**Report**

I have implemented 3 kinds of naïve bayes classifier. Below are the details regarding the same.

**Naïve Bayes classifier:**

This works on the principle as mentioned below by the equation.



But I have chosen the alpha value to be 0.2 instead of 1

**Binarized Naïve Bayes classifier:**

This model makes sure that there are no repetitions of words in each document. As dictated by the below equation. Alpha is set to 0.2



**Naïve Bayes using bigrams as features:**

This model takes bigrams of sentences of each speaker as features and calculates the accuracy. And alpha, is set to 0.2

**Performance:**

Accuracy of Naive Bayes is 59.0 %

Accuracy of Binarized Naive Bayes is 62.0 %

Accuracy of Naive Bayes using Bigrams as features 61.0 %

**Challenges faced:**

* Handling numerical underflows.
* Preventing math range error, caused when values were close to 0.
* Optimizing the code to reduce the execution time.
* Apart from these, trivial coding challenges were faced, which were appropriately handled.

**Graph plots:**

The attached R file generates the graph for top 20 words of each speaker