**Week 1 Project Summary**

**1. Importing Libraries**

We started by importing the necessary libraries for data handling, natural language processing, and model building, including pandas, nltk, and tensorflow.

**2. Data Collection**

* **Data Sourcing**: The dataset was sourced and provided by **Mostafa Elhashmy**.

**3. Data Reading**

* **Reading and Loading Data**: The data was loaded into a DataFrame using pandas for further analysis. This task was completed by **Eyad Bassem** and **Ahmed Abd Elnasser**.

**4. Data Preprocessing**

The following tasks were handled by **Mohamed Mokhtar**, **Mohamed Magdy**:

* **Text Cleaning**: Several preprocessing steps were applied:
  + **Stop Words Removal**: Common, non-informative words were removed.
  + **Punctuation Removal**: Standardized the text by eliminating punctuation.
  + **Repeating Characters Removal**: Handled excessive repetition of characters.
  + **Tokenization**: Converted the cleaned text into word tokens.
* **Lemmatization**: Lemmatized words to reduce inflectional forms and improve generalization.

**5. Separating Features and Labels**

We separated the input feature (text) and the target label (label). The features were assigned to X, and labels to y.

**6. Preparing Input Features for Training**

The following tasks were handled by **Ali Nabil**, and **Mohamed Reda El Esh**:

* **Text Tokenization**: Tokenized the tweets using a Tokenizer to transform them into numerical sequences.
* **Padding**: Ensured each sequence was padded to a uniform length of 500 words.

**7. Data Splitting**

Finally, we split the dataset into training and testing sets to ensure the model would be evaluated on unseen data after training.