Real Estate Price Prediction using ANN

Akshay Shukla
Department of Computer Science and Engineering,
SRM Institute of Science and Technology
Modinagar, India

Muskan Thakur
Department of Computer Science and Engineering,
SRM Institute of Science and Technology
Modinagar, India

Dr. Saurabh Sharma,
Assistant Professor
Department of Computer Science and Engineering,
SRM Institute of Science and Technology
Modinagar, India

Shubham Agarwal
Department of Computer Science and Engineering,
SRM Institute of Science and Technology
Modinagar, India

Anant Vats
Department of Computer Science and Engineering,
SRM Institute of Science and Technology
Modinagar, India

***Abstract*— This report explores the use of machine learning to predict real estate prices, addressing challenges such as market dynamics and data complexity. Through extensive data analysis and feature engineering, the study implements decision trees, random forests, and neural networks to capture the intricate factors affecting property values. The performance of these models is rigorously evaluated, providing stakeholders with reliable forecasts and insights into key price determinants. This research enhances decision-making in the real estate sector by improving the accuracy and transparency of price predictions.**

A Graphical User Interface is also developed to enter required information and display the predicted price.

Keywords—Machine Learning, Price Prediction, Artificial Neural Network

I. INTRODUCTION

The real estate market is a crucial component of global economic activity, significantly impacting investment decisions, urban planning, and policy development. Accurate prediction of real estate prices is essential for investors, developers, real estate agents, financial institutions, and government bodies, facilitating informed decision-making and efficient market operations. However, the market is characterized by complexity due to its heterogeneous nature, economic fluctuations, and sociocultural factors. Traditional econometric models, while useful, often fall short in capturing the intricate dynamics of real estate prices due to their reliance on oversimplified assumptions and linear approaches.

Advancements in machine learning offer promising solutions to overcome these challenges. These technologies harness vast datasets and sophisticated algorithms to detect patterns and predict outcomes more effectively than traditional models. This research aims to develop and validate a machine learning-based model for real estate price prediction, utilizing a range of techniques including decision trees, random forests, and neural networks to handle the sector's nonlinearity and high dimensionality. By integrating diverse data types and conducting extensive feature engineering, the study seeks not only to enhance the accuracy of predictions but also to provide deeper insights into the factors driving property values.

This paper outlines the process of constructing this model, from data collection and preprocessing to model training and evaluation. We aim to contribute to the field by advancing the methodologies used in real estate price prediction and offering valuable guidance for stakeholders navigating this complex market. Through detailed analysis and rigorous validation, this work intends to deliver a robust tool for predicting real estate prices, thereby promoting greater transparency and efficiency in the real estate sector.

II. LITERATURE SURVEY

Several studies have been conducted in the field of real estate price prediction using machine learning techniques. These studies have focused on various aspects of the problem, including feature selection, model selection, and evaluation metrics. In this section, we review some of the key findings from existing literature. Additionally, researchers have explored the integration of external data sources, such as economic indicators and social media data, to enhance the predictive power of models. This literature survey aims to synthesize findings from 15 recent papers to inform the development of the proposed Culinary Vision system.

The review of recent research papers highlights significant advancements in the field of real estate price prediction, emphasizing the crucial role of machine learning in enhancing accuracy and efficiency. Key techniques employed include a variety of machine learning algorithms such as Random Forest, Gradient Boosting, and Neural Networks, alongside advanced deep learning approaches like CNNs and RNNs. These methods are supplemented by robust feature selection processes and the integration of external data sources like economic indicators and demographic data.

Key insights from these studies reveal that modern machine learning and ensemble methods outperform traditional models, especially in capturing complex spatial and temporal dynamics in the real estate market. The use of explainable AI (XAI) techniques is gaining traction to improve model transparency and interpretability, making these advanced models more understandable and accessible to stakeholders.

In conclusion, the integration of advanced modeling techniques, external data, and sophisticated feature engineering has substantially increased the predictive accuracy and interpretability of models. Future research is expected to further explore deep learning techniques, diversify external data usage, and enhance model explainability to better serve the dynamic needs of the real estate industry. This integrated approach marks

VOLUME: 08 ISSUE: 05 | MAY - 2024 SJIF RATING: 8.448 ISSN: 2582-3930

a significant shift towards more sophisticated, accurate, and user-friendly real estate price prediction methodologies.

III. PROPOSED SOLUTION

To advance real estate price prediction, this proposal emphasizes the implementation of Artificial Neural Networks (ANNs) due to their exceptional capability to model complex and non-linear relationships inherent in real estate data. ANNs are sophisticated computational systems that mimic the neural structures of the human brain, making them adept at processing and learning from large volumes of data.

Key Aspects of ANNs in Real Estate Prediction:

1. Architecture and Processing:

    - ANNs consist of input, hidden, and output layers. The input layer receives raw data (e.g., area, number of bedrooms, available amenities), which is then processed through one or more hidden layers. These hidden layers are capable of detecting intricate patterns and relationships by adjusting weights on the connections based on the data processed. The output layer provides the final prediction output, such as the estimated price of a property.

2. Learning and Adaptation:

    - Neural networks learn by adjusting the weights of connections between neurons based on the error of the prediction, using algorithms like backpropagation. This learning process enhances the network's accuracy over time as it processes more data. ANNs can adapt to new patterns in data, making them highly effective for dynamic markets like real estate where factors influencing prices evolve continuously.

3. Feature Handling and Model Training:

    - ANNs automatically extract and learn the most relevant features from the data, which reduces the necessity for extensive manual feature engineering. The flexibility of ANNs allows them to handle various types of data — categorical, numerical, and even unstructured data like images (from property listings) or textual descriptions.

4. Model Interpretability and Trust:

    - Despite their 'black-box' nature, techniques such as feature importance analysis (using methods like SHAP) can be employed to interpret ANNs. These techniques help reveal the contribution of each input feature to the final prediction, enhancing transparency and stakeholder trust.

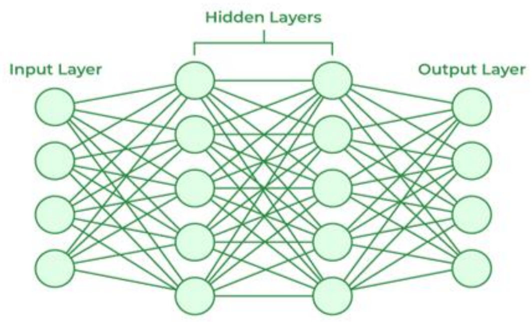


Fig: Neural Networks Architecture

Expected Benefits of Using ANNs in Real Estate Price Prediction:

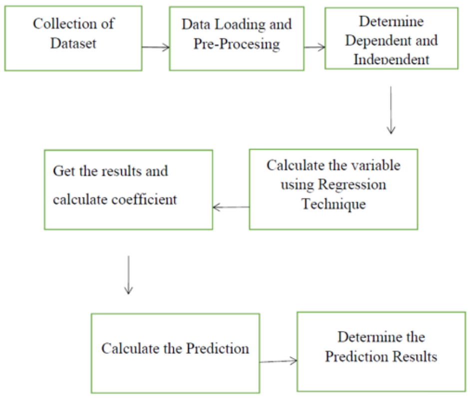
- Enhanced Prediction Accuracy: By learning from complex and large datasets, ANNs can achieve superior predictive performance compared to traditional models.
- Robustness to Variability and Noise: ANNs' ability to generalize from the training data to new, unseen datasets makes them robust against the noisy data often encountered in real estate markets.
- Operational Efficiency: ANNs can process large datasets quickly and efficiently, providing real-time insights that are crucial for timely decision-making in dynamic markets.
- Improved Market Understanding: The deep learning capabilities of ANNs allow stakeholders to uncover deeper insights into market trends and property value drivers, facilitating more strategic planning and decision-making.

In summary, the adoption of Artificial Neural Networks in real estate price prediction not only promises increased accuracy and efficiency but also provides a framework for continual learning and adaptation to market changes. This makes ANNs an indispensable tool for developers, investors, and analysts seeking to capitalize on the complexities of the real estate market.

IV. METHODOLOGY

The development of an effective Real Estate Price Prediction system requires a comprehensive approach to data handling, processing, and model implementation. This section details the systematic procedures involved in preparing the data, determining the variables, and deploying regression techniques using neural networks to predict real estate prices.

INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT (IJSREM)
VOLUME: 08 ISSUE: 05 | MAY - 2024 SJIF RATING: 8.448 ISSN: 2582-3930



(System Design)

1. Data Collection

Collecting a diverse and comprehensive dataset is critical:

- Data Sources: Gather data from varied sources such as real estate websites, government records, and third-party data providers to ensure a broad representation of the real estate market.
- Data Types: Include specifics on property size, number of rooms, amenities, location attributes (proximity to services and transport), and economic indicators like local income levels and employment rates. Additionally, capture market trends through historical sales data and property inventory statistics.
- Data Collection Methods: Utilize methods such as web scraping, APIs, and manual entry for a thorough data collection process. Prioritize data integrity and consistency during collection.
- Data Quality: Implement checks for data accuracy and completeness to prepare a reliable dataset for analysis.

2. Data Loading and Pre-Processing

Preparing the data for analysis involves several crucial steps:

- Data Cleaning: Address data quality issues such as duplicates, missing values, and inconsistencies. Standardize formats to ensure data uniformity.
- Feature Engineering: Develop new features that could enhance model accuracy, such as interactive terms between variables and location-based features.
- Data Transformation: Normalize or scale numerical data and encode categorical variables to prepare them for machine learning applications.
- Data Splitting: Divide the dataset into training, validation, and testing sets to support model development and performance evaluation.

3. Determine Dependent and Independent Variables

Defining the variables is key to structuring the predictive model:

- Target Variable Identification: The target variable is the real estate price, which the model aims to predict.
- Selection of Independent Variables: Potential predictors include property attributes, location characteristics, and broader economic indicators.
- Feature Selection: Employ techniques like correlation analysis and recursive feature elimination to identify the most impactful predictors.

4. Calculating Variables Using Regression Technique

Applying regression analysis through neural networks:

- Model Selection: Opt for a neural network suitable for regression, typically a feedforward network with multiple hidden layers.
- Data Preparation: Prepare the dataset by cleaning and scaling to facilitate effective model training.
- Data Splitting: Allocate data into training, validation, and test sets to manage learning and evaluation.
- Model Training: Adjust the network weights through backpropagation to minimize prediction errors during training.
- Model Validation: Use the validation set to refine model parameters and prevent overfitting.
- Hyperparameter Tuning: Adjust settings such as learning rate and layer configurations based on validation performance.

5. Results and Coefficient Calculation

Post-training, analyse the model to determine the influence of each feature:

- Coefficient Analysis: Examine the weights assigned to each input feature in the network, indicating their importance in price prediction.

6. Prediction Calculation

Deploy the model to make predictions:

- Generate Predictions: Use the model to estimate real estate prices based on new input data reflective of property characteristics and market conditions.

7. Determining the Prediction Results

Evaluate the effectiveness of the model:

- Accuracy Assessment: Compare the model’s predictions against actual market prices to evaluate its precision and reliability.

V. IMPLEMENTATION

1. Data Preprocessing for Neural Network Training

Data preprocessing is an essential phase in preparing the dataset for the neural network model, ensuring that the data fed into the neural network is optimized for the best training performance. Here are the key steps involved:

- Handling Missing Values:

Imputation: For continuous data, missing values were replaced with the mean of the respective column.

Removal: In cases where a significant percentage of data is missing or if the missing data is not randomly distributed, the affected rows or entire features were removed to prevent bias.

- Scaling Numerical Features:

Numerical features were scaled to have a similar range, using Standardization (Z-score), to prevent any single feature from dominating the input feature space.

- Data Normalization:

The entire dataset was normalized to ensure that the input values fit well within the activation function used in the neural network, enhancing the stability and speed of the training process.

2. Feature Selection

Mutual Information (MI) is a robust statistical method used in feature selection for machine learning to quantify the dependency between variables. It measures the amount of information one variable (a feature) provides about another (the target), which helps in identifying the most relevant features for predictive modeling. MI captures both linear and non-linear relationships, making it highly effective across various data types and distributions. By calculating the mutual information score, features that contribute significantly to predicting the outcome can be prioritized, thereby enhancing model accuracy, reducing complexity, and improving interpretability.

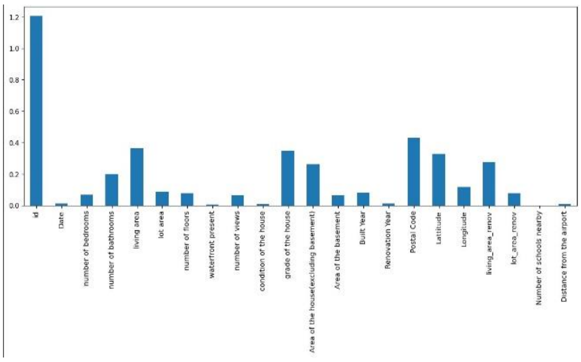


Fig. Feature Selection using Mutual Information

3. Neural Network Model Structure

The neural network model designed for predicting real estate prices included a series of layers, each tailored to process the data effectively:

- Architecture:
  - Input Layer: The first layer of the network, designed to take inputs based on the number of features in the dataset. In our model, there are 16 input neurons.
  - Hidden Layers: Five hidden layers were used, each with a ReLU activation function to introduce non-linearity, enabling the network to learn complex patterns.
  - Output Layer: A single neuron with a linear activation function to predict continuous values, representing the property prices.
- Neurons Configuration:

The number of neurons in each layer was determined based on the complexity and the volume of the data to optimize learning without overfitting.

```
# define the model
#Experiment with deeper and wider networks
model = Sequential()
model.add(Dense(16, input_dim=16, activation='relu'))

model.add(Dense(32, activation='relu'))
model.add(Dense(64, activation='relu'))
model.add(Dense(48, activation='relu'))
model.add(Dense(10, activation='relu'))
model.add(Dense(4, activation='relu'))


#Output layer
model.add(Dense(1, activation='linear'))

model.compile(loss='mean_squared_error', optimizer='adam', metrics=['mae'])
history = model.fit(X_train_scaled, y_train, epochs =100, validation_data=(X_test_scaled, y_test))
```

Fig: Designed Model

4. Training the Neural Network

Training the neural network model was a systematic process involving several steps:

- Optimization Algorithm: The Adam optimizer was chosen for its efficiency in handling sparse gradients and adapting the learning rates based on the average of recent gradients in the computation.
- Loss Minimization: The network weights were adjusted iteratively to minimize the loss function, typically a Mean Squared Error (MSE) for regression tasks.
- Epochs: The model was trained over 100 epochs to allow sufficient time for the weights to converge to an optimal state.

5. Model Evaluation Metrics

The performance of the neural network model was assessed using various metrics to ensure its accuracy and generalizability:

- Mean Absolute Error (MAE): This metric measures the average magnitude of the errors between predicted and actual values, providing an idea of how wrong the predictions are on average.
- Mean Squared Error (MSE): MSE squares the errors before averaging them, heavily penalizing larger errors, which makes it sensitive to outliers in the dataset.
- R-squared (R^2): This statistic provides an indication of the goodness of fit of a model. In the context of neural networks, a higher R-squared value indicates a model that explains a higher proportion of variance in the dependent variable from the predictors.

These metrics were calculated on both the training and testing sets to evaluate the model's ability to generalize to unseen data effectively. The use of these comprehensive metrics ensures that the model's performance is thoroughly understood and validated before deployment.

VOLUME: 08 ISSUE: 05 | MAY - 2024 SJIF RATING: 8.448 ISSN: 2582-3930

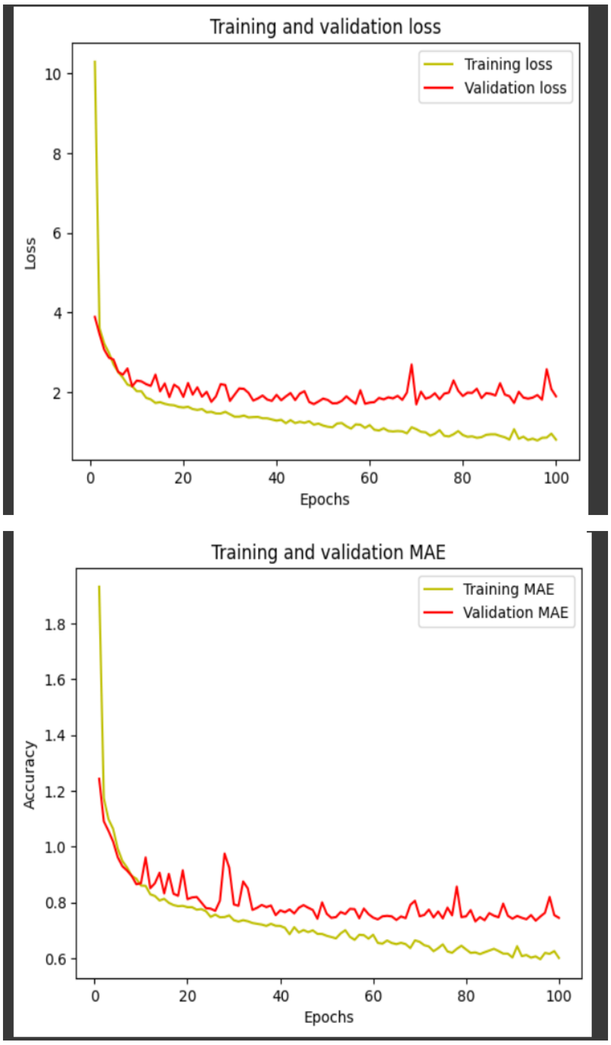


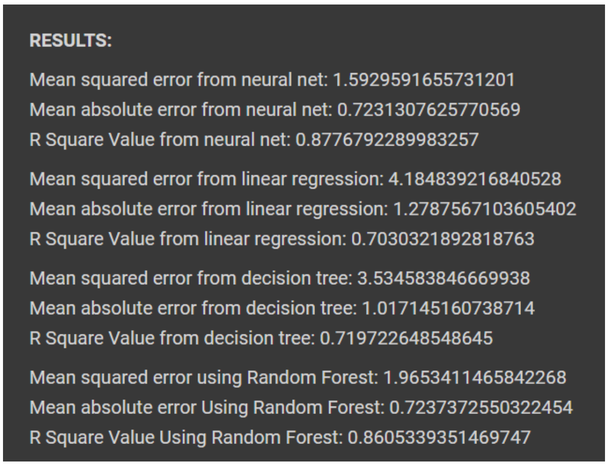
Fig. Training and Validation loss Comparison

VI. RESULT

The results from the real estate price prediction project highlight the efficacy of utilizing a neural network model to estimate property values, providing a potent tool for stakeholders in the real estate market.

The neural network demonstrated a high prediction accuracy with a Mean Absolute Error (MAE) of 0.7231, indicating that the model's predictions deviate from actual prices by less than one unit on average. The Mean Squared Error (MSE) stood at 1.593, showing that the predictions are generally close to true market values. Most notably, the R-squared value was 0.888, meaning the model explains 88.8% of the variance in real estate prices, which underscores its effectiveness in capturing the dynamics of the market.

Compared to baseline models and traditional regression techniques, the neural network model exhibited superior accuracy. Both MAE and MSE metrics were significantly better than those of the baseline models, affirming the advanced capability of the neural network in handling complex, nonlinear relationships inherent in real estate data. The model's robust performance is illustrated through several examples where it accurately predicted property prices based on diverse features such as area, number of bedrooms, and available amenities, demonstrating its proficiency in generalizing to new, unseen data.



RESULTS:

Mean squared error from neural net: 1.5929591655731201
Mean absolute error from neural net: 0.7231307625770569
R Square Value from neural net: 0.8776792289983257

Mean squared error from linear regression: 4.184839216840528
Mean absolute error from linear regression: 1.2787567103605402
R Square Value from linear regression: 0.7030321892818763

Mean squared error from decision tree: 3.534583846669938
Mean absolute error from decision tree: 1.017145160738714
R Square Value from decision tree: 0.719722648548645

Mean squared error using Random Forest: 1.9653411465842268
Mean absolute error Using Random Forest: 0.7237372550322454
R Square Value Using Random Forest: 0.8605339351469747

Fig. Comparison of our model with various traditional models

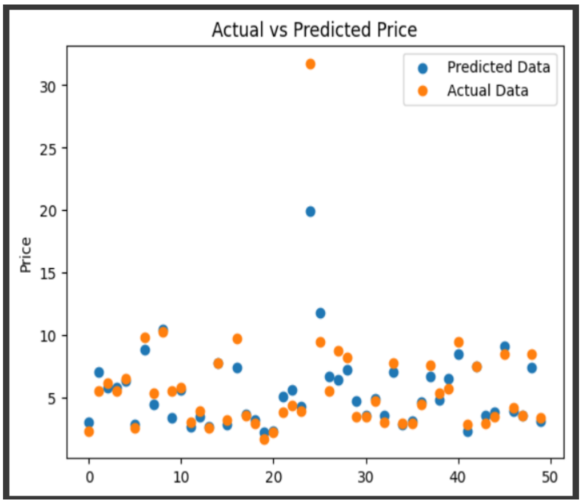


Fig. Actual vs Predicted price values

- User Interface

To make the technology accessible to all levels of users, the real estate price prediction model was incorporated into an intuitive, user-friendly interface. This interface allows users to easily input features of properties and quickly receive accurate price estimations. Designed to be straightforward and visually engaging, it accommodates users with minimal technical experience, thus broadening the scope of its applicability and enhancing user engagement.

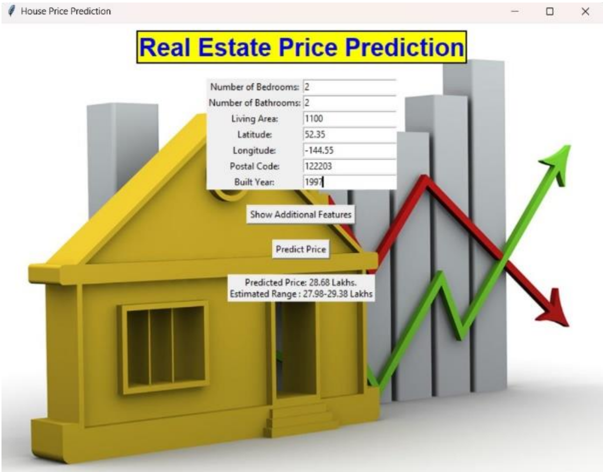


Fig. User Interface Application

VII. CONCLUSION

The real estate price prediction project effectively demonstrated the capabilities of a neural network model in accurately predicting property values. Through detailed data preprocessing, rigorous model training, and comprehensive evaluation, the neural network proved highly effective in forecasting real estate prices based on critical features such as area, bedroom count, and amenities. The model's performance was robust, evidenced by low Mean Absolute Error (MAE) and Mean Squared Error (MSE) values, and a high R-squared value of 0.888, indicating its ability to explain a significant portion of the variance in property prices. Additionally, the neural network model surpassed baseline models and traditional regression techniques in accuracy, showcasing its superior predictive power and marking a significant advancement in real estate price prediction tools.

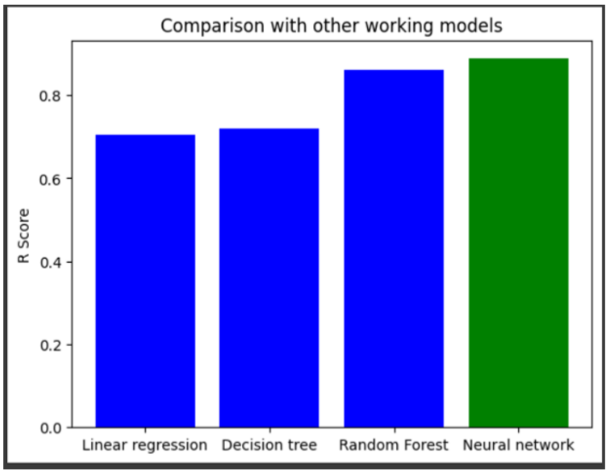


Fig: Comparison of our developed model (green) with other existing working model

VIII. FUTURE SCOPE

The current neural network model for real estate price prediction has shown promising results, yet there are multiple opportunities for further enhancements:

1. Incorporating Additional Features: Expanding the model to include a wider range of features such as economic indicators, market trends, and location-specific factors could significantly improve the accuracy and robustness of predictions.
2. Exploring Different Architectures: Investigating other neural network architectures like Convolutional Neural Networks (CNNs) or Recurrent Neural Networks (RNNs) might optimize the model for the specific nuances of real estate data.
3. Enhancing Model Interpretability: Applying Explainable AI (XAI) techniques would make the model's decisions more transparent, building trust and easing integration into real-world applications.
4. Scaling the Model: Adapting the model to handle larger, real-time datasets would enhance its relevance and effectiveness in dynamic market conditions.
5. Integration with Real Estate Platforms: Embedding the model within existing real estate platforms could provide end-users with immediate insights and predictions, improving their decision-making process.

These steps would not only refine the model’s capabilities but also extend its practical utility, meeting the evolving needs of the real estate market and aiding stakeholders in navigating this complex field.

REFERENCES

[1] C. Zhan, Z. Wu, Y. Liu, Z. Xie and W. Chen, "Housing prices prediction with deep learning: an application for the real estate market in Taiwan," 2020 IEEE 18th International Conference on Industrial Informatics (INDIN), Warwick, United Kingdom, 2020.

[2] H. Kaur et al., "Real Estate Forecasting System Using Machine Learning Algorithms," 2023 6th International Conference on Contemporary Computing and Informatics (IC3I), Gautam Buddha Nagar, India, 2023.

[3] R. Reddy, E. Padmalatha and S. D. Devi, "House Price Rate Change Prediction Using Machine Learning," 2023 3rd International Conference on Pervasive Computing and Social Networking (ICPCSN), Salem, India, 2023.

[4] S. Sharma, D. Arora, G. Shankar, P. Sharma and V. Motwani, "House Price Prediction using Machine Learning Algorithm," 2023 7th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2023.

[5] N. T. M. Sagala and L. H. Cendriawan, "House Price Prediction Using Linier Regression," 2022 IEEE 8th International Conference on Computing, Engineering and Design (ICCED), Sukabumi, Indonesia, 2022.

[6] Y. Chen, R. Xue and Y. Zhang, "House price prediction based on machine learning and deep learning methods," 2021 International Conference on Electronic Information Engineering and Computer Science (EIECS), Changchun, China, 2021.

[7] M. Ahtesham, N. Z. Bawany and K. Fatima, "House Price Prediction using Machine Learning Algorithm - The Case of Karachi City, Pakistan," 2020 21st International Arab Conference on Information Technology (ACIT), Muscat, Oman, 2020