



COMP 6721
Applied Artificial Intelligence
Project Assignment 2 Report

Submitted to: Dr. Rene Witte

Team members:

Group: HA-G04	
Dhruv Goyani	40121588
Sai Krishna Metpalli	40124552

Analysis

Introduction:

After the generation of “result.txt”, we measure the performance of the classifier by calculating the accuracy (*instances of the test set the algorithm correctly classifies*) of the given responses of the testing set. We also measure the Precision (*What proportion of instances labelled with each class are correct?*), Recall (*What proportion of the instances in each class are labelled correctly?*) and the F1-measure (*used to seek a balance between Precision and Recall and check if there is an uneven class distribution*) of individual class.

$$\text{Accuracy} = \frac{\text{Number of correctly classified mail}}{\text{Total number of classified mail}}$$

$$\text{Precision} = \frac{tp}{tp+fp}$$

$$\text{Recall} = \frac{tp}{tp+fn}$$

$$\text{F1-Measure} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

Approach:

From the “result.txt”, the actual and expected classification labels are compared for each text file (*i.e., ham or spam file*). We maintained the count for each class (*i.e., ham or spam*) of actual and expected classification labels. These values are used for generation of confusion matrix as shown in the below Table 1.

Predicted	Actual	
	ham	spam
ham	394	64
spam	6	336

Table 1: Confusion Matrix

Using the confusion matrix, the accuracy, precision, recall and f1-measure is calculated using the above formulas. The results are shown in below Table 2.

Accuracy = 0.91			
	Precision	Recall	F1- Score
ham	0.86	0.98	0.92
spam	0.98	0.84	0.90

Table 2: Analysis Result

Discussion:

The accuracy on the given test result is identified as 91.25% which shows good model performance, but this cannot be the only measure for assessing the model. The precision for ham is 0.86 while the spam is 0.98 which indicates the proportion of instances labelled with class spam are higher than ham. When we consider recall, ham has 0.98 while the spam has 0.84 which indicates the percentage of class ham identified by model is higher than the spam. Following the f1-measure of ham i.e., 0.92 is slightly higher than spam i.e., 0.90 but both strikes good balance of precision and recall. If we want to create a balanced classification model with the optimal balance of precision and recall, then we try to maximize the f1-measure score.

References

1. Towards Data Science. (2019). Beyond Accuracy: Precision and Recall. [online] Available at: <https://towardsdatascience.com/beyond-accuracyprecision-and-recall-3da06bea9f6c>
2. <https://web.stanford.edu/class/cs124/lec/naivebayes.pdf>
3. Datascienceassn.org. (2019). Natural Language Processing With Python.[online] Available at: <http://www.datascienceassn.org/sites/default/files/Natural%20Language%20Processing%20with%20Python.pdf>
4. Towards Data Science. (2019). Applied Text-Classification on Email Spam Filtering. [online] Available at: <https://towardsdatascience.com/applied-textclassification-on-email-spam-filtering-part-1-1861e1a83246>
5. Kumar, V. (2019). Email Spam Filtering: A python implementation with scikit-learn. [online] Machine Learning in Action. Available at: <https://appliedmachinelearning.blog/2017/01/23/email-spam-filter-pythonscikit-learn/>