**Cyber Threat Detection Based on Artificial Neural Networks Using Event Profiles**

In this paper author is describing concept to detect threats using AI-SIEM (Artificial Intelligence-Security Information and Event Management) technique which is a combination of deep learning algorithms such as FCNN, CNN (Convolution Neural Networks) and LSTM (long short term memory) and this technique works based on events profiling such as attack signatures. Author evaluating propose work performance with conventional algorithms such as SVM, Decision Tree, Random Forest, KNN and Naïve Bayes. Here I am implementing CNN and LSTM algorithms.

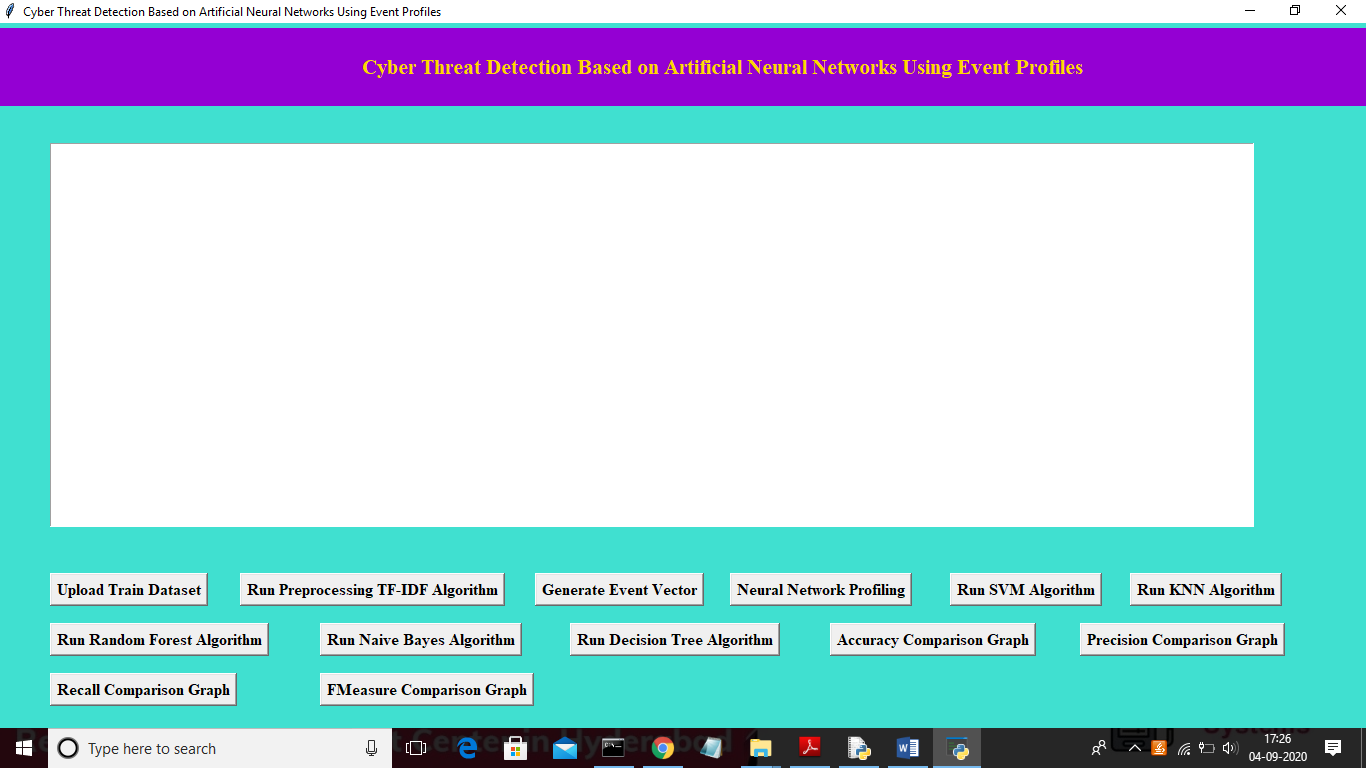
Propose algorithms consists of following module

1. Data Parsing: This module take input dataset and parse that dataset to create a raw data event model
2. TF-IDF: using this module we will convert raw data into event vector which will contains normal and attack signatures
3. Event Profiling Stage: Processed data will be splitted into train and test model based on profiling events.
4. Deep Learning Neural Network Model: This module runs CNN and LSTM algorithms on train and test data and then generate a training model. Generated trained model will be applied on test data to calculate prediction score, Recall, Precision and FMeasure. Algorithm will learn perfectly will yield better accuracy result and that model will be selected to deploy on real system for attack detection.

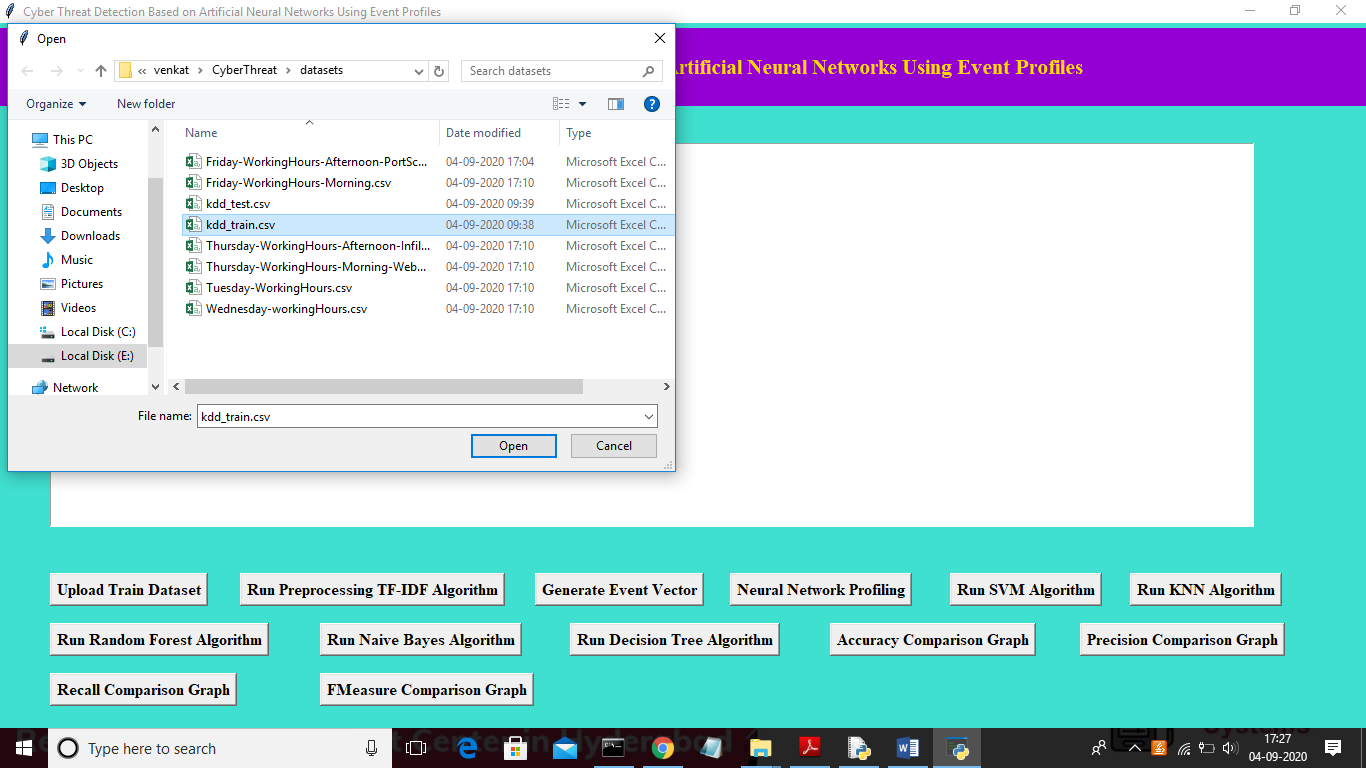
Datasets which we are using for testing are of huge size and while building model it’s going to out of memory error but kdd\_train.csv dataset working perfectly but to run all algorithms it will take 5 to 10 minutes. You can test remaining datasets also by reducing its size or running it on high configuration system.

Screen shots

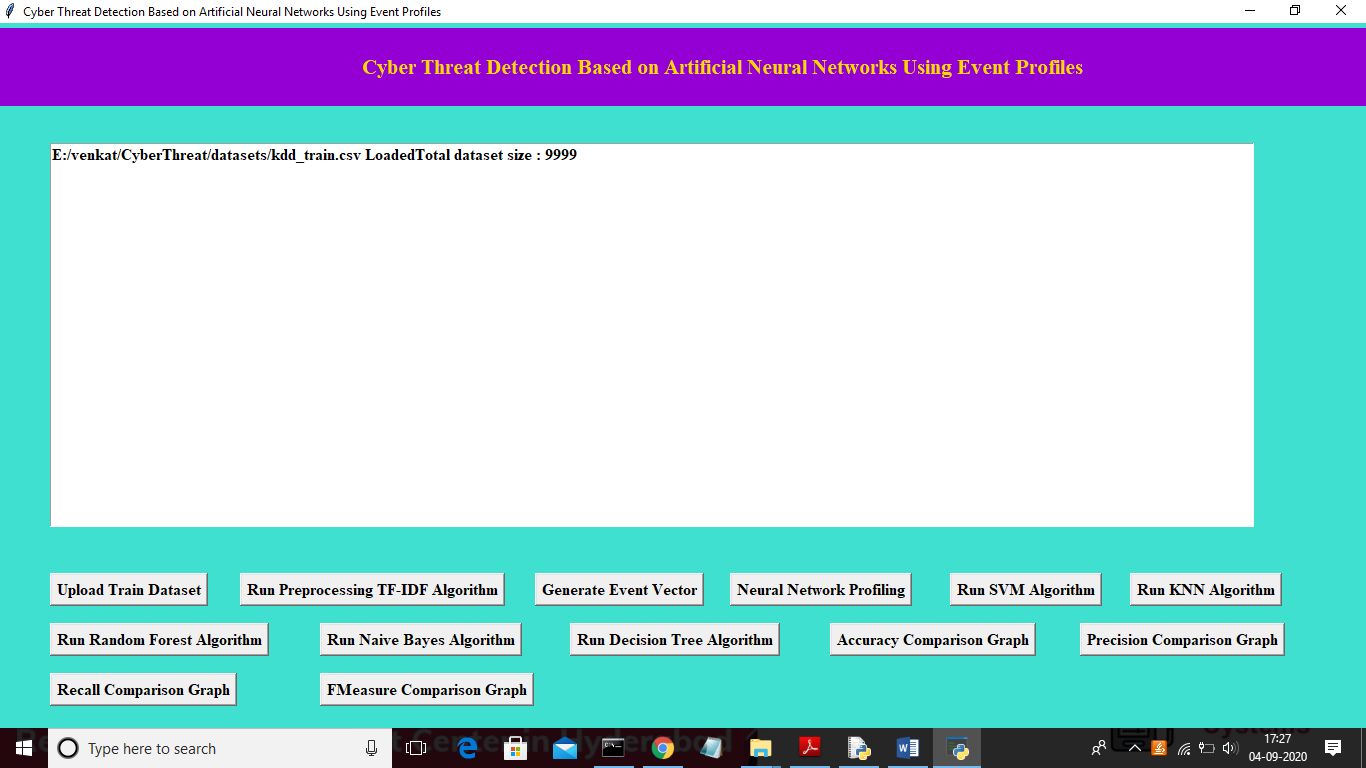
To run project double click on ‘run.bat’ file to get below screen



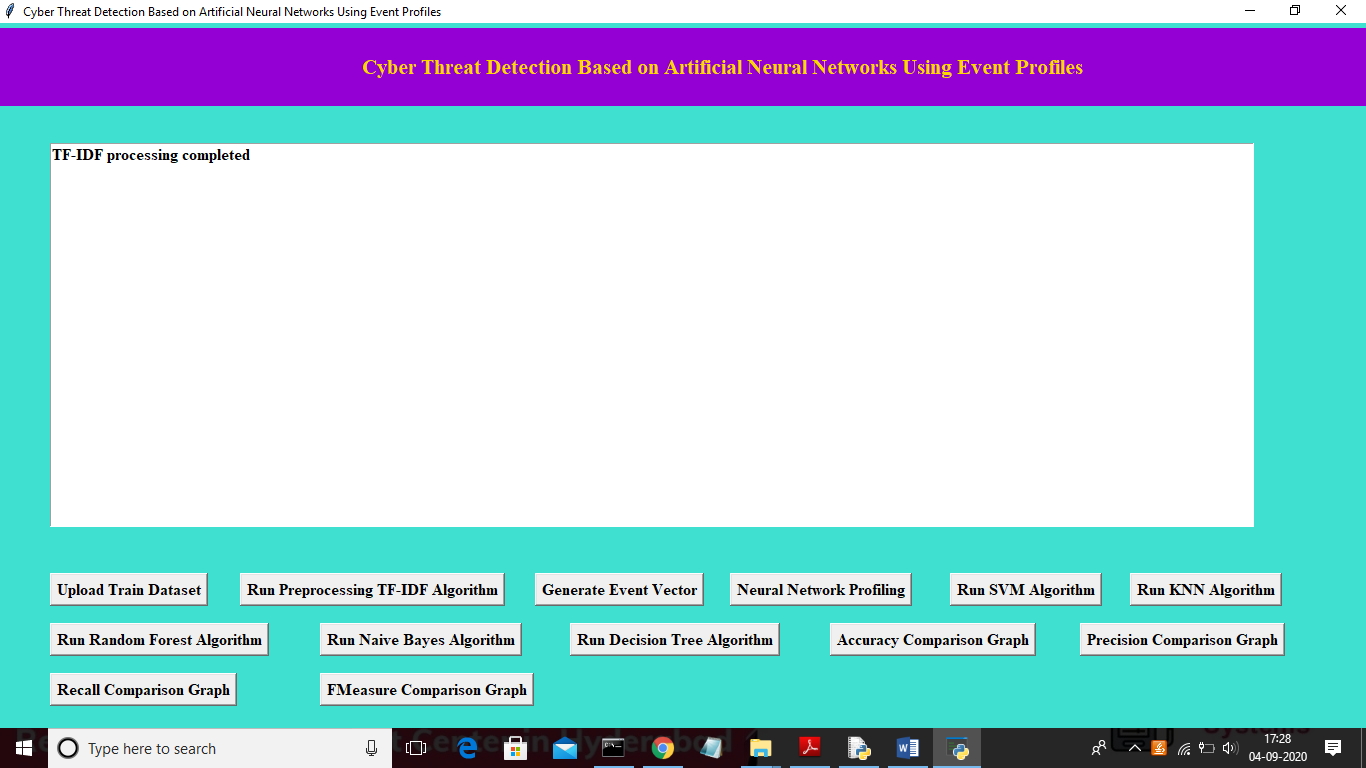
In above screen click on ‘Upload Train Dataset’ button and upload dataset



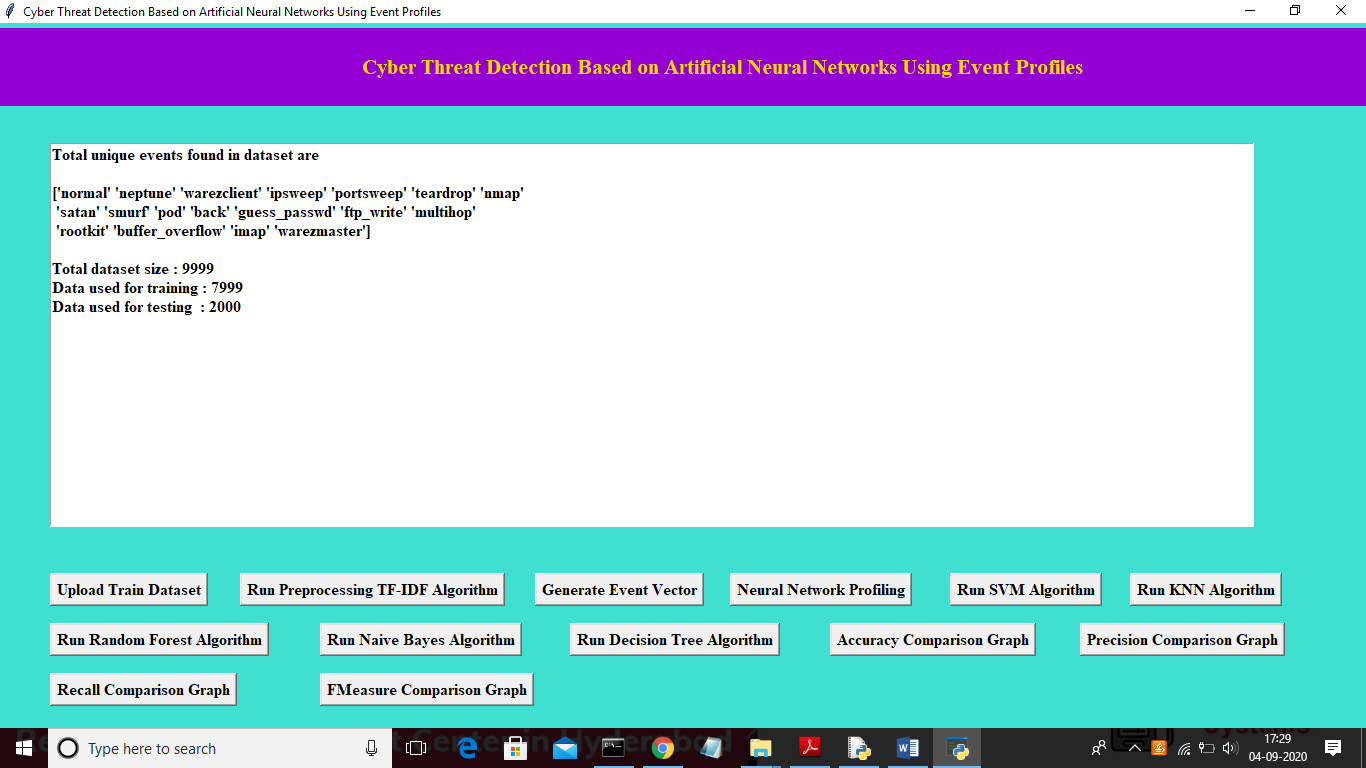
In above screen uploading ‘kdd\_train.csv’ dataset and after upload will get below screen



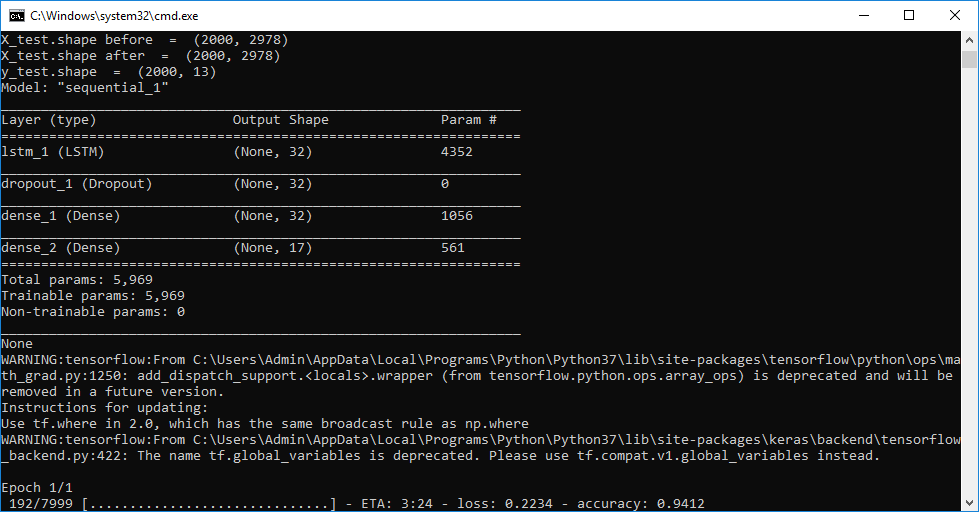
In above screen we can see dataset contains 9999 records and now click on ‘Run Preprocessing TF-IDF Algorithm’ button to convert raw dataset into TF-IDF values



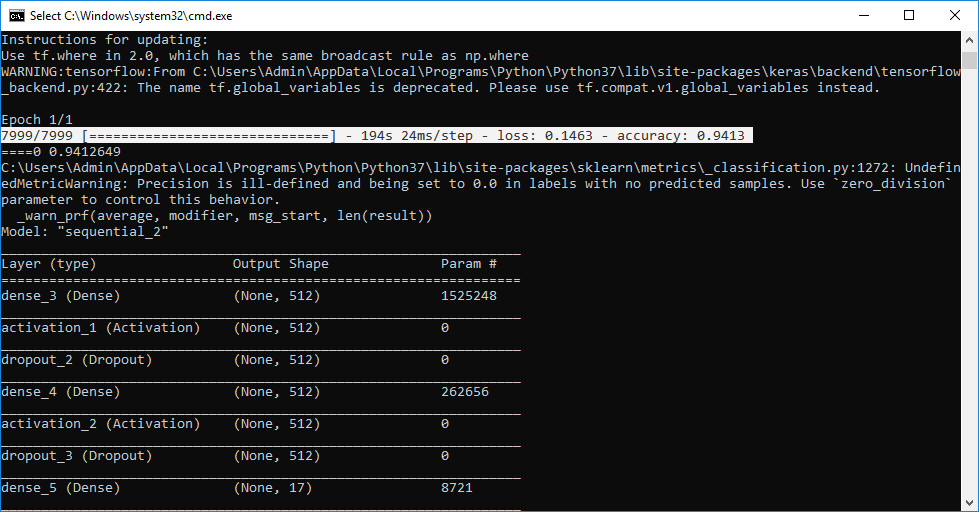
In above screen TF-IDF processing completed and now click on ‘Generate Event Vector’ button to create vector from TF-IDF with different events



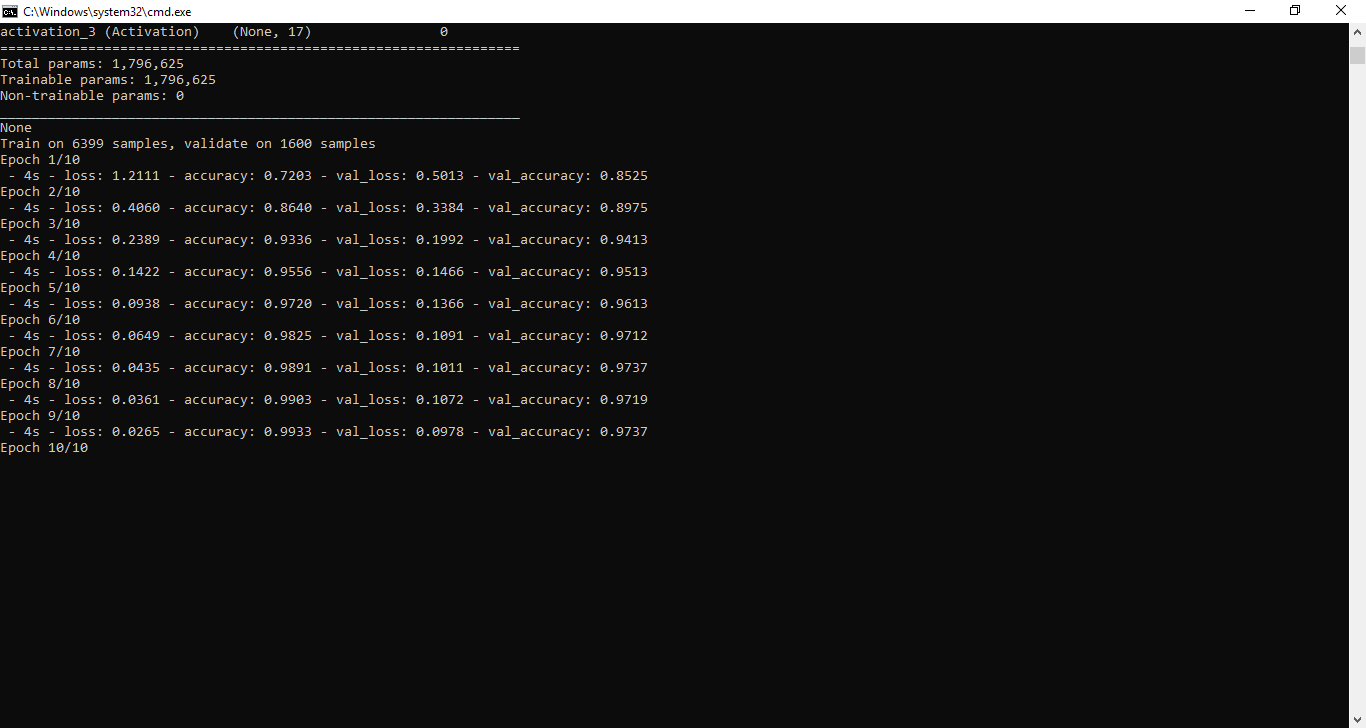
In above screen we can see total different unique events names and in below we can see dataset total size and application using 80% dataset (7999 records) for training and using 20% dataset (2000 records) for testing. Now dataset train and test events model ready and now click on ‘Neural Network Profiling’ button to create LSTM and CNN model



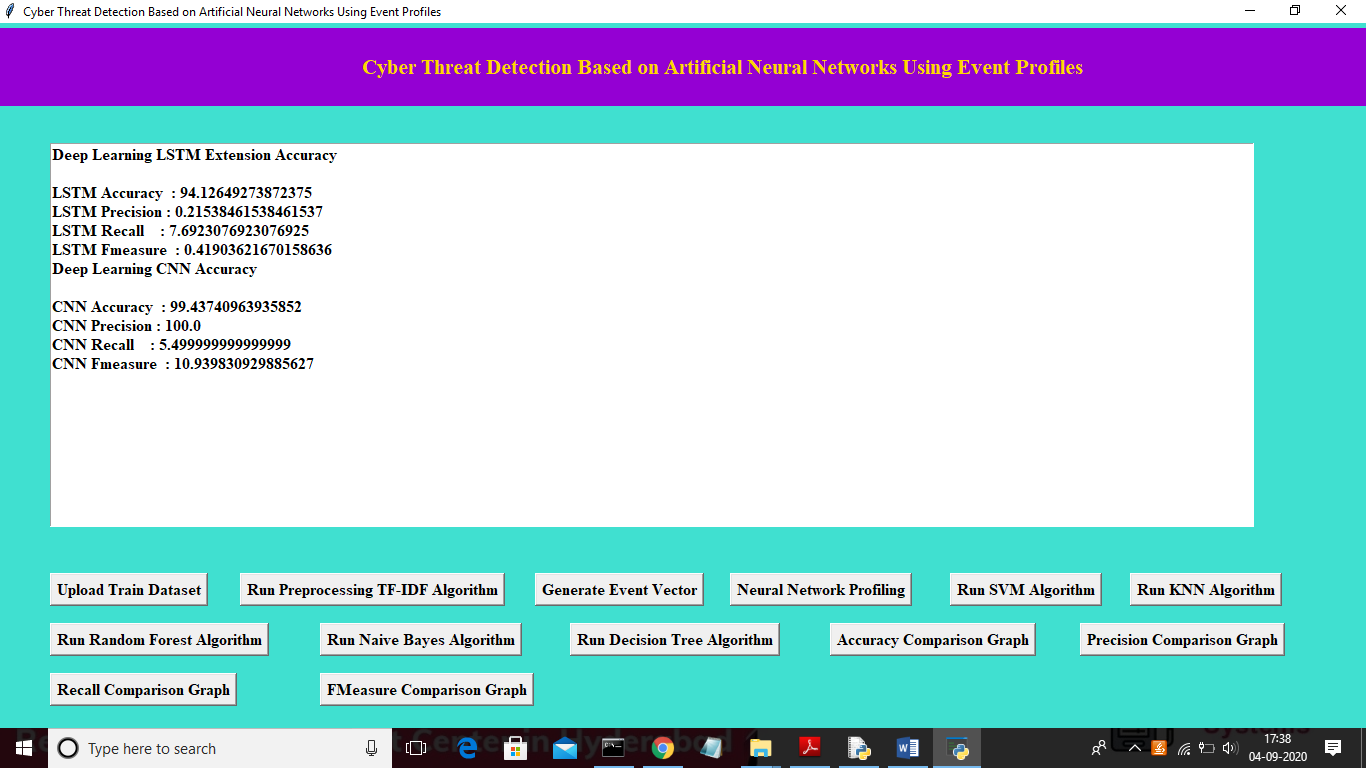
In above screen LSTM model is generated and its epoch running also started and its starting accuracy is 0.94. Running for entire dataset may take time so wait till LSTM and CNN training process completed. Here dataset contains 7999 records and LSTM will iterate all records to filter and build model.



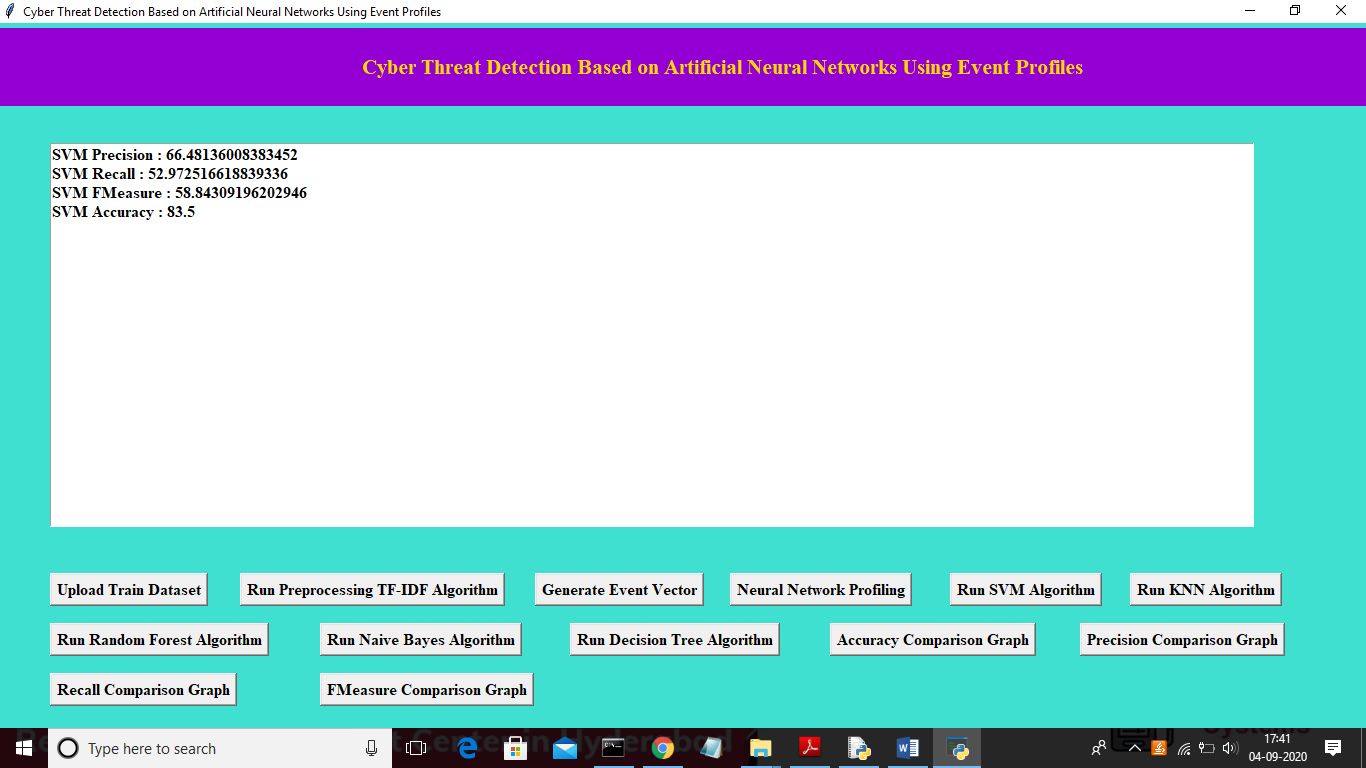
In above selected text we can see LSTM complete all iterations and in below lines we can see CNN model also starts execution



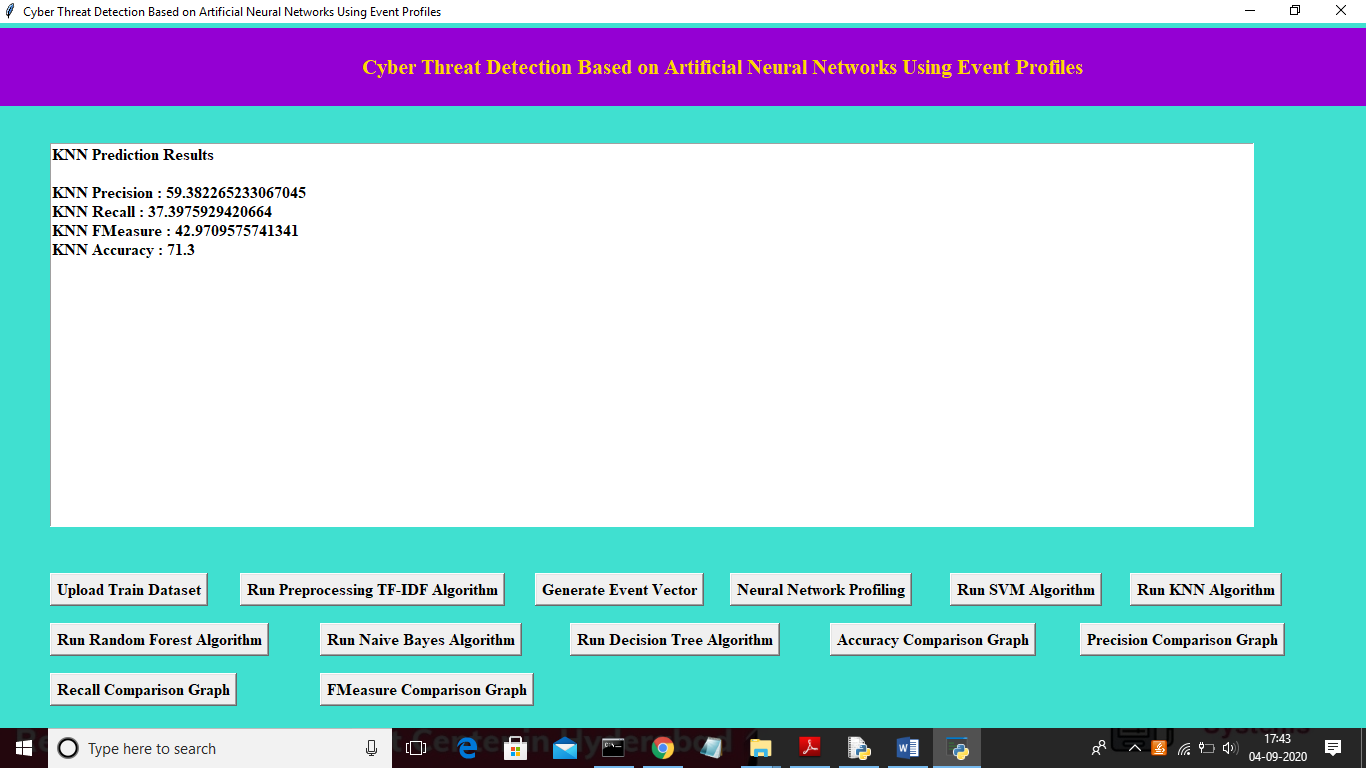
In above screen CNN also starts first iteration with accuracy as 0.72 and after completing all iterations 10 we got filtered improved accuracy as 0.99 and multiply by 100 will give us 99% accuracy. So CNN is giving better accuracy compare to LSTM and now see below GUI screen with all details



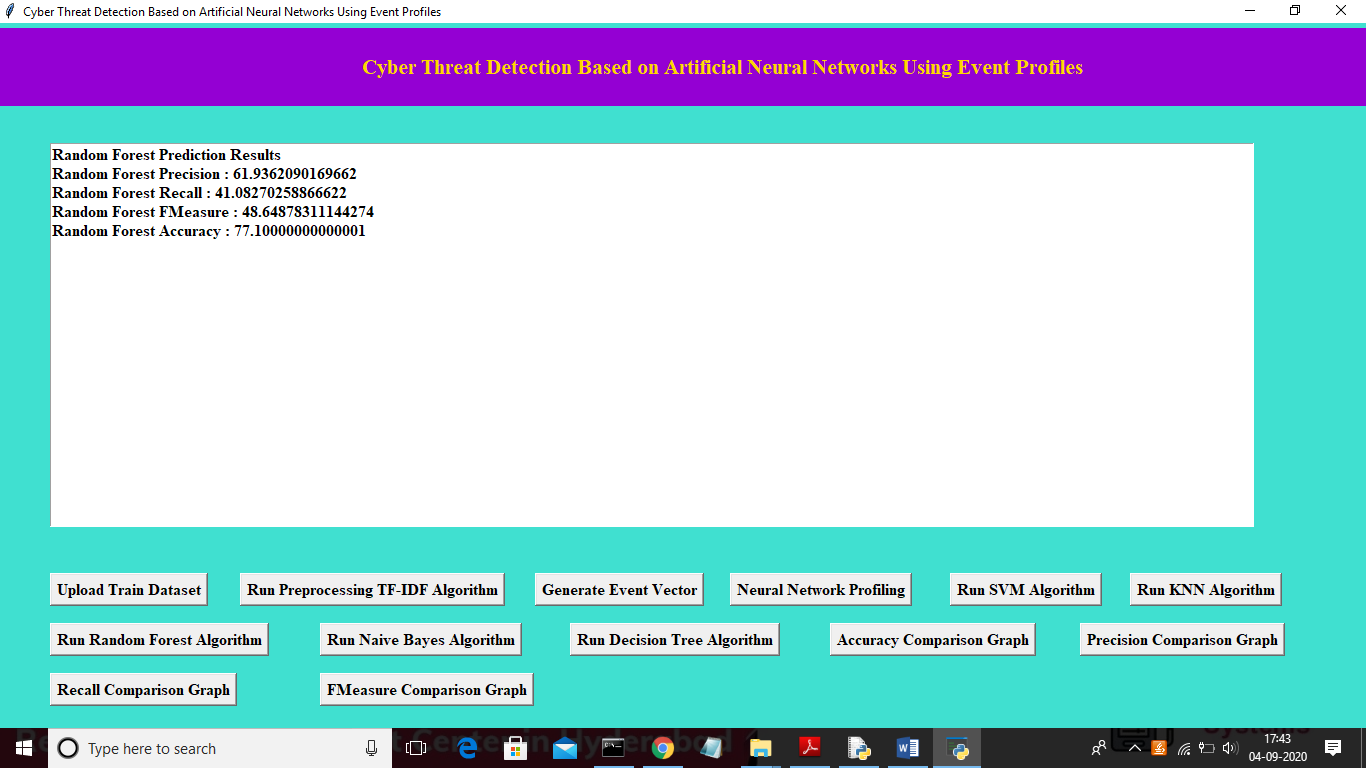
In above screen we can see both algorithms accuracy, precision, recall and FMeasure values. Now click on ‘Run SVM Algorithm’ button to run existing SVM algorithm



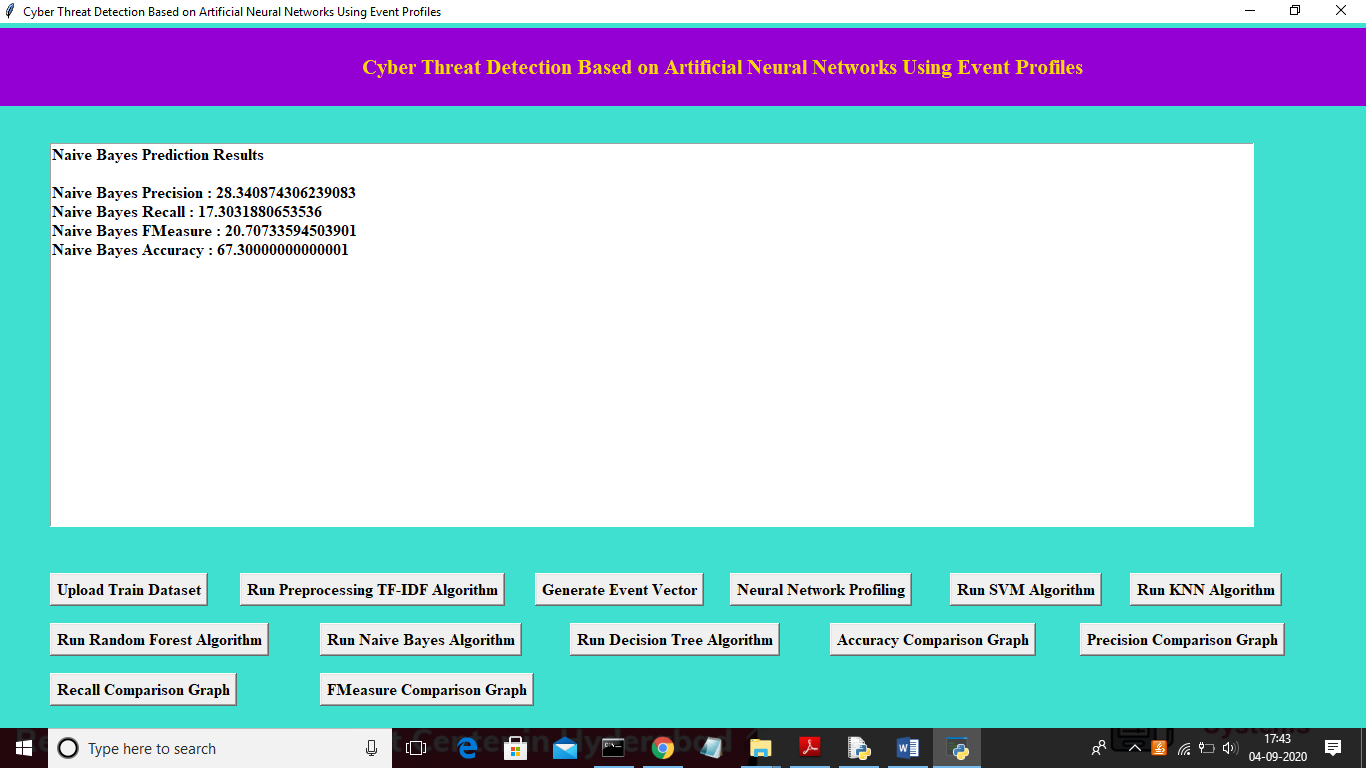
In above screen we can see SVM algorithm output values and now click on ‘Run KNN Algorithm’ to run KNN algorithm



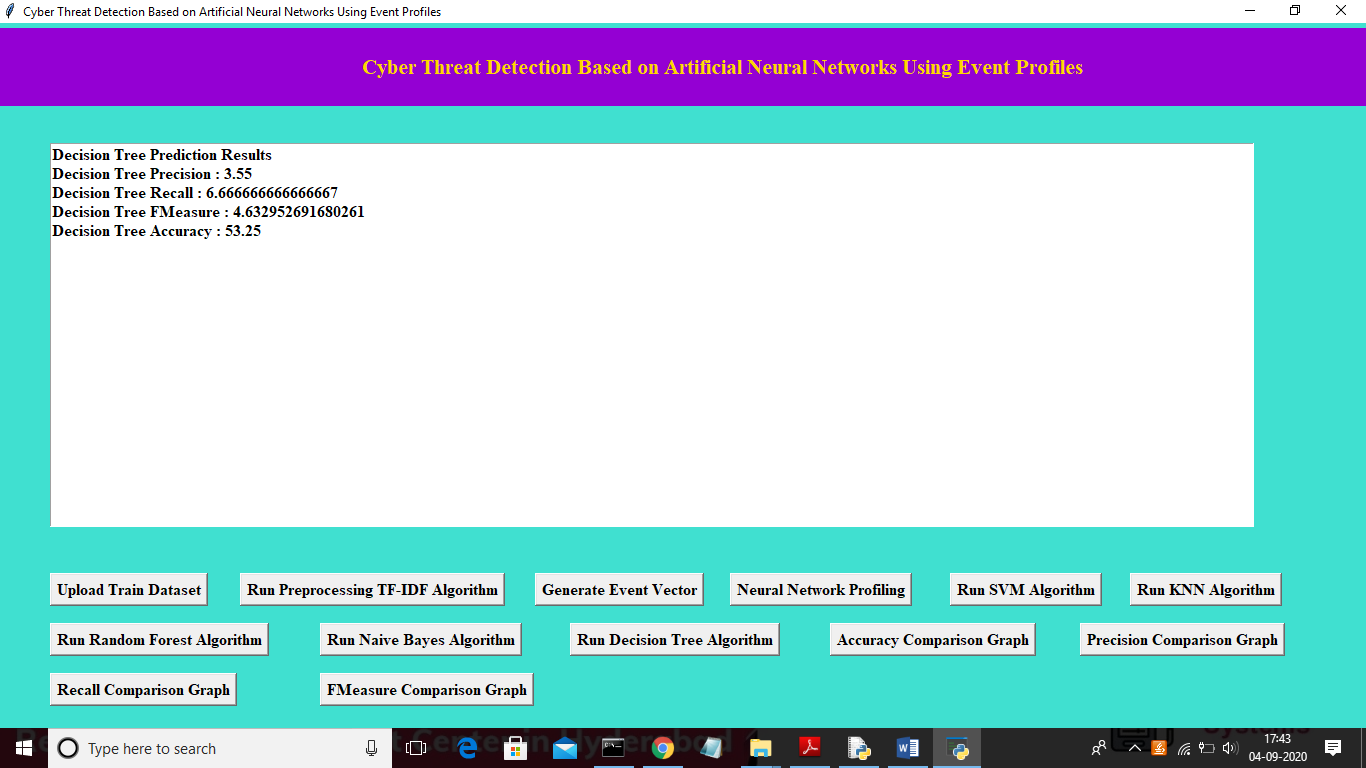
In above screen we can see KNN algorithm output values and now click on ‘Run Random Forest Algorithm’ to run Random Forest algorithm



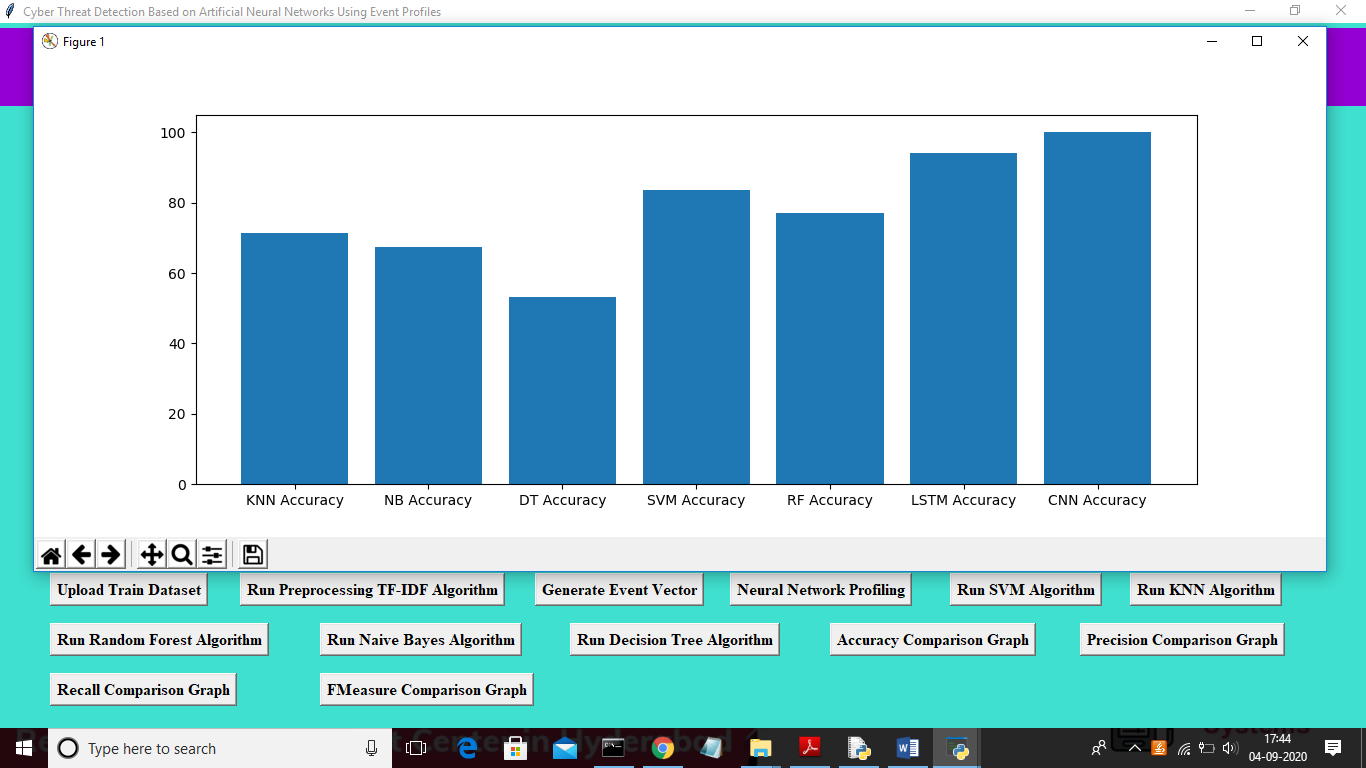
In above screen we can see Random Forest algorithm output values and now click on ‘Run Naïve Bayes Algorithm’ to run Naïve Bayes algorithm



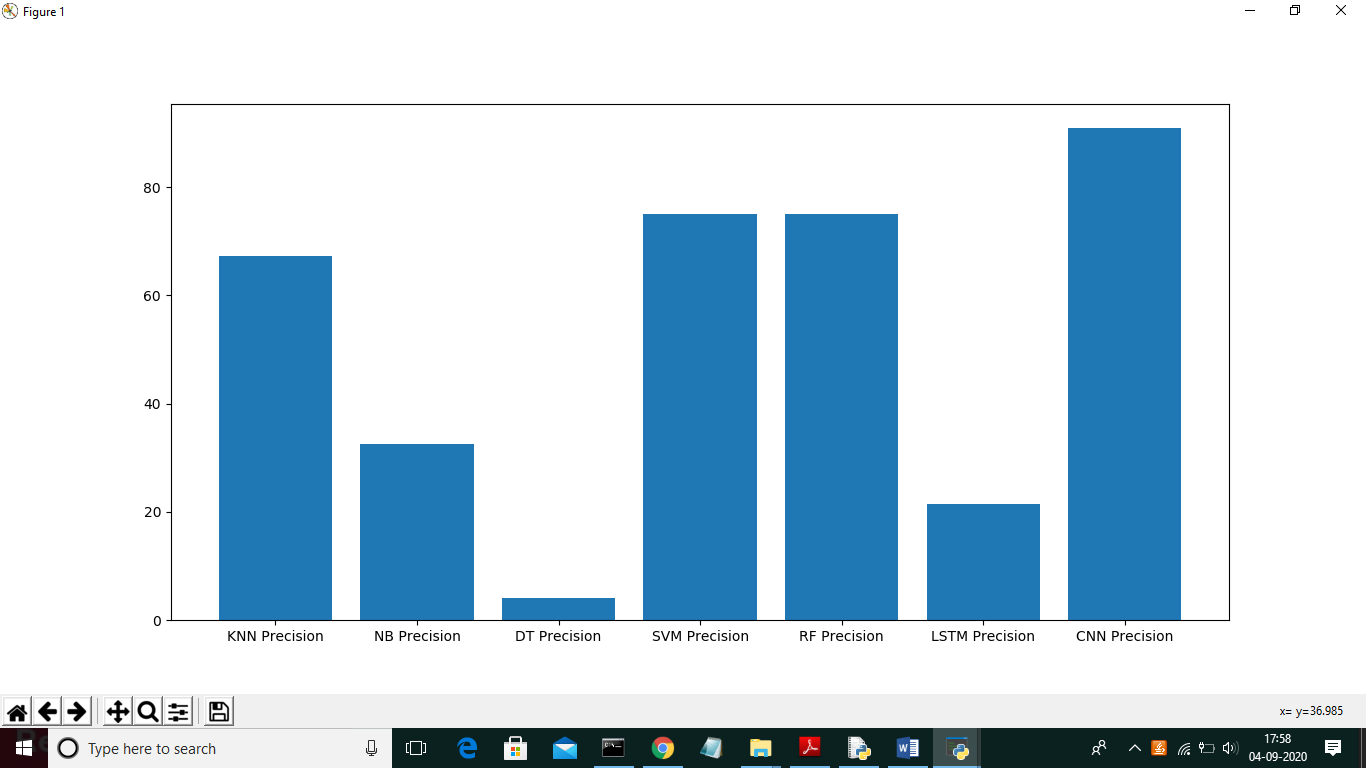
In above screen we can see Naïve Bayes algorithm output values and now click on ‘Run Decision Tree Algorithm’ to run Decision Tree Algorithm



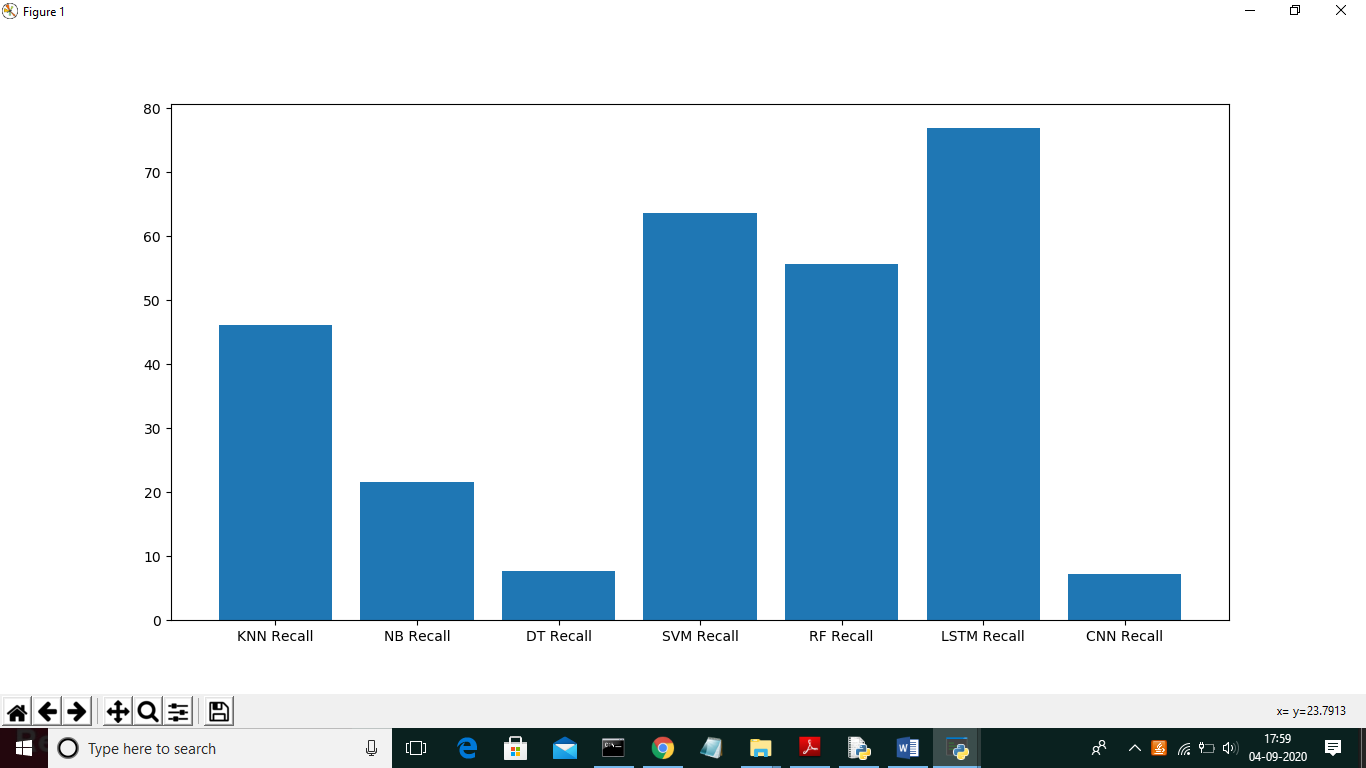
Now click on ‘Accuracy Comparison Graph’ button to get accuracy of all algorithms



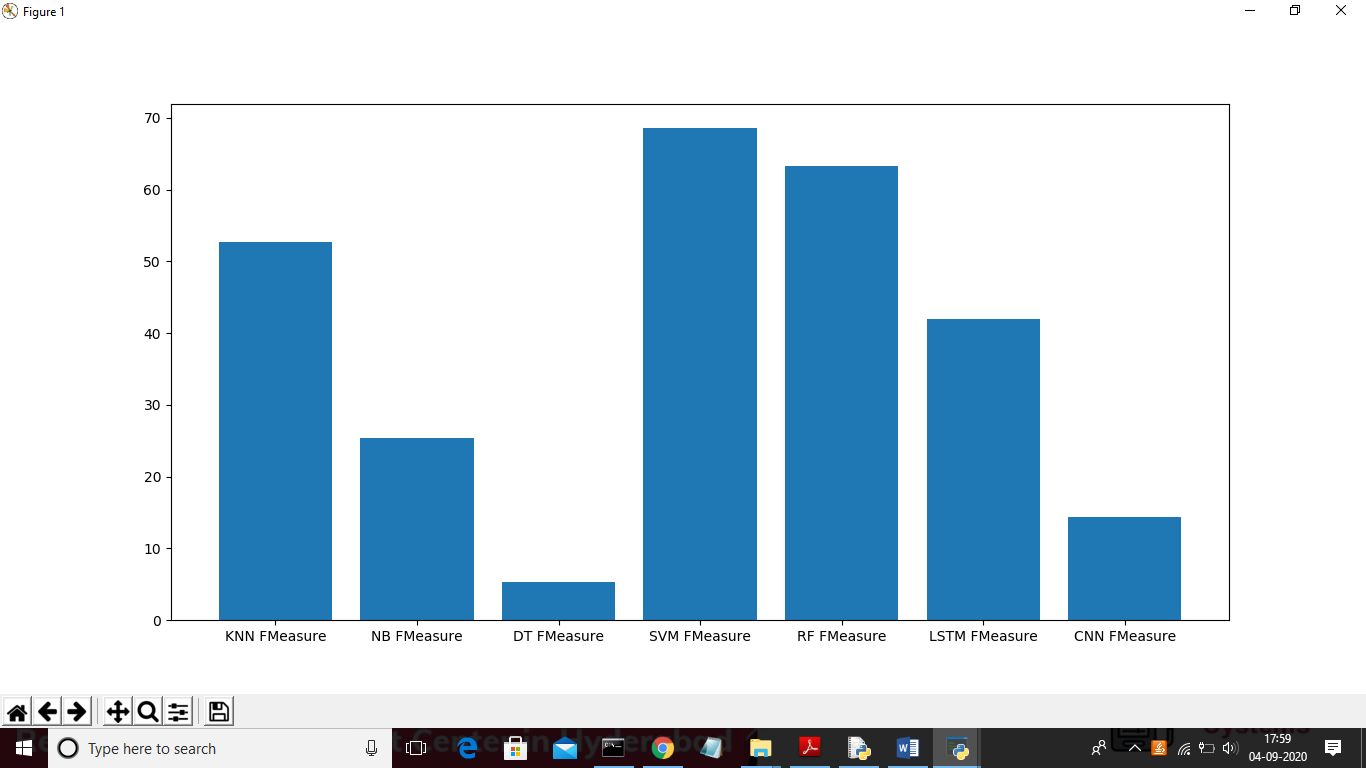
In above graph x-axis represents algorithm name and y-axis represents accuracy of those algorithms and from above graph we can conclude that LSTM and CNN perform well. Now click on Precision Comparison Graph’ to get below graph



In above graph CNN is performing well and now click on ‘Recall Comparison Graph’



In above graph LSTM is performing well and now click on FMeasure Comparison Graph button to get below graph



From all comparison graph we can see LSTM and CNN performing well with accuracy, recall and precision.