Case Study on "IMDB reviews Sentiment Analysis"

Analyse how travellers in February 2015 expressed their feelings on Twitter



Tasks (Each task 10 marks):

- 1. Load the dataset "Tweets.csv" and explore. Select the required columns and pre-process if needed.
- 2. Perform text pre-processing and tokenization
 - a. Display the size of vocabulary.
 - b. Print a tweet and it corresponding encoded doc(vector).
- 3. Build the text classifier model using LSTM
 - a. Print the model summary
- 4. Train the model for 5 epoch and validation split as 20%
 - a. Plot the accuracy and loss
 - b. Add a dropout layer of 20%.
 - i. Plot the accuracy and loss
 - c. Comment on the accuracy and loss changes.
- 5. Define a function that takes a text as input and outputs its prediction label.