Performance Analysis of House Price Prediction Models

Gnana Kousik (23B1000)

March 29, 2025

1 Performance Metrics

The following table summarizes the Performance of each model based on the performance metrics Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-Squared (R^2) score.

Model	**RMSE (Lower Better)**	**MAE (Lower Better)**	R^2 **Score (Higher Better)**
Linear Regression	**11,719.55**	**10,334.13**	**0.99896**
Decision Tree	88,013.64	71,358.00	0.94262
Neural Network	1,369,368.16	1,323,044.05	-8266.79

Table 1: Performance Comparison of Different Models

2 Analysis

2.1 Linear Regression (best performer)

Linear Regression has achieved the best results with lowest RMSE and MAE errors , indicating that the predicted prices are very close to the actual prices. It also has the highest R^2 score, indicating that the model explains almost all the variance of the data. Its advantages include :

- High interpretability and high computational efficiency.
- Suitable for data with Linear Relationships.
- Minimal overfitting as compared to Complex Models.

However. it's does not perform well with complex , non-linear, high variance datasets. In these datasets it underfits the data. It is sensitive to outliers

2.2 Decision Tree (moderate performer)

The decision model also performed resonably well with an R^2 score of 0.9426. However it has significantly higher RMSE and MAE values compared to Linear regression model. The likely reason that it performed worse than linear regression may be due strong linearity of the data. It's advantages are as follows :

- Can capture non-linear relationships and complex feature interactions.
- Decision tree provides a more clearer understanding of which features are more important than others.
- Works well directly with categorical data without encoding and also works well even without standardizing the data(unlike linear regression and neural network).
- It's more robust to outliers.

However, it's prone to overfitting leading to more predicition errors, if depth of the tree is large. It is more computationally expensive for larger datasets. It performs poorly on continuous data if sufficient splits are not done.

2.3 Neural Network (worst performer)

The neural network model exhibited the worst performance with an extremely high RMSE and MAE, and a negative R^2 score indicating that it performed worse that a naive mean based predictor. The likely reasons for poor performance :

- Insufficient data for deep learning, it requires larger datasets to generalize well. The model overfitted the training data for the current dataset.
- Poor hyperparameter tuning (e.g., hidden layers, learning rate).

It's advantages are as follows :

- Can model complex relationships.
- When trained with larger datasets it usually outperforms the traditional methods.
- Unlike decision and linear regression, Neural networks automatically learns feature interactions without requiring manual selection.

However it has limitations like Needs larger datasets, computationally expensive, difficult to interpret, prone to overfitting, requires hyperparameter tuning, Sensitive to outliers and local minima.

3 Conclusion

- Linear Regression is the best model for this dataset, as it provided the most accurate predictions with the lowest error.
- Decision Tree could be more useful if more complex, non linear relationships exists in the dataset.
- Neural Networks are not recommended for this dataset, as they performed poorly due to overfitting, data limitations, and computational inefficiency.