House Price Prediction using Neural Networks

Aditya Sajith
B.E Biotechnology
BITS Pilani Dubai Campus
Dubai, UAE
f20220124@dubai.bits-pilani.ac.in

Nihal Muhammedali
Computer Science Engineering
BITS Pilani Dubai Campus
Dubai, UAE
f20220270@dubai.bits-pilani.ac.in

Mohammed Farhaan
Computer Science Engineering
BITS Pilani Dubai Campus
Dubai, UAE
f20220040@dubai.bits-pilani.ac.in

Yuvan Karthik
Computer Science Engineering
BITS Pilani Dubai Campus
Dubai, UAE
f20220258@dubai.bits-pilani.ac.in

Abstract*—This project represents a strong method for house price prediction through a neural network model that is based on TensorFlow and Streamlit. The main goal is to design a user-friendly web application that is capable of producing price predictions for residential properties from multiple input features like area, no. of bedrooms, bathrooms, stories etc. some other categorical fields such as furnishing status, air conditioning, various etc. are also used in this prediction. The model is trained using a pre-processed dataset that makes use of some state-of-the-art data transformation technique including standard scaling on numeric features, one-hot encoding on categorical features. The neural network architecture is designed for regression tasks, and this guarantees effective learning while delivering correct predictions. Evaluation metrics used for the model's performance comprise the Mean Squared Error (MSE), Root Mean Squared Error (RMSE) and R-squared (R2). Accurate house price prediction plays a crucial role in real estate analysis, aiding investors, homeowners, and policymakers in decision-making. Traditional statistical techniques such as hedonic regression and autoregressive integrated moving average (ARIMA) often struggle with high-dimensional and nonlinear real estate data, leading to limitations in predictive accuracy. Artificial Neural Networks (ANNs) have emerged as a powerful alternative, capable of recognizing complex patterns and improving forecast reliability.

***Keywords*—House Price Prediction, Artificial Neural Networks (ANNs), Deep Learning, Streamlit Application, Regression Modeling, Structured Data Prediction, Property Valuation, Data Preprocessing**

1. Introduction

Accurately predicting house prices plays a key role in the real estate market, especially for investors, banks, and policy planners who rely on such forecasts for important financial decisions. Housing prices are influenced by a wide mix of factors, including economic trends, population growth, and changes in local market behavior. For many years, traditional methods like linear regression, ARIMA, and hedonic pricing models were commonly used to estimate property values. While these models laid the groundwork for price prediction, they often struggle to handle the complex and non-linear nature of real estate data, especially when the number of features grows or when relationships between them are not clearly defined (Kershaw and Rossini, 1999).

As a result, researchers began turning to artificial intelligence tools to improve forecasting accuracy. Among these, artificial neural networks have gained attention for their ability to identify patterns in large and complex datasets. Unlike traditional models, ANNs can learn relationships that are not immediately obvious, making them particularly effective in predicting housing prices with greater precision (Nguyen and Cripps, 2001). Recent work has shown that ANNs can outperform standard models in both flexibility and accuracy, especially when working with data that includes a mix of numerical values and categorical features (Huang et al., 2020). In more advanced applications, deep learning models such as convolutional neural networks and long short-term memory networks have also been used. CNNs are especially useful when capturing location-related trends, while LSTM models help in analyzing how prices change over time (Borovykh et al., 2017). Some researchers have gone a step further by combining multiple techniques or adding broader economic indicators—like inflation, unemployment, and GDP growth—to enhance the accuracy of their predictions (Deng et al., 2021).

However, these advanced models also come with their own challenges. Neural networks typically require large volumes of clean and balanced data, and missing values or inconsistencies can weaken their performance (Nguyen et al., 2021). Deep learning systems also demand more computational power, which can make them difficult to run on standard machines. Another ongoing issue is transparency—because many AI models function like black boxes, it can be hard for users to understand how predictions are made, which is a serious drawback when clear reasoning is needed in high-stakes decisions like home buying (Zhang et al., 2022).

This project builds on these ideas by designing a user-friendly house price prediction system powered by a neural network. The model is built using TensorFlow and runs inside a web application created with Streamlit. It allows users to enter details about a property—such as size, number of rooms, and available amenities—and instantly receive an estimated selling price. The model is trained on well-prepared data and uses a strong preprocessing pipeline to handle both numerical and categorical information. Along with its predictions, the app also shows performance metrics and helpful graphs, making it useful for both real estate professionals and everyday users. By combining practical tools with solid machine learning techniques, this project aims to highlight both the strengths and limits of using AI for real estate forecasting in a way that's accurate, transparent, and accessible.

2. Dataset Description

This project uses a dataset which has included information about different residential properties with their corresponding selling prices. Each entry denotes one house and contains a list of features describing the physical property, the amenities, and some locational or quality attributes. The numerical features include the built-up area in square feet, number of bedrooms /bathrooms and the number of floors (stories) and availability of parking. Besides, the dataset contains several categorical features, namely, whether the house is situated at the main road, has a guest room, contains a basement, has hot water heating, has air conditioning, is situated in a preferred residential area, and furnishing status of the house (furnished, semi-furnished or unfurnished).

A pre-processing stage was completed on the data before training on the model to ensure the data was in the correct format for a machine learning algorithm to use. Categorical variables were converted to numerical representation by means of one-hot encoding, which enables model to read them without affecting its decision with bias on the invented label values. At the same time, the numerical features were scaled in a way so they would all contribute proportionally during the training process. This change would make sure that no single feature would emerge superior to all the rest by virtue of different scales. The data was then split into training and validation sets, approximately 80% for training and 20% for validation. It was this structure that gave the underlying basis for the formulation and validation of the predictive model.

3. Proposed Model and Algorithm

4.1. Data Preprocessing

This project uses a dataset which has included information about different residential properties with their corresponding selling prices. Each entry denotes one house and contains a list of features describing the physical property, the amenities, and some locational or quality attributes. The numerical features include the built-up area in square feet, number of bedrooms /bathrooms and the number of floors (stories) and availability of parking. Besides, the dataset contains several categorical features, namely, whether the house is situated at the main road, has a guest room, contains a basement, has hot water heating, has air conditioning, is situated in a preferred residential area, and furnishing status of the house (furnished, semi-furnished or unfurnished).

A pre-processing stage was completed on the data before training on the model to ensure the data was in the correct format for a machine learning algorithm to use. Categorical variables were converted to numerical representation by means of one-hot encoding, which enables model to read them without affecting its decision with bias on the invented label values. At the same time, the numerical features were scaled in a way so they would all contribute proportionally during the training process. This change would make sure that no single feature would emerge superior to all the rest by virtue of different scales. The data was then split into training and validation sets, approximately 80% for training and 20% for validation. It was this structure that gave the underlying basis for the formulation and validation of the predictive model.

4.2. Model Architecture

The predictive model applied in this research is a feedforward artificial neural network developed to perform regression tasks with structured tabular data. The network architecture includes an input layer that takes the entire set of processed features, and three hidden layers. There are 128, 64 and 32 neurons in these hidden layers respectively, and they all have the use of the ReLU(Rectified Linear Unit) activation function. This decision of activation prevents the model from learning complex non-linearities without introducing vanishing gradients. To finish it off is a single neuron with a linear activation function that gives us the predict house price. The depth and size of the layers, which were selected in order to create a good balance between complexity and computational efficiency, allowed the model to learn well from the data that were available, without overfitting.

4.3. Compilation and Training

After the architecture of the model was ready, it was compiled based on the Adam optimization algorithm, which is suitable for deep learning due to its ability of setting up the adaptive learning rate. The mean squared error (MSE) loss function was used to train the model, MSE being the standard loss function choice that ramps up heavier penalties on larger errors for regression problems. Training took 200 epochs and gave the network enough time to adjust its weights and biases by backpropagation. A proportion of the training data (~10%) was reserved as a validation set to keep track of the model's performance during training and try to catch, in case, some signs of overfitting. This strategy led to a model that fit the training data very well, but it preserved the capacity to generalise to new unseen data.

4.4. Algorithm Workflow

The development process was made up of clear steps. First was the loading of the raw data and determination of which columns, if numerical, and which, if categorical. The preprocessing pipeline took care of encoding and scaling such that the feature matrix was ready for modeling input. Following the data split into training and testing, a neural network was developed through TensorFlow's Keras API. The model was then trained on the prepared data and the results tested using key performance measures including mean squared error and R-squared. Lastly, the trained model was incorporated into a Streamlit web application that enables an interactive approach to entering details about a house by allowing users to input house details interactively and making real-time price predictions. This interface improves the usability of the model and shows its practical relevance for real-world decision making at the housing market.

4. Simulation Results and Discussion

After training the artificial neural network on the preprocessed housing dataset, the model's performance was evaluated using a separate test set. The primary evaluation metrics used were Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and the coefficient of determination (R-squared or R^2). The results showed that the model achieved an MSE of approximately $\mathbf{1.81 \times 10^{12}}$, with an RMSE of around **1,343,619**, and an R^2 score of **0.64**.

These results suggest that the model is able to capture a significant portion of the variance in housing prices based on the given features. The R^2 value of 0.64 indicates a moderately strong fit, which means that the model explains about 64% of the variation in house prices from the selected input variables. While not perfect, this level of accuracy is promising, especially considering the presence of both numerical and categorical features, which can introduce complexity into the learning process.

The relatively high RMSE indicates that there is still some error margin in the predictions, likely due to factors such as hidden variables not present in the dataset, price outliers, or slight inconsistencies in categorical encoding. Additionally, real estate markets are often influenced by external, macroeconomic factors—like recent policy changes, buyer sentiment, or local development—which may not be captured in the current dataset.

The trained model was deployed in an interactive Streamlit application that allows users to input specific details about a house, such as the number of bedrooms, presence of a basement, furnishing status, and total area. Upon entering the inputs, the application instantly returns a price prediction. This interface not only demonstrates the practical usability of the model but also enhances user accessibility, making it suitable for both real estate professionals and everyday users who are exploring property valuations.

The model's visual outputs, such as the scatter plot of actual versus predicted prices, help in understanding how closely the predictions align with real values. While some scatter is expected due to the nature of regression problems, the plot overall confirms the model's ability to follow the general pricing trend.

In summary, the simulation confirms that the proposed ANN-based approach, supported by a well-designed preprocessing pipeline and an intuitive user interface, provides a practical solution for house price prediction. Although there is room for improvement in terms of precision, the current results lay a solid foundation for further refinement and real-world application.

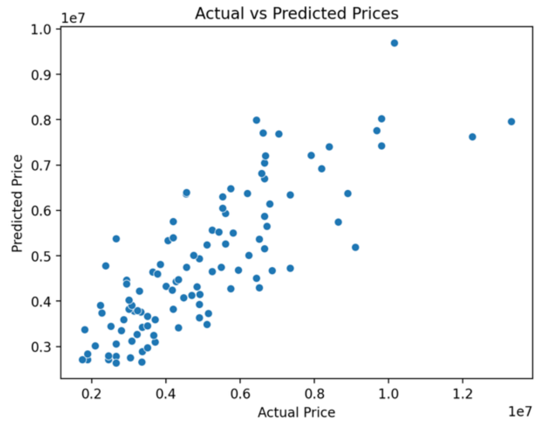

Fig. Actual vs. Predicted Prices

Conclusion

This project set out to create an effective and accessible solution for predicting house prices using artificial neural networks. By combining structured data preprocessing with a multi-layer neural network, the model was able to learn complex patterns in housing data and provide fairly accurate predictions. The achieved R-squared value of 0.64 reflects the model's ability to explain a good portion of the variation in prices based on the available features. Although not perfect, the results demonstrate that neural networks offer a strong alternative to traditional regression models, especially in cases where data relationships are non-linear and high-dimensional.

The deployment of the model through a user-friendly Streamlit application further emphasizes the project's practical value. The interactive interface allows users to input key property features and receive instant price predictions, making the tool useful for real estate professionals, investors, and everyday buyers and sellers alike. The visual outputs, such as scatter plots of predicted versus actual prices, also help users understand the model's behavior and build trust in its predictions.

There is, however, plenty of room for growth and refinement in future versions of this project. One potential improvement would be to expand the dataset by including more detailed features, such as neighborhood quality, proximity to schools or transportation, and historical pricing trends. Integrating the model with publicly available real estate databases could also provide access to real-time data, enhancing both accuracy and relevance.

From a modeling perspective, exploring more advanced neural network architectures like convolutional neural networks (CNNs) could help capture spatial relationships, while recurrent neural networks (RNNs) would be useful for analyzing trends over time if temporal data is introduced. The model's performance could also benefit from hyperparameter tuning using techniques such as grid search or Bayesian optimization.

On the deployment side, the application could be scaled further through containerization using tools like Docker and deployed to cloud platforms such as AWS, Azure, or Google Cloud. Features like user authentication could also be added to allow for personalized predictions and saved sessions. Additionally, enhancing the user interface with interactive visualizations and charts could make the application more engaging and informative. Introducing multilingual support would also increase the accessibility of the tool for users in different regions around the world.

In summary, this project demonstrates how artificial neural networks, combined with thoughtful data engineering and accessible design, can provide a strong foundation for house price prediction. With continued development, it has the potential to evolve into a comprehensive and intelligent tool for use across the real estate industry.

References

[1] A. Khalafallah, "Neural network-based model for predicting housing market performance," in Tsinghua Science and Technology, vol. 13, no. S1, pp. 325-328, Oct. 2008, doi: 10.1016/S1007-0214(08)70169-X.

[2] W. T. Lim, L. Wang, Y. Wang and Q. Chang, "Housing price prediction using neural networks," *2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*, Changsha, China, 2016, pp. 518-522, doi: 10.1109/FSKD.2016.7603227.

[3] L. Wang, F. F. Chan, Y. Wang and Q. Chang, "Predicting public housing prices using delayed neural networks," *2016 IEEE Region 10 Conference (TENCON)*, Singapore, 2016, pp. 3589-3592, doi: 10.1109/TENCON.2016.7848726.

[4] M. F. Mukhlishin, R. Saputra and A. Wibowo, "Predicting house sale price using fuzzy logic, Artificial Neural Network and K-Nearest Neighbor," *2017 1st International Conference on Informatics and Computational Sciences (ICICoS)*, Semarang, Indonesia, 2017, pp. 171-176, doi: 10.1109/ICICOS.2017.8276357.

[5] J. J. Wang *et al*., "Predicting House Price With a Memristor-Based Artificial Neural Network," in *IEEE Access*, vol. 6, pp. 16523-16528, 2018, doi: 10.1109/ACCESS.2018.2814065.

[6] A. Varma, A. Sarma, S. Doshi and R. Nair, "House Price Prediction Using Machine Learning and Neural Networks," *2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)*, Coimbatore, India, 2018, pp. 1936-1939, doi: 10.1109/ICICCT.2018.8473231.

[7] Y. Piao, A. Chen and Z. Shang, "Housing Price Prediction Based on CNN," *2019 9th International Conference on Information Science and Technology (ICIST)*, Hulunbuir, China, 2019, pp. 491-495, doi: 10.1109/ICIST.2019.8836731.

[8] Z. Jiang and G. Shen, "Prediction of House Price Based on The Back Propagation Neural Network in The Keras Deep Learning Framework," *2019 6th International Conference on Systems and Informatics (ICSAI)*, Shanghai, China, 2019, pp. 1408-1412, doi: 10.1109/ICSAI48974.2019.9010071.

[9] L. Xiao and T. Yan, "Prediction of House Price Based on RBF Neural Network Algorithms of Principal Component Analysis," *2019 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS)*, Shanghai, China, 2019, pp. 315-319, doi: 10.1109/ICIIBMS46890.2019.8991474.

[10] 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, pp. 719-724, doi: 10.1109/INDIN45582.2020.9442244.

[11] J. Xu, "A Novel Deep Neural Network based Method for House Price Prediction," *2021 International Conference of Social Computing and Digital Economy (ICSCDE)*, Chongqing, China, 2021, pp. 12-16, doi: 10.1109/ICSCDE54196.2021.00012.

[12] 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, pp. 699-702, doi: 10.1109/EIECS53707.2021.9587907.

[13] Y. Ni, "A housing price prediction method based on neural network," *2022 International Conference on Big Data, Information and Computer Network (BDICN)*, Sanya, China, 2022, pp. 592-595, doi: 10.1109/BDICN55575.2022.00114.

[14] C. Chee Kin, Z. Arabee Bin Abdul Salam and K. Batcha Nowshath, "Machine Learning based House Price Prediction Model," *2022 International Conference on Edge Computing and Applications (ICECAA)*, Tamilnadu, India, 2022, pp. 1423-1426, doi: 10.1109/ICECAA55415.2022.9936336.

[15] H. V, D. B. B, S. R, V. R and R. D, "A Comparative Study of House Price Prediction Using Machine learning and Deep learning Techniques," *2024 Third International Conference on Smart Technologies and Systems for Next Generation Computing (ICSTSN)*, Villupuram, India, 2024, pp. 1-6, doi: 10.1109/ICSTSN61422.2024.10671079.

Contributions By Each Student

Aditya Sajith - Proposed method and architecture, Results and Conclusion
Yuvan Karthik - Implementation
Nihal Ali - Dataset and implementation
Mohammed Farhaan - Abstract, Introduction, Future scope