ACTION PLAN FOR HOUSE PRICE PREDICTION PROJECT

-Surya Prabha V P

Project Goal

The objective of this project is to build and evaluate machine learning models to accurately predict house prices using a public dataset. This project will serve as a comprehensive demonstration of data science skills, including data analysis, model building, and documentation.

Project Timeline and Key Steps

Week 1: Data Collection & Preprocessing

- **Task 1:** Download and load the house price dataset into a Jupyter Notebook.
- **Task 2:** Perform initial data inspection (.info(), .describe(), etc.) to understand the dataset's structure and identify missing values.
- **Task 3:** Clean the data by handling missing values appropriately for both numerical and categorical features.

Week 2: Exploratory Data Analysis (EDA) & Feature Engineering

- **Task 1:** Conduct univariate and bivariate analysis using visualizations such as histograms, scatter plots, and box plots to understand key feature distributions and relationships with SalePrice.
- **Task 2:** Identify and handle outliers, specifically in the GrLivArea feature, to improve model performance.
- **Task 3:** Perform hypothesis testing (e.g., T-test) to statistically validate key insights from the EDA, such as the relationship between OverallQual and SalePrice.
- **Task 4:** Apply a log transformation to the skewed SalePrice variable to normalize its distribution.

Week 3: Model Building & Evaluation

- **Task 1:** Prepare the data for machine learning by performing one-hot encoding on all categorical features.
- **Task 2:** Split the data into a training set and a testing set.
- **Task 3:** Train a **Linear Regression** model to establish a baseline performance.

- **Task 4:** Train and evaluate three advanced regression models: **Lasso Regression**, **XGBoost**, and **Random Forest Regressor**.
- **Task 5:** Calculate key performance metrics (RMSE, R^2, MAE) for all models.

Week 4: Finalization & Documentation

- **Task 1:** Compare the performance of all four models to determine the best-performing one.
- **Task 2:** Draft the main **Project Report** (PDF/DOCX), including all sections from the Abstract to the Conclusion.
- **Task 3:** Draft this Action Plan Document.
- **Task 4:** Create a GitHub repository and upload all project files, including the Jupyter Notebook, reports, and a requirements.txt file.
- **Task 5:** Write a detailed README.md file for the GitHub repository to provide a clear project overview.