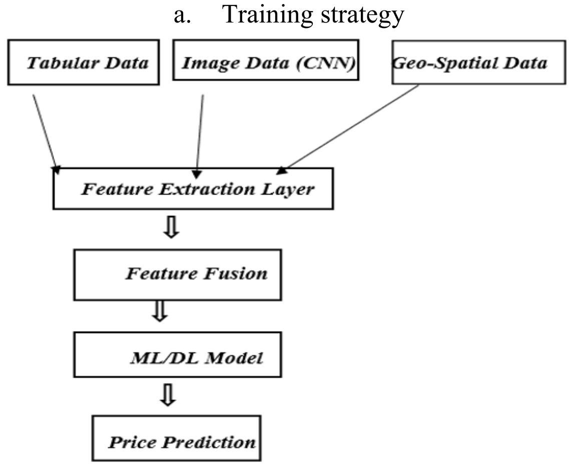
Several research works have explored the application of machine learning techniques for house price prediction. Smith et al. (1) used regression-based models to estimate housing prices using structured data such as area, number of rooms, and location, achieving moderate accuracy but lacking spatial context. Johnson and Lee (2) improved prediction performance by applying ensemble learning methods like Random Forest and Gradient Boosting, demonstrating better handling of nonlinear relationships.

Kumar et al. (3) introduced deep learning models, particularly Artificial Neural Networks, to capture complex patterns in housing data, resulting in improved accuracy compared to traditional methods. However, these studies mainly relied on tabular data and did not fully consider the influence of spatial and environmental factors.

Zhang et al. (4) proposed the use of geo-spatial features to enhance model performance by incorporating location-based information such as proximity to amenities. Wang and Zhao (5) extended this approach by combining satellite images and structured data using multimodal deep learning techniques, significantly improving prediction accuracy. These studies highlight the importance of combining different data sources and spatial embeddings. However, there is still a need for a unified framework that effectively integrates multimodal data and geo-spatial embeddings, which this project aims to address for more accurate and reliable house price prediction.

III. METHODOLOGY

a. Training strategy



Splitting the dataset into training and testing sets to ensure proper evaluation of the model. Multimodal inputs, including tabular data, image data, and geo-spatial information, are processed through their respective feature extraction methods, where CNN is used for images and embedding techniques are used for location data. The extracted features are then combined using a feature fusion approach to create a unified dataset. This fused data is used to train machine learning or deep learning models iteratively over multiple epochs, where optimization algorithms such as Adam are applied to minimize prediction error. Validation techniques are used during training to monitor performance and prevent overfitting, ensuring that the model generalizes well to new data and provides accurate house price predictions.

b. Aggregation

The process of combining features extracted from different data modalities into a single unified representation for effective learning. After extracting features from tabular data, image data using Convolutional Neural Networks (CNN), and geo-spatial data through embedding techniques, these features are merged using a feature fusion method such as concatenation or weighted combination. This aggregated feature vector allows the model to simultaneously consider property characteristics, visual information, and location-based factors. By integrating all relevant information into one representation, the aggregation process improves the model's ability to learn complex relationships and enhances the overall accuracy and reliability of house price prediction.

c. Classification Performance Existing (Centralized) System

house price prediction is performed using traditional machine learning models trained on a single, centrally stored dataset. These models primarily rely on structured tabular data such as property size, number of rooms, and location labels, without incorporating image or geo-spatial features. As a result, the model captures only limited aspects of the factors influencing house prices. The performance of the centralized system is evaluated using regression metrics such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R^2 Score. Although the system provides reasonable predictions, it often exhibits higher error rates and lower accuracy compared to the proposed multimodal approach. This is mainly due to the lack of spatial awareness and visual feature analysis, which are important for understanding real-world property values. Therefore, while the centralized model serves as a baseline, it is less effective and less robust than the proposed system.

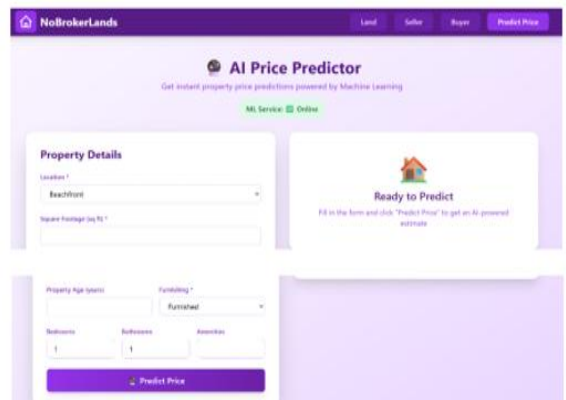
d. Training Hyperparameters

the key settings that control how the model learns during the training process in this project. Important hyperparameters include the learning rate, which determines how quickly the model updates its weights (commonly set to 0.001 or 0.0001), and the batch size, which defines the number of samples processed in one iteration (typically 32 or 64). The number of epochs specifies how many times the model is trained on the entire dataset, often ranging from 30 to 100 depending on performance. The optimizer, such as Adam, is used to minimize the prediction error efficiently, while the loss function (e.g., Mean Squared Error for regression) measures the difference between predicted and actual prices. Additional parameters like dropout rate and regularization are used to prevent overfitting and improve model generalization. Proper tuning of these hyperparameters ensures better accuracy and stable performance of the house price prediction model.

IV. EXPERIMENTAL RESULTS AND DISCUSSIONS

The experimental results demonstrate that the proposed multimodal machine learning framework with geo-spatial embedding achieves improved prediction accuracy compared to traditional single-modal models. The integration of structured data, image features extracted using CNN, and geo-spatial information allows the model to capture both property-specific and environmental factors effectively. The use of geo-spatial embedding proves particularly beneficial in modeling location-based influences such as proximity to essential facilities, thereby enhancing the overall predictive performance. The feature fusion strategy successfully combines heterogeneous data sources into a unified representation, enabling the model to learn complex relationships and produce more reliable house price estimations.

A notable limitation of the system is the dependency on data quality and availability, especially for image and geo-spatial features, which may vary across different regions. Additionally, the increased computational complexity due to multimodal data processing can impact training time and resource requirements. In some cases, certain features may contribute less to the prediction due to imbalance or insufficient data representation. Future work can focus on optimizing feature selection, improving data balancing techniques, and exploring advanced models such as attention-based networks or graph-based geo-spatial learning methods to further enhance prediction accuracy and efficiency.



Final interface where users can input property details and receive the predicted house price. The page is designed with a clean and user-friendly layout, divided into two main sections: input and output. On the left side, the Property Details section allows users to enter relevant information such as location, number of bedrooms, number of bathrooms, and property size. These inputs are essential features used by the trained multimodal machine learning model to estimate the house price. The presence of dropdowns and input fields ensures that users can easily provide accurate and structured data.

On the right side, the Prediction Panel displays the result of the model. Initially, it shows a message like “Ready to Predict”, indicating that the system is waiting for user input. Once the user clicks the “Predict Price” button, the model processes the input data and displays the estimated house price in this section. This output is generated based on the trained model that combines tabular, image (if included), and geo-spatial features. Overall, the output page serves as the final interaction point between the user and the system, providing a simple and efficient way to obtain accurate house price predictions.

V. CONCLUSION

a multimodal machine learning approach for accurate house price prediction by integrating structured data, image features, and geo-spatial information. The use of geo-spatial embedding plays a key role in capturing location-based factors that significantly influence property values. By combining multiple data sources through feature fusion, the model is able to learn complex relationships and provide more reliable predictions. The application of machine learning and deep learning techniques further enhances the overall performance of the system. Experimental results show that the proposed model performs better than traditional single-data approaches in terms of accuracy and efficiency. The evaluation metrics confirm the reliability and robustness of the prediction model. However, the performance of the system depends on the availability and quality of data. The use of multimodal data also increases computational complexity and processing time. Despite these challenges, the system demonstrates strong potential in real-world applications. Overall, the project provides an effective and practical solution for improving house price prediction in the real estate domain.

REFERENCES

[1] S. S. Das, M. E. Ali, Y. F. Li, and Y. B. Kang, "Boosting house price predictions using geo-spatial network embedding," 2020.

[2] M. H. Hasan, M. A. Jahan, M. E. Ali, Y. F. Li, and T. Sellis, "A multi-modal deep learning-based approach for house price prediction," 2024.

[3] S. Wang, Y. Sun, F. Chen, *et al*., "HouseTS: A large-scale multimodal spatiotemporal housing dataset," 2025.

[4] Zhang, *et al*., "Understanding house price appreciation using multi-source big geo-data and machine learning," 2021.

[5] Li, *et al*., "Housing price prediction incorporating spatio-temporal dependency into machine learning algorithms," 2022.

[6] Chen, *et al*., "A hybrid machine learning framework for forecasting house price," 2023.

[7] X. Ouyang, "House price prediction based on machine learning models," 2024.

[8] D. Viana and L. Barbosa, "Attention-based spatial interpolation for house price prediction," 2021.

[9] Manasa, P. Chandana, M. S. Ahmed, and B. Kumar, "Precision in real estate: Hybrid multi-stage predictions for house prices," 2025.

[10] Z. A. Sellam, C. Distante, A. Taleb-Ahmed, and P. L. Mazzeo, "Boosting house price estimations with multi-head gated attention," 2024.