2020146 Hamna Iftikhar

Report : Models

After implementation of different models including Naive Bayes, logistic Regression & Random forest etc on both the datasets we compared the performance of the models based on evaluation metrics such as accuracy, precision, recall and F1-score.

In the case of 20 Newsgroup dataset we found that the best performance model was logistic regression with accuracy 0.82 %, Naive Bayes with 0.78 % accuracy, and Random forest with the least of all i.e 0.70% accuracy.

In the case of Reuters dataset, the best performing model was SVM with accuracy 0.90 %, then Naive Bayes with 0.88 % accuracy, then Logistic regression and random forest with 0.85 and 0.84 % accuracy respectively.

The reason of difference in performance of all these models lie in the handling of type of the data. So naive Bayes and SVM performed well in Reuters dataset well because of high dimensionality of dataset & independence b/w features. so, the choice of model with best performance depends on certain characteristics of data & the goals of classification tasks.