CSST 102

Mangulabnan Jasper
BSCS 3A

The objective of this project is to predict the price of the houses based on the various features such as size, number of bedrooms, age of the house and proximity to downtown. The aim of this model is to assist real states in estimating house prices more accurately.

The data was loaded into the pandas dataframe for inspection and preprocessing.

The missing values are processed to clean the data that will be used.

To ensure all features were on a similar scale, the normalization was applied to the numerical features.

To split the data, I used the train and test split with 70/30 ratio to evaluate model performance.

Recursive Feature Elimination or RFE was used to identify the most significant features.

The linear regression model was trained using the training set.

The model was evaluated using Mean Square Error or MSE, R-Squared and adjusted R-Squared.

A Scatter Plot was created to visualize the relationship between predicted prices and actual prices.

Some of the challenge I faced is handling missing data, I implemented mean imputation for numerical columns to handle missing data effectively, ensuring the model was trained on complete data.

Other challenge is feature scaling, I applied normalization to bring all features to a similar scale, which improves the model's performance and stability.

Feature selection, I determine which feature significantly impact the house price. I used the RFE on this to automate feature selection, which helped in identifying the most relevant predictors and improving model interpretability.

In conclusion, the multiple regression model effectively predicts house prices based on the available features. With the R-Squared and Adjusted R-Squared values suggest that the model explains a substantial portion of the variance in house prices, making it a valuable tool for real estate professionals.

This report and model provides a comprehensive overview of the steps taken to develop, evaluate and interpret a multiple regression model for predicting house prices, along with insights into its applicability and limitations in real-world scenario