Machine Learning Task 1 - House Price Prediction

Internship Task Report

Name: Harshali Ramesh Upase

Internship: Credora - Machine Learning Internship

Task Title: House Price Prediction using Linear Regression

Objective:

To build a regression model that accurately predicts house prices based on real-world housing data.

Dataset Source:

Kaggle - House Prices: Advanced Regression Techniques

https://www.kaggle.com/c/house-prices-advanced-regression-techniques/data

Tools & Technologies:

- Python
- Pandas, NumPy
- Matplotlib, Seaborn
- Scikit-learn
- Google Colab

Workflow:

1. Data Preprocessing:
   - Handled missing values
   - Encoded categorical features

- Feature scaling where necessary

2. Feature Selection:

- Selected key features such as:
  * GrLivArea (Above ground living area)
  * BedroomAbvGr (Number of bedrooms)
  * FullBath (Number of full bathrooms)

3. Model Building:

- Used Linear Regression model from scikit-learn
- Split the dataset into training and testing sets

4. Model Evaluation:

- Evaluation metrics such as MAE, RMSE, R^2 score were used
- Compared actual vs predicted prices using graphs

Output:

- Predicted house prices for test data
- Visualization: Actual vs Predicted values (tabular & graphical)

Live Demo:

Google Colab Link - To be added

Conclusion:

Successfully implemented a regression model using Linear Regression to predict house prices. The

project helped in understanding the process of feature selection, preprocessing, model evaluation, and visualization.