Shahjalal University of science &
technology,sylhet

Assignment on Text Classification

Submitted To :

Engineering(IICT),SUST
Ms Sayma
Sultana Chowdhury
Assistant Professor
Dept. of
Software Engineering
(IICT),SUST

Submitted By :

Maraz Mia
Reg.No. : 2017831036

12 February, 2021

1.
Classification Report

1 Naive Bayes Classification:

The cross accuracy of Multinomial Native Bayes Model is shown in the below picture. Here is the accuracy, macro average & weighted average are given.

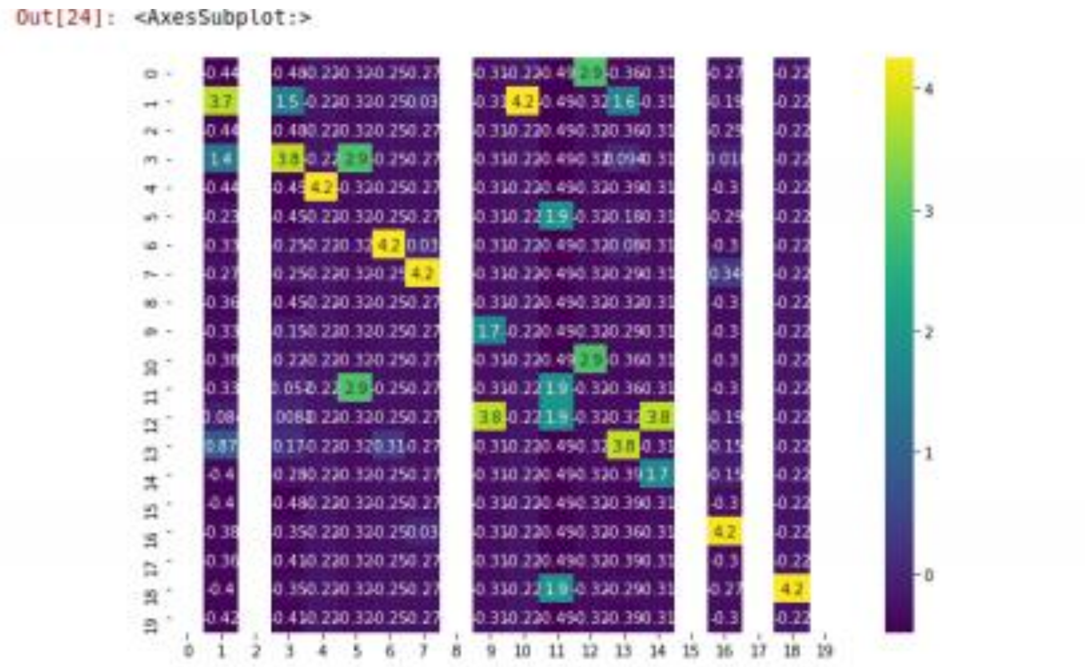
```
Cross Accuracy: 0.31 (+/- 0.05)
              precision    recall  f1-score   support

           0       0.00      0.00      0.00         2
           1       0.29      1.00      0.45       336
           2       0.00      0.00      0.00         3
           3       0.55      0.02      0.05       248
           4       0.00      0.00      0.00         2
           5       0.00      0.00      0.00        27
           6       0.00      0.00      0.00        31
           7       0.00      0.00      0.00        61
           8       0.00      0.00      0.00         5
           9       0.00      0.00      0.00        16
          10       0.00      0.00      0.00        15
          11       0.00      0.00      0.00        24
          12       0.00      0.00      0.00        33
          13       0.67      0.01      0.02       226
          14       0.00      0.00      0.00        26
          16       0.00      0.00      0.00         2
          17       0.90      0.57      0.70       269
          18       0.00      0.00      0.00         4
          19       0.00      0.00      0.00        22
          20       0.00      0.00      0.00         3

    accuracy                           0.37      1355
   macro avg       0.12      0.08      0.06      1355
weighted avg       0.46      0.37      0.26      1355
```

The confusion matrix of Multinomial Native

Bayes Model.



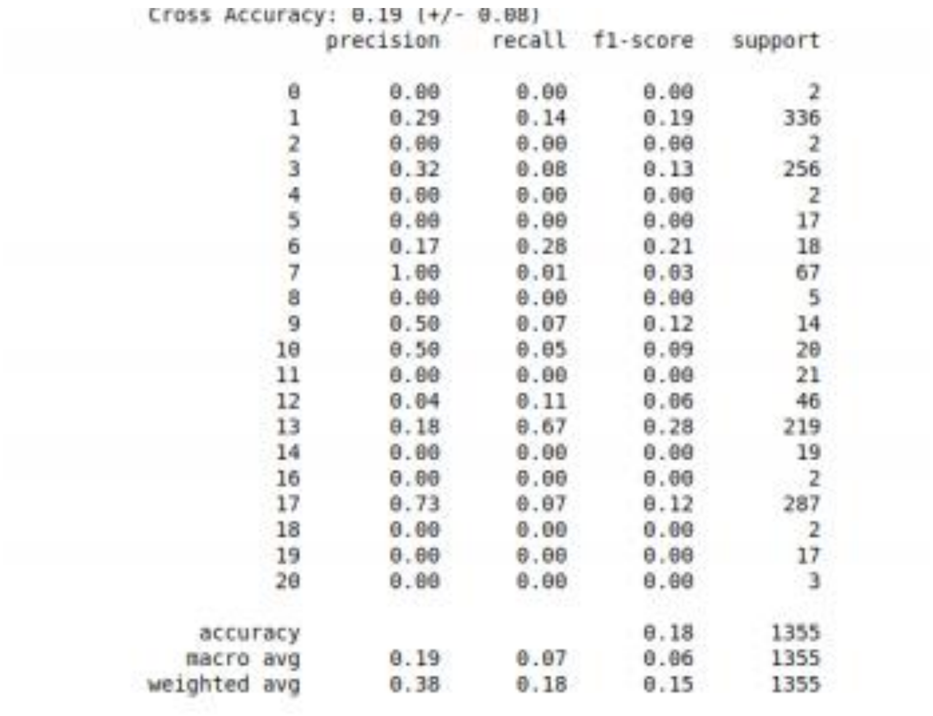
```
Out[24]: <AxesSubplot:>
```

2 k-Nearest Neighbors :

The cross accuracy of KNeighborsClassifier Model is shown in the be low picture.

Here is the accuracy,macro average & weighted average are

given.

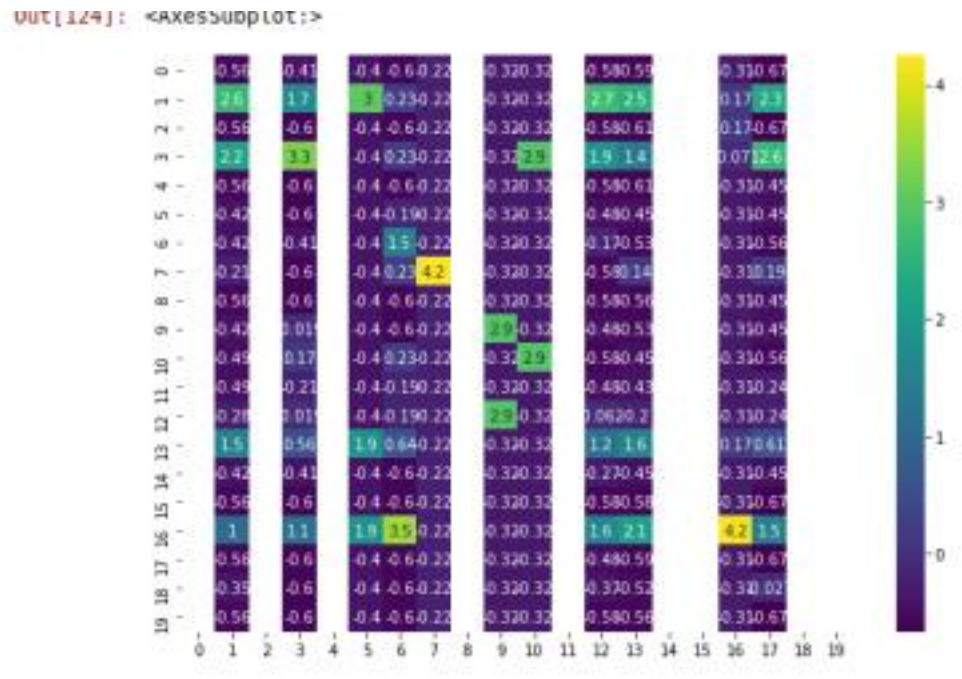


```
Cross Accuracy: 0.19 (+/- 0.08)
              precision    recall  f1-score   support

           0       0.00      0.00      0.00         2
           1       0.29      0.14      0.19       336
           2       0.00      0.00      0.00         2
           3       0.32      0.08      0.13       256
           4       0.00      0.00      0.00         2
           5       0.00      0.00      0.00        17
           6       0.17      0.28      0.21        18
           7       1.00      0.01      0.03        67
           8       0.00      0.00      0.00         5
           9       0.50      0.07      0.12        14
          10       0.50      0.05      0.09        20
          11       0.00      0.00      0.00        21
          12       0.04      0.11      0.06        46
          13       0.18      0.67      0.28       219
          14       0.00      0.00      0.00        19
          16       0.00      0.00      0.00         2
          17       0.73      0.07      0.12       287
          18       0.00      0.00      0.00         2
          19       0.00      0.00      0.00        17
          20       0.00      0.00      0.00         3

    accuracy                           0.18      1355
   macro avg       0.19      0.07      0.06      1355
weighted avg       0.38      0.18      0.15      1355
```

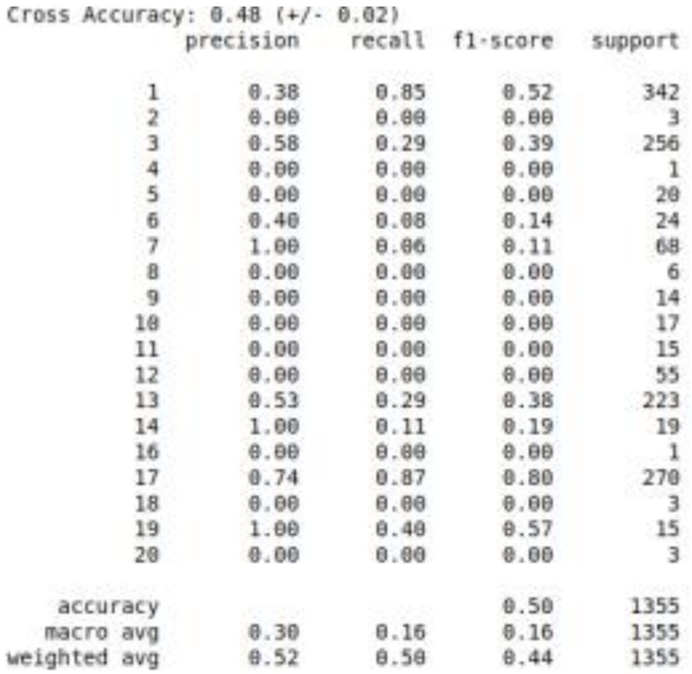
The confusion matrix of KNeighborsClassifier Model.



```
Out[124]: <AxesSubplot:>
```

3 Random Forest Classifier :

The cross accuracy of RandomForestClassifier Model is shown in the below picture. Here is the accuracy, macro average & weighted average are given

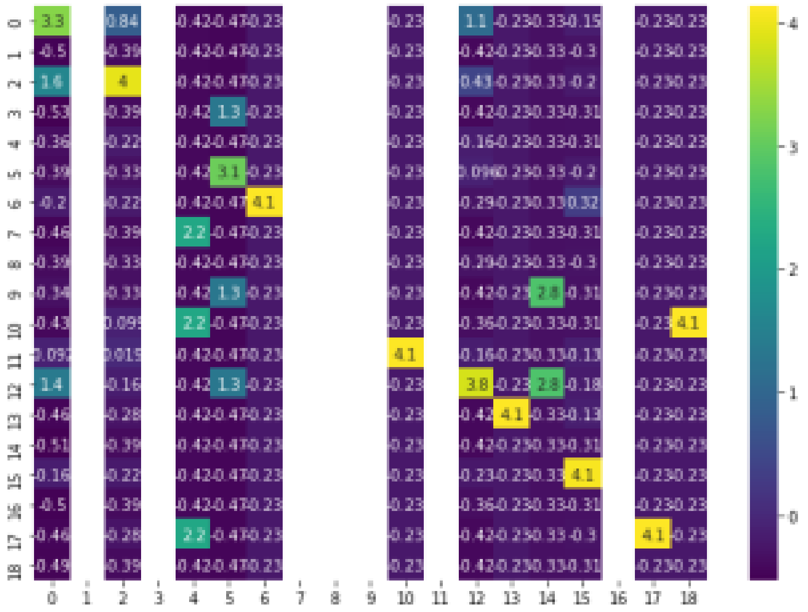


```
Cross Accuracy: 0.48 (+/- 0.02)
              precision    recall  f1-score   support

           1       0.38      0.85      0.52       342
           2       0.00      0.00      0.00         3
           3       0.58      0.29      0.39       256
           4       0.00      0.00      0.00         1
           5       0.00      0.00      0.00        20
           6       0.40      0.08      0.14        24
           7       1.00      0.06      0.11        68
           8       0.00      0.00      0.00         6
           9       0.00      0.00      0.00        14
          10       0.00      0.00      0.00        17
          11       0.00      0.00      0.00        15
          12       0.00      0.00      0.00        55
          13       0.53      0.29      0.38       223
          14       1.00      0.11      0.19        19
          16       0.00      0.00      0.00         1
          17       0.74      0.87      0.80       270
          18       0.00      0.00      0.00         3
          19       1.00      0.40      0.57        15
          20       0.00      0.00      0.00         3

    accuracy                           0.50      1355
   macro avg       0.30      0.16      0.16      1355
weighted avg       0.52      0.50      0.44      1355
```

The confusion matrix of RandomForestClassifier Model



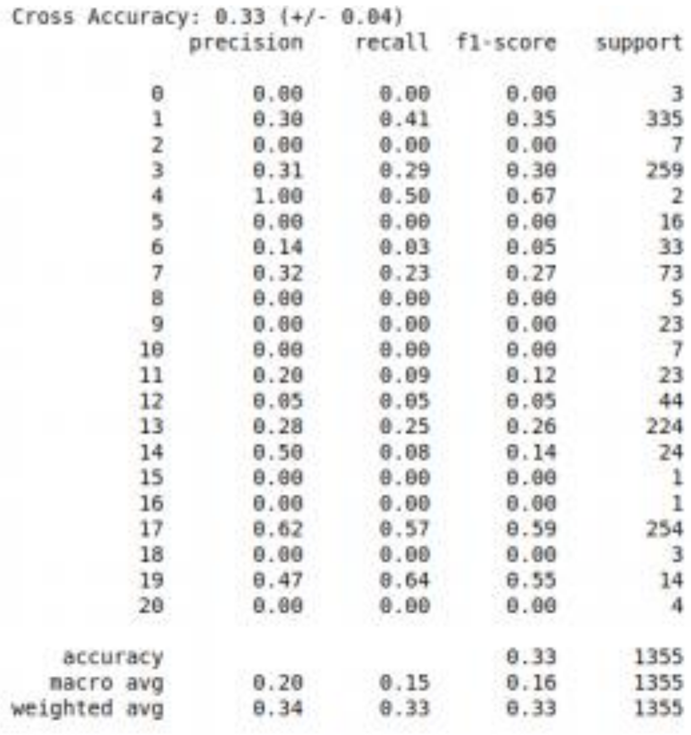
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
0	3.3		0.84		-0.42	-0.47	-0.23				-0.23		1.1	-0.23	-0.33	-0.15		-0.23	-0.23
1	-0.5		-0.39		-0.42	-0.47	-0.23				-0.23		-0.42	-0.23	-0.33	-0.3		-0.23	-0.23
2	1.6		4		-0.42	-0.47	-0.23				-0.23		0.43	-0.23	-0.33	-0.2		-0.23	-0.23
3	-0.53		-0.39		-0.42	1.3	-0.23				-0.23		-0.42	-0.23	-0.33	-0.31		-0.23	-0.23
4	-0.36		-0.22		-0.42	-0.47	-0.23				-0.23		-0.16	-0.23	-0.33	-0.31		-0.23	-0.23
5	-0.39		-0.33		-0.42	3.1	-0.23				-0.23		0.096	-0.23	-0.33	-0.2		-0.23	-0.23
6	-0.2		-0.22		-0.42	-0.47	4.1				-0.23		-0.29	-0.23	-0.33	0.32		-0.23	-0.23
7	-0.46		-0.39		2.2	-0.47	-0.23				-0.23		-0.42	-0.23	-0.33	-0.31		-0.23	-0.23
8	-0.39		-0.33		-0.42	-0.47	-0.23				-0.23		-0.29	-0.23	-0.33	-0.3		-0.23	-0.23
9	-0.34		-0.33		-0.42	1.3	-0.23				-0.23		-0.42	-0.23	2.8	-0.31		-0.23	-0.23
10	-0.43		0.095		2.2	-0.47	-0.23				-0.23		-0.36	-0.23	-0.33	-0.31		-0.23	4.1
11	0.092		0.015		-0.42	-0.47	-0.23				4.1		-0.16	-0.23	-0.33	-0.13		-0.23	-0.23
12	1.4		-0.16		-0.42	1.3	-0.23				-0.23		3.8	-0.23	2.8	-0.18		-0.23	-0.23
13	-0.46		-0.28		-0.42	-0.47	-0.23				-0.23		-0.42	4.1	-0.33	-0.13		-0.23	-0.23
14	-0.51		-0.39		-0.42	-0.47	-0.23				-0.23		-0.42	-0.23	-0.33	-0.31		-0.23	-0.23
15	-0.16		-0.22		-0.42	-0.47	-0.23				-0.23		-0.23	-0.23	-0.33	4.1		-0.23	-0.23
16	-0.5		-0.39		-0.42	-0.47	-0.23				-0.23		-0.36	-0.23	-0.33	-0.31		-0.23	-0.23
17	-0.46		-0.28		2.2	-0.47	-0.23				-0.23		-0.42	-0.23	-0.33	-0.3		4.1	-0.23
18	-0.49		-0.39		-0.42	-0.47	-0.23				-0.23		-0.42	-0.23	-0.33	-0.31		-0.23	-0.23

4 Decision Tree Classifier :

The cross accuracy of DecisionTreeClassifier Model is shown in the below picture.

Here is the accuracy,macro average & weighted average are

given.

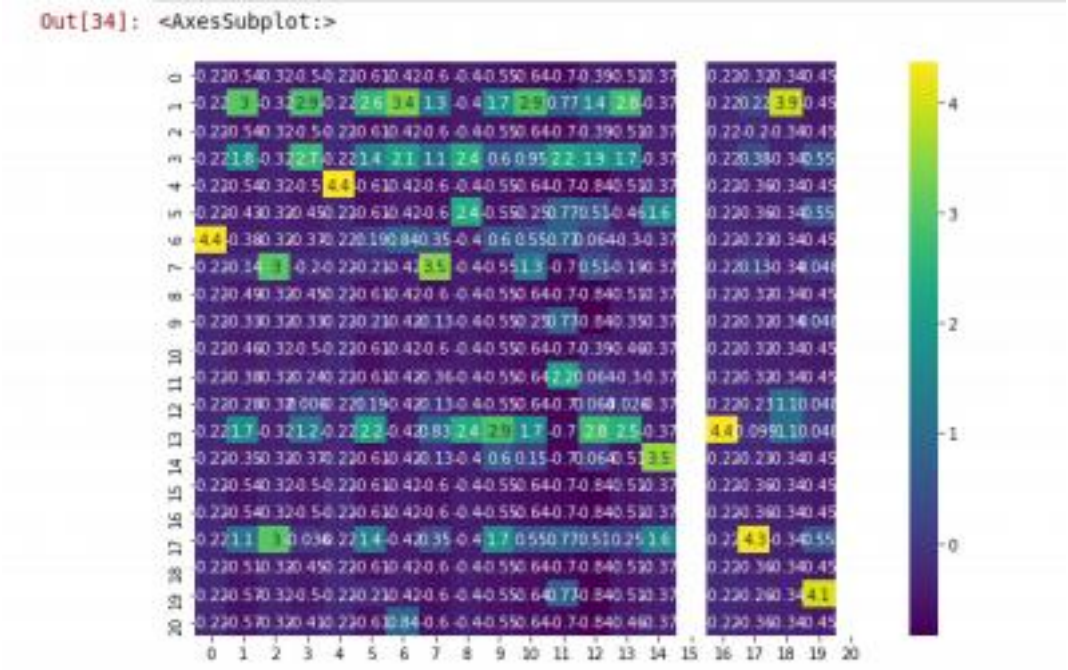


```
Cross Accuracy: 0.33 (+/- 0.04)
              precision    recall  f1-score   support

           0       0.00      0.00      0.00         3
           1       0.30      0.41      0.35       335
           2       0.00      0.00      0.00         7
           3       0.31      0.29      0.30       259
           4       1.00      0.50      0.67         2
           5       0.00      0.00      0.00        16
           6       0.14      0.03      0.05        33
           7       0.32      0.23      0.27        73
           8       0.00      0.00      0.00         5
           9       0.00      0.00      0.00        23
          10       0.00      0.00      0.00         7
          11       0.20      0.09      0.12        23
          12       0.05      0.05      0.05        44
          13       0.28      0.25      0.26       224
          14       0.50      0.08      0.14        24
          15       0.00      0.00      0.00         1
          16       0.00      0.00      0.00         1
          17       0.62      0.57      0.59       254
          18       0.00      0.00      0.00         3
          19       0.47      0.64      0.55        14
          20       0.00      0.00      0.00         4

    accuracy                           0.33      1355
   macro avg       0.20      0.15      0.16      1355
weighted avg       0.34      0.33      0.33      1355
```

The confusion matrix of DecisionTreeClassifier Model

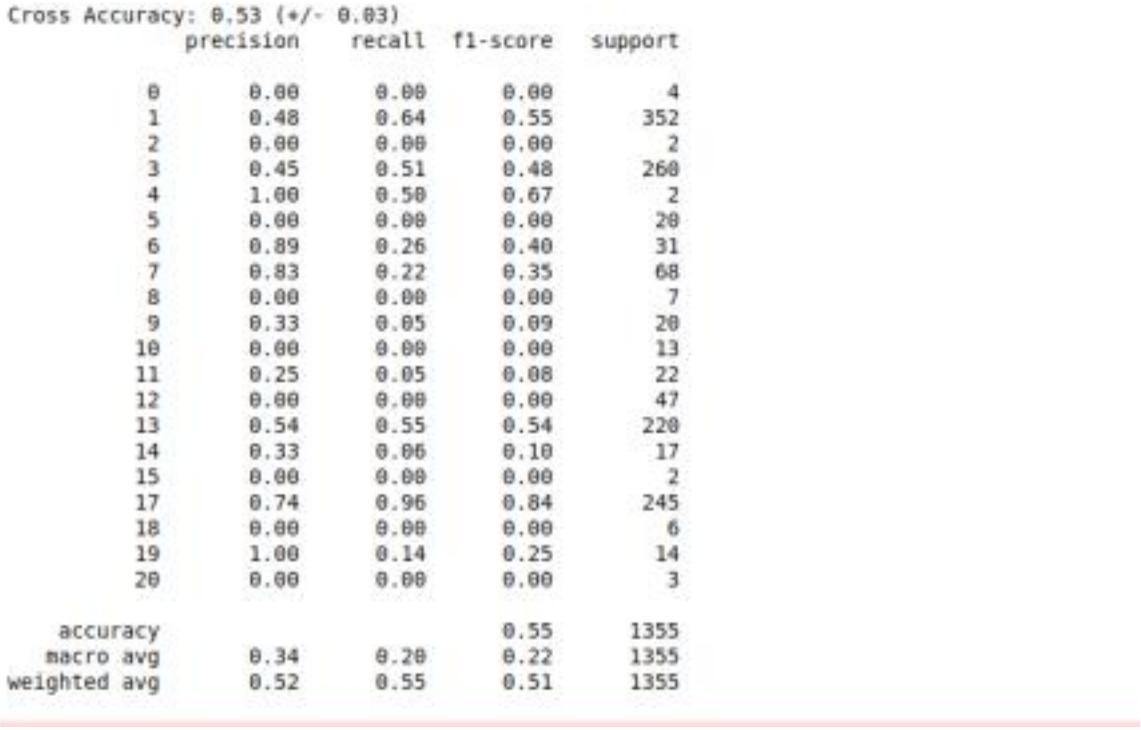


Out[34]: <AxesSubplot:>

5 Artificial neural networks :

The cross accuracy of TfidfTransformer Model is shown in the below picture.

Here is the accuracy,macro average & weighted average are given.

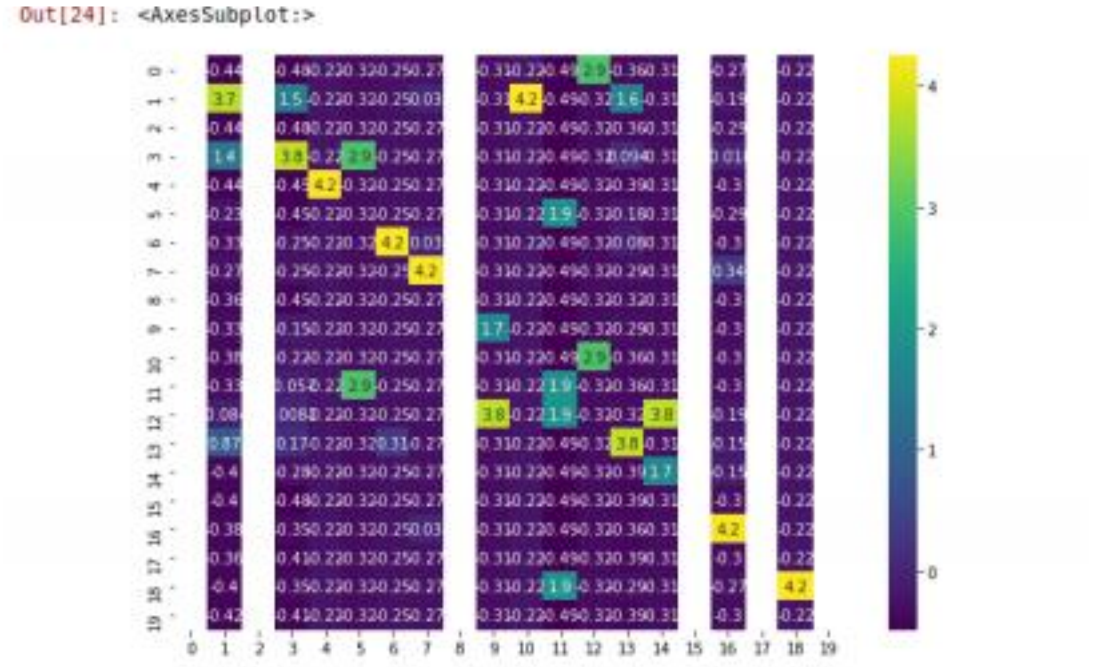


```
Cross Accuracy: 0.53 (+/- 0.03)
              precision    recall  f1-score   support

           0       0.00      0.00      0.00         4
           1       0.48      0.64      0.55       352
           2       0.00      0.00      0.00         2
           3       0.45      0.51      0.48       260
           4       1.00      0.50      0.67         2
           5       0.00      0.00      0.00        20
           6       0.89      0.26      0.40        31
           7       0.83      0.22      0.35        68
           8       0.00      0.00      0.00         7
           9       0.33      0.05      0.09        20
          10       0.00      0.00      0.00        13
          11       0.25      0.05      0.08        22
          12       0.00      0.00      0.00        47
          13       0.54      0.55      0.54       220
          14       0.33      0.06      0.10        17
          15       0.00      0.00      0.00         2
          17       0.74      0.96      0.84       245
          18       0.00      0.00      0.00         6
          19       1.00      0.14      0.25        14
          20       0.00      0.00      0.00         3

    accuracy                           0.55      1355
   macro avg       0.34      0.20      0.22      1355
weighted avg       0.52      0.55      0.51      1355
```

The confusion matrix of DTfidfTransformer Model.



```
Out[24]: <AxesSubplot:>
```

2.
The explanation behind ”The model gives very low f1 score for some classes but not the same for others”is given below :

We know that the F1/F Score is a measure of how accurate a model is by using Precision and Recall following the formula of:

F1 Score = 2 * ((Precision * Recall) / (Precision + Recall))

Precision is commonly called positive predictive value. It is also inte resting to note that the PPV can be derived using Bayes' theorem as well.
Precision = True Positives / (True Positives + False Positives)

Recall is also known as the True Positive Rate and is defined as the following:
Recall = True Positives / (True Positives + False Negatives)

If the precision is very low and recall value gets very high then the F1 score will become very low. But it should become the average of precision and recall. The alternative situation also behave the same. So, In the end, We can say. Some model gives the high precision and high recall value ,which are made the F1 score high. But if one's score gets very low then the F1 score also become very low.

3.
The low f1 score issue is tried to fix in below :

If the F1-score is the figure of merit, I would try to tune the class weights. It should be pretty easy, if we have a binary

classification problem. We can feed class weight a dictionary with the weights for each class.
Here’s a little example.

```
clf = RandomForestClassifier()
params = {’class weight’:[{0:neg weight, 1:1} for neg weight in
np.arange(1.0, 5.0, 0.5)]}
gs = GridSearchCV(estimator= clf, param grid = params, cv =
5) gs.fitX train, y train
```

The End