Tester Document

Table of Contents

Team Members

1. Shravani R S
2. Himaja S
3. Karanbir Singh
4. Tarunaa A C
5. Anushka Dutta

Date: 5th December 2024
Prepared By: Shravani R S

1. Load and Explore the Dataset

Objective

- Verify that the dataset is loaded correctly.
- Check for data integrity (missing values, duplicates, column names).
- Explore the dataset's structure and key statistics.

Test Cases

Test id	Test Case	Expected Outcome	Pass/Fail
TC1	Load the data file	File loads successfully with correct data type	Pass
TC2	Verify dataset dimensions	Shape matches the expected number of row/columns	Pass
TC3	Check for missing or duplicate values.	Accurate counts of missing/duplicate entries are returned.	Pass
TC4	Display basic statistics	Summary statistics are generated correctly	Pass
TC5	Visualize data distribution.	Distribution plots are created without errors.	Pass

2. Feature Engineering

Objective

- Create new features, handle categorical data, and check correlations.
- Ensure engineered features improve model performance**.**

Test Cases

Test ID	Test Case	Expected Outcome	Pass/Fail
TC6	Generate polynomial or interaction features.	New features are added and verified correctly.	Pass
TC7	Encode categorical variables.	Categorical features are encoded without errors.	Pass
TC8	Drop irrelevant or low-variance features.	Unnecessary features are removed.	Fail
TC9	Analyze correlations between features and target.	Heatmaps or correlation matrices display accurate results.	Pass

3. Preprocessing and Splitting Data

Objective

- Prepare data for training by handling outliers, scaling, and splitting.

Test Cases

Test ID	Test Case	Expected Outcome	Pass/Fail
TC10	Handle missing values.	Missing values are replaced or removed.	Pass
TC11	Scale numerical features.	Data scaling (standardization or normalization) is accurate.	Fail
TC12	Split dataset into training and test sets.	Splits match specified ratios.	Pass

4. Implement Linear Regression

Objective

- Train and test a baseline linear regression model.
- Analyze the coefficients and residuals.

Test Cases

Test ID	Test Case	Expected Outcome	Pass/Fail
TC13	Train the linear regression model.	Model trains successfully with no errors.	Pass
TC14	Extract and analyze coefficients.	Coefficients are displayed and interpretable.	Pass
TC15	Visualize residuals.	Residuals are distributed randomly around zero.	Fail

5. Evaluate the Model

Objective

Validate model evaluation using metrics.

Test Cases

Test ID	Test Case	Expected Outcome	Pass/Fail
TC16	Calculate MSE, RMSE, and R^2 scores.	Metrics are computed accurately.	Pass
TC17	Perform cross-validation.	Model performance is consistent across folds	Pass

6. Linear Regression Alternatives

Objective

Test alternative models for performance improvements.

Test Cases

Test ID	Test Case	Expected Outcome	Pass/Fail
TC18	Train and evaluate a decision tree model.	Model performs better or similarly to baseline.	Pass
TC19	Train and evaluate a random forest model.	Random forest improves prediction accuracy.	Pass
TC20	Train and evaluate an SVM model.	SVM runs successfully and outputs valid results.	Pass
TC21	Train and evaluate a KNN model	KNN predictions vary with different k values.	Pass

7. Time Series Prediction

Objective

Test time series models for sequential data.

Test Cases

Test ID	Test Case	Expected Outcome	Pass/Fail
TC22	Analyze trends and seasonality.	Trends and seasonality are visualized correctly	Fail
TC23	Train and test ARIMA model.	Model fits well and forecasts accurately.	Pass
TC24	Train and test an LSTM model.	Model captures sequential dependencies properly.	Pass
TC25	Evaluate time series predictions.	Metrics like RMSE and MAE are computed correctly.	Pass

Testing Guidelines

1. Code Inclusion

- Include code snippets for key tasks like model training, preprocessing, and evaluation.
- Use graphs (e.g., distributions, feature importance, trends) for validation and comparison.

2. Graphs and Visualizations

- Use distribution plots for data exploration.
- Add heatmaps for feature correlation.
- Provide residual plots and prediction error charts for model evaluation.
- Time series data should include trend and forecast plots.