Jaiv Doshi
jpdoshi@usc.edu

CAIS++ Winter Project

---

2. Since the valence was separated into discrete values {0, 2, 4}, I chose to use classification to separate the tweets into those three categories. Although regression could be used to...

3. I based my data pre-processing on the RNN Notebook. Therefore, I followed the steps to assign the data to tweets and valence, tokenize the data by converting the sentences to a sequence of words, pad sequences to ensure the samples are the same size, load the pre-trained word embeddings, and prepare the word embedding matrix.

4. By my understanding, the features (post-processing) were the 118 spaces for words. The targets for post preprocessing are the valence choices {0,2,4}.

5. I used the naive bayes classifier as my model. I tested different classifier models such as K-nearest neighbors, Decision Trees, Logistic Regression, and the naive bayes returned the best result.

7. Similarly, I tested different values of splitting training and chose 0.70 as the most accurate number.

8. I chose accuracy as the evaluation metric; however, my accuracy was not optimal as it was around 50% over every run.