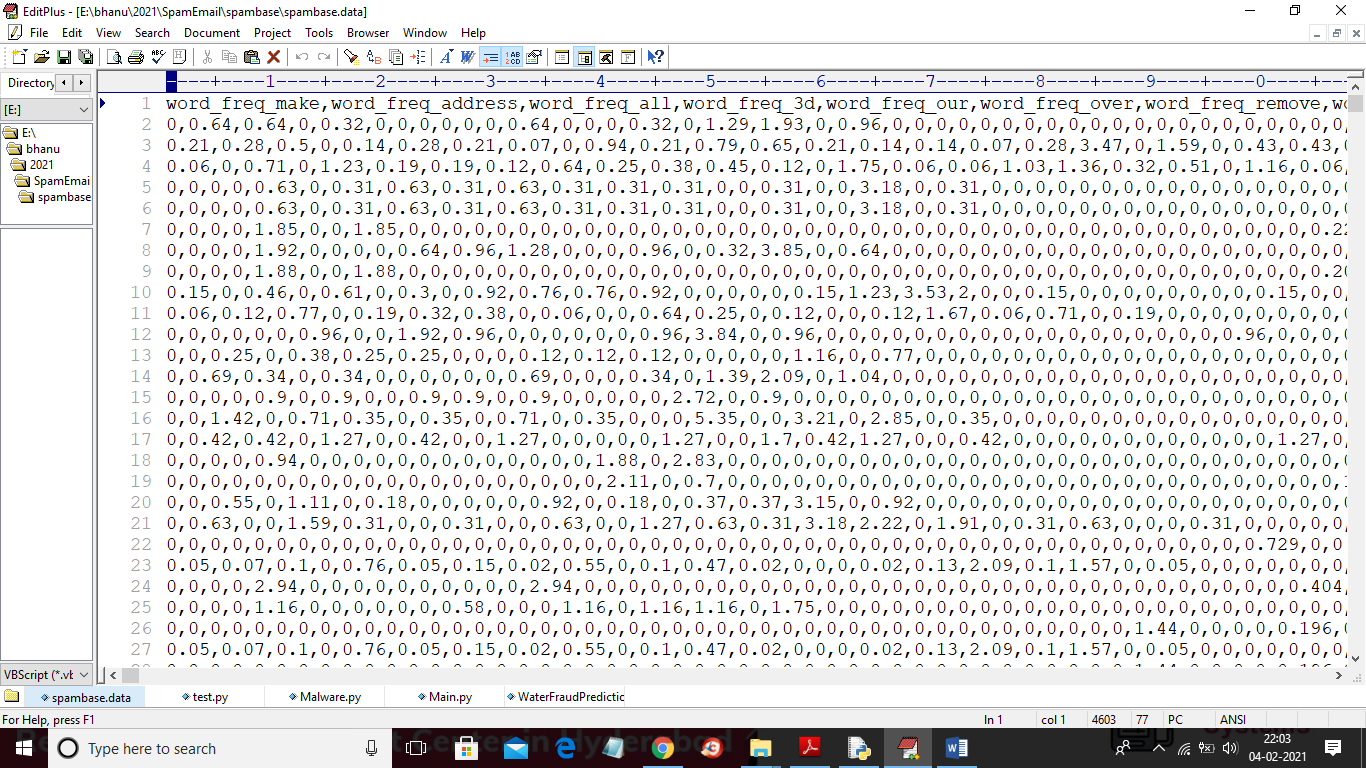
Machine Learning for Email Spam Filtering: Review, Approaches and Open Research Problems

In this paper author is giving brief review on various machine learning algorithms such as SVM, Random Forest, Decision Tree, CNN, KNN, MLP and many more to predict spam emails. Author has done experiments with above algorithms by using various SPAM datasets such as SPAM ARCHIVE, SPAMBASE, LINGSPAM, PU1 and many more but we are using SPAMBASE dataset to evaluate performance of above algorithms in terms of accuracy, precision and recall.

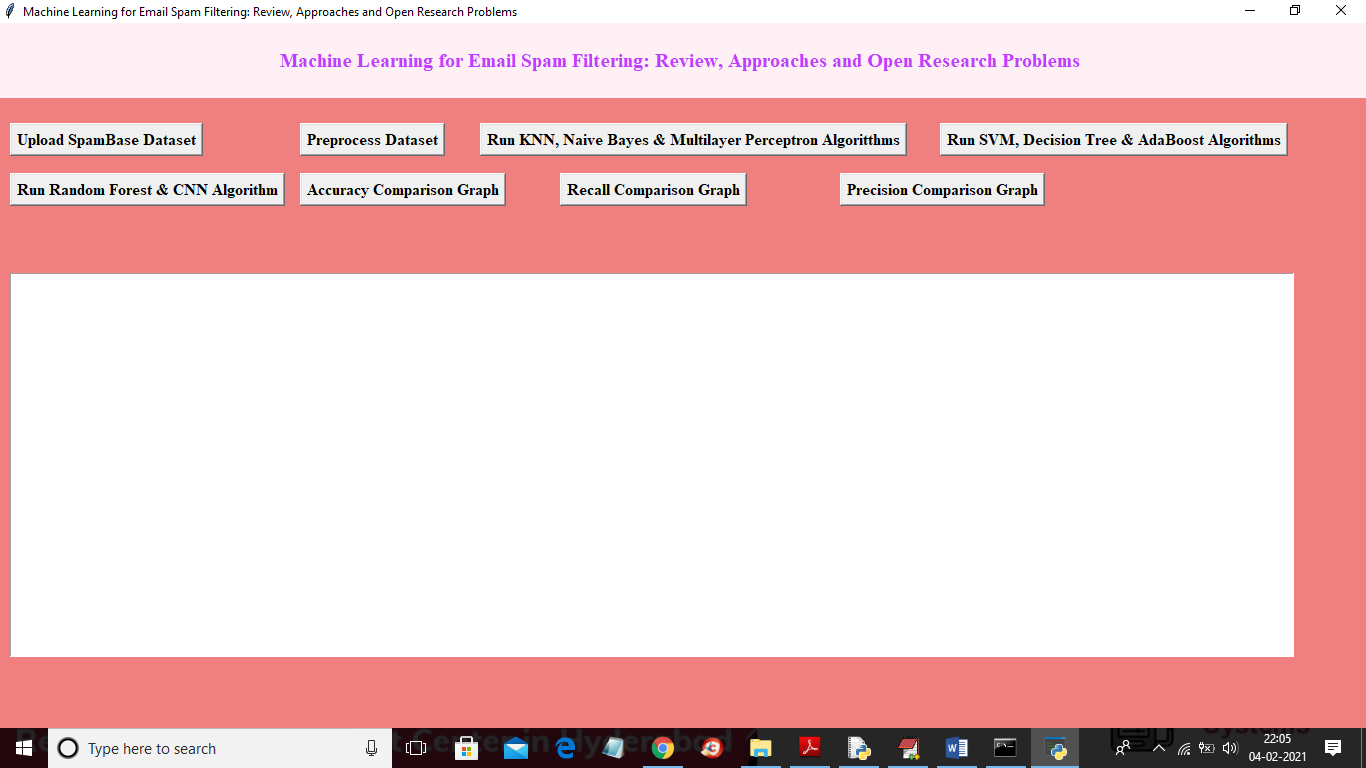
SPAMBASE dataset saved inside spambase folder and below screen shots showing details of dataset



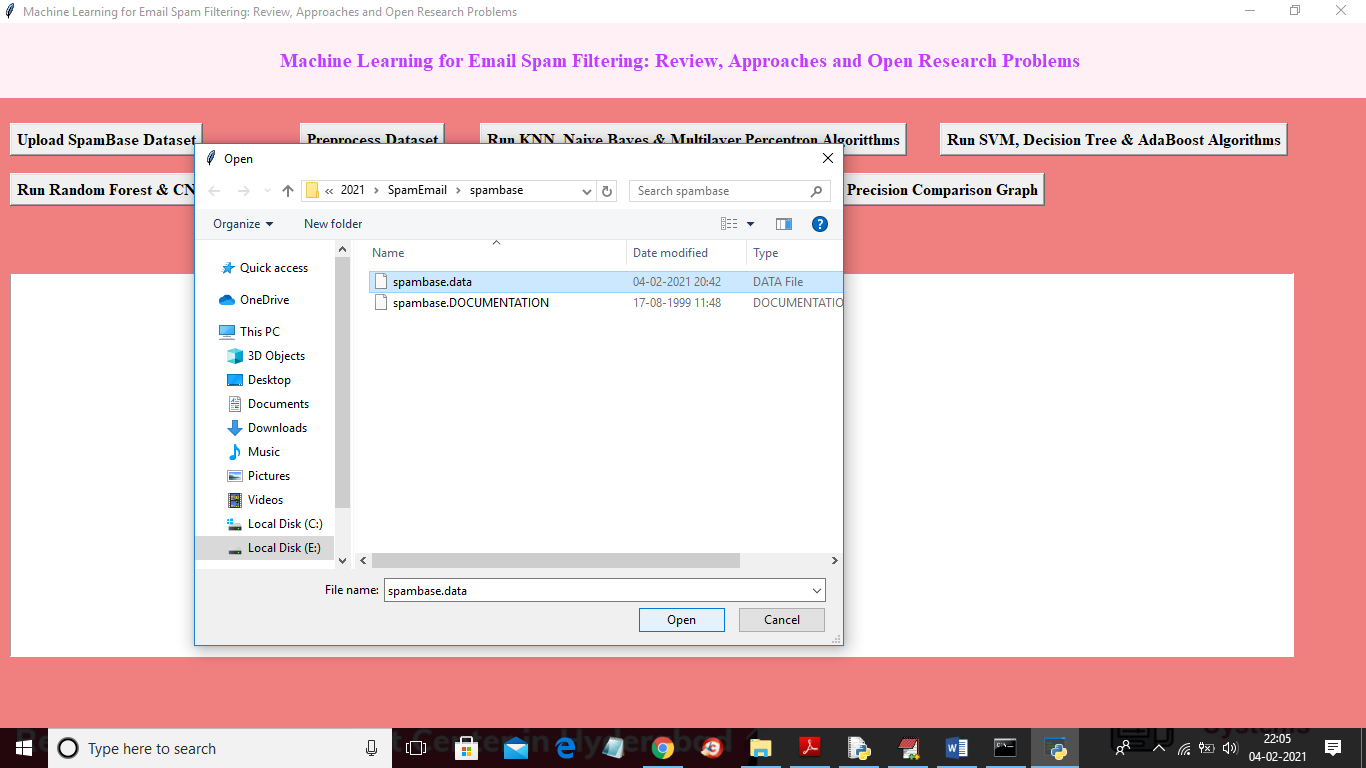
In above dataset first row contains column names as email word frequency and other rows contains probability of that word occur in email and in last column we have value as 0 or 1 where 0 indicate normal email and 1 indicate SPAM email. We will use above dataset to train all ML algorithms.

SCREEN SHOTS

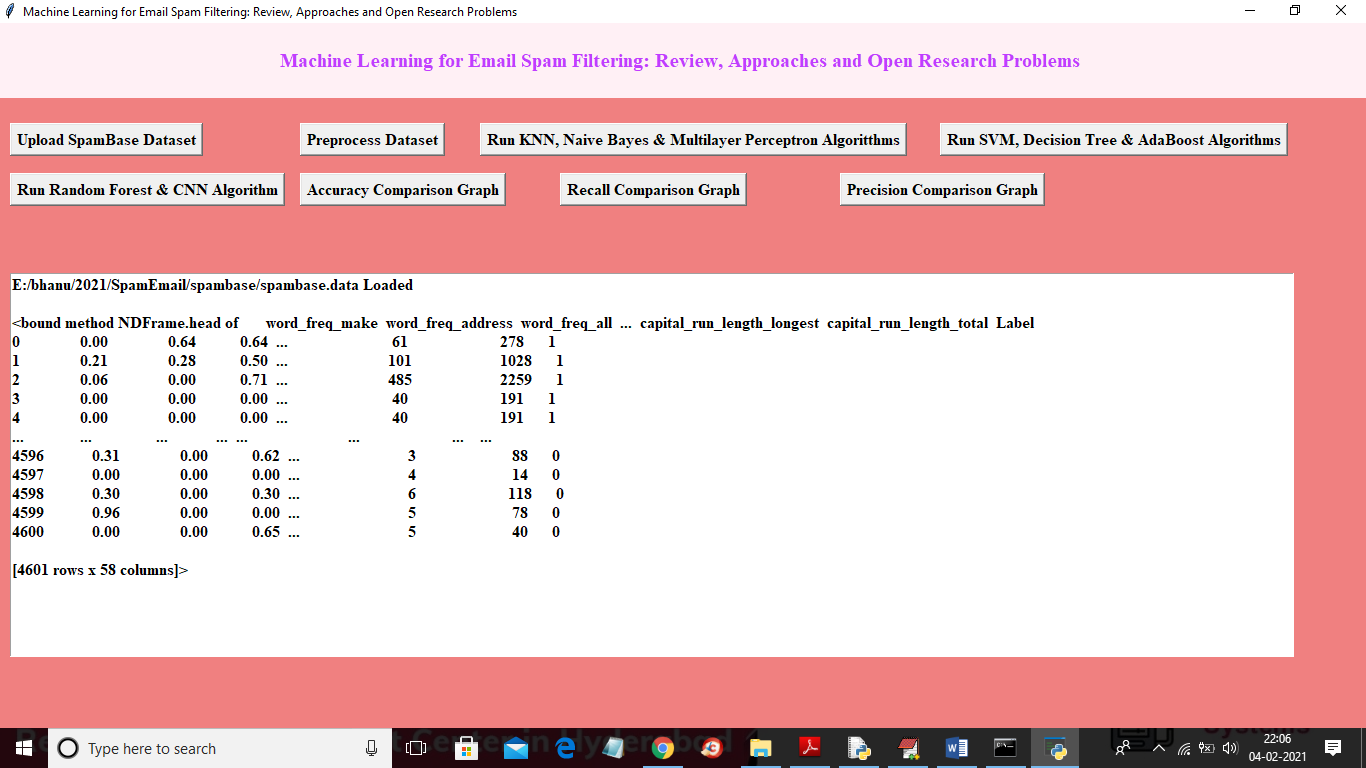
To run project double click on ‘run.bat’ file to get below screen



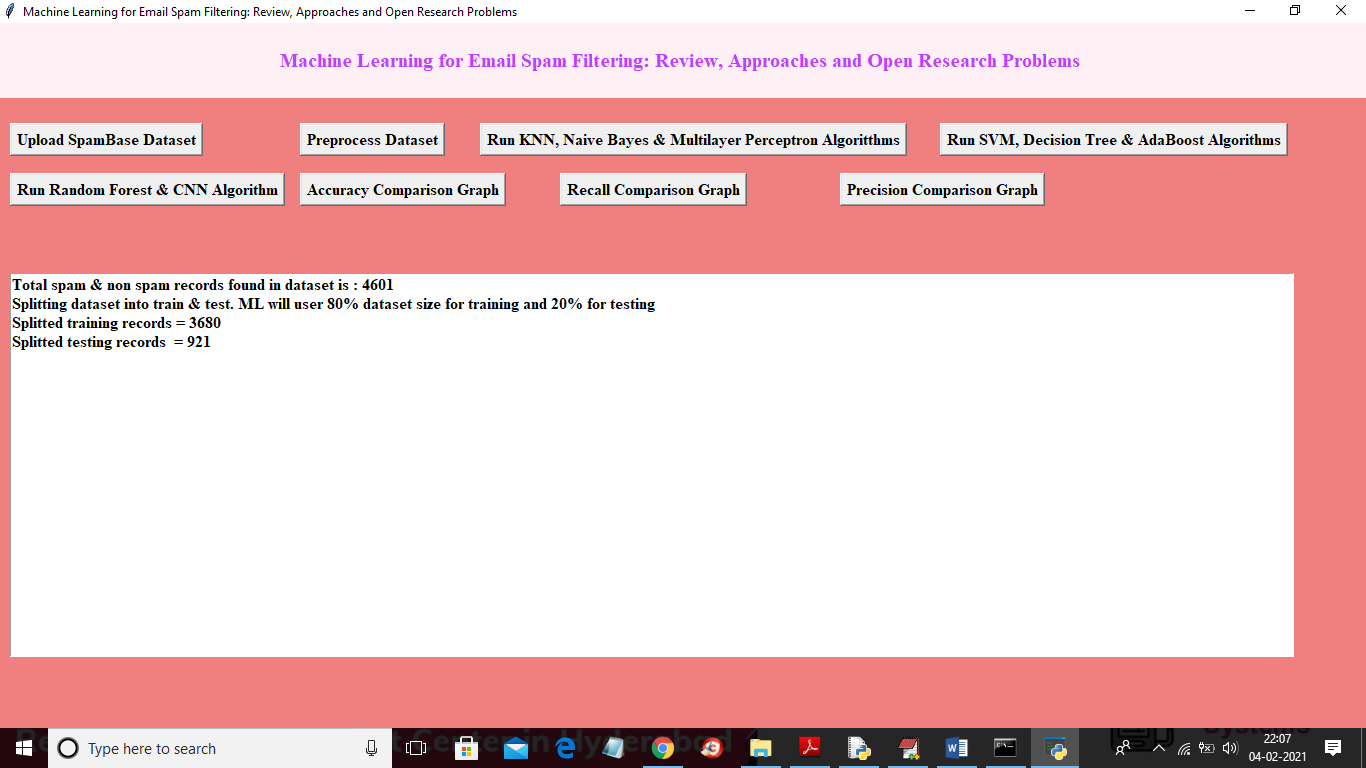
In above screen click on ‘Upload SpamBase Dataset’ and upload dataset



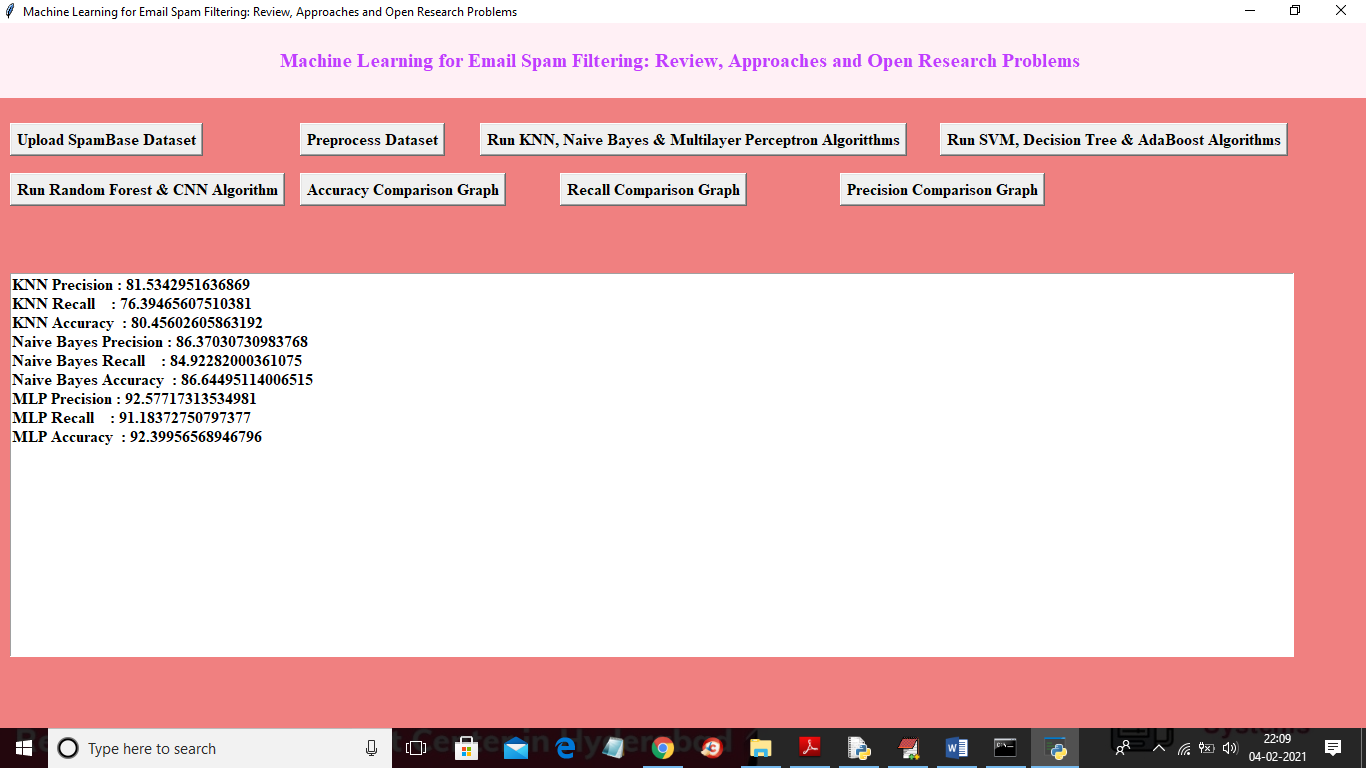
In above screen selecting and uploading ‘spambase.data’ dataset and then click on ‘Open’ button to load dataset and to get below screen



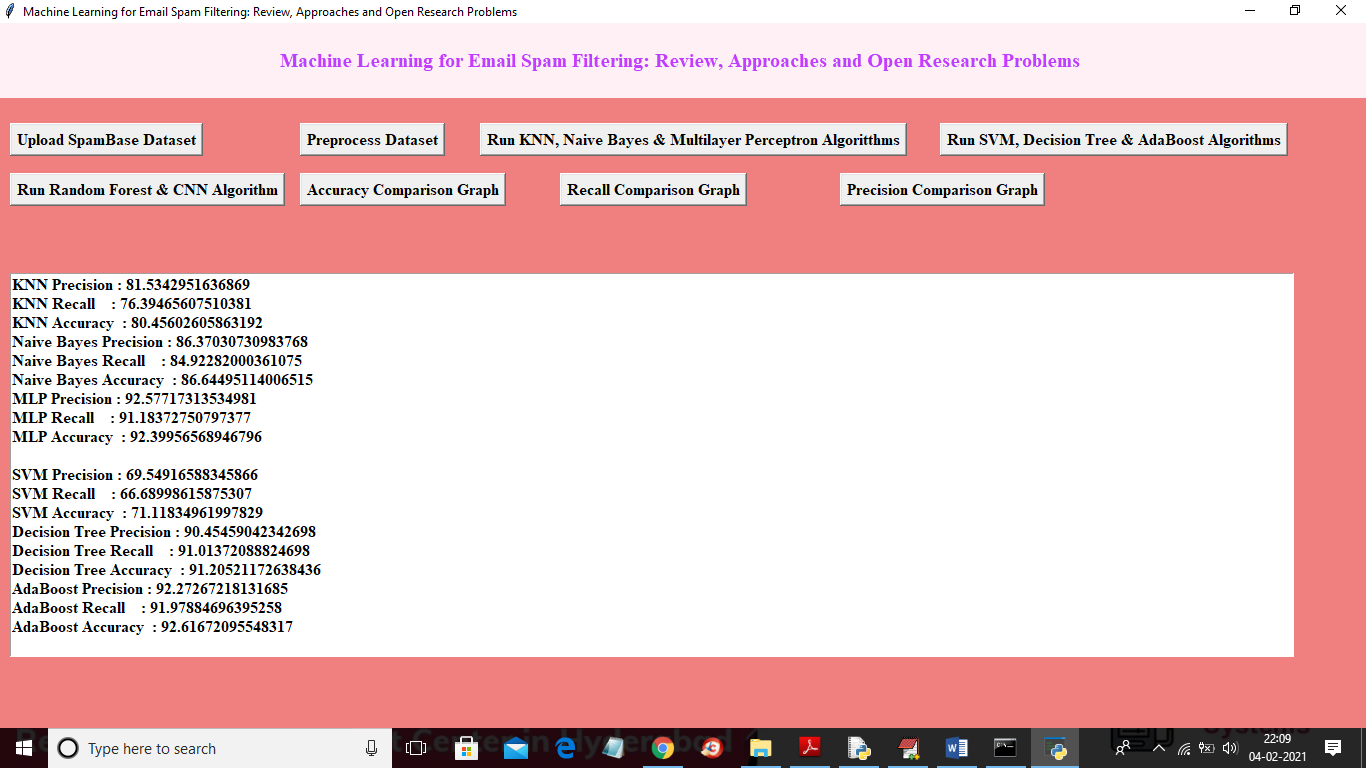
In above screen dataset loaded and we can see some records from the dataset and now click on ‘Preprocess Dataset’ button to read all values from dataset and then split data into train and test part where application used 80% dataset for training and 20% dataset for testing



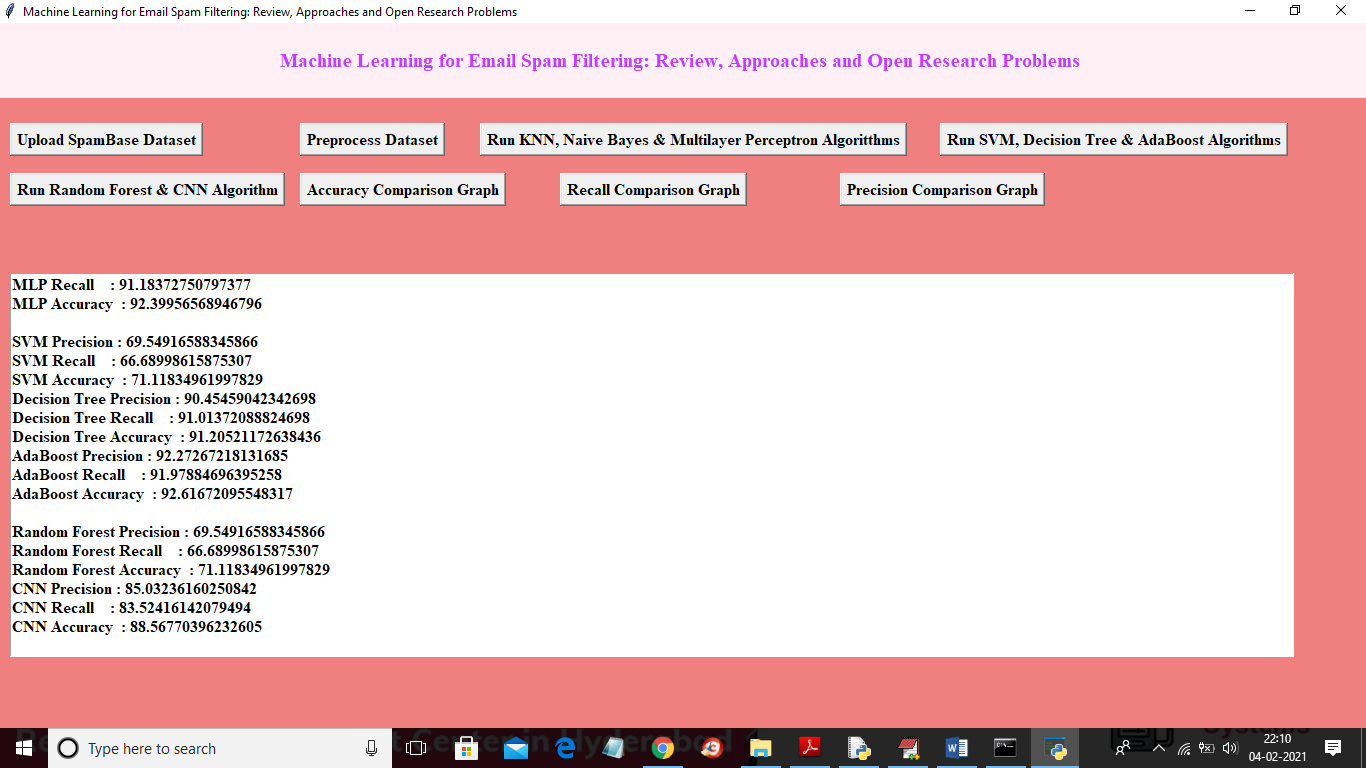
In above screen we can see dataset contains 4601 records and application using 3680 for training and 921 records for testing and now dataset is ready and now click on ‘Run KNN, Naive Bayes & Multilayer Perceptron Algorithms’ button to run all 3 algorithms and get there prediction metrics



In above screen we got evaluation metrics such as accuracy, recall and precision for all 3 algorithms and now click on ‘Run SVM, Decision Tree & AdaBoost Algorithms’ button to run this 3 algorithms also



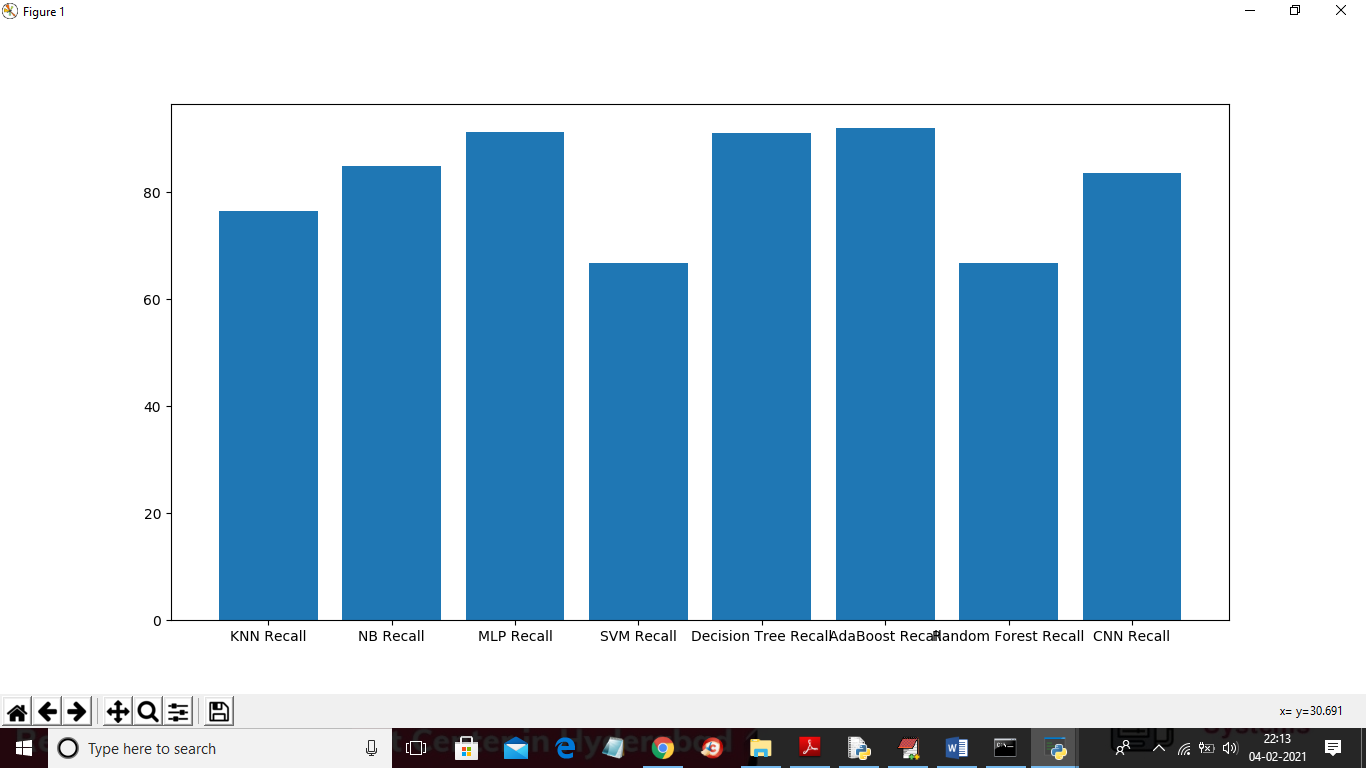
In above screen we got metrics for SVM, decision tree and AdaBoost algorithms and now click on ‘Run Random Forest & CNN Algorithm’ button to run both algorithms and then get below result



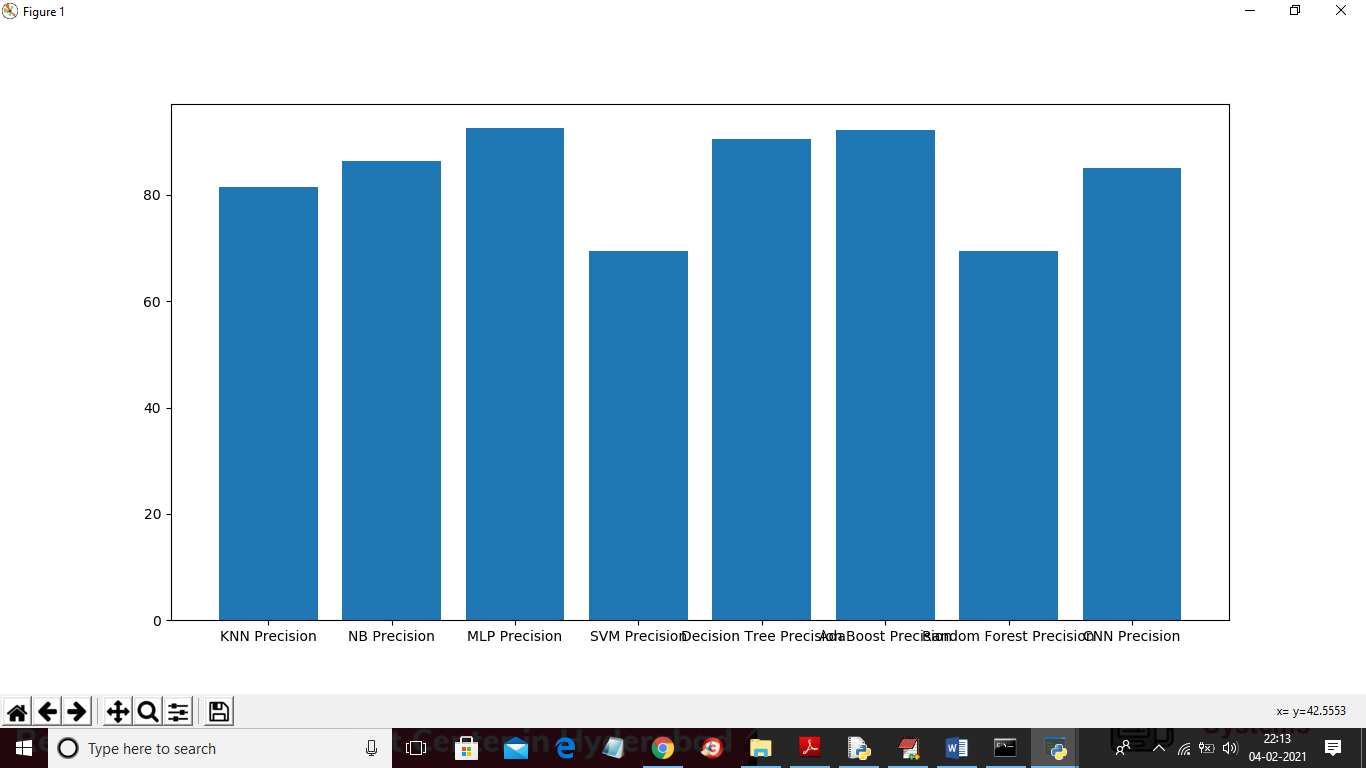
In above screen we got accuracy for CNN and random forest algorithms also and now click on ‘Accuracy Comparison Graph’ button to get below accuracy comparison between all algorithms



In above screen x-axis represents algorithm name and y-axis represents accuracy of all those algorithms and from above graph we can conclude that MLP neural network give better prediction accuracy compare to all other algorithm. Now click on ‘Recall Comparison Graph’ button to get below recall graph



Now click on ‘Precision Comparison Graph’ button to get below precision graph



In all above 3 graph MLP give better accuracy, precision and recall