Assignment 2: Sentiment Classification

The dataset was divided into 3 sets:

a. Train dataset (80% of dataset)
b. Dev dataset (10% of dataset)
c. Test dataset (10% of dataset)

The Bi-Lstm model was trained on a **Train dataset** which gave an accuracy of **97.62%** and F1-Score of **0.9761**. When analyzed on the **dev dataset** it gave an accuracy of **89.38%** and F1-Score of **0.89**.

Finally the model and the tokenizer was saved in the google drive, the link is attached here

On loading the saved model and testing it on the **test dataset** it gave an accuracy of **89%** and F1-Score of **0.89**

How to run the code?
This google drive link contains the **dataset**, the **tokenizer** as well as the **Bi-Lstm Model**.

First put all the files in your drive folder and then follow these steps:

Enter the path of the IMDB dataset in the variable **ImdbDatasetPath** (here note that the path of the IMDB dataset should contain the name of the file as well like: path/IMDB dataset.csv)

Enter the path of the location where you want to save the model in the **ModelPath** variable

Enter the path of the location where you want to save the Tokenizer in the **TokenizerPath** variable (note to include the file name as well as in the case of IMDB dataset)