Documentation of Emotion Detection in Text

# 1. Install Required Libraries

!pip install pandas numpy scikit-learn nltk  
  
Installs the necessary libraries for the project:  
- pandas: For data manipulation.  
- numpy: For numerical operations.  
- scikit-learn: For machine learning algorithms.  
- nltk: For natural language processing (NLP).

# 2. Import Libraries

Imports the required libraries for data handling, NLP, and machine learning.

# 3. Download NLTK Data

Downloads necessary NLTK resources like tokenizers and stopwords.

# 4. Load and Preprocess the Dataset

Loads the dataset from a CSV file, renames columns, and handles missing values.

# 5. Preprocess Text Data

Cleans and tokenizes text by removing stopwords, lowercasing, and keeping only alphanumeric tokens.

# 6. Reduce Dataset Size

Reduces the dataset size by selecting 500 random samples for faster training.

# 7. Split the Data into Training and Testing Sets

Splits the data into 80% training and 20% testing sets.

# 8. Model Training with Naive Bayes

Trains a Naive Bayes model using a TF-IDF vectorizer for text transformation.

# 9. Model Evaluation

Evaluates the model with accuracy and classification report metrics.

# 10. Save the Model

Saves the trained model as 'emotion\_model.pkl' using pickle.

# 11. Additional Analysis and Cleaning

Removes duplicate rows from the dataset.

# 12. Hyperparameter Tuning with GridSearchCV

Uses GridSearchCV to find the best hyperparameters for the Naive Bayes model.

# 13. Final Model Evaluation

Evaluates the model using accuracy, precision, recall, and F1-score on the test set.