COMP6721 (Winter 2020)

Project 2 Report

Python-based spam detector using the Naive Bayes approach

Submitted To: Dr. René Witte

By

Zankhanaben Patel

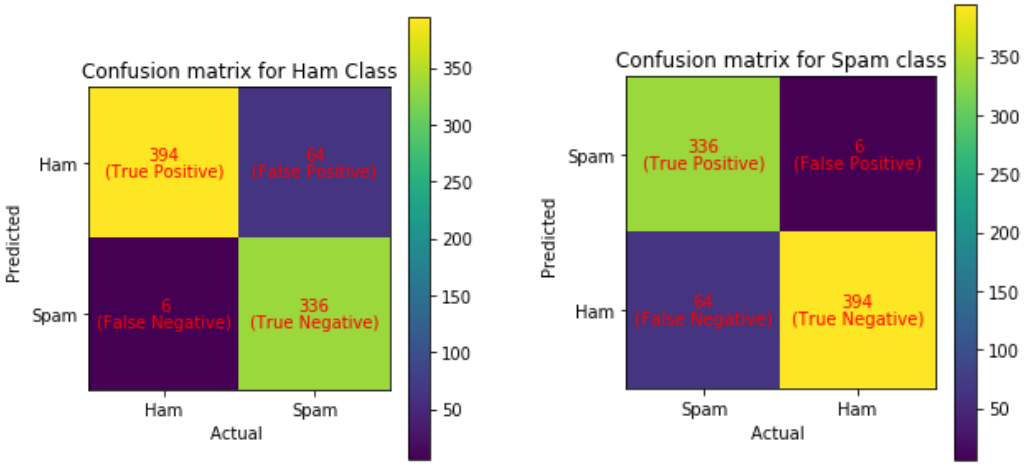
Arjuman Mansuri

Deep Patel

Analysis :

Explanation of how the results were generated.

- Confusion matrix was constructed from the True Positives, False Positives, False Negatives and True Negatives, obtained while generating 'result.txt'.[3]



- The following table was constructed, resulting in the 1st row generated by keeping Ham as the positives while 2nd by keeping Spam as positives. [4]
- **Accuracy of each class** = true_positive / (true_positive + false_negative) * 100
- **Precision of each class** = true_positive / (true_positive + false_positive) * 100
- **Recall of each class** = true_positive / (true_positive + false_negative) * 100
- **F1 Score of each class** = 2 * (Recall * Precision) / (Recall + Precision)
- **Accuracy of the model** = (true_positive + true_negative) / (true_positive + true_negative + false_negative + false_positive) * 100

Class	Accuracy	Precision	Recall	F1-measure
Ham	98.5	0.860	0.985	0.918
Spam	84.0	0.982	0.84	0.906
Model	91.25			

Discuss these results.

- Since the ham-spam classification problem at hand has balanced classes, accuracy would be enough to comment on the performance of the model.
- Accuracy of the model for identifying Ham files is better than for Spam files.
- Also, F1-measure states that the model is favorable for detecting Ham files.
- So, overall it is evident that the trained model is better at distinguishing Ham files as compared to Spam files.
- Nevertheless, it is visible that the model is not that bad of a fit for Spam classification.

References :

1. Lecture slides and worksheets of Dr. René Witte for this course (COMP6721 Winter-2020).
2. https://scikit-learn.org/stable/modules/generated/sklearn.metrics.confusion_matrix.html
3. https://medium.com/@shivangisareen/confusion-matrix-3ac02a1719ba
4. https://blog.exsilio.com/all/accuracy-precision-recall-f1-score-interpretation-of-performance-measures/
5. https://stackoverflow.com/questions/17232683/creating-tables-in-matplotlib/17237728