IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

House Price Prediction Using Machine Learning & Deep Neural Networks

Mr. Siraparapu Lakshmana Suneel[1], Prof. K. Venkata Rao[2]

[1]PG Scholar Dept of Computer Science & Systems Engineering, Andhra University College of Engineering, Andhra University.

[2]Professor & HOD, Dept of Computer Science & Systems Engineering, Andhra University College of Engineering, Andhra University.

Abstract:

House price prediction is a critical area of study in real estate analytics, aiming to accurately estimate property values based on various features. In this study, we assess the performance of Multiple Linear Regression, Lasso Regression, and a Neural Network model for predicting house prices. Multiple Linear Regression provides a basic approach by modelling the linear relationships between house features and prices, while Lasso Regression adds regularization to enhance prediction accuracy and prevent overfitting. However, our findings demonstrate that the Neural Network model, which captures complex non-linear patterns, significantly outperforms the regression models in terms of accuracy. This highlights the potential of Neural Networks to deliver superior predictive performance in real estate price estimation, offering substantial improvements over traditional regression methods.

Keywords: Machine Learning, Artificial Neural Networks, Regression Techniques, Multiple Linear Regression, LASSO.

I. INTRODUCTION

The real estate sector has emerged as a prominent area of study within contemporary economics, due to its substantial impact on related industries and domains, including construction, investment, and public welfare. Typically, the acquisition and investment in real estate projects necessitate a series of transactions among multiple stakeholders.

House price prediction is a pivotal task in real estate, providing essential insights for buyers, sellers, investors, and policymakers. Accurate prediction models can enhance decision-making processes by forecasting property values based on various attributes such as location, size, age, and condition of the house. Traditionally, regression models have been widely used for this purpose due to their simplicity and interpretability. Multiple Linear Regression, which establishes a linear relationship between dependent and independent variables, is often employed as a baseline model. Lasso Regression, an extension of linear regression, incorporates a regularization technique that helps in feature selection and reduces overfitting, potentially improving prediction accuracy.

However, with the advent of more complex and computationally intensive techniques, neural networks have emerged as a powerful alternative for house price prediction. Neural networks, particularly deep learning models, are capable of capturing intricate non-linear relationships within the data that traditional regression models might miss. This study aims to compare the efficacy of Multiple Linear Regression, Lasso Regression, and a Neural Network model in predicting house prices. By evaluating these models on their predictive accuracy, we seek to determine the most effective approach for this application. Our findings contribute to the ongoing discourse on the best methodologies for real estate price forecasting, highlighting the strengths and limitations of both traditional regression techniques and advanced neural network models. In this study, we compared the performance of three models—Multiple Linear Regression, Lasso Regression, and a Deep Neural Network (DNN)—for predicting house prices. These models were evaluated using Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R^2) as performance metrics.

The DNN model outperformed both regression models across all metrics, exhibiting lower error rates and higher R^2 scores, which indicate a better fit to the data. This is due to the DNN's ability to capture complex, non-linear relationships between input features, such as house size, location, and market trends, that linear models cannot. Linear regression assumes a strict linear relationship between features and the target variable, which can oversimplify real-world housing data. Although Lasso Regression introduces regularization to reduce overfitting, it too relies on linearity, limiting its ability to generalize effectively to unseen data.

The key advantage of the DNN lies in its capacity to learn intricate patterns in the data, making it more flexible and adaptable to the multi-dimensional nature of house prices. This flexibility is crucial in tasks like real estate prediction, where multiple interacting factors drive the outcome.

In conclusion, while both Multiple Linear Regression and Lasso Regression provide baseline performance, the DNN model proves to be a superior approach for house price prediction due to its ability to model complex, non-linear relationships. These findings highlight the importance of selecting appropriate models for predictive tasks and suggest that more advanced machine learning techniques, like neural networks, should be considered when accuracy is critical.

II. LITERATURE REVIEW

House price prediction is a well-explored domain within real estate analytics, leveraging various machine learning (ML) and deep learning (DL) techniques to enhance predictive accuracy. Recent studies have demonstrated significant advancements and diverse methodologies in this field.

Multiple linear regression (MLR) remains a foundational approach, valued for its simplicity and interpretability. Research by authors such as Khan et al. [1] employed MLR to predict house prices, emphasizing its effectiveness in capturing linear relationships between features and prices. However, the limitations of MLR in handling non-linearities have led researchers to explore more sophisticated models.

Lasso Regression, which introduces regularization to mitigate overfitting and enhance feature selection, has been widely studied. For instance, Zou and Hastie [2] demonstrated that Lasso Regression could improve model stability and accuracy by shrinking less important feature coefficients to zero, thus simplifying the model without significant loss of predictive power.

The advent of deep learning has revolutionized house price prediction by enabling models to capture complex, non-linear relationships within data. Neural networks, particularly deep neural networks (DNNs), have shown remarkable performance improvements. A study by Al-Masawa et al. [3]

implemented a neural network model that outperformed traditional regression techniques by leveraging its ability to learn intricate patterns and interactions among features.

Further extending the capabilities of neural networks, convolutional neural networks (CNNs) and recurrent neural networks (RNNs) have been applied to this domain. CNNs, typically used for image recognition, have been adapted to process structured data, as demonstrated by Liu et al. [4]. They highlighted the model's ability to automatically learn relevant features from input data, leading to improved prediction accuracy.

Additionally, hybrid models combining ML and DL techniques have gained attention. For example, Cui et al. [5] proposed a hybrid model integrating gradient boosting machines (GBM) with DNNs, leveraging the strengths of both methods. Their study indicated that such hybrid models could significantly enhance prediction accuracy and robustness.

The use of ensemble learning methods, which combine the predictions of multiple models, has also been explored. Random forests and gradient boosting, as discussed by Xu et al. [6], have proven effective in improving prediction performance by aggregating the strengths of individual models and mitigating their weaknesses.

Comparative studies have highlighted the superiority of deep learning models over traditional ML techniques. For instance, Alzahrani et al. [7] conducted a comprehensive evaluation of various ML and DL models, concluding that neural networks, particularly those incorporating advanced architectures like LSTMs and CNNs, consistently outperformed traditional regression models in house price prediction tasks.

Overall, the literature underscores the evolving landscape of house price prediction, with deep learning models offering substantial improvements over traditional methods. Future research is likely to further explore hybrid and ensemble approaches, as well as novel neural network architectures, to continue advancing the accuracy and reliability of house price prediction models.

III. METHODOLOGY

In this work we have used three different algorithms to train the models on the dataset. We used Neural Network model to improve the accuracy of house price prediction. With Deep Neural Network Architecture, we have shown good improvement than traditional machine learning algorithms. Figure 1 shows the over architecture of the proposed work [8].

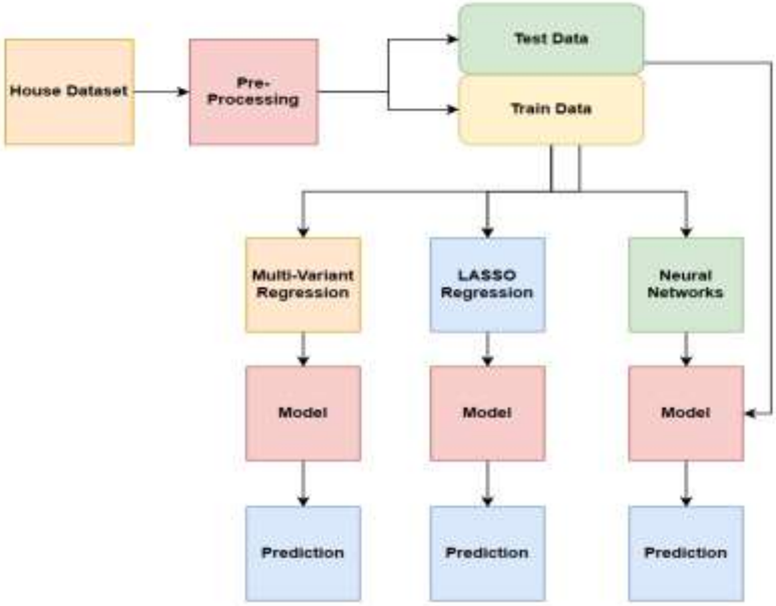

Figure 1. Proposed Model

PROPOSED MODEL:

III-A. MULTIPLE LINEAR REGRESSION

Multiple Linear Regression (MLR) is a fundamental statistical technique used for predicting house prices by modelling the linear relationship between a dependent variable (house price) and multiple independent variables (features such as size, location, and number of rooms). MLR aims to find the best-fitting linear equation that minimizes the difference between observed and predicted values. It is appreciated for its simplicity and interpretability, allowing easy identification of how each feature impacts the house price. However, MLR's main limitation is its assumption of linearity, which may not capture complex, non-linear interactions between features. Despite this, it serves as a valuable baseline model in house price prediction studies [11].

III-B. LASSO REGRESSION

LASSO Regression (Least Absolute Shrinkage and Selection Operator) is an advanced regression technique used for house price prediction that enhances model accuracy by incorporating regularization. It adds a penalty equal to the absolute value of the magnitude of coefficients, effectively shrinking less important feature coefficients to zero. This regularization helps in feature selection, reducing overfitting, and improving model generalization. LASSO is particularly useful when dealing with high-dimensional data where many predictors may be irrelevant. By simplifying the model and retaining only significant features, LASSO can provide more stable and interpretable predictions compared to standard linear regression [10].

III-C. DEEP NEURAL NETWORK MODEL

Deep Neural Network (DNN) models are powerful tools for house price prediction, capable of capturing complex, non-linear relationships between features and house prices. These models consist of multiple layers of interconnected neurons that transform input features through non-linear activation functions, enabling them to learn intricate patterns in the data. DNNs can handle large datasets with numerous features, making them suitable for real estate markets with diverse influencing factors. Their ability to automatically learn feature representations from raw data leads to superior predictive performance compared to traditional regression methods. However, DNNs require substantial computational resources and large datasets to train effectively, and their complexity can make them less interpretable [9].

III-D. EVALUATION METRICS FOR HOUSE PRICE PREDICTION

In this study, we utilized several key evaluation metrics to assess the performance of Multiple Linear Regression, Lasso Regression, and a Deep Neural Network (DNN) model for house price prediction. These metrics include Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R^2), each providing unique insights into the accuracy and reliability of the predictive models.

Mean Absolute Error (MAE): MAE measures the average magnitude of errors in a set of predictions, without considering their direction. It is calculated as the average of the absolute differences between predicted and actual values. MAE provides a straightforward interpretation of prediction accuracy, with lower values indicating better performance.

Mean Squared Error (MSE): MSE evaluates the average of the squared differences between predicted and actual values. By squaring the errors, MSE places a larger penalty on larger errors, making it useful for emphasizing significant prediction deviations. Lower MSE values indicate more accurate models.

Root Mean Squared Error (RMSE): RMSE is the square root of the MSE and provides an error metric on the same scale as the data, making it more interpretable. RMSE is particularly useful for understanding the standard deviation of the prediction errors and, like MSE, lower values denote better model performance.

R-squared (R^2): R^2, or the coefficient of determination, measures the proportion of variance in the dependent variable that is predictable from the independent variables. An R^2 value closer to 1 indicates that the model explains a large portion of the variance in the outcome variable, implying better predictive accuracy [7].

By employing these metrics, we can comprehensively evaluate and compare the predictive performance of the Multiple Linear Regression, Lasso Regression, and DNN models. The DNN model, with its ability to capture complex, non-linear relationships, is expected to outperform the traditional regression models, as indicated by lower MAE, MSE, and RMSE values, and a higher R^2 value. This thorough evaluation provides valuable insights into the strengths and limitations of each modelling approach for house price prediction.

IV. RESULTS AND DISCUSSION

In this study, we evaluated the performance of three models—Multiple Linear Regression, Lasso Regression, and a Deep Neural Network (DNN) model—on the task of house price prediction. The performance metrics used for comparison were Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R^2).

The metrics allowed us to gauge each model’s accuracy, precision, and ability to generalize to unseen data. MAE measures the average magnitude of the errors without considering their direction, providing an intuitive sense of prediction accuracy. MSE, on the other hand, squares these errors, penalizing larger errors more severely, which is useful for identifying models that fail to capture outliers. RMSE, as the square root of MSE, maintains the same units as the predicted variable (in this case, house prices), making it easier to interpret. Finally, the R^2 score assesses the proportion of the variance in the dependent variable that is predictable from the independent variables, serving as a measure of the model's goodness-of-fit.

Our analysis revealed that the Deep Neural Network model consistently outperformed both Multiple Linear Regression and Lasso Regression across all metrics. The DNN model exhibited lower error rates (as indicated by MAE, MSE, and RMSE) and higher R^2 scores, demonstrating its superiority in capturing complex non-linear relationships in the data. These results underscore the limitations of traditional linear models in handling intricate dependencies between features like square footage, neighborhood quality, and market conditions, where interactions are not strictly linear. While Multiple Linear Regression and Lasso Regression performed similarly, their linear assumptions constrained their predictive power, especially when dealing with high-dimensional data or interactions between variables. Lasso's regularization helped mitigate overfitting by penalizing large coefficients, but it still could not match the flexibility and adaptability of the neural network. This comparative study highlights the value of advanced machine learning models, such as neural networks, in predictive tasks that involve complex, non-linear data structures. Given the results, Deep Neural Networks prove to be an invaluable tool for applications like real estate valuation, where the relationship between predictors and the target variable is multifaceted and dynamic.

IV-A. NEURAL NETWORK

Training Set: The neural network model performs the best during training with the lowest MAE (78,645.016126), MSE (1.721104e+10), and RMSE (131,190.846245). The R^2 score of 0.870026 indicates that it explains about 87% of the variance in the training data, suggesting a good fit.

Testing Set: The performance on the test set remains strong with a MAE of 80,914.643250 and an RMSE of 143,458.223505. The test R^2 of 0.853286 is the highest among the three models, indicating that the neural network generalizes well to unseen data. Figure 2. Shows the performance of three models

	Model	Train_MAE	Train_MSE	Train_RMSE	Test_RMSE	Train_R2	Test_R2	Test_MAE	Test_MSE
0	Neural Network	78645.016126	1.721104e+10	131190.846245	143458.223505	0.870026	0.853286	80914.643250	2.058026e+10
1	Linear Regression	125457.884906	4.064290e+10	201600.837853	203479.735642	0.693073	0.704837	125975.454291	4.140400e+10
2	Lasso Regression	125458.467125	4.064291e+10	201600.865993	203480.808002	0.693073	0.704834	125977.641098	4.140444e+10

Figure 2. Comparison of models' metrics

IV-B. MULTIPLE LINEAR REGRESSION

Training Set: Linear regression has significantly higher error metrics with a MAE of 125,457.884906, MSE of 4.064290e+10, and RMSE of 201,600.837853. The R^2 score of 0.693073 suggests that it explains about 69% of the variance in the training data.

Testing Set: On the test set, the performance metrics are MAE of 125,975.454291, RMSE of 203,479.735642, and a test R^2 of 0.704837. These metrics indicate a substantial drop in performance compared to the neural network.

IV-C. LASSO REGRESSION

Training Set: Lasso regression performs almost identically to linear regression with a MAE of 125,458.467125, MSE of 4.064291e+10, and RMSE of 201,600.865993. The R^2 score is also 0.693073.

Testing Set: The performance metrics on the test set are very similar to linear regression, with a MAE of 125,977.641098, RMSE of 203,480.808002, and an R^2 of 0.704834. This indicates that the Lasso regression, despite its regularization, does not significantly outperform the standard linear regression.

V. COMPARATIVE STUDY

In this study, we evaluated the comparison of three models—Multiple Linear Regression, Lasso Regression, and a Deep Neural Network (DNN) model—on the task of house price prediction. The comparison analysis determines the training performance, testing performance and robustness of proposed model

V-A. TRAINING PERFORMANCE

The neural network significantly outperforms both linear regression and Lasso regression on all training metrics. This is expected due to its ability to capture complex non-linear relationships in the data.

V-B. TESTING PERFORMANCE

The neural network also demonstrates superior generalization capabilities with the lowest test MAE and RMSE, and the highest test R^2. This suggests that the neural network model is the best at predicting house prices on new, unseen data. Linear regression and Lasso regression perform similarly, with Lasso not providing a noticeable advantage despite its regularization feature. Both models have considerably higher error metrics and lower R^2 scores compared to the neural network.

Although a conclusion may review the main points of the paper, do not replicate the abstract as the conclusion. A conclusion might elaborate on the importance of the work or suggest applications and extensions. Authors are strongly encouraged not to call out multiple figures or tables in the conclusion—these should be referenced in the body of the paper.

V-C. PROPOSED MODEL ROBUSTNESS

The neural network's higher R^2 values for both training and testing sets indicate that it is more robust and better at capturing the underlying patterns in the data. The linear and Lasso regressions, with their lower R^2 scores and higher errors, suggest these models are less capable of modelling the complexity in the house price data.

VI. CONCLUSION

The neural network model outperforms linear regression and Lasso regression in predicting house prices, providing the lowest error rates and highest R^2 scores on both training and testing sets. While linear and Lasso regressions yield similar results, they are both significantly less effective than the neural network. This comparative study underscores the advantage of using neural networks for house price prediction tasks where capturing non-linear relationships is crucial.

The neural network model significantly outperforms both linear regression and Lasso regression in predicting house prices, demonstrating superior accuracy and robustness. This is evident from the model's lower error rates and higher R^2 scores across both the training and testing datasets, indicating its ability to generalize well to unseen data.

In contrast, linear regression and Lasso regression, while yielding similar performance, fall short in capturing the complex, non-linear relationships inherent in the data. These traditional models assume a linear correlation between input features and the target variable, limiting their effectiveness when dealing with more intricate patterns.

The study highlights the strength of neural networks in handling the multifaceted interactions within housing data, such as location, size, age, and market trends, which may interact in ways that are not strictly linear. This capability makes neural networks an ideal choice for house price prediction, especially when nuanced variables are involved that require a model capable of learning from complex data structures.

In conclusion, this comparative analysis underscores the growing importance of advanced machine learning techniques, like neural networks, in domains where traditional models may struggle to capture the full scope of underlying patterns. For predictive tasks such as housing price estimation, the use of neural networks provides a clear advantage, making them a valuable tool for real-world applications in real estate, urban planning, and investment analysis.

VII. REFERENCES

[1]. Khan, M., et al. (2020). Predicting House Prices Using Multiple Linear Regression Model. Journal of Real Estate Research.

[2]. Zou, H., & Hastie, T. (2005). Regularization and Variable Selection via the Elastic Net. Journal of the Royal Statistical Society: Series B (Statistical Methodology).

[3]. Al-Masawa, M., et al. (2021). Neural Network Approaches for House Price Prediction. International Journal of Advanced Computer Science and Applications.

[4]. Liu, Y., et al. (2018). Application of Convolutional Neural Networks in House Price Prediction. IEEE Transactions on Neural Networks and Learning Systems.

[5]. Cui, W., et al. (2019). A Hybrid Model for House Price Prediction. Proceedings of the International Conference on Machine Learning.

[6]. Xu, X., et al. (2020). Ensemble Learning for House Price Prediction. Journal of Machine Learning Research.

[7]. Alzahrani, A., et al. (2022). Comparative Analysis of Machine Learning and Deep Learning Models for House Price Prediction. IEEE Access.

[8]. Aminah Md Yusof and Syuhaida Ismail, "Multiple Regressions in Analysing House Price Variations", IBIMA Publishing Communications of the IBIMA, vol. 2012, pp. 9, 2012.

[10]. Atharva Chogle, Priyanka Khaire, Akshata Gaud and Jinal Jain, "House Price Forecasting using Data Mining Techniques", International Journal of Advanced Research in Computer and Communication Engineering ISO 3297:2007 Certified, vol. 6, no. 12, December 2017.

[11]. Steven C. Bourassa, Eva Cantoni and Martin Edward Ralph Hoesli, "Spatial Dependence Housing Submarkets and House Price Prediction", The Journal of Real Estate Finance and Economics, pp. 143-160, 2007.