

Thoughts

On looking at the problem, I assumed it to be a simple nlp problem which i have already practiced. I have tried tougher problems than this. So i was confident of doing it. First of all i have encoded all the Target column to its respective number(category). Then with help of 'stopwords' i have tried to clean the text and stored in a list. Then with the help of Count Vectorizer text's words have been converted to its respective value. Later i have tried different classification models. Random forest classifier gave me the best result among them. Since the accuracy score was less i thought of trying ANN but then the accuracy score was too bad .

Accuracy Score

Train Accuracy Score is 99%

Test Accuracy Score is 61.7%

Limitations

The accuracy score is less.

Model is slow.

Large number of trees.

