

The top-2 scores obtained are 72.274 (For Multinomial Naive Bayes) and 67.015(In KNN for k=10 and set weights parameter to "distance")

Reasons for Multinomial Naive Bayes giving me the best result as it works best when features are discrete like the output of the CountVectorizer. Whereas Gaussian Naive Bayes assume a normal distribution with decimal value. This data might not fit into that distribution as a result of which I got an accuracy of 23 percent.

In case of KNN, as the number of dimensions are high, it couldn't perform as well as the Naive Bayes but when trying the various values of 'k' , found k=10 to work better than others.

In case of Decision Tree, we could have got a better answer than the one obtained but it needs fine tuning of the parameters which are large in number. I tried out a few parameters that I understood and got these results.

Some of the other params I tried are list below with their accuracy along with the parameters tuned for it)

Decision trees:

Basic: 64.26

max_depth=80 and max_features=0.99 is : 64.48

Max_depth

10: 48.8

20: 56.7

30: 61.1

40: 62.2

50: 63.4

60: 63.7

70: 63.9

75: 63.8

80: 63.7

Max_features (max_depth as 70):

0.9 : 63.6

0.95: 63.9

0.98: 64.0

0.99: 64.2

KNN:

for k and weights="distance"

3: 62.4---weights:64.5

5: 64.8---weights:66.5

9: 65.3---weights:66.8

10: 65.7---weights:67.015

11: 65.6---weights: 66.8

12: 65.6---weights:66.8

13: 65.3---weights: 66.8

15: 65.07---weights: 66.4

Gaussian Naive Bayes(default): 23.04

Multinomial Naive Bayes: alpha=1(default):72.274

with alpha=0 : 70.223

alpha=2: 71.6

alpha=1.5: 71.983

Bernoulli Naive Bayes(default): 71.44