**Chapter 7**

**Experimental Analysis and Results**

In this chapter, the performance, experimental analysis of the algorithm implemented in the project and the results obtained are presented.

**7.1 Evaluation Metric**

The algorithm implemented is evaluated for the following three metrics

* Accuracy
* Precision
* Recall
* F measure

**Accuracy:** Accuracy or Accuracy rate (or percent correct), is deﬁned as the number of correct cases divided by the total number of cases.

**Precision:** Precision (also called positive predictive value) is the fraction of retrieved instances that are relevant or it is the percentage of selected items that are correct

**Recall:** Recall (also known as sensitivity) is the fraction of relevant instances that are retrieved or it is the percentage of correct items that are selected.

**F Measure:** A metric that combines precision and recall metrics, it is the weighted harmonic mean or can be considered as a combined measure that assesses the precision recall trade off.

Let us consider a 2-by2 contingency table 7.1 show the analysis.

**Table 7.1-Predicted class**

|  |  |  |
| --- | --- | --- |
|  | **actual class (observation)** | |
|  |  |  |
| **predicted class (expectation)** | **tp** (true positive) Correct result | **fp** (false positive) Unexpected result |
| **fn** (false negative) Missing result | **tn** (true negative) Correct absence of result |

Using the above table we can calculate the above mentioned measures by using the formulas discussed below.

Accuracy = (tp + tn) / (tp + fp + fn + tn)

Precision = tp / (tp + fp)

Recall = tp / (tp + fn)



In the above formula P is precision, R is recall and α is factor that controls the trade off between precision and recall (0 =< α <= 1).

If we substitute for α = 0.5 (β = 1) we get the F1 measure

F1 = 2 \* *P \* R* / (*P* + *R*)

**7.2 Performance Analysis**

To compute the performance of the Naïve Bayes algorithm implemented the test data is extracted from the training data.

30 tweets from each of the positive and negative training data and 10 tweets from neutral training data was taken as the test data, hence the test data consisted of 70 tweets.

The algorithm was implemented using test data as the input

Now precision, recall and f measure is calculated as follows

* Case1: Compute precision, recall and F1 measure for positive test data by comparing the program prediction and the actual true result.
* Case2: Compute precision, recall and F1 measure for negative test data by comparing the program prediction and the actual true result.
* Case3: Compute precision, recall and F1 measure for neutral test data by comparing the program prediction and the actual true result.

Table 7.2 and figure 7.1 explain the performance analysis of all the three cases.

**Table 7.2 Performance Analysis Results**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Class** | **Accuracy** | **Precision** | **Recall** | **F1-measure** |
| Positive | 51.43% | 30% | 8.33% | 13.0394% |
| Negative | 48.57% | 44.44% | 70.59% | 54.5426% |
| Neutral | 81.43% | 14.29% | 1.75% | 3.1181% |

**Figure 7.1 Performance Analysis**

The graph 7.1 shows the precision, recall and F1-measure for the test set comprising of positive, negative and neutral tweets taken from the training set. The project implemented uses Naïve Bayes classifier for analyzing sentiments in tweets, however sentiment analysis of tweets has been carried out by using different algorithms and different training data sets. [17] uses Support Vector Machines for classifying the twitter data, the training corpus used is obtained from Stanford twitter sentiment data. The accuracy claimed is 75.39% if the tweets are classified as to belonging to one of the 2 classes (positive or negative class) and 60.83% if the tweets are classified as to belonging to one of the 3 classes (positive or negative or neutral class). [20] uses Support Vector Machines but is focused on a specific target, the training corpus used contains tweets specific to a particular subject { Obama, Google, iPad, Lakers, Lady Gaga}. The accuracy claimed is 85.6% if the tweets are classified as to belonging to one of the 2 classes (positive or negative class) and 68.3% if the tweets are classified as to belonging to one of the 3 classes (positive or negative or neutral class).