**Comparing google, vader and flair sentiment analysis of the stock tweets.**

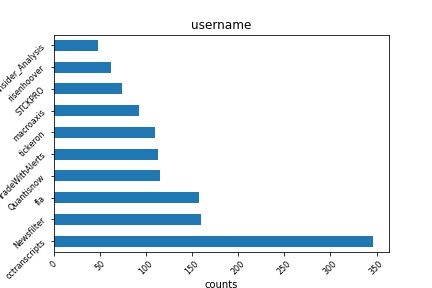
Natural language processing(NLP) has been a separate field of data science and machine learning. In NLP text processing is a huge field, most of the data we get is unstructured text format. Among different text processing techniques, a sentiment analysis, a way to find if this is making the positive sentiment or negative sentiment a hot topic. We can see the global market is able to gather tons of data from customer reviews and comments and also the tweets from social media.

In my Capstone 3 project I choose more than 2500 stock tweets. These tweets are from 100 different stock tickers. It consists of tweets for different usernames on different dates.

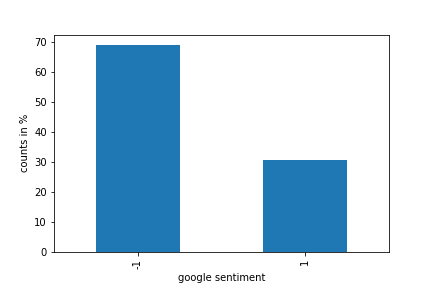
I used the pretrained model which is built by google, flair and vader to get the sentiment analysis. I mainly discuss the difference and compare the results of these three different models.

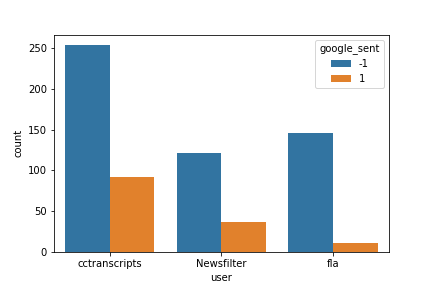
**Part I.Observe and compare three pretrained models**

**I.a Compare for different highest count user**

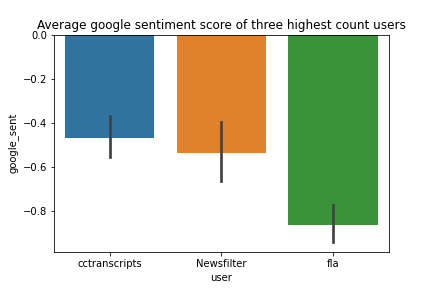
350

This plot shows the count of username. The **cctranscripts** has a maximum number of tweets around 350 , this is followed by the **Newsfilter** and **fla** 150. I used the pretrained model which is built by google, flair and vader to get the sentiment analysis.

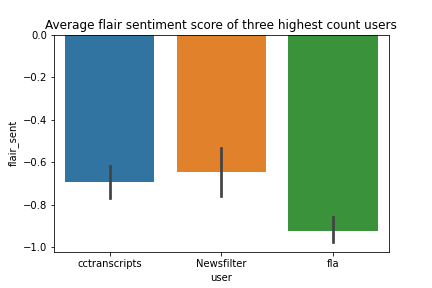


Here we can clearly see the bias in the data, more than 70% are negative sentiment. I was interested to see the distribution of the sentiment score of the tweets of the highest count users. 

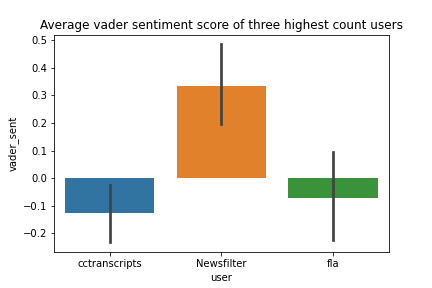
The highest count users also have significantly negative sentiment.



This plot shows the average of the google sentiment for the highest user and it clearly shows the average sentiment is negative and minimum for **fla** compared to two others.

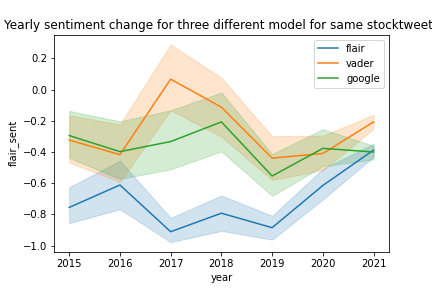


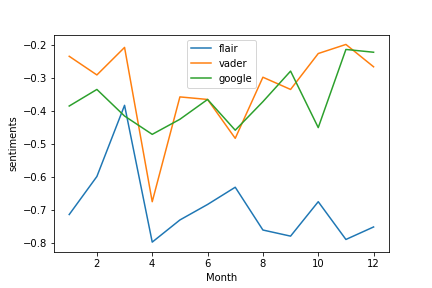
We can see a huge difference in the analysis of sentiment of Vader compared to flair and google. The user **Newsfilter** has average sentiment positive for Vader while it’s negative for google and flair.



**I.a Time series analysis of sentiment trends**

The plot below shows the average sentiment score for three different models for different years. The three models show completely different trends.





The monthly average also doesn’t follow a similar trend. All models seem to have the dip in the months of April but the flair shows completely different after April.

**Part II.Build the supervised Naive Bayes model**

We got the label from three different pretrained models. From these labelled tweets, I build the supervised Multinomial Naive Bayes(MNB) model.

I follow the steps, like text cleaning ,removing stop words, tokenizing and finally fitting with TfidfVectorizer. Finally I split the data into train, test and fit with the MNB model. Here is the summary:

**Google:**

The confusion\_matrix for MNB is

[[470 158]

[ 10 47]]

The precision score for MNB is 0.22926829268292684

The recall score for MNB is 0.8245614035087719

The f1 score for MNB is 0.35877862595419846

The accuracy score for MNB is 75.47445255474453

**Flair:**

The confusion\_matrix for MNB is

[[507 96]

[ 26 56]]

The precision score for MNB is 0.3684210526315789

The recall score for MNB is 0.6829268292682927

The f1 score for MNB is 0.47863247863247865

The accuracy score for MNB is 82.18978102189782

**Vader:**

The confusion\_matrix for MNB is [[402 86]

[ 27 170]]

The precision score for MNB is 0.6640625

The recall score for MNB is 0.8629441624365483

The f1 score for MNB is 0.7505518763796909

The accuracy score for MNB is 83.5036496350365

The accuracy score for the flair and vader is almost the same but greater than google. The precision score for google is highly smaller than the recall score. But we don’t see such a difference in vader. According to the fit summary, vader can give the best result to train the MNB model .