DS 7337 Final:

PART I

Given the GOP twitter dataset (a dataset of tweets from 2012 with 3 sentiments—see the attached file):

1: Build a model the predict the sentiment of the tweet based on a sequence of characters and a second model based on a sequence of bi-grams (2-letter sequences).

Please include ALL the code you used to develop this model.

2. What is the vector you learned for the following emoji.

If you have trouble figuring out which face it is, the total count in the data set is 95

😂

Send your model to [rslater@smu.edu](mailto:rslater@smu.edu) (Probably to big to submit to 2DS)

# Save the model  
model.save('path\_to\_my\_model.h5')

3. What is the most similar character for the above emoji

PART II

Given the following paper: <https://arxiv.org/pdf/1802.05365.pdf>

Write an essay (min 500 words) Describing this model, how it is trained and on how this method captures each method described in the NLP pyramid. Please cite specific passages where the paper mentions different levels.

