

Caleb Klenda

CSE 590

HW 4

Part 1 – Cross Fold Validation

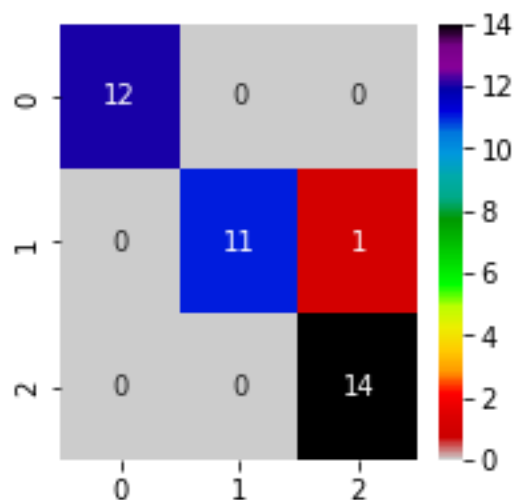
K-Closest Neighbors, Stochastic Gradient Descent and Gaussian Naive Bayes were used on the Iris dataset. Below are the results when run on the iris dataset.

```
With k = 2, KNeighborsClassifier: mean accuracy=95.33%; standard deviation=0.67%
With k = 2, SGD: mean accuracy=76.00%; standard deviation=12.00%
With k = 2, GaussianNB: mean accuracy=95.33%; standard deviation=3.33%
With k = 20, KNeighborsClassifier: mean accuracy=96.70%; standard deviation=5.74%
With k = 20, SGD: mean accuracy=81.79%; standard deviation=20.30%
With k = 20, GaussianNB: mean accuracy=94.91%; standard deviation=7.39%
```

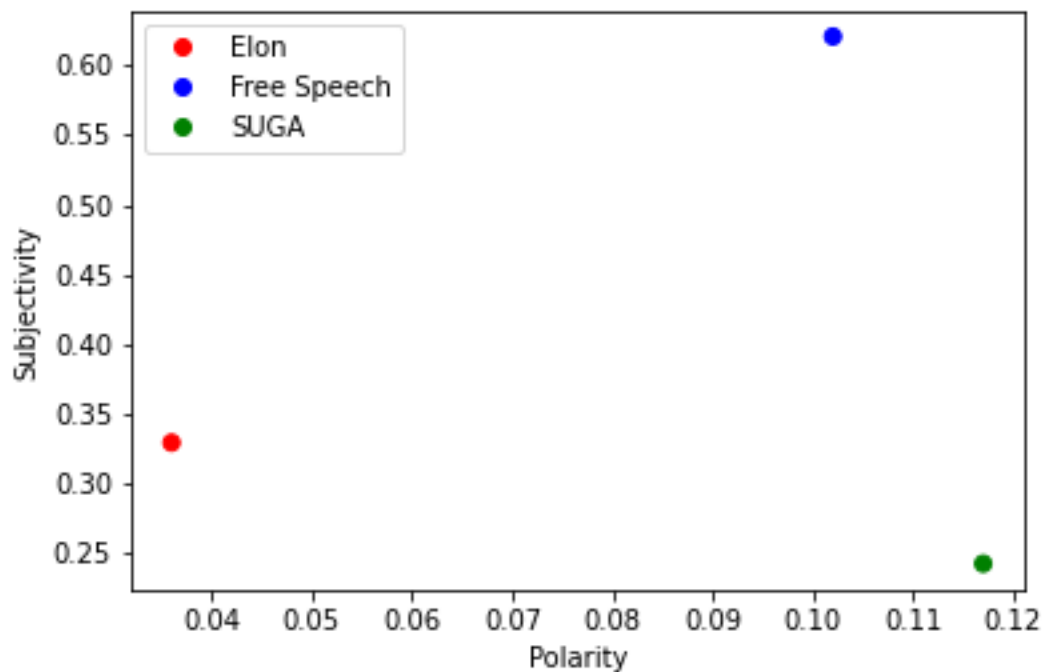
For K-Closest Neighbors and SGD, the 20 fold improved upon the 2 fold in terms of mean accuracy. However, in the GaussianNB case, there was a very slight decrease. Of all the methods, the standard deviation for SGD was the highest, and the standard deviation for all classifier was higher with k = 20

Part 1 – Confusion Matrix

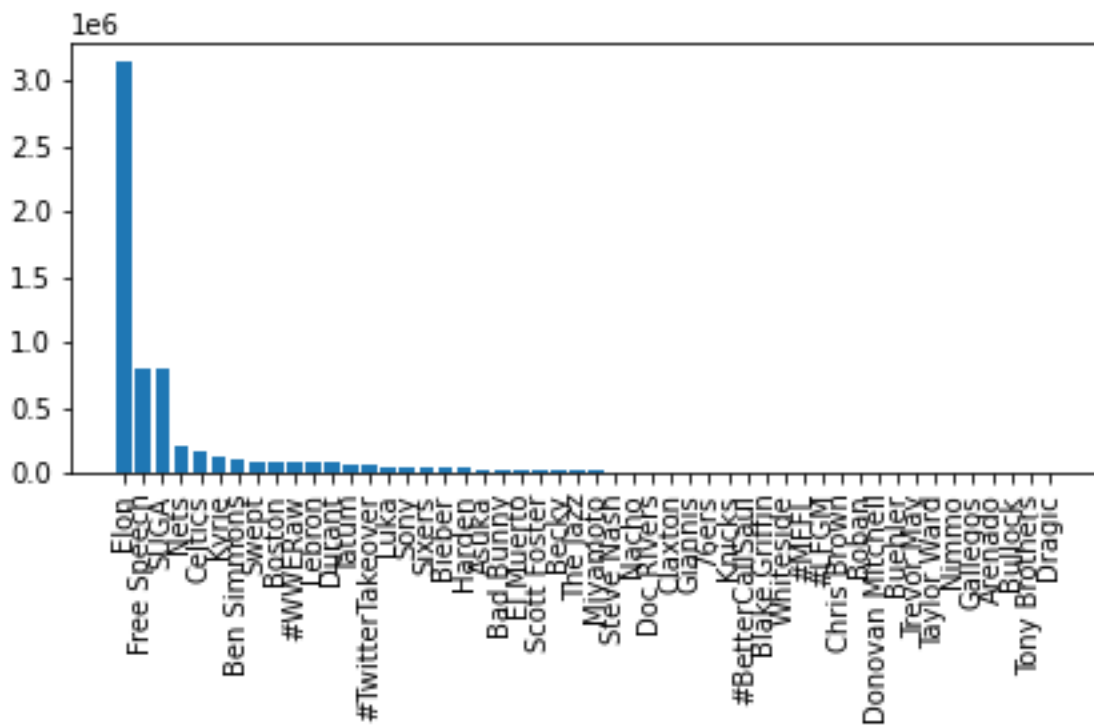
The best estimator for this dataset was the K-Closest Neighbors classifier with 20-fold cross validation. Below is the confusion matrix generated for that classifier. As seen, class 1 was the most likely to be confused (as class 2), although with the small dataset this only occurred once. The other two classes (1 & 3) were both accurately identified all times.



Part 2 – Sentiment Analysis



Sentiment Analysis of 20 Tweets from Top 3 Trends (Avg)



Number of tweets for trends in Cincinnati

- i.) In general, all tweets had a positive sentiment (as the value is above 1 for polarity in all cases). However, The tweets about Elon Musk were most divided as they are barely positive and hover around 0.0. Of note, when running this a few times, the results varied based on which 20 tweets were grabbed, and the polarity varied widely for the Elon trend. Overall, it appears to be a very divided trend.
- ii.) While not shown in the graph, the responses for individual tweets varied widely, indicating that there was little agreement on any three of the trends. This is typical of social media, which tends to push people towards extreme opinions. Also, since no sarcasm is being detected, it is hard to analysis the value of this conclusion from such a small sample of tweets.