REPORT
Task 01: House Price Prediction using Linear Regression

1. Introduction

House price prediction is a common real-world Machine Learning problem used in the real estate industry to estimate the selling price of a property based on different features. Accurate prediction helps buyers, sellers, and businesses make better decisions.

In this project, a **Linear Regression model** is developed to predict house prices using basic property-related features such as **square footage**, **number of bedrooms**, and **number of bathrooms**.

2. Objective

The objective of this task is to build a Machine Learning model that can **predict the house sale price** based on:
• Square Footage
• Number of Bedrooms
• Number of Bathrooms

3. Dataset Description

The dataset used is taken from Kaggle:
House Prices: Advanced Regression Techniques
(https://www.kaggle.com/c/house-prices-advanced-regression-techniques/data)

For this task, the following columns were used:
• GrLivArea (Square Footage)
• BedroomAbvGr (Bedrooms)
• FullBath (Bathrooms)
Target: SalePrice

4. Methodology

This project follows a standard Machine Learning workflow:

4.1 Data Loading: Loaded train.csv using Pandas.
4.2 Data Preprocessing: Selected required columns and removed missing values.
4.3 Train-Test Split: Split into 80% training and 20% testing.
4.4 Model Training: Trained Linear Regression model using scikit-learn.
4.5 Prediction: Predicted house prices on test data.

5. Model Evaluation

Model performance was evaluated using:
• MAE (Mean Absolute Error)
• MSE (Mean Squared Error)
• RMSE (Root Mean Squared Error)
• R^2 Score

6. Results and Visualization

A scatter plot of **Actual vs Predicted Prices** was created to visually analyze the model performance. A residual distribution plot was also generated to analyze prediction errors.

7. Sample Prediction

The model was tested on a sample input:
• 2000 sqft
• 3 bedrooms
• 2 bathrooms
The model outputs a predicted sale price based on learned patterns.

8. Conclusion

This task demonstrates how a Linear Regression model can predict house prices using simple features. The model was trained and evaluated with proper metrics and visualization.

9. Future Improvements

Possible improvements include:
• Adding more relevant features
• Using Ridge/Lasso regression
• Trying advanced models like Random Forest or XGBoost
• Hyperparameter tuning using GridSearchCV
• Deploying the model using Streamlit

10. Tools and Libraries Used

Python, Pandas, NumPy, Matplotlib, Seaborn, scikit-learn, Jupyter Notebook

Author

Pratham Agarwal
Machine Learning Intern – Prodigy InfoTech
GitHub: https://github.com/1234pratham2k6k1234-glitch