**Sentiment Analysis Assignment**

**Introduction**

This task involves text classification using deep learning techniques. The dataset used consists of tweets labeled with sentiment categories. The workflow includes data preprocessing, model training, and evaluation.

**Dataset Description**

* **Source:** Twitter training dataset (CSV file)
* **Columns:**
  + id: Unique identifier (removed during preprocessing)
  + country: Country of origin (removed during preprocessing)
  + Label: Sentiment category (Target variable)
  + Text: Tweet content (Feature variable)
* **Data Cleaning:**
  + Removed unnecessary columns (id, country)
  + Handled missing values by dropping null entries

**Data Preprocessing**

**Text Cleaning**

* **Lowercasing:** Converted all text to lowercase
* **Removing HTML tags:** Used BeautifulSoup to strip HTML elements
* **Removing URLs and special characters:** Applied regular expressions for text normalization
* **Tokenization:** Used TensorFlow's Tokenizer to convert text into sequences
* **Padding:** Applied padding to ensure uniform input size

**Model Training**

* **Neural Network Architecture:**
  + Embedding Layer (Pre-trained word embeddings used)
  + LSTM (Long Short-Term Memory) layers for sequence processing
  + Dense layers with activation functions for classification
  + Dropout layers to prevent overfitting
* **Training Process:**
  + Optimizer: Adam
  + Loss Function: Categorical Crossentropy
  + Evaluation Metrics: Accuracy

**Model Evaluation**

* **Confusion Matrix:** Analyzed class-wise prediction performance
* **Classification Report:** Included Precision, Recall, and F1-Score
* **Accuracy Metric:** Used to measure overall performance

**Results & Conclusion**

* Successfully trained an LSTM-based text classification model
* Achieved reasonable accuracy in sentiment classification
* Preprocessing steps significantly improved model performance