Claire Chemutai

12102019

NLP Lab report

**Introduction**

The aim of the project is to evaluate two different classifiers, which include the naïve bayes and logistic regression classifiers. The accuracy from testing the model will prove which classifier is better.

**Objectives and Results**

The two main libraries I used for classification include the sklearn.linear model for logistic regression and sklearn for naïve bayes.

from sklearn.linear\_model import LogisticRegression

from sklearn import naive\_bayes

The train data used for evaluation was from amazon. The test data was from given test sentences in class. The data was first unnormalized, trained and tested with both classifiers. The results were as below.

Logistic Regression classifier accuracy with unnormalized data is 0.68

Naive Bayes classifier accuracy with unnormalized data is 0.73

The data was then normalized, train and tested. The results were as below.

Logistic Regression classifier accuracy with normalized data is 0.71

Naive Bayes classifier accuracy with normalized data is 0.74

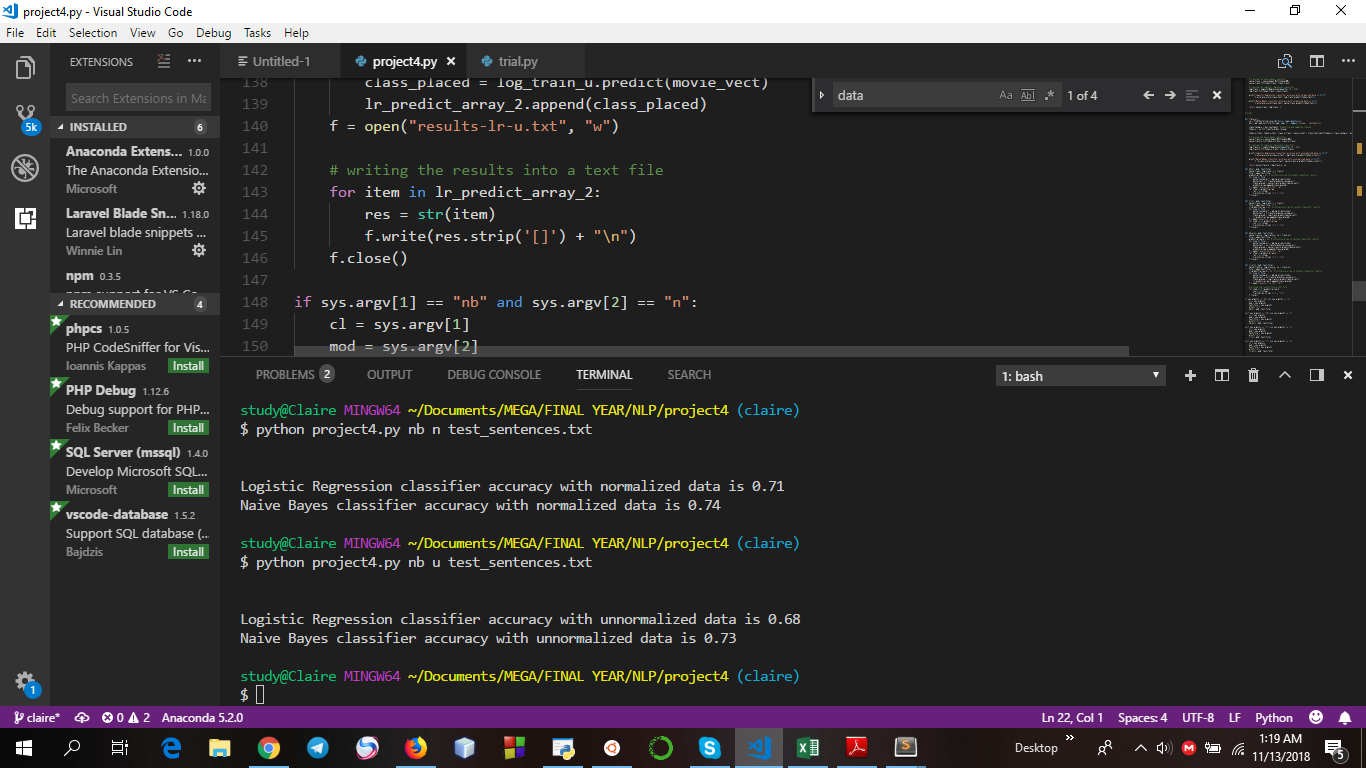


Figure 1: Accuracy of both normalized and unormalized naive bayes and logistic regression classifiers

|  |  |
| --- | --- |
| Figure 2: Result of naive bayes normalized classifier | Figure 3: Results of naïve-bayes unnormalized classifier |

The probability of the given test sentences were also printed out and the results were as above.

**Conclusion**

Normalized data give better results as compared to unnormalized data as it has removed stopwords and has most used data.

In both cases of normalized and unomarlized data, naïve bayes performs better than logistic regression.