The dataset consisted of 2 columns one was sentences and other str words.

So Basically I used KNN classifier as the classification model for splitting the data into training and testing, I had to use nltk libraries for the y dataset as just using standardization or label encoding was not working, had to use stopwords and tokenizer for the same. Applied conditions and exceptions according to the dataset.

The accuracy was coming around 0.29 for the model.

The accuracy score and classification report is attached below.

```
accuracy = metrics.accuracy_score(y_test, y_pred)
print('Accuracy: {:.2f}'.format(accuracy))
```

```
Accuracy: 0.29
```

```
#classification report
from sklearn.metrics import classification_report
print(classification_report(y_test,y_pred))
```

```
                precision    recall  f1-score   support

       Bigdata       0.40      0.40      0.40       761
    Blockchain       0.10      0.11      0.11       439
 Cyber Security      0.19      0.20      0.19       862
 Data Security       0.04      0.03      0.04       115
       FinTech       0.42      0.42      0.42      2852
 Microservices       0.07      0.06      0.06       352
      Neobanks       0.17      0.17      0.17       347
      Reg Tech       0.32      0.36      0.34       699
 Robo Advising       0.03      0.03      0.03       249
 Stock Trading       0.05      0.04      0.04       251
credit reporting     0.13      0.12      0.12       566

      accuracy                           0.29      7493
     macro avg       0.17      0.18      0.18      7493
  weighted avg       0.28      0.29      0.28      7493
```

This was model was specifically designed for the dataset provided hence it might not work for other data sets where datatypes are different.