

### **Result-**

```
('Confusion matrix:\n', array([[919, 463],  
      [ 29, 890]]), dtype=int64))  
( 'Accuracy Value: \t', 0.7861799217731421)  
( 'Precision Value: \t', 0.6577974870657798)  
( 'Recall Value: \t', 0.9684439608269858)
```

The results show us that the accuracy of Naïve Bayes is not good. By looking at confusion matrix we can say which emails were classified correctly and incorrectly. We can see that the recall value is higher than the precision value.

### **Do you think the attributes here are independent, as assumed by Naïve Bayes?**

Yes the attributes are independent here but not 100%. Here we can see dependence because one word can be dependent with other word nearly 90% of time. Like “Jackpot” and “Win” and “prizes” can be in the same email.

### **Does Naïve Bayes do well on this problem despite the independence assumption?**

### **Speculate on other reasons Naïve Bayes might do well or poorly on this problem.**

Here the result shows that the Naïve Bayes does not perform well in spite of independence assumption. If we had achieved higher accuracy we could have said so.

The reason it didn't do well is because the attributes which are statistically meaningful and more important are used only. If we use only important feature the performance would drastically improve.