**Ergool App Store Insight Challenge**

Overview

The task in hand required to explore the comments of 5 apps from the google play store. The 5 apps were named jobswipe, jobget, harri candidate, poachedjobs, jobseeker. The total number of comments found from each of these apps are –

|  |  |
| --- | --- |
| **App** | **Number of Comments** |
| Candidate | 143 |
| Jobget | 3374 |
| Jobseeker | 8909 |
| Jobswipe | 703 |
| Pocahed jobs | 105 |

The total number of usable comments were – 13225

Using keyBERT for top keywords. For a three word keyword search we got –

* 'job apps good'
* 'job apps great'
* ‘job app better'
* ‘better job apps'
* 'suggest app jobemployment'

5 word phrase top 5 keywords –

* 'find job good great app'
* 'find job need excellent app'
* 'good app find job good'
* 'good app opportunity get job'
* 'find job good app great'

1 word keywords using KeyBERT Maximal Marginal Relevance –

* 'jobsearching'
* 'applay'
* 'confidence'
* 'ovfyqqqefywfqwwfjsowakmzmsjzmzjsjssjissijssjsjjsjx'
* 'coronavirus'

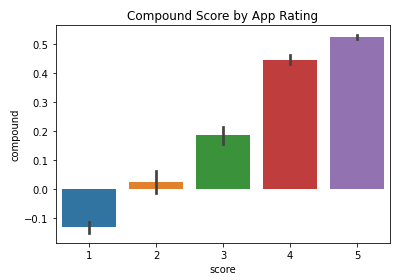
1 word keywords using KeyBERT Max Sum –

* jobswip
* 'appsuper'
* 'jobemployment'
* 'jobsearch'
* 'apps'

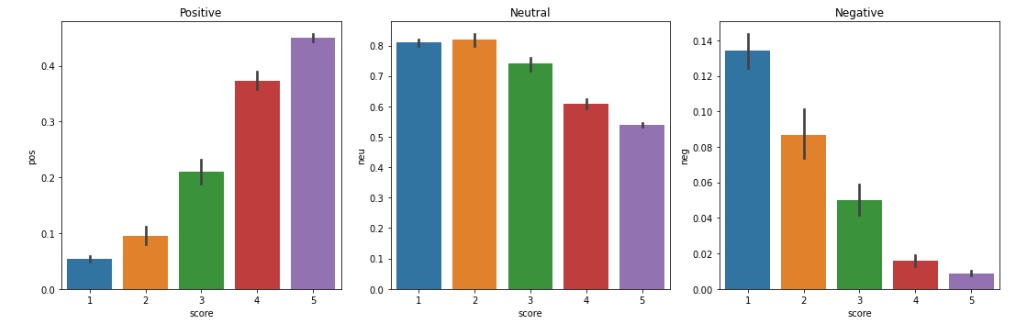
Average negative aptitude of all the comments is – 0.031

Average neutral aptitude of all the comments is – 0.60

Average positive aptitude of all the comments is – 0.37

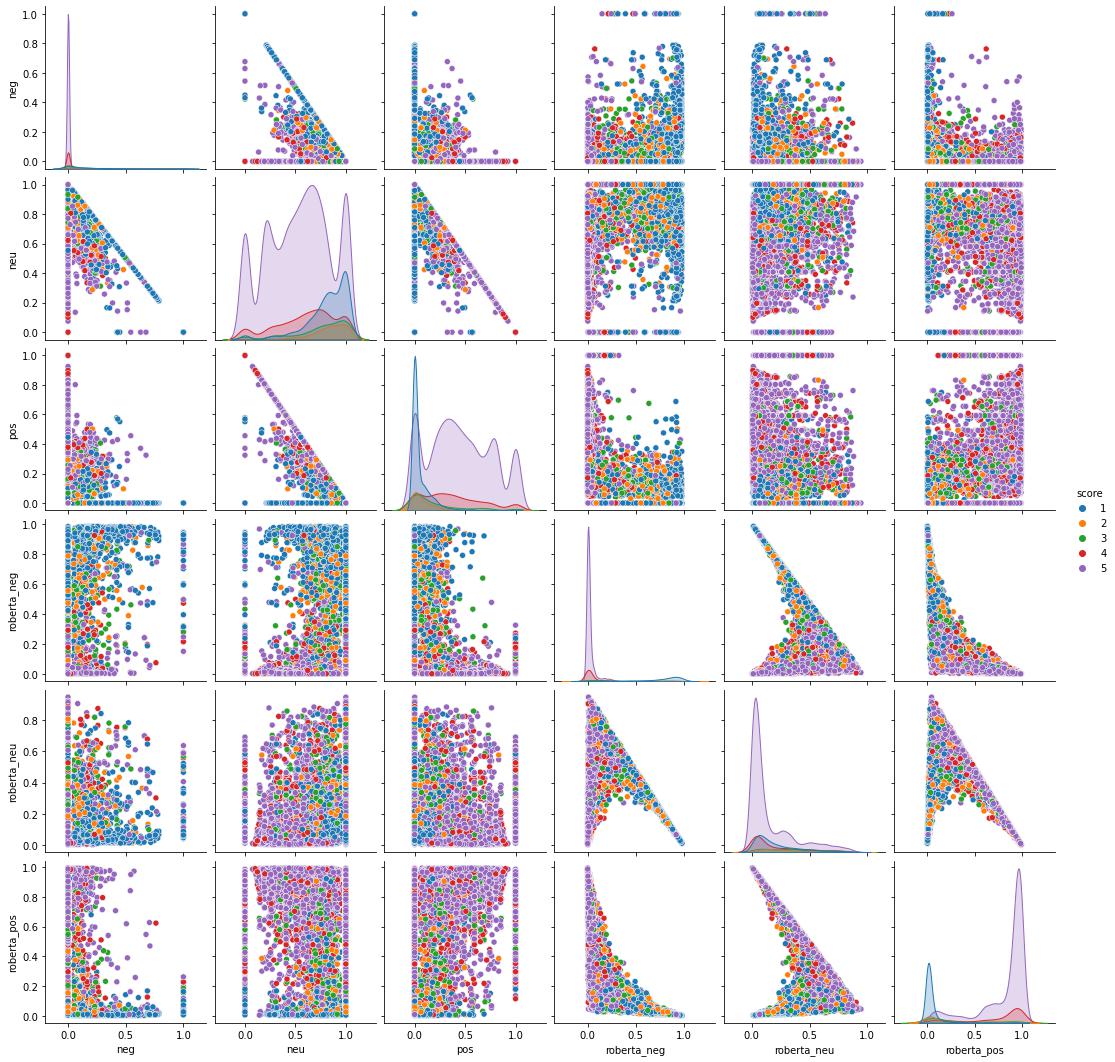


The compound score shows that majority of the comments are 5 rated with positive aptitude. And all the 1 rated comments have a negative aptitude close to -0.1.



Looking at the score to aptitude relationship. We can see that the higher the rating the more positive the comment is. Which is a linear relationship and is expected. On the other hand it is interesting to see that more neutral comments have come from rating 2. Which is below average but the model found the comments to be more on the neutral side. Finally, following the trend the negative comments have an inverse relationship with the rating. The higher the rating the less negative the comments are.

A comparison between Vader results and Roberta Model Results:



Found 4934 different words in the corpus. And the highest length of a sentence was 78.

The LSTM model achieved 97.7% training accuracy and on average 92% validation accuracy which shows how robust the model is. The model also had a very minimal loss 6% training loss and 36% validation loss. With the test data the model was successful in classifying the sentiment.