**A LAB REPORT ON THE COMPARISON IN PERFORMANCE BETWEEN NAÏVE BAYES AND LOGISTIC REGRESSION**

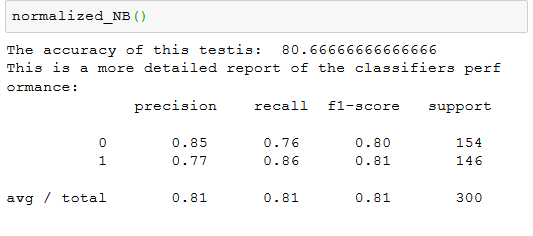
# LIBRARIES USED

In order to conduct this study I needed to use various libraries in python. The libraries I took advantage of when modeling the various classifiers were the scikit-learn library, Natural Language Toolkit(NLTK), and re. I used **re** which is a regex library to help preprocess my data and by that I mean clean all punctuation so I was only left with characters. **The NLTK library** was also used for text processing such as stemming and lemmatization in an attempt to improve the performance of my classifier. **The scikit-learn Library** was the main library that was used to model both my naïve\_bayes and logistic regression classifiers. It is very user friendly, including many preprogrammed algorithms that make machine learning easy especially for novices. In addition, there was a lot of well written documentation for this particular library which made it far easier to understand and therefore use.

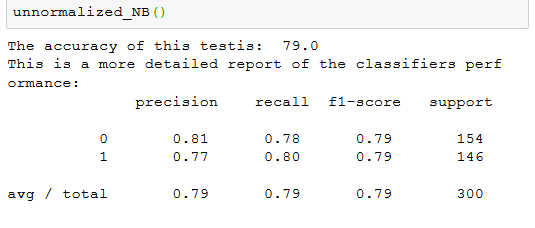
# *EVALUATION OF RESULTS*

In order to evaluate both the normalized and unnormalized classifiers. I decided to use accuracy, precision, recall, F-meausre. Although accuracy is a bit flawed, it in combination with the other factors would have helped me get a better picture of how well my classifier was doing. Thankfully employing the sci kit learn classification library made this process much easier as all this was calculated for me.

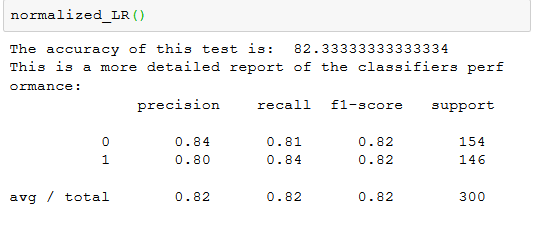
**Result for Normalized Naïve Bayes**



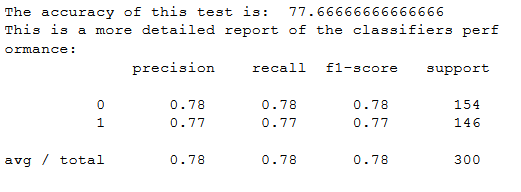
**Result for Unormalized Naïve Bayes**



**Result for Normalized Logistic Regression**



**Result for Unnormalized Logistic Regression**



# Discussion of Results.

# DISCUSSION OF RESULTS

The results from both classifiers showed that normalized classifiers perform better than unnormalized classifiers no matter the type of classifier. This performance was measured using recall, precision and the F1 score predominantly with accuracy supporting. This is the case because normalization

0 and 1 in the table below are the labels referring to positive and negative sentiment

**Comparison of Results**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Classifier | Accuracy(%) | Precision | Recall | F1 score |
| Normalized Naïve bayes | 80.6 | 0: 0.85  1: 0.77 | 0: 0.76  1: 0.86 | 0: 0.80  1: 0.81 |
| Unnormalized Naïve bayes | 79 | 0: 0.81  1: 0.77 | 0: 0.78  1: 0.80 | 0: 0.79  1: 0.79 |
| Normalized Logistic Regression | 82.3 | 0: 0.84  1: 0.80 | 0: 0.81  1: 0.84 | 0: 0.82  1: 0.82 |
| Unnormalized Logistic | 77.6 | 0: 0.78  1: 0.77 | 0: 0.78  1: 0.77 | 0: 0.78  1: 0.77 |