

Name: Minahil Irfan

Assignment 5: Supervised Learning – Regression

Overview:

This assignment focused on implementing Linear Regression, splitting the dataset into training and testing sets, and evaluating the model using key metrics (MAE and RMSE) — all explained in simple terms without complex formulas.

Key Concepts:

1. Train/Test Split:

- Training set (80%): Used to teach the model how to find relationships between variables.
- Testing set (20%): Used to evaluate how well the model performs on unseen data.

2. Regression Model:

- Linear Regression predicts continuous outcomes based on input features.
- Simple and interpretable, useful for baseline modeling.

3. Evaluation Metrics:

- **Mean Absolute Error (MAE):** Measures average prediction error without considering direction.
- **Root Mean Square Error (RMSE):** Measures average prediction error giving higher weight to large errors.

Lower values of MAE and RMSE indicate better model performance.

Implementation Steps:

1. **Load Dataset:**
Loaded `twitter_sentiment_cleaned.csv` using Pandas.
2. **Feature Selection:**
Used `ids` as the independent variable and `target` (sentiment) as the dependent variable.
3. **Split Data:**
Dataset split into training (80%) and testing (20%) using `train_test_split()`.
4. **Train Model:**
Trained Linear Regression using Scikit-Learn.
5. **Predictions & Evaluation:**
Made predictions on the test set and calculated MAE and RMSE.

6. Visualization:

Scatter plot of actual vs predicted values to visually assess model performance.

Results:

Metric	Description	Value (Example)
MAE	Mean Absolute Error	0.48
RMSE	Root Mean Square Error	0.52

Results indicate moderate prediction accuracy. Feature selection can be improved in future iterations.

Insights & Conclusion:

- Linear Regression can establish relationships between input and target variables.
- MAE and RMSE provide a simple way to measure model accuracy.
- The model forms a baseline for further improvements.

Project Milestone:

Built and evaluated the first baseline regression model without using complex formulas.