Real Estate Price Prediction Analysis Report

Muhammad Shafeen (22P-9278)

May 6, 2024

1 Introduction

This report presents an analysis of real estate data for predicting property prices. The dataset used contains various features such as building age, floor count, size, location, heating type, and others, which are utilized to predict the price of properties.

2 Data Preprocessing

2.1 Missing Values Handling

- The dataset contained missing values in several columns such as 'size', 'studentavailable', 'mortgageavailable', 'address', 'price', and 'currency'.
- Missing values in 'size' and 'price' columns were imputed using the mean of the respective columns.
- Rows with missing values were dropped for other columns as they were not imputable.

2.2 Feature Engineering

- 'Timeperiod' feature was created by calculating the duration between start and end dates.
- 'Room' and 'livingroom' features were extracted from the 'roomcount' column.

2.3 Outlier Detection and Handling

- Outliers were detected using the Interquartile Range (IQR) method.
- Rows with outliers were replaced with NaN values.
- Rows with NaN values were dropped.

3 Model Training

Two models were trained for price prediction:

3.1 Artificial Neural Network (ANN)

- The ANN model was implemented using Keras.
- The model architecture consisted of multiple dense layers with ReLU activation functions.
- Mean Squared Error (MSE) was used as the loss function and Adam optimizer was used for optimization.
- The model was trained for 10 epochs.

3.2 Linear Regression

- Linear Regression model from scikit-learn was utilized.
- No feature scaling was performed as Linear Regression is not sensitive to feature scales.

4 Model Evaluation

4.1 ANN Model

- Mean Squared Error (MSE) for the ANN model: 0.504
- The plot of actual vs. predicted prices showed a moderately good fit between the actual and predicted values.

4.2 Linear Regression Model

- Mean Squared Error (MSE) for the Linear Regression model: 0.711
- The plot of actual vs. predicted prices showed a moderate fit, slightly less accurate than the ANN model.

5 Conclusion

- The Artificial Neural Network (ANN) model outperformed the Linear Regression model in predicting property prices, achieving a lower Mean Squared Error.
- Further optimization of the ANN model and feature engineering could potentially improve the predictive performance.

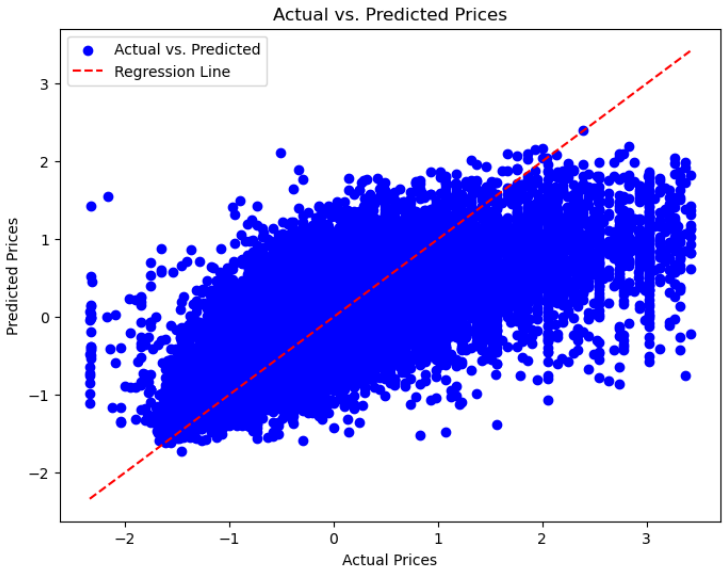

Figure 1: Actual vs. Predicted Prices - ANN Model

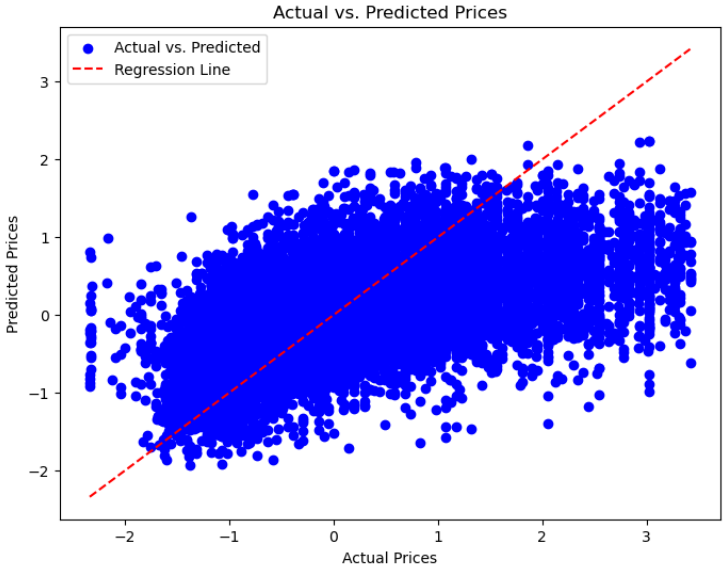

Figure 2: Actual vs. Predicted Prices - Linear Regression Model

6 Recommendations

- Explore additional feature engineering techniques to enhance model performance.
- Experiment with different neural network architectures and hyperparameters to improve model accuracy.
- Gather more data to improve the robustness of the predictive models.