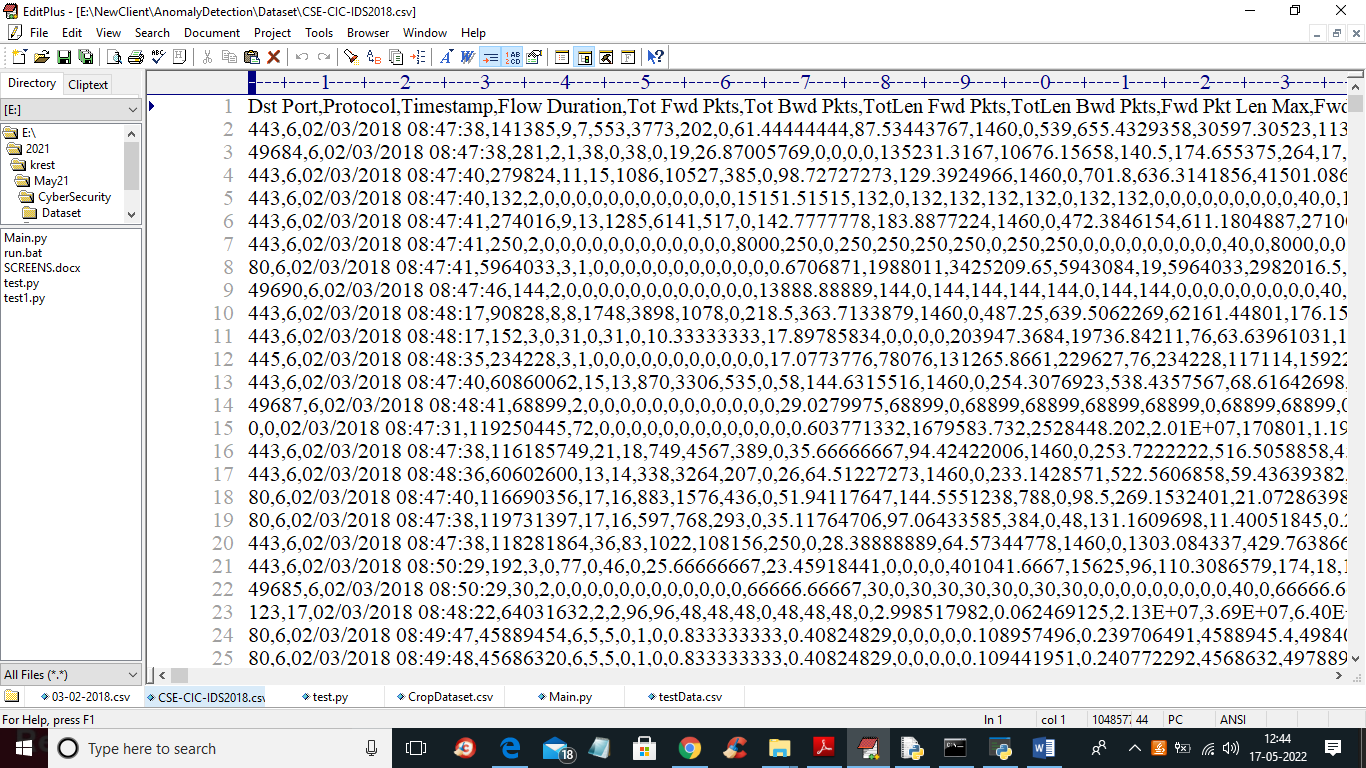
A Data Analytics Approach to the Cybercrime Underground Economy

Now-a-days 90% peoples are dependent on ONLINE Internet services to perform their daily activities such as banking transaction, online health consultation, online road traffic status and many more but this facility raised a problem like cyber theft and attack where malicious coders can intercept network communication to steal data or to add fake data which causes user to lost money or control over their system. Attackers may use brute force technique to guess user password and may BOT attack to raise fake ratings. So on internet various types of attacks are available and this attacks can be developed by paid programmers by using Black Market money.

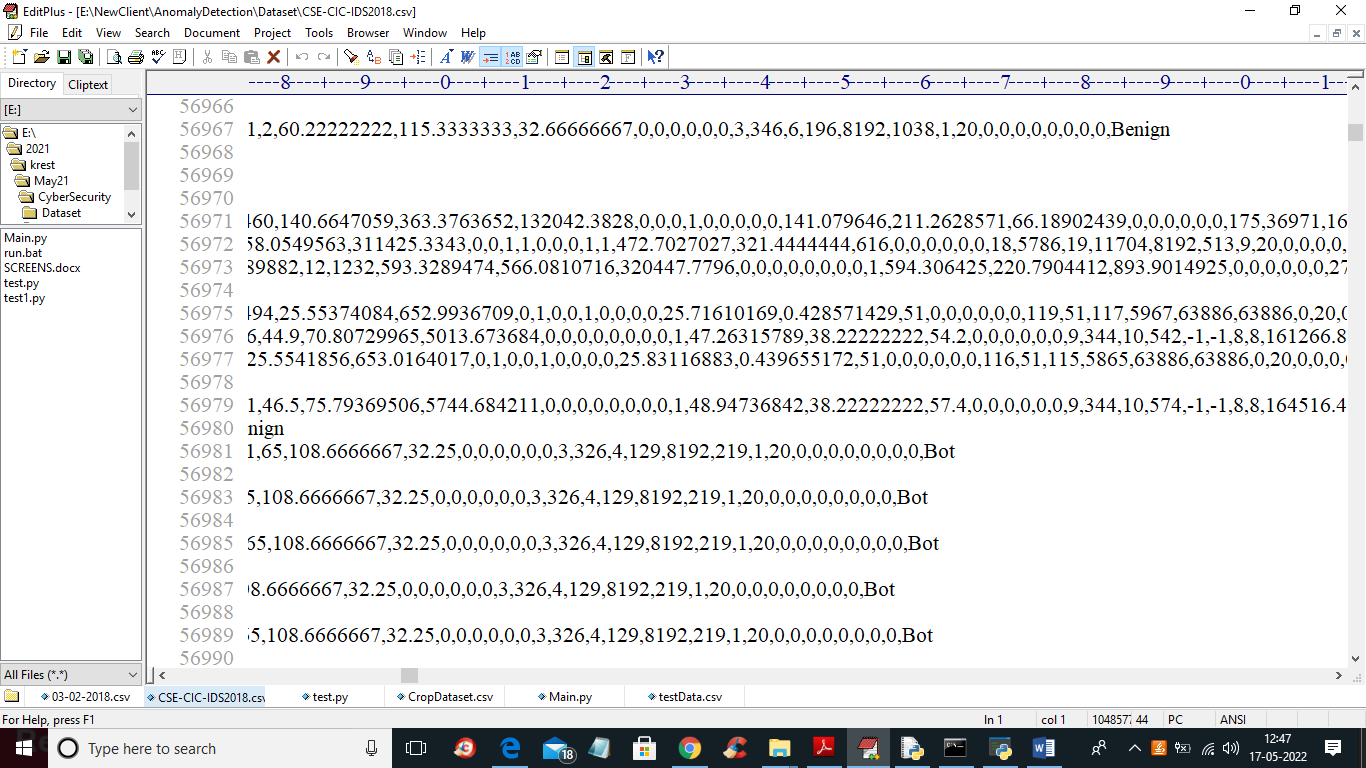
Author of this paper describing such several attacks and then introducing data analytics model to know in which attacks more economy or money is using and then introducing classification model by employing Naïve Bayes algorithm. To train classification model author is using MALWARE attacks dataset from Hacking companies and this dataset contains signature of various attacks such as SPAM, Brute force, SQL injection, BOT and many more.

To know more about various cybercrime details you can read base paper.

To implement this project we have used below malware dataset



In above screen first row contains dataset column names and remaining rows contains dataset value and in last column we can see the attack names showing in below screen



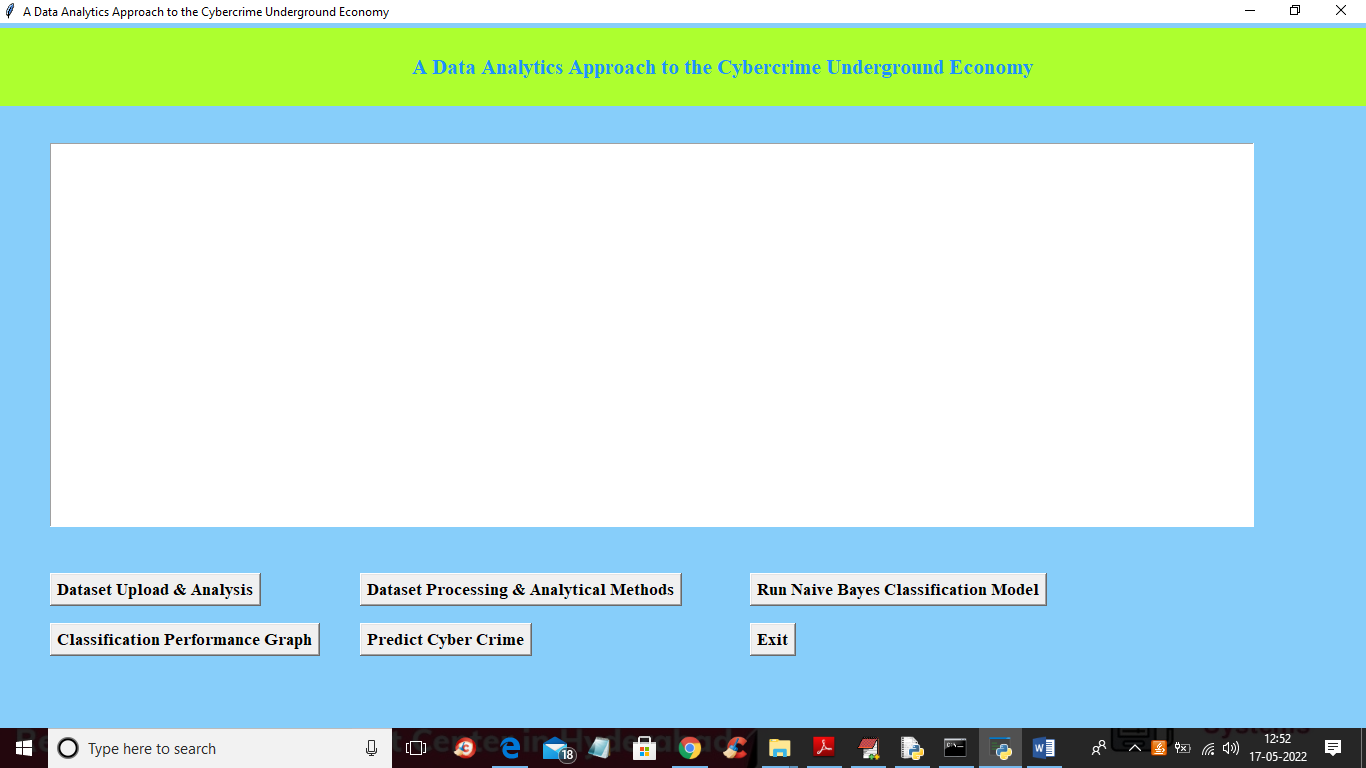
In above screen in last columns you can see the names as BOT or BENIGN so each records contains network communication data and associate with attack class label.

To implement this project we have designed following modules

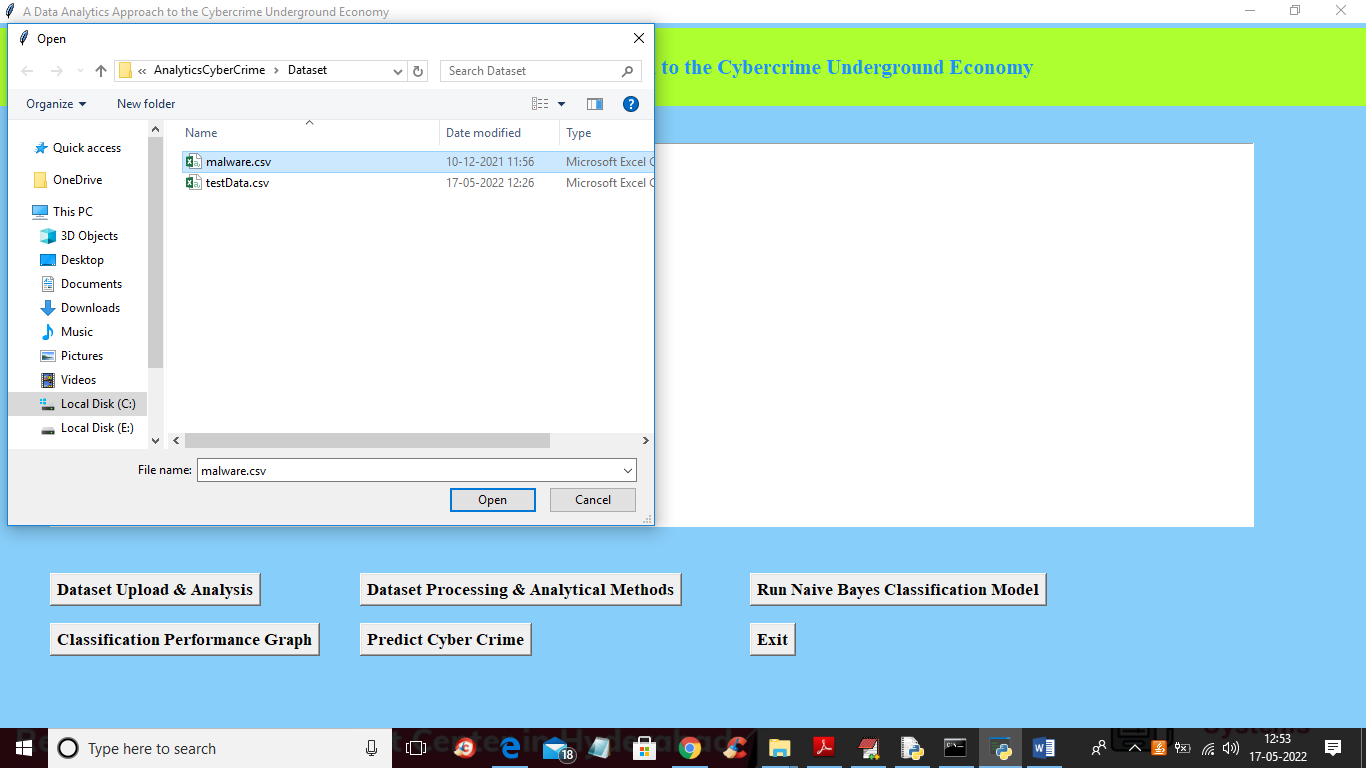
1. Dataset Upload & Analysis: using this module we will upload dataset and then perform analysis methods such as finding various cybercrime and its count and then clean dataset by removing missing values
2. Dataset Processing & Analytical Methods: using this module we will encode attack labels with integer ID and then split dataset into train and test where application used 80% dataset to train classification Naïve Bayes algorithm and 20% to test its prediction performance
3. Run Naive Bayes Classification Model: using this module we will trained classification algorithm with above 80% dataset and then build a prediction model
4. Classification Performance Graph: using this module we will plot naïve Bayes accuracy and precision graph to know its performance accuracy
5. Predict Cyber Crime: using this module we will upload test cybercrime network dataset and then classification model will predict weather test data contains any cybercrime signature.

SCREEN SHOTS

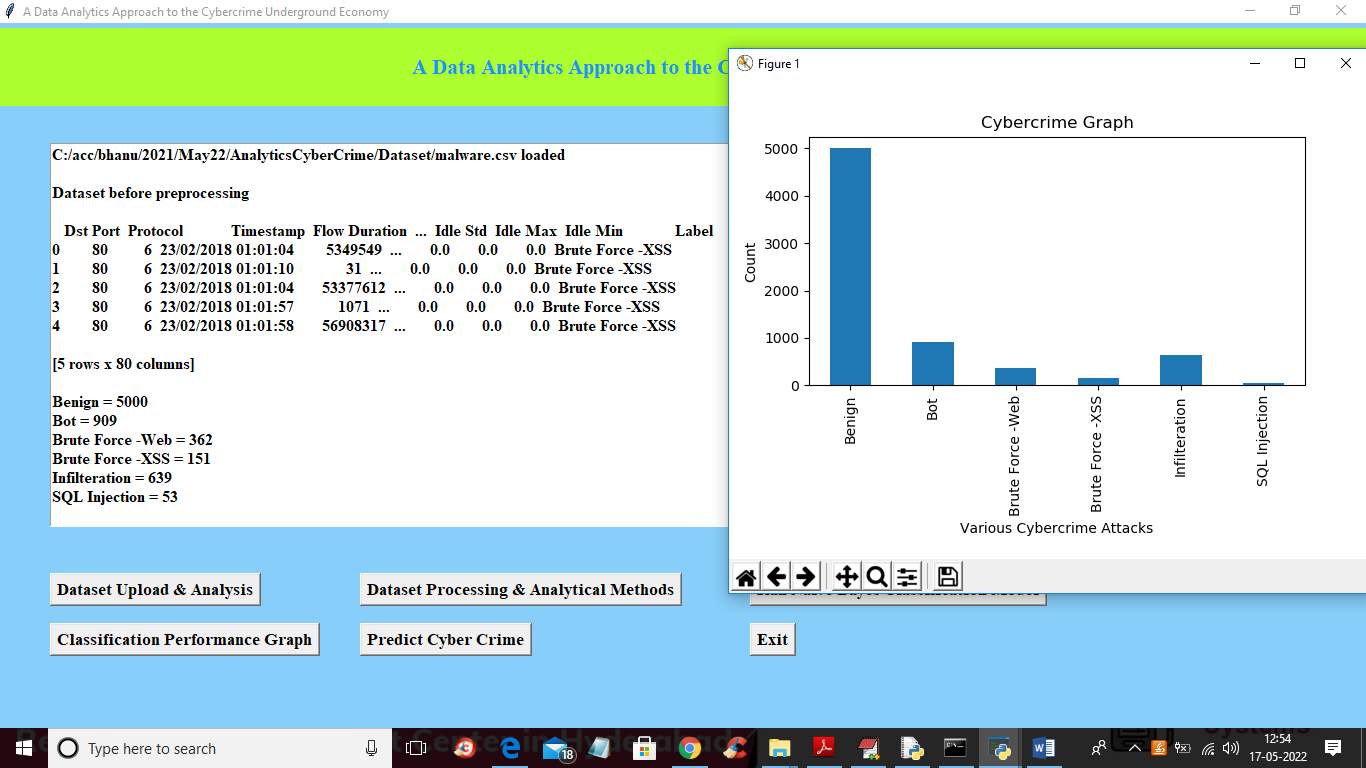
To run project double click on ‘run.bat’ file to get below output



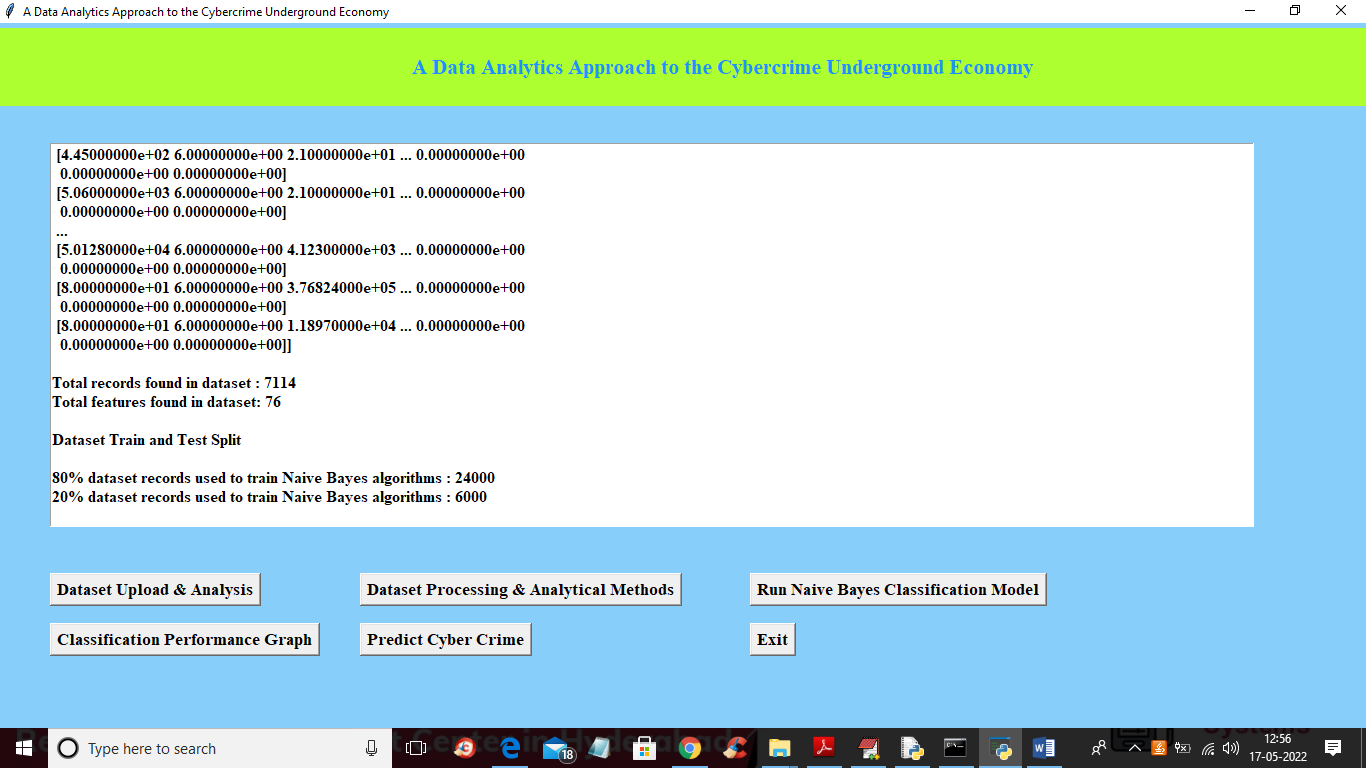
In above screen click on ‘Dataset Upload & Analysis’ button to upload dataset and get below output



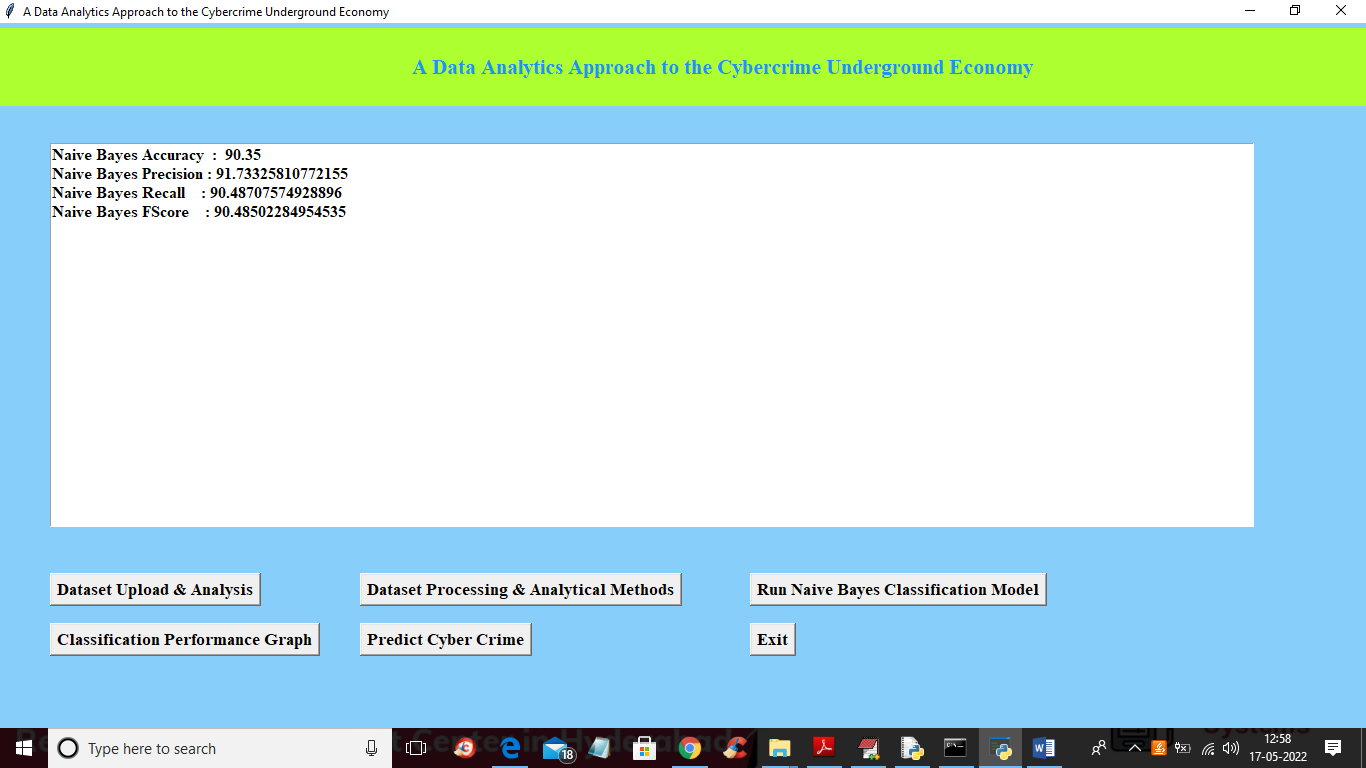
In above screen selecting and uploading ‘malware.csv’ file and then click on ‘Open’ button to load dataset and get below analysis output



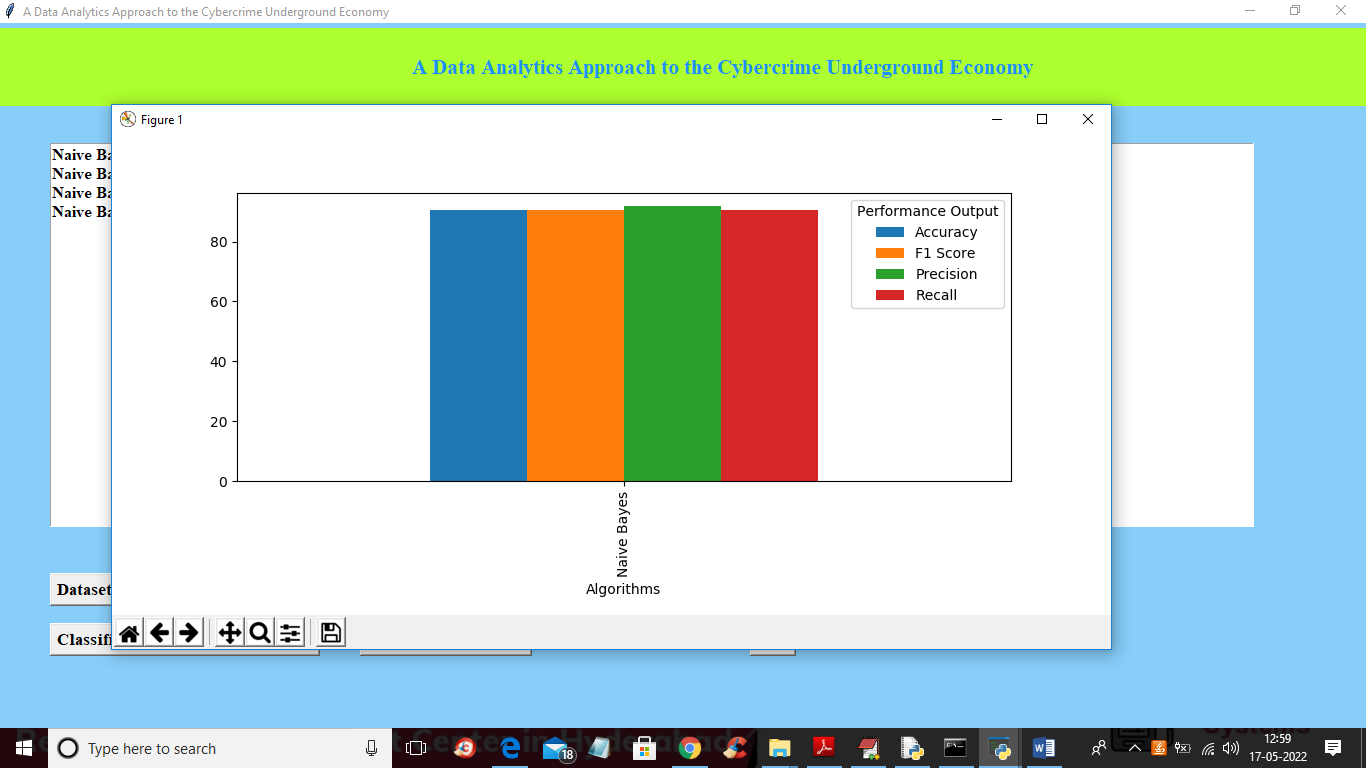
In above screen we can see dataset loaded and dataset contains some non-numeric characters and algorithm will not take such non-numeric data so we need to convert it into integer ID and in above screen we can various cybercrime and its count so by seeing this we can analyse on which cybercrime most money is using and in above graph we can see that analysis output. Now close above graph and then click on ‘Dataset Processing & Analytical Methods’ button to clean dataset and then split data into train and test



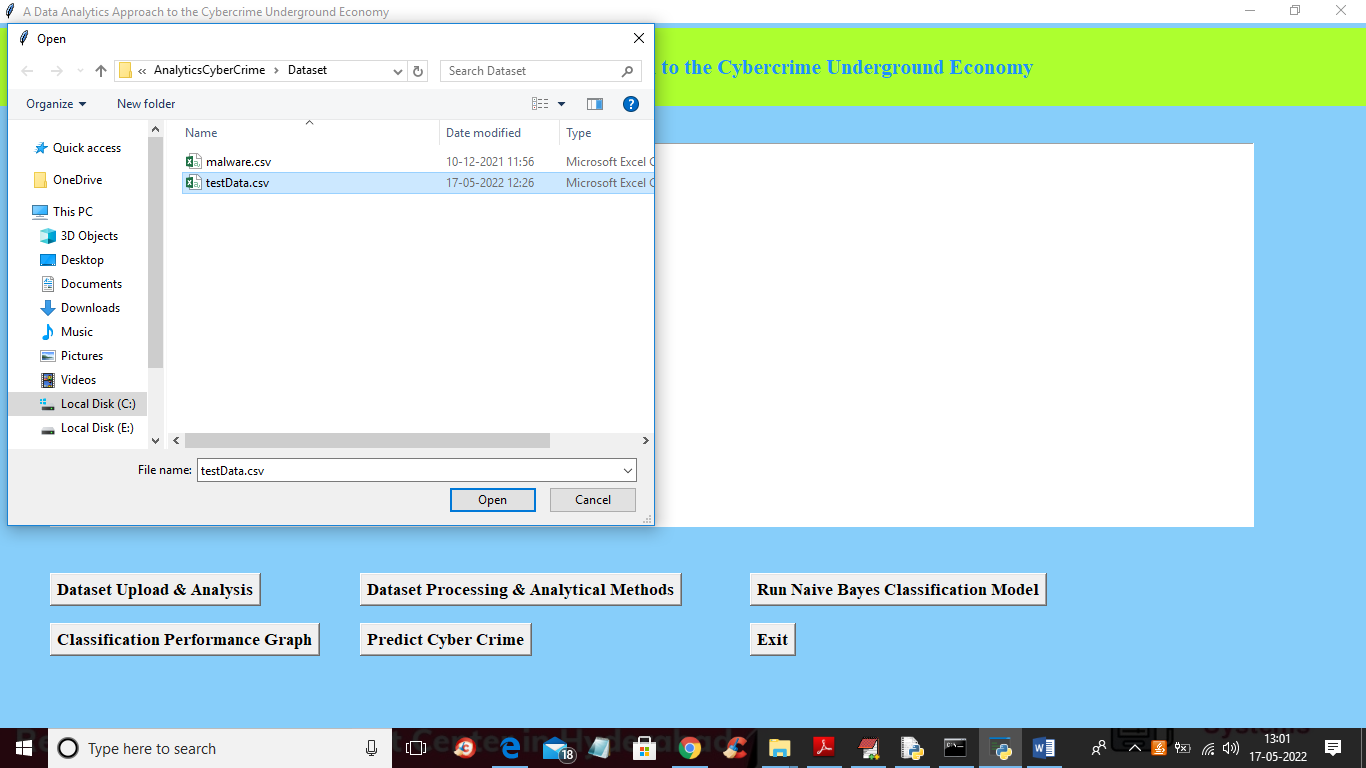
In above screen we can see all data is converted to numeric format and then we can see total dataset and then we can see 80% dataset used for training and 20 for testing and now dataset is ready and now click on ‘Run Naive Bayes Classification Model’ button to train classification model and get below output.



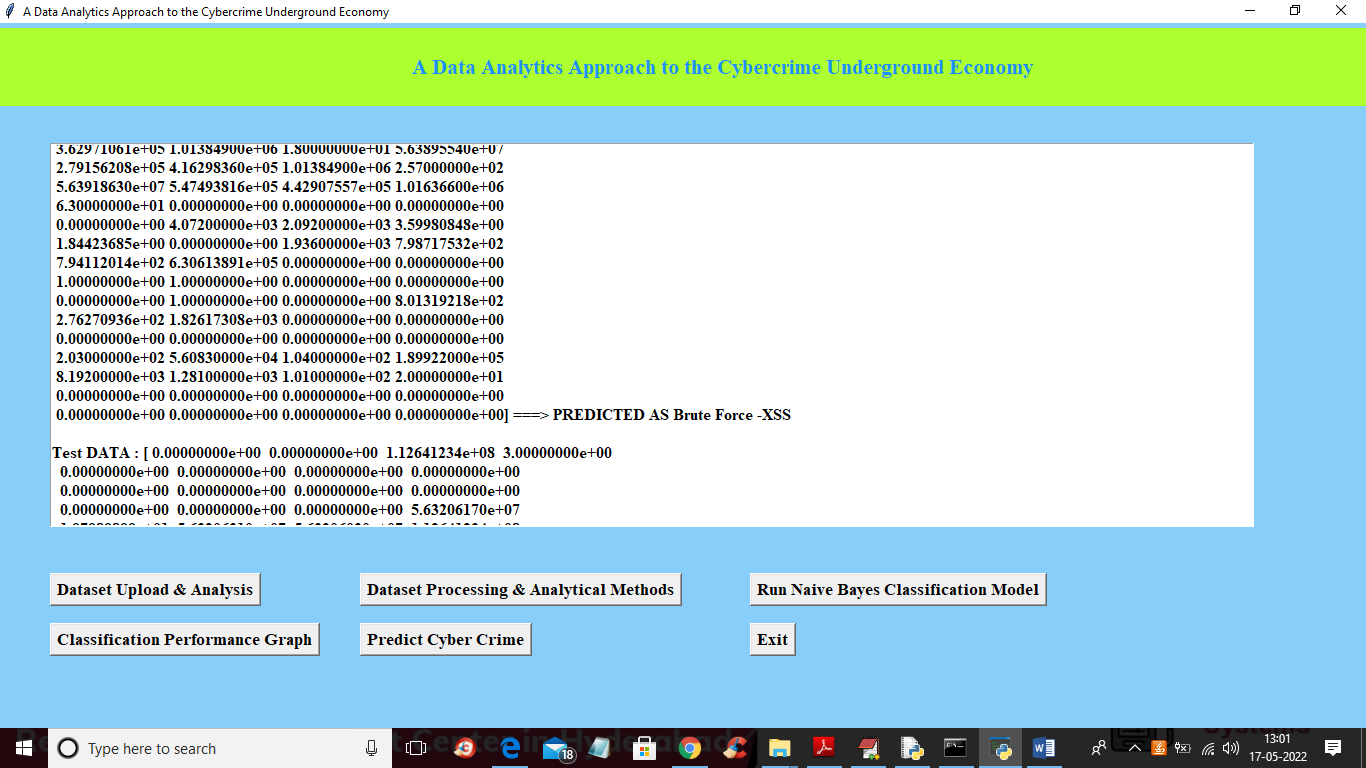
In above screen Naïve Bayes training is completed and we got its prediction accuracy as 90% and now click on ‘Classification Performance Graph’ button to get below graph



In above graph x-axis represents algorithm names and each different colour bar represents different metrics such as Accuracy, precision etc. in above graph we can see all metrics performance is above 90% and now close above graph and then click on ‘Predict Cyber Crime’ button to upload test data and get below output



In above screen selecting and uploading ‘testData.csv’ file and then click on ‘Open’ button to get below output



In above screen in square bracket we can see network traffic data and after arrow =🡺 symbol we can see the type of cybercrime attack prediction. Scroll down above screen to view all cybercrime prediction

