

DATS_6312_NLP

Team 3: Jonathan Schild, Thomas Stanton, Ashish Telukunta

Bert-based multiclass classifier is trained on approximately 90,000 tweets to a .95 validation accuracy and .95 F1 score. This model is then applied to a separate dataset of unlabeled comments. The results are filtered to only retain the reddit comments that contained an emotion with above a .6 "confidence" score. These comments and their new labels are then combined with out original twitter dataset, and used to train a new model. This second model achieves a validation accuracy of .91 and F1 of .91. The apply file loads the model and applies it to some input'ed text.

Implement a scraper to get comments from Reddit posts (live) to be fed to the model. Display the sentiments (emotions, pos/neg) of the text specified by the user of the application. The application will also display the general change in sentiments for a given subreddit over X amount of time.

This application is intended to illustrate changes in sentiments of a given subreddit over time. It can assist moderators identify times when intervention might be warranted, or predict and prepare for events that could have longlasting negative impacts on the direction of their subreddit.