

| Name | SBU ID |
|--------------|-----------|
| Apurva Kumar | 109751733 |
| Gaurav Gawde | 110008637 |

AI Assignment 05 Spam Project

Aim:

Implementing Naive Bayes algorithm to classify emails as spam or ham.

Classes:

| Class Name | Description |
|------------------|--|
| Classifier Class | This is the main class which implements various methods to achieve functionality of the project. |
| Email Class | This class creates email object. |
| NaiveBayes Class | This class creates NaiveBayes object to classify emails as spam or ham. |

Class and Function Details:

1.Classifier Class Functions

| | |
|-----------|---|
| Functions | <ul style="list-style-type: none"> • parseEmail() • computeAccuracy() • main() |
|-----------|---|

2.Email Class Funtions

| | |
|-----------|--|
| Functions | <ul style="list-style-type: none"> • Email() • getEmail() • getLabel() • addWord() • getWords() |
|-----------|--|

3.NaiveBayes Class Functions

| | |
|-----------|---|
| Functions | <ul style="list-style-type: none"> • NaiveBayes() • Train() • Predict() • computeLabelCounts() • computePriors() • computeWordProbs() |
|-----------|---|

Statistics:

[A] Smoothing Parameter $k=1$

Test-Data Prediction Statistics

Matches: 902

Accuracy: 90.20%

[B] Smoothing Parameter $k=2$

Test-Data Prediction Statistics

Matches: 903

Accuracy: 90.30%

[C] Smoothing Parameter $k=3$

Test-Data Prediction Statistics

Matches: 903

Accuracy: 90.30%

[D] Smoothing Parameter $k=4$

Test-Data Prediction Statistics

Matches: 903

Accuracy: 90.30%

[E] Smoothing Parameter $k=5$

Test-Data Prediction Statistics

Matches: 904

Accuracy: 90.40%

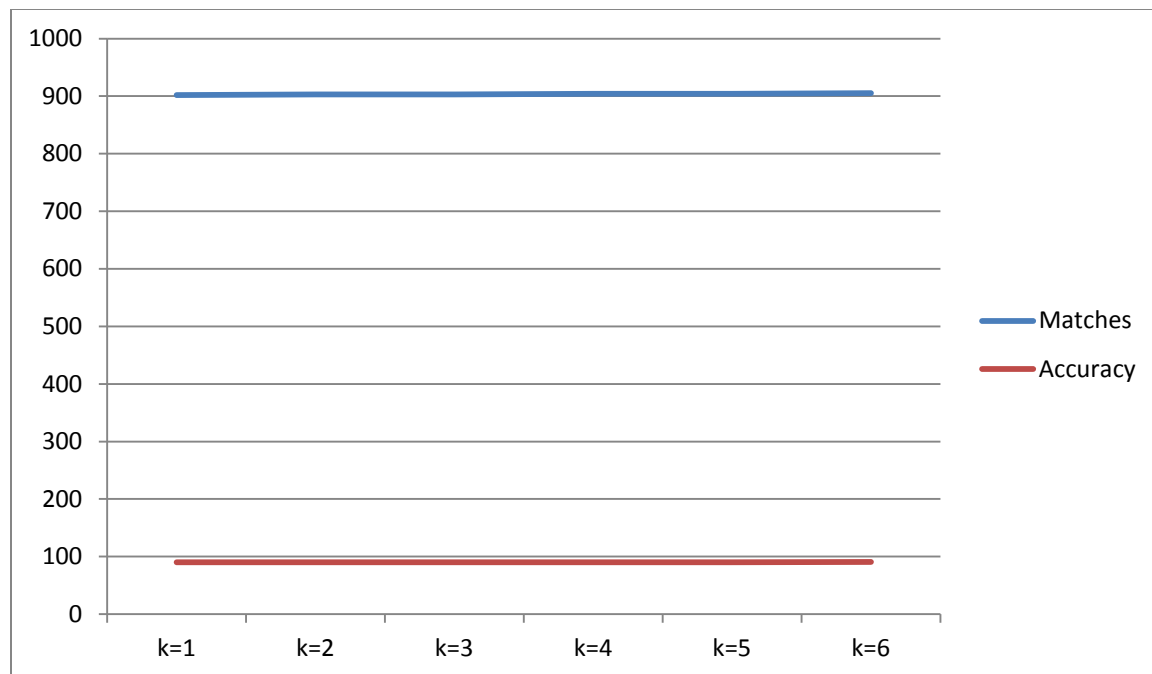
[F] Smoothing Parameter $k=10$

Test-Data Prediction Statistics

Matches: 905

Accuracy: 90.50%

Diagrams:

**Conclusion:**

Thus we can conclude that as we go on increasing the smoothing parameter value, number of matches increases and hence the accuracy of the data also increases when we use Naïve Bayes algorithm with smoothing parameter.