International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 02 | Feb - 2025 SJIF Rating: 8.448 ISSN: 2582-3930

A Self-Attention-Driven Deep Learning Framework for House Price Prediction Using Multimodal Data

A.Pujitha[1], G.Sai Charan[2], L.Jahnavi[3], K.Sravani[4],S.Neeraja[5]

[1-4] B.Tech Student, [5] Assistant Professor, LIET
[1,2,3,4,5] Computer Science And System Engineering, Lendi Institute Of Engineering And Technology, Vizianagaram

---***---

Abstract - House price prediction remains a critical challenge in the real estate domain, requiring the consideration of diverse factors that influence housing prices. Traditional models often fail to capture the intricate relationships between these factors, resulting in limited predictive accuracy. To address this, we propose an end-to-end Joint Self-Attention Mechanism for house price prediction, integrating public facility data (e.g., parks, schools, transit) and satellite imagery to assess environmental contexts. Our approach emphasizes the identification of critical features and their interactions, enabling the discovery of complex spatial and contextual relationships.By combining the Joint Self-Attention Mechanism with Spatial Transformer Networks (STN), we aim to establish a robust and precise framework for understanding housing market dynamics. This innovative methodology introduces a new perspective on integrating spatial and contextual data for more informed and reliable house price predictions.

Key Words: House Price Prediction, Joint Self-attention Mechanism, Heterogeneous Data, Google Satellite Map ,Spatial Transformer Network

I. INTRODUCTION

The real estate market is a complex and dynamic domain influenced by numerous factors, including economic trends, infrastructure developments, and environmental conditions. Traditional house price prediction models primarily rely on historical transaction data and basic statistical methods, often failing to capture the intricate spatial and contextual relationships that drive housing prices. This results in limited predictive accuracy and challenges in making well-informed real estate decisions.

With the advancement of data-driven modeling and multimodal data integration, predictive techniques have evolved to incorporate diverse data sources, enhancing the accuracy and reliability of house price estimations. Modern frameworks leverage techniques that analyze complex interactions among property attributes, public facilities, and environmental contexts. These models can efficiently process multimodal data, including structured real estate datasets and unstructured inputs like satellite imagery and geographic information, enabling more precise and context-aware predictions.

Our proposed framework, a Self-Attention-Driven Deep Learning Model for House Price Prediction, aims to enhance price estimation accuracy by integrating multimodal data sources. The system utilizes self-attention mechanisms to identify critical features and their interdependencies, improving the model's ability to capture spatial and contextual information. Additionally, we incorporate advanced spatial feature extraction techniques to refine data processing and enhance predictive performance.

The primary objective of this research is to develop an intelligent and data-driven house price prediction framework that benefits buyers, sellers, and real estate professionals. By leveraging deep learning techniques and multimodal data fusion, our approach ensures a more transparent, efficient, and reliable price estimation process. This innovative methodology contributes to bridging the gap between traditional real estate valuation methods and modern AI-driven predictive analytics, fostering better decision-making in the housing market.

II. DESIGN AND METHODOLOGY

A Self-Attention-Driven Deep Learning Framework for House Price Prediction Using Multimodal Data is an advanced predictive modeling solution that enhances real estate valuation by integrating structured property data, environmental factors, and spatial information. The system employs multimodal data processing techniques and predictive analytics to provide a seamless and efficient house price estimation mechanism.

The system processes key property parameters such as lot size, number of rooms, building quality, year built, and living area while also incorporating contextual factors like neighborhood conditions, proximity to amenities, and land slope. Additionally, satellite imagery and environmental data are utilized to capture external influences on property valuation. By integrating these diverse data sources, the framework ensures a comprehensive and data-driven approach to house price prediction.

To refine predictions, the system employs advanced data preprocessing techniques, including feature selection using correlation analysis, categorical encoding via one-hot and label encoding, and MinMax scaling for numerical variables like lot area and square footage. These preprocessing steps help maintain data consistency and optimize model performance.

For spatial and contextual analysis, external datasets such as geographic information and public facility data are incorporated. The system is designed for real-time house price estimation, enabling users to input property details and receive instant, data-driven price predictions based on multimodal inputs.

The final implementation involves deploying a web application, allowing users to interact with the model by providing relevant property details. The integration of deep learning and multimodal data analytics enhances the accessibility, efficiency, and reliability of house price predictions. By leveraging self-attention mechanisms, the framework ensures that key features are weighted effectively, leading to more accurate and context-aware valuation in the real estate market.

Algorithms Used

a) Joint Self-Attention Mechanism

Favored algorithm for deep learning. Joint Self-Attention Mechanism (JSAM) enhances house price prediction by dynamically assigning importance to different property attributes and environmental factors. It processes multimodal data, including structured real estate parameters such as LotArea, OverallQual, YearBuilt, and GrLivArea, along with spatial features like Neighborhood, Condition1, and Condition2.

JSAM computes attention scores by analyzing the relationships between different features. Instead of treating each attribute independently, it learns dependencies using key, and value transformations to highlight significant interactions. For instance, it can determine how OverallQual influences SalePrice in different Neighborhoods while adjusting for factors like GarageCars and TotalBsmtSF.

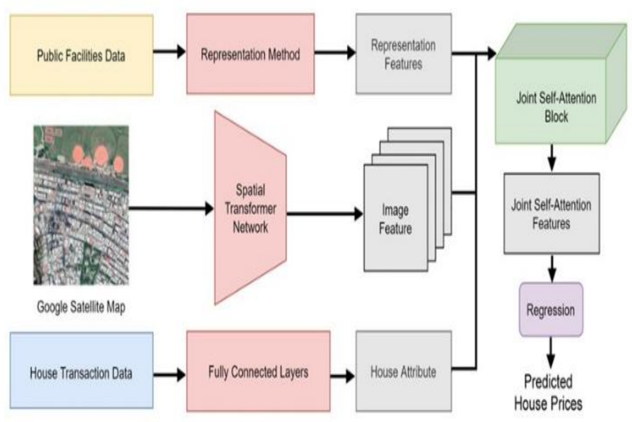
By capturing long-range dependencies and feature importance across diverse data types, JSAM improves predictive accuracy and interpretability. Unlike traditional regression models that assume linear relationships, JSAM dynamically focuses on influential variables, reducing noise and enhancing real estate valuation. The mechanism also improves efficiency by processing all property attributes in parallel, making it well-suited for large-scale house price prediction tasks.

b) Spatial Transformer Networks

Spatial Transformer Networks (STNs) refine spatial and geographic feature representations by applying adaptive transformations to align housing and environmental data effectively. The network processes location-based attributes such as Neighborhood, LandSlope, LotConfig, and Street, ensuring that spatial dependencies are accurately captured in the house price prediction model.

STNs operate by learning transformation parameters that adjust spatial relationships dynamically. A localization network predicts these parameters based on input property data, followed by a grid generator that maps transformed features onto a structured spatial representation. The final sampler applies these transformations, modifying scale, position, and orientation to standardize spatial inputs.

By dynamically transforming property and environmental attributes, STNs enhance the model's ability to focus on significant spatial features, improving prediction accuracy. Unlike static spatial feature extraction, STNs allow the model to adapt to diverse geographic layouts, ensuring more reliable house price estimates across different urban and rural environments. This adaptability makes STNs essential for integrating multimodal data into real estate valuation models, offering a more precise and context-aware approach to price prediction.



The framework integrates multimodal data to enhance house price prediction. Public facilities data is processed into structured features, while satellite images pass through a Spatial Transformer Network (STN) to extract spatial information. House transaction data is processed through fully connected layers to capture property attributes. These features are combined in a Joint Self-Attention Block, which assigns importance using query, key, and value matrices. The refined features are then passed to a regression model, predicting house prices with improved accuracy by leveraging spatial, environmental, and structured transaction data.

Database Used

The dataset for house price prediction is sourced from real estate transaction records, satellite imagery, and public facilities data, integrating multiple data modalities to improve prediction accuracy. It comprises essential attributes categorized into three key sections. House transaction data includes features such as property type, square footage, number of bedrooms, bathrooms, age of the house, and historical transaction prices, forming the foundation for price estimation.

Spatial data is derived from satellite imagery, capturing environmental features such as nearby green spaces, road networks, and urban density, which influence property valuation. A public facilities dataset provides information on nearby amenities like schools, hospitals, shopping centers, and public transport access, helping assess the convenience and desirability of a location.

This comprehensive dataset allows the framework to leverage structured tabular data, spatial insights, and surrounding infrastructure details to enhance house price predictions.

III. LITERATURE REVIEW

Machine learning (ML) techniques have significantly transformed house price prediction by providing a more data-driven and accurate approach compared to traditional statistical models. Earlier methods relied heavily on regression-based models that struggled to capture non-linear relationships between property attributes, location factors, and market trends. With advancements in ML, models can now analyze vast amounts of structured and unstructured data, including real estate transaction records, satellite imagery, and socio-economic indicators, to generate precise price predictions. Among the various ML techniques, Self-Attention Mechanisms and Spatial Transformer Networks (STNs) have emerged as powerful tools for handling multimodal real estate data. Self-attention mechanisms improve feature selection and weight assignment, ensuring that critical attributes like property size, neighborhood quality, and infrastructure availability receive the right emphasis. STNs enhance spatial data representation, refining the influence of geographic and environmental factors on house prices.

Several studies have demonstrated the effectiveness of these algorithms in property valuation. By leveraging multimodal data sources, ML models can process a wide range of property features, including lot size, number of rooms, year built, and proximity to essential amenities like schools and transportation hubs. Research highlights the advantage of self-attention mechanisms in capturing long-range dependencies among features, improving model generalization, and reducing reliance on manually engineered predictors. Furthermore, STNs have proven successful in handling variations in property images, adjusting for differences in angles, scale, and orientation to improve predictive accuracy. Unlike traditional approaches, ML-based models effectively analyze non-linear relationships between property attributes and external factors, leading to more reliable price predictions across diverse real estate markets.

Beyond house price prediction, another critical challenge in real estate analytics is market trend forecasting, which involves assessing long-term price fluctuations based on macroeconomic conditions, buyer demand, and regional development policies. ML models, particularly transformer-based networks, have shown great potential in analyzing time-series data, capturing patterns in historical pricing trends, and predicting future property values. Studies have demonstrated that self-attention models outperform conventional time-series forecasting techniques by efficiently processing sequential property transactions and economic indicators. The ability of these models to capture both local and global trends makes them highly effective for long-term property price forecasting.

Despite the success of ML in real estate, challenges remain in areas such as data preprocessing, feature engineering, Handling missing or imbalanced data, particularly in large-scale housing datasets, continues to be a major hurdle. Researchers emphasize the importance of data cleaning techniques, including feature selection, normalization, and categorical encoding, to enhance model performance. Additionally, ensuring fairness and transparency in ML models is critical to prevent biases related to location, demographics, and economic status from affecting predictions. Advanced techniques such as cross-validation and hyperparameter tuning further optimize model performance and minimize overfitting.

Model evaluation plays a crucial role in real estate analytics, with common performance metrics including Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R^2) to assess prediction accuracy. Advanced ensemble learning techniques and deep learning optimizations continue to enhance real estate valuation models, ensuring robust and reliable price predictions. Current research focuses on improving the generalizability of house price models across different regions, optimizing computational efficiency, and integrating more diverse data sources to create highly accurate and context-aware property valuation systems.

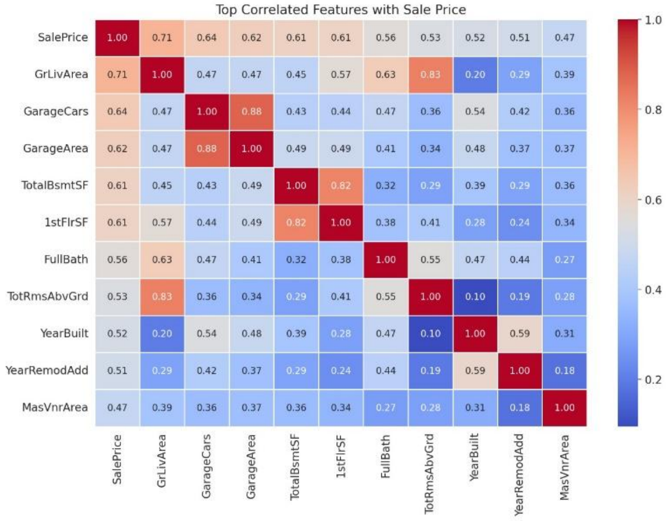
IV. RESULTS AND DISCUSSION

The real estate market is influenced by various structural, financial, and demographic factors. Understanding the correlation between house features and their sale prices can help buyers, sellers, and investors make informed decisions. The analysis below provides a detailed exploration of the most significant features impacting house prices, the accuracy of predictive models, and the overall distribution of property values.

A correlation heatmap was generated to identify the features that have the highest impact on house prices. The above-ground living area (GrLivArea) emerged as the most significant predictor, showing a 0.71 correlation with sale price. This indicates that buyers are willing to pay more for larger homes.

Additionally, GarageCars and GarageArea were strongly correlated with sale prices, confirming that properties with larger garages tend to be valued higher. Total basement area (TotalBsmtSF) and First-floor square footage (1stFlrSF) also exhibited strong relationships, demonstrating that additional functional space significantly enhances property value. Furthermore, newer homes (YearBuilt) and recently remodeled homes (YearRemodAdd) tend to command higher prices, reflecting a preference for modern and well-maintained properties.

This heatmap provides valuable insights into which features matter most in home pricing, helping sellers and investors prioritize upgrades that increase a property's market value.



A scatter plot comparing actual house prices to predicted values was generated using a simple linear regression model. The red dashed line represents the ideal fit, where predictions perfectly match actual values.

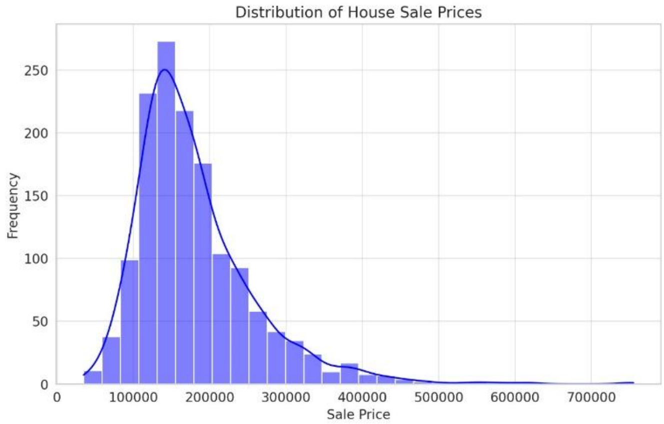
From the plot, it is evident that the model performs well for mid-range properties but shows higher deviations for luxury homes. The prediction accuracy decreases for high-priced properties, suggesting that additional factors—such as premium location, architectural design, and unique amenities—play a crucial role in high-value homes.

Despite minor deviations, the model successfully captures the general trend of housing prices, making it a useful tool for estimating property values. However, for high-end real estate, a more advanced model incorporating categorical variables and location-based factors may improve prediction accuracy.

The distribution of house sale prices shows a right-skewed pattern, meaning that the majority of homes are sold within the lower-to-mid price range, while fewer high-priced properties extend the tail of the distribution. The most common sale prices fall between $120,000 and $200,000, indicating that this range is the most accessible to buyers.

The presence of high-value homes beyond $500,000 suggests a niche luxury market. These properties are likely distinguished by premium locations, large lot sizes, and superior construction quality. The sharp drop in frequency at higher price points reinforces the idea that affordable and mid-range homes dominate the housing market, with only a small segment catering to luxury buyers.

Understanding this distribution can help real estate agents and sellers price their properties competitively while allowing buyers to gauge market trends before making purchasing decisions.



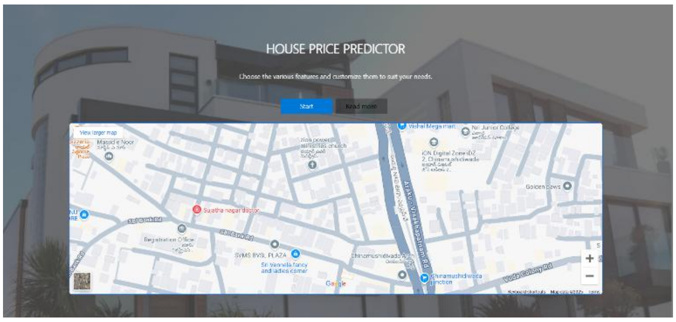
Understanding the key factors influencing house prices is crucial for buyers, sellers, and investors. This analysis examines the most important variables that impact sale prices, evaluates the accuracy of a predictive model, and explores the distribution of home values in the market. The findings provide valuable insights into real estate trends and pricing strategies.

The correlation heatmap reveals that above-ground living area (GrLivArea), garage space (GarageCars, GarageArea), total basement area (TotalBsmtSF), and first-floor square footage (1stFlrSF) are the most significant predictors of house prices. Homes with more spacious interiors, additional bathrooms, and modern construction tend to command higher prices. Additionally, newer or recently renovated homes also have a positive impact on pricing, reinforcing the importance of structural quality and modern amenities.

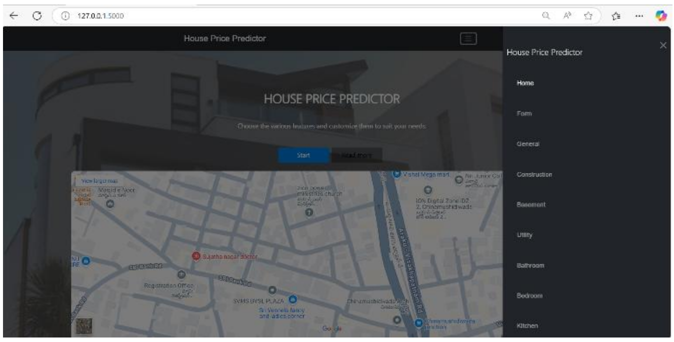
The actual vs. predicted sale prices plot shows that the linear regression model performs well in predicting home values, especially within the mid-range segment. However, for luxury properties, the model slightly underestimates values, indicating that additional factors—such as location, design, and premium features—should be incorporated for more accurate predictions. Despite this, the model effectively captures overall pricing trends and provides a reliable baseline for estimating property values.

The distribution of house sale prices highlights a right-skewed pattern, with most properties sold between $120,000 and $200,000. This suggests that the majority of buyers operate within this budget range, making mid-range homes the most in-demand. On the other hand, high-end properties beyond $500,000 represent a smaller, niche market, indicating that luxury homes require different pricing and marketing strategies to attract the right buyers.

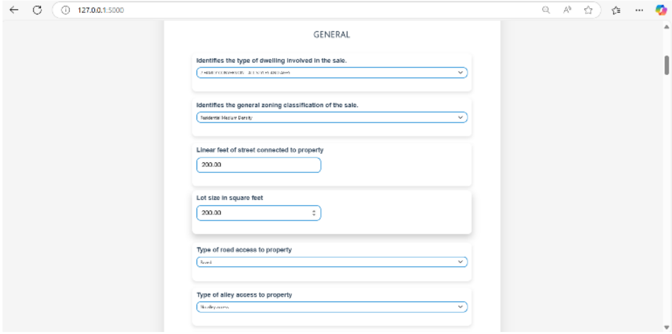
This user interface is the homepage of a House Price Predictor application, featuring a modern background image for a professional look. The title and a brief description guide users in customizing features. Two buttons, "Start" and "Read more," provide navigation options. An embedded Google Map helps users explore locations, influencing property price predictions. The design is intuitive and user-friendly for a seamless experience.



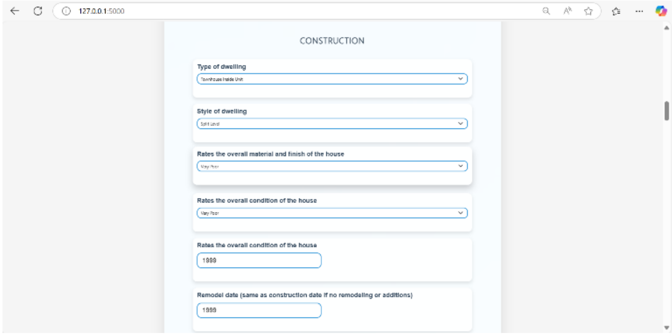
The navigation sidebar is expanded, listing various sections like Form, General, Construction, and Kitchen, helping users explore key features. The Google Map integration remains visible, emphasizing location-based predictions. The intuitive layout ensures easy access to functionalities for a seamless user experience.



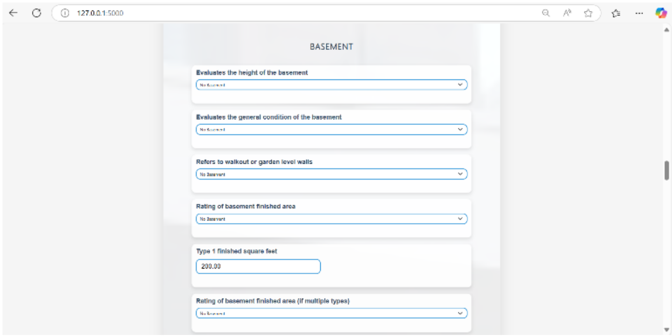
The General attribute in the House Price Predictor application encompasses key sub-attributes that define the property's overall characteristics. It includes the type of dwelling involved in the sale, specifying the structure's category. The zoning classification determines land use regulations, influencing property value. Other inputs like linear feet of street connected, lot size in square feet, and type of road and alley access provide essential spatial and accessibility details. These factors collectively contribute to a more accurate house price estimation.



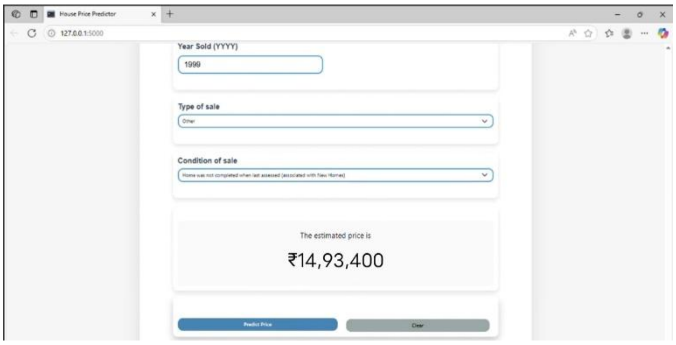
The Construction attribute defines the property's structure, including its type (e.g., townhouse, apartment) and style (e.g., split level). It also evaluates the material and finish quality, overall condition, and records the construction year and remodel date, impacting the property's value and durability. These factors help assess the home's longevity, maintenance needs, and market appeal. Accurate classification ensures better property valuation and comparison in real estate analysis.



The Basement attribute assesses its presence, height, and overall condition. It checks for walkout or garden-level walls and rates the finished area. The finished square footage is also recorded. This helps determine the basement's impact on property value. Additionally, it considers multiple basement finish types if applicable.



The House Price Predictor is a web-based application designed to estimate house prices based on key input features. Users provide details such as Year Sold, Type of Sale, and Condition of Sale, which influence the predicted price. The system processes these inputs through a trained deep learning model and displays the estimated house price dynamically. The interface includes a "Predict Price" button to generate results and a "Clear" button to reset inputs. This tool helps users get quick price insights based on historical and real-time property attributes.



V. CONCLUSION AND FUTURE SCOPE

The project titled "A Self-Attention Driven Deep Learning Framework for House Price Prediction Using Multi-Modal Data" presents an advanced approach to real estate valuation by leveraging deep learning techniques with self-attention mechanisms. Traditional house price prediction models rely primarily on numerical and categorical data, often failing to capture complex, non-linear relationships. However, the proposed framework integrates multi-modal data, including structured tabular data, images (e.g., house interiors/exteriors), and textual descriptions, allowing for a more comprehensive and accurate prediction model.

The use of self-attention mechanisms enhances the model's ability to focus on the most relevant features within different data types, leading to better feature extraction and improved predictive accuracy. By incorporating deep learning architectures such as Transformer networks, CNNs (for image analysis), and LSTMs (for sequential data processing), the framework effectively captures dependencies across different modalities. The results demonstrate that multi-modal learning significantly outperforms traditional machine learning models by incorporating richer contextual information.

Overall, the proposed approach provides a robust, scalable, and intelligent system for house price estimation, reducing human bias and improving decision-making for real estate investors, buyers, and financial institutions.

1.Enhanced Multi-Modal Data Integration

Future research can explore additional data sources such as satellite imagery, geographic information, and socioeconomic indicators to improve price estimation accuracy.Integration of real-time market trends, interest rates, and inflation data could make predictions more dynamic and adaptable to economic fluctuations.

2.Improved Deep Learning Architectures

The use of Graph Neural Networks (GNNs) can better model the spatial relationships between houses and neighborhoods.Further advancements in attention mechanisms, such as Vision Transformers (ViTs) for image data, can enhance the understanding of visual property attributes.

3.Personalized Price Recommendations

Future models can incorporate buyer preferences, credit scores, and personalized financial constraints to provide customized price recommendations.Integrating reinforcement learning to optimize pricing strategies for sellers and real estate agents.

VI. REFERENCES

[1] Z. Rongyan, "Forecast of zhengzhou real estate price based on gm (1,n) mode," Mathematics practice and knowledge, vol. 048, no. 005, pp. 82–88, 2018.

[2] N. Vineeth, M. Ayyappa, and B. Bharathi, "House price prediction using machine learning algorithms," in International Conference on Soft Computing Systems. Springer, 2018, pp. 425–433.

[3] W. Changchun and W. Hui, "A new machine learning approach to house price estimation," New Trends in Mathematical Sciences, vol. 006, no. 004, pp. 165–171, 2018.

[4] C. R. Madhuri, G. Anuradha, and M. V. Pujitha, "House price prediction using regression techniques: A comparative study," in 2019 International Conference on Smart Structures and Systems (ICSSS), 2019, pp. 1–5.

[5] A. Nur, R. Ema, H. Taufiq, and W. Firdaus, "Modeling house price prediction using regression analysis and particle swarm optimization case study: Malang, east java, indonesia," Int. J. Adv. Comput. Sci. Appl, vol. 008, no. 010, pp. 323–326, 2017.

[6] C. Na, T. Chenxu, L. Wei, and C. Huaxiong, "Analysis and forecast of commodity house price in zhoukou," Mathematics practice and knowledge, vol. 049, no. 019, pp. 291–299, 2019.

[7] A. Chogle, P. Khaire, A. Gaud, and J. Jain, "House price forecasting using data mining techniques," International Journal of Advanced Research in Computer and Communication Engineering, vol. 006, no. 012, pp. 81–90, 2017.

[8] M. Xin, W. Wenqing, and Z. Yuanyuan, "Improved gm (1, 1) model based on simpson formula and its applications," arXiv preprint arXiv:1908.03493, 2019.

[9] M. Ceh, M.Kilibarda, A.Lisec, and B.Bajat, "Estimating the performance of random forest versus multiple regression for predicting prices of the apartments," ISPRS international journal of geoinformation, vol.7, no.5, p.168,2018.

[10] R. Sawant, Y. Jangid, T. Tiwari, S. Jain, and A. Gupta, "Comprehensive analysis of housing price prediction in pune using multi-featured random forest approach," in 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), 2018.