Naive Bayes Algorithm

4138 sentences with 3 different classes

-->Max features 1000 (Dictionary size 1000)

Training set=%80 Test set=%20

Accuracy = 0.18478

Class Statistics:

	Menfi	Müspet	Uyarı
Miss Rate:	0.845588	0.965517	0.0267857
F1 Score:	0.185022	0.066335	0.263923

-->Max features 2000 (Dictionary size 2000)

Training set=%80 Test set=%20

Accuracy = 0.18115

Class Statistics:

	Menfi	Müspet	Uyarı
Miss Rate:	0.845588	0.965517	0.0267857
F1 Score:	0.185022	0.066335	0.263923

-->Max features 4000 (Dictionary size 4000)

Training set=%80 Test set=%20

Accuracy = 0.18719

Class Statistics:

	Menfi	Müspet	Uyarı
Miss Rate:	0.0147059	0.965517	0.991071
F1 Score:	0.285106	0.066335	0.0176991

K Nearest Neighbor Algorithm

-->Max features 1000 (Dictionary size 1000)

Training set=%80 Test set=%20

Accuracy = 0.702898

Class Statistics:

	Menfi	Müspet	Uyarı
Miss Rate:	0.926471	0.0137931	1
F1 Score:	0.12987	0.823022	0

-->Max features 2000 (Dictionary size 2000)

Training set=%80 Test set=%20

Accuracy = 0.700483

Class Statistics:

	Menfi	Müspet	Uyarı
Miss Rate:	0.926471	0.0172414	1
F1 Score:	0.128205	0.821326	0

-->Max features 4000 (Dictionary size 4000)

Training set=%80 Test set=%20

Accuracy = 0.704106

Class Statistics:

	Menfi	Müspet	Uyarı
Miss Rate:	0.926471	0.012069	1
F1 Score:	0.130719	0.823868	0

Decision Tree Algorithm

-->Max features 1000 (Dictionary size 1000)

Training set=%80 Test set=%20

Accuracy = 0.69927536

Class Statistics:

	Menfi	Müspet	Uyarı
Miss Rate:	0.985294	0.005172	1
F1 Score:	0.0283688	.0.822523	0

-->Max features 2000 (Dictionary size 2000)

Training set=%80 Test set=%20

Accuracy = 0.69927536

Class Statistics:

	Menfi	Müspet	Uyarı
Miss Rate:	0.985294	0.005172	1
F1 Score:	0.0283688	. 0.822523	0

-->Max features 4000 (Dictionary size 4000)

Training set=%80 Test set=%20

Accuracy = 0.69927536

Class Statistics:

	Menfi	Müspet	Uyarı
Miss Rate:	0.985294	0.005172	1
F1 Score:	0.028368	0.822523	0

Neural Network Algorithm

Data Length: 2624 (complaints)

Train Part: %80 Test Part: %20

Accuracy: %79.8

→ We can't calculate f1 score

Results

We try four different machine learning algorithm with three different dictionary sizes to train on data part that we separate from our dataset as %80 of it. And then we calculate the accuracy of these algorithms by predict the classes of data part that we separate from dataset as %20 of it, by comparing them with real classes.

Our first algorithm was naive bayes algorithm and this algorithm gives the worst accuracy like 0.18, then as we increase the dictionary size the accuracy is increase very little but it was still the worst accuracy. And most of its accuracy came from the 'Uyarı' class which is not least but very few in dataset.

Second algorithm was K-nearest-neighbor algorithm and its accuracy was over %70 which is far better that naive bayes algorithm. The increase in dictionary size also increase the accuracy very little again. And most of its accuracy came from the 'Müspet' class which is very much in dataset.

Third algorithm was Decision tree algorithm. Its accuracy is also better than naive bayes algorithm but very little worse than k-nearest neighbor algorithm and it was under %70. As we increase the dictionary size, the accuracy is doesn't change at all. Most of its accuracy came from 'Müspet' class and this algorithm may be the most overfitted algorithm because its miss rate in Müspet has the smallest value among all algorithms.

Our last algorithm was a neural network algorithm and it has the best accuracy which was 79.8. But we can't calculate the f1 score because the code was a little homemade code which didn't use a machine learning algorithm from library so functions that calculates f1 score are didn't work properly.