Hemali Chotalia
23BHI10073

GDG_WTM_TASK2

Project: Sentiment Analysis of Product Reviews

Step 1: I downloaded a Twitter Sentiment Analysis dataset from Kaggle, which contains tweets and their corresponding sentiments (Positive/Negative).

Step 2: Installed and imported the necessary libraries for the analysis

Step 3: Data Preprocessing

- Cleaned Text Data:
  Removed noise (punctuation, stopwords, etc.) and tokenized the text into useful features.
- Handled Missing Values:
  Ensured that both the feature matrix `X` and target labels `y` were free of `NaN` values and aligned in size.

Step 4: Feature Extraction (TF-IDF)

- TF-IDF (Term Frequency-Inverse Document Frequency):
  Converts text data into numerical vectors. It reflects how important a word is in a document relative to the entire dataset.

Step 5: Splitting the Data

- Train-Test Split:
  - Training Set (80%): Used to train the model.
  - Test Set (20%): Used to evaluate model performance.

Step 6: Model Training

- Logistic Regression:
  Fitted the model using the training data (`X_train` and `y_train`).

Step 7: Predictions:
Used the test data (`X_test`) to predict sentiments.

Evaluation Metrics:
Calculated accuracy, precision, recall, and F1-score to assess performance.