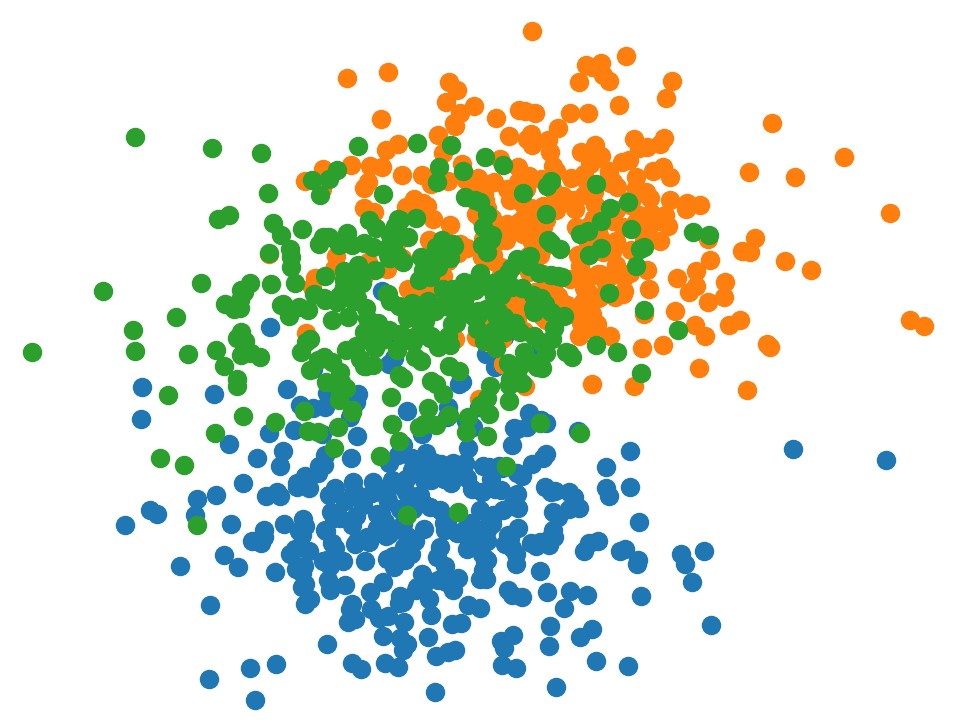
ROOT2AI MODEL REPORT

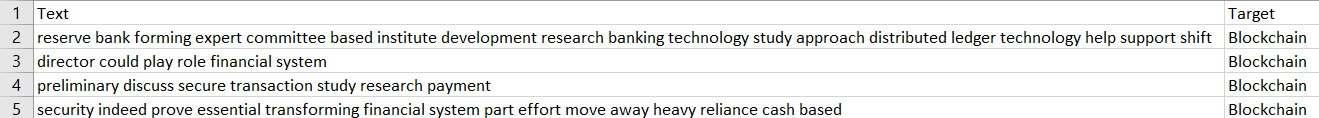
**MULTICLASS CLASSIFICATION**:

Multiclass or multinomial classification is the problem of classifying instances into one of three or more classes.



**DATASET:**

* Our dataset from blockchain domain contains tow columns and 22704 rows.
* The 1st column ‘Text’ contains the data collected from the domain as text.
* The 2nd column ‘Target’ contains the label of the text in the 1st column respectively.



**APPROACH TO SOLVE THE PROBLEM:**

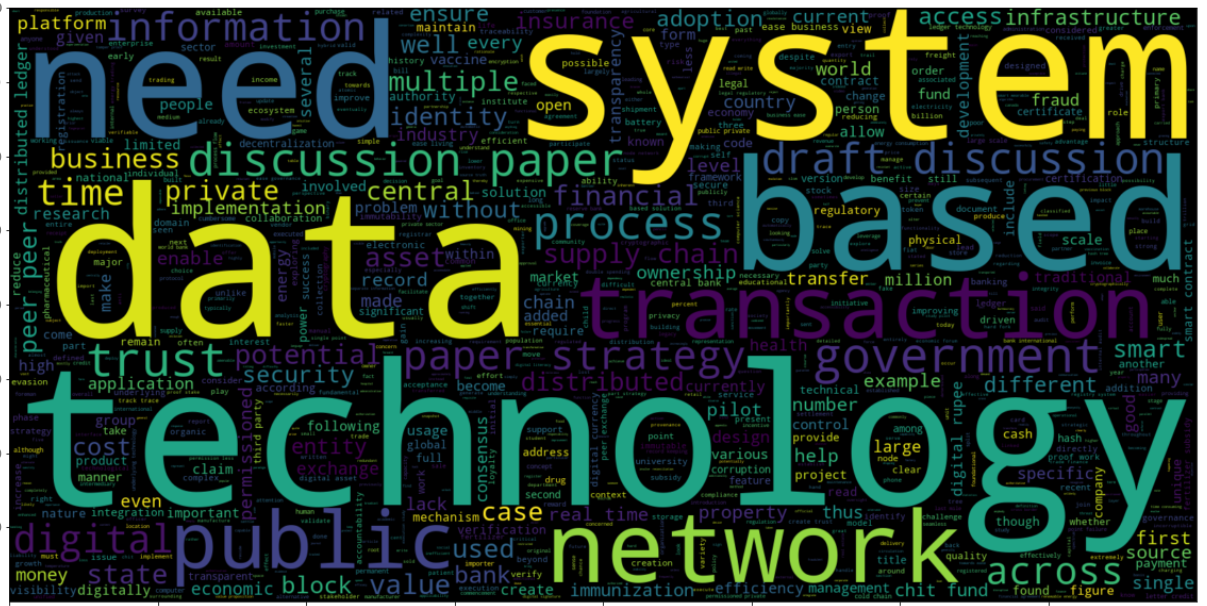
* Data Cleaning
* Vectorizing the text data
* Training the model
* Evaluation
* Multiclass Classification problems can be solved by various machine learning algorithms like SVM, Naïve Bayes etc.
* Indeed, we will be using Naïve Bayes’s MultinomialNB for solving the problem.

**DATA CLEANING:**

We can’t feed the raw input to our model. Thus, we will be performing Data Cleaning.

**TEXT PREPROCESSING:**

* At first, we will remove the regular expressions and replace it with white space.
* Then we will convert the words in the text column of dataset into lower letters.
* Then we will split each word in the text column of dataset.
* Now we will remove the stop words like ‘is’, ‘are’, ‘was’ etc., in English.
* Then we will join all the words with a space.
* Now we got our preprocessed text data which can be fed to our model.



**VECTORIZING THE TEXT DATA:**

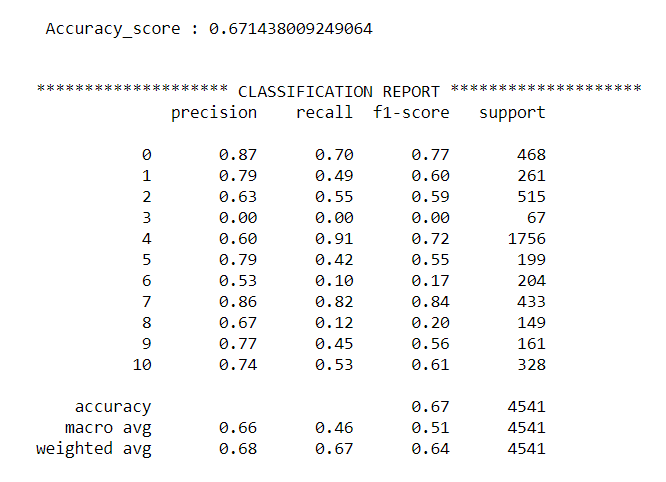
* + We are using the CountVectoizer function of sklearn for our purpose.
  + Now after initialization, we will fit transform our Text data in vectors and then to array.
  + Then encode the target class.

**TRAINING:**

* + - Now split the dataset using train\_test\_split. We are ready to train the model.
    - Import the Multinomial Naïve Bayes from sklearn and initialize it.
    - Then fit the model on the training set.
    - Predict the classification for the data in X\_test

**EVALUATING OUR MODEL:**

Using the classification metrics like accuracy\_score and confusion\_matrix evaluate our model with the y\_test.



**LIMITATIONS OF OUR MODEL:**

* Of course, the accuracy of our model is low in terms of a good classification model, but we have done some augmentation steps like removing stop word and stemming etc.
* Indeed, we can use some deep learning models like LSTM for better accuracy, but due to less computational power, the time consumption for training the model will be more.