IST736 Text Mining  
HW2 Linear Classifiers

Instruction

1. Download the [Kaggle Financial Sentiment Data](https://www.kaggle.com/datasets/sbhatti/financial-sentiment-analysis?resource=download)
2. Use a randomized sample of 80% data for training, and the rest 20% for testing
3. Build a linearSVC classifier using unigrams. You can decide on the other vectorization options.
   1. Report the top 20 positive features and negative features.
   2. Report the f1 and accuracy results.
   3. Examine up to 25 FP and FN errors and report linguistic patterns.
4. Similarly, build a logistic regression classifier using pre-trained fastext embedding. Repeat 3b and 3c and report result.
5. Write one paragraph conclusion, summarizing the strengths and weakness of using these two methods for financial sentiment analysis.
6. Submit your python notebook script with explanatory comments.