

Group id: 1801cs08_1801cs11_1801cs53

Group Members:

1. Ammaar Ahmad - 1801cs08
2. Ayush Pandey- 1801cs11
3. Sumit Raj- 1801cs53

Subject: **Artificial Intelligence Lab 3**

Date: 14/09/2021

The naive Bayes classifier has been implemented as a MULTINOMIAL model and a MULTIVARIATE model both with add-one smoothing.

The multinomial model is better. It has higher accuracy(86.48%) than the multivariate model (40.467%).

reason: The multinomial model takes into account the frequency of each word while multivariate model takes only the occurrence of a word into account.

The models are compared on the basis of 5-fold cross-validation.

The results are as follows –

Iteration-1:

Size of training set = 4458
Size of testing set = 1114
Length of vocabulary = 7806
Accuracy of the multinomial model = 0.8491921005385996

Length of vocabulary = 7806
Accuracy of the multivariate model = 0.32046678635547576

Iteration-2:

Size of training set = 4458
Size of testing set = 1114
Length of vocabulary = 7762
Accuracy of the multinomial model = 0.8734290843806104

Length of vocabulary = 7762
Accuracy of the multivariate model = 0.23159784560143626

Iteration-3:

Size of training set = 4458
Size of testing set = 1114
Length of vocabulary = 7775
Accuracy of the multinomial model = 0.8752244165170556

Length of vocabulary = 7775
Accuracy of the multivariate model = 0.414721723518851

Iteration-4:

Size of training set = 4458
Size of testing set = 1114
Length of vocabulary = 7789
Accuracy of the multinomial model = 0.8572710951526032

Length of vocabulary = 7789
Accuracy of the multivariate model = 0.5385996409335727

Iteration-5:

Size of training set = 4456
Size of testing set = 1116
Length of vocabulary = 7809
Accuracy of the multinomial model = 0.8691756272401434

Length of vocabulary = 7809
Accuracy of the multivariate model = 0.517921146953405

Mean accuracy of multinomial model = 0.8648584647658024

Mean accuracy of multivariate model = 0.4046614286725482