Module 9: Introduction to NLP
- Assignment

Problem Statement:

You are a Data Scientist in a big firm. You have to develop a deep learning model to perform sentiment analysis on a dataset of tweets related to various candidates.

Tasks to be Performed:

Data Loading and Preprocessing:

- Load the tweet data from a CSV file.

- Filter out the relevant columns: 'candidate', 'sentiment', and 'text'.

- Preprocess the text data by removing stop words, punctuation, converting to lowercase, and other cleaning steps.

Text Vectorization:

Convert the preprocessed text data into numerical format using tokenization and padding, so that it can be fed into a deep learning model.

Model Development:

Develop a deep learning model using TensorFlow and Keras. The model includes an Embedding layer, a SpatialDropout1D layer to prevent overfitting, an LSTM layer for sequence data processing, and a Dense layer for output. It aims to classify the sentiment of each tweet into one of the three categories.

Model Training and Evaluation:

- Train the model on the processed text data, using categorical cross-entropy as the loss function, and accuracy as the evaluation metric.

- Use a validation split to evaluate the model's performance and prevent overfitting.