NLP Assignment 2_AMJ: Last Date of Submission: 11-04-2020

Objective:

In this exercise we will build an analyser to process the performance of banks with respect to the recent government decisions. Given a tweet corpus, this model computes the performance of each bank and assigns a score.

Steps:

(a) Dataset

i. You are provided a dataset in csv form that has tweet versus label. The labels are Displeasure, Compliment and Miscellaneous

ii. You are required to use the @xxx at the beginning of the tweets in order to identify the target bank to which the tweet is addressed.

iii. Create a test dataset that contains randomly selected 500 tweets per bank. This will be used later to compute the bank's performance

(b) Building Classifier

i. You are required to build the bank performance analyzer using 3 classifiers:

(a) MaxEnt (b) Deep learning classifier (c) Hybrid

ii. Train the system with the labelled corpus. Write suitable feature functions for the MaxEnt, use word2vec for the deep learning and combine these features to train a hybrid model (You should think through the design yourself, I am not providing step by step instructions so that you can think through with more freedom)

iii. Use the held out dataset and compute a performance index for each bank. Rank order these and identify best and worst performers

(c) As the tweets may have spelling errors, perform your study using both word2vec as well as lexicalized word2vec

(d) Before training the system, you may perform necessary pre-processing such as tweets cleaning etc.

Deliverables

1. Source code of classifiers, application

2. Test results: Who is the topper among banks?

3. Explain in the document what went right and where you could have taken a better approach to getting better results. Optionally you can include any graphics, visualization etc

Submit the deliverables only at amj.nlp2019@gmail.com

Submission on ashwinimjoshi@pes.edu will not be evaluated.

All the Best!!