**Chapter 5**

**Experimental Results**

We have tested our system with extensive experiment. In this section, we first introduce how data are collected for our Conspiracy Detection Model. Then we will present the performance of the system and compare it with existing systems.

**5.1 Data Collection**

For Conspiracy detector we collect the data from the real environment. First of all finding the email data is not so easy. There is only some of real email dataset available in this world that are free for general research. We uses Enron Dataset. So many researches has been done successfully with this set of data.

**5.1.1** **Green data collection**

As we are going to use data of two classes. So far our plan was to collect the official email like office affair, work related email, deal related, client related, gratitude related, personal email, internal component operation, legal advice, humor, friendship affection related, jokes, forwarding email, Logistic arrangement etc. We have uses the data with perfectly classified with these sort of classes. These sort of data or email are frequently exchanged between the employees of the company. So we have to classify these data as a green data.

We collect the email from that dataset and labeled it with class 0(zero). These are our Green dataset. We have stored these data in a csv file with two rows. Row one is for the body of the email. And the second row is for the sentiment or class. Class row holds the zero as the sentiment.

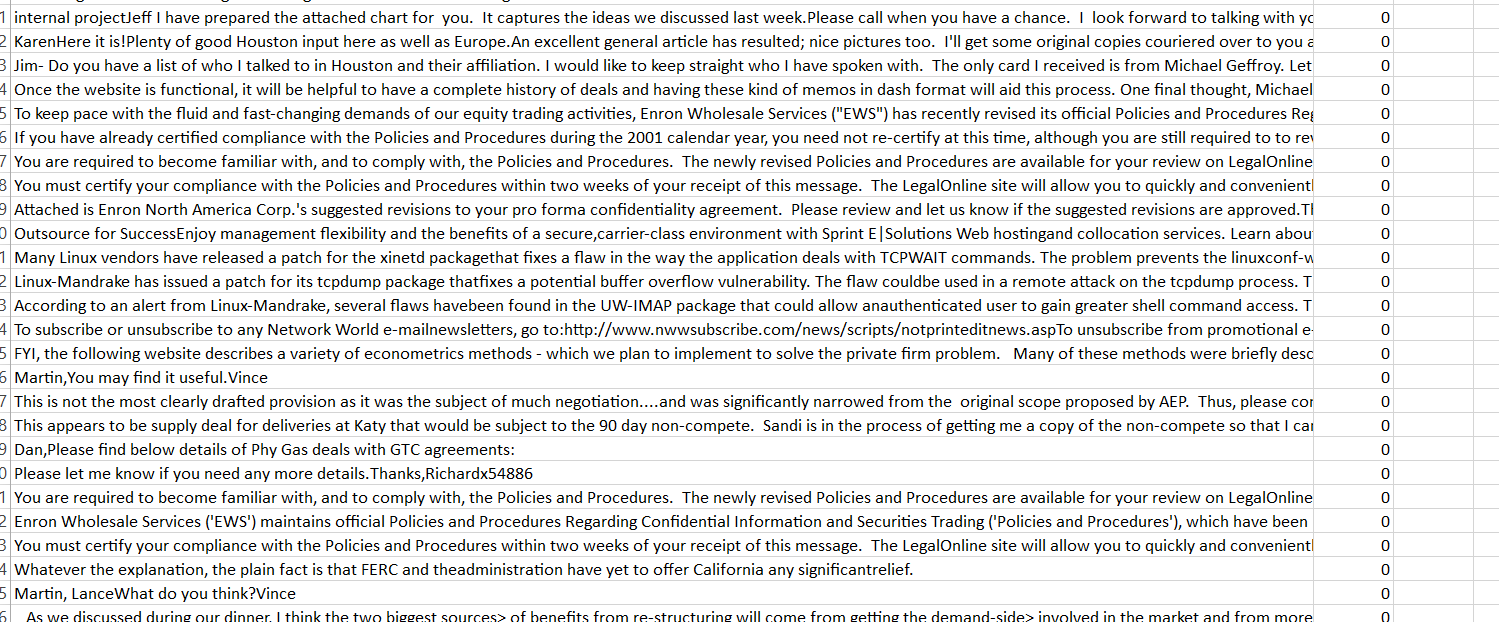


Fig 5.1: Green Data CSV file.

**5.1.2 Red data collection**

In this section we are going to explain the process of collecting the red data. It was not as easy as collecting the green data. As we are designing a model that can easily detect or predict the conspiracy in the email data. We have to learn the machine about the sentiment and psychology behind the concept of conspiracy. We have to frame the related word and concept clear about the theory.

As conspiracy is a psychological concept in human life. First we had to study through the concept of the conspiracy theory and about the working place conspiracy theory. There are too many conspiracy theory over the world. But we have just look through the working place conspiracy theory and the study over the theory. We have collected the consequences of conspiracy theory and the reason behind it. After all that study we made a list of situation based on the theory. We have come to the concept that there could be 3 possible angle of conspiracy in a working place that may causes the after effect that we have mentioned earlier. So we have sorted some point in which we will give our focus to create the real time environment and find the email with the concept of conspiracy.

Here we came out with the three angles of conspiracy, they are:

1. **Financial conspiracy**
2. **Organizational conspiracy**
3. **Reputational conspiracy.**

**5.1.2.1 Financial Conspiracy**

It reflects the concept of harming a company financially planning with the employees of that company in several way. It could be with direct fraud in financial account, could be investing in any dead project, missing the proper paper work in every financial transaction in the office place

**5.1.2.2 Organizational Conspiracy**

This concept reflects the view of overpower the company from the current management or owner, chairperson. This means the organizational change as well as the leadership changes in between the company. It could be in various scale of changes.

**5.1.2.3 Reputational Conspiracy**

It means the overall reputation of the company such as any rumor about the management, pricing, strategy, working condition, and financial statement so many fake news can harm the reputation of the company. This types of conspiracy can effect a company over night and destroy the socio-economic value of that company

After reviewing these concept more and more we select some people who can visualize things in real life. We have given them the knowledge about the theory in every angle with some example of conspiracy related conversation. We have cleared the concept of that individual with every way they asked. After some grooming, these person was monitored by us during making the data as form of email. We monitor the data about the reflection of the concept through the email properly. And after a certain period of time we became successful to make some conspiracy reflecting emails.

That was a great success for us as we at last have something to teach the machine. Then we collectively do this process in various environment. And collect these data. We have used almost 30 different persons to collect these data. So that was the most challenging part in collecting data set.

As we are trying to teach the machine a purely psychological concept of human nature, it was a tough job for sorting out the reflection of conspiracy throughout the email. After collecting these data we have made a csv file leveled with one in the sentiment column, the file looks like the figure 5.2

Fig 5.2 will reflect the Red email in a csv file:

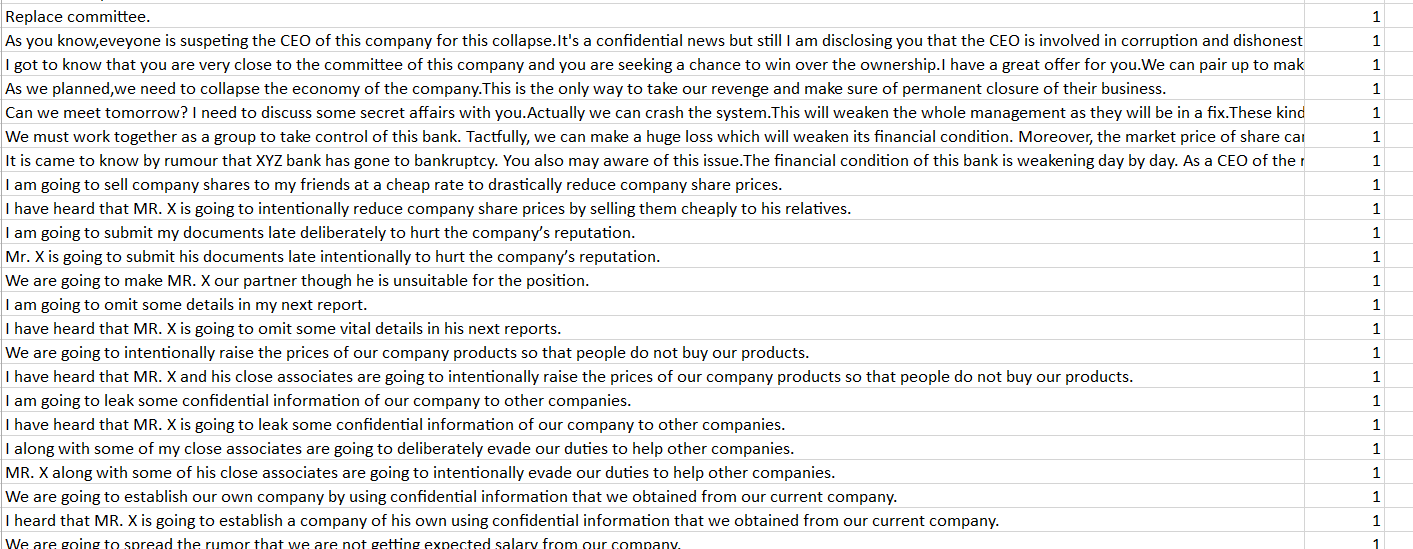


Fig 5.2: Red Data set

**5.2 Evolution of the System**

The collected data are used to evaluate the system. The system is evaluated on the basis of two ways:

1. Evaluates from the mail dataset by splitting them into training and testing
2. Evaluate the system in real time.

**5.2.1 Evaluates from the mail dataset**

We have done this work in out algorithm during training the model. In that period we have separated the dataset with training set 80% and testing 20%. And after that we have a certain accuracy of ~68%

We have used 450 class 0 tagged data to evaluate the detecting percentage of the model. It gives us that 324 email with green detection and 126 email with false detection.

Thus the accuracy over detecting the green data is =72

So the accuracy in the Green data detection is 72%

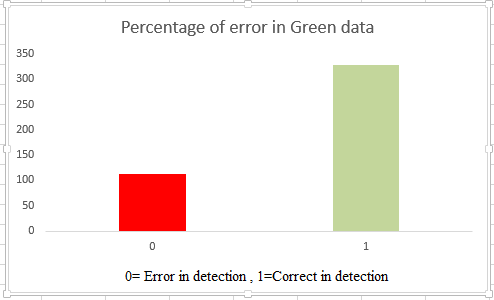


Fig 5.3: Percentage of error in Green Data

Again in the Red data we continued the process of detection. Here we also uses 450 email data from the dataset that is leveled with one. And after processing these data out model detect conspiracy successfully from 286 number of mail. And it predict 164 number of data as wrong detection.

Thus the accuracy of the model in the Red data set is = =65

So the accuracy for the Red data is 65%.

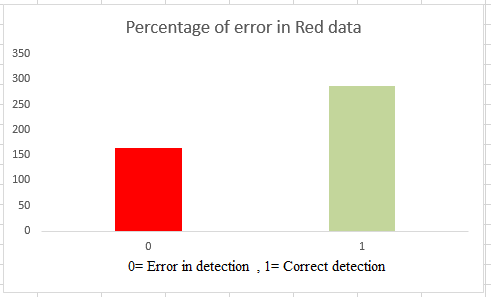


Fig 5.4: Percentage of error in Red Data

**5.2.2 Evaluates from the real time mail**

In the other way we can check the accuracy of our system in real time. For this way we can predict the accuracy of the system for real time email. In here we tested 100 mixed data send from one account to another account and count the number in four different ways. They are

1. True positive
2. True negative
3. False positive
4. False negative

True positive means the right detection of a Green email, true negative means predict a green mail as a red email, false positive means detect the Red data accurately, and false negative means incorrect detection of the Red data.

Here we are giving the table of this ratio for detection by the module in real time. Table 5.1 gives us the proper understanding of that concept

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Total Email | True Positive | True negative | False positive | False negative |
| 100 | 42 | 18 | 26 | 14 |

Table 5.1: Real time accuracy of mail data

So the overall real time accuracy of my system is = =68

Now we can say the real time accuracy of this system is 68%

So the overall accuracy of this system can be found by merging both the real time and the train set data is = = 67.8

Thus we can say the accuracy is 67.8% overall.