CYBER SECURITY

PBL- 1

Harshil Gupta -17BCE1112

Shekhar Gaur – 17BCE1183

Shikhar Chaudhary – 17BCE1238

Mohika Thampi -17BCE1079

Pranjal Sachan -17BCE1027

Classification Algorithm Used:

- Decision Tree
- Gaussian Process
- Gradient Boosting

DECISION TREE:-

Tree based learning algorithms are considered to be one of the best and mostly used supervised learning methods. Tree based methods empower predictive models with high accuracy, stability and ease of interpretation.

A decision tree is a flowchart-like structure in which each internal node represents a “test” on an attribute and each branch represents the outcome of the test, and each leaf node represents a class label (decision taken after computing all attributes). The paths from root to leaf represent classification rules.

Dataset is divided into two sets of training and test dataset and each of the leaf node represent two classes to which test data belong to.

GAUSSIAN PROCESS:-

This classifier classifies new data into classes which has been made by training and test data initially. for a given new data(x), we want to estimate p (y = 1|x) and p (y = 2|x). X is assigned to any class which has the highest probability. There are two variants of Gaussian classifier, depending on whether covariance matrices of classes are assumed to be equal or not. Covariance matrix assumption has an impact on the class boundary. Shared covariance matrix leads to the linear boundary while separate covariance matrices lead to the quadratic boundary.

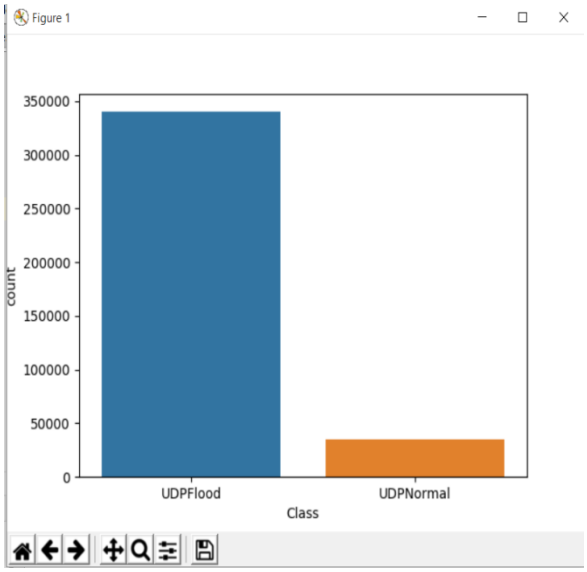
Our Dataset of different attacks has been classified in two various classes , where we predict for each attribute of data to belong in class 1 or class 2.

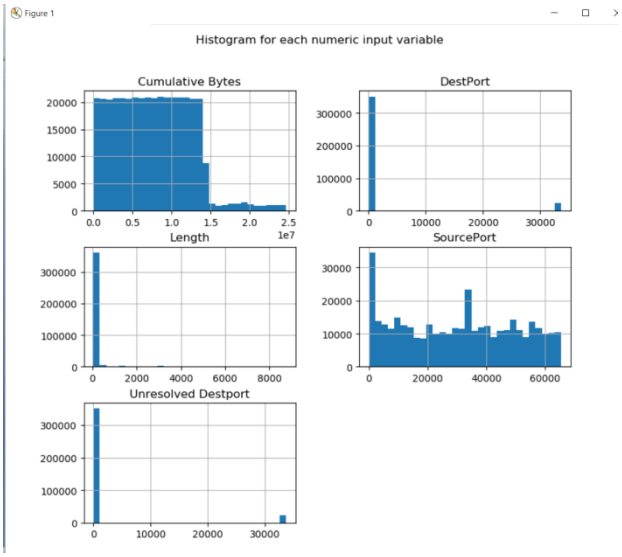
GRADIENT BOOSTING:-

Gradient boosting is a machine learning technique for regression and classification problems, which produces a prediction model in the form of an ensemble of weak prediction models, typically decision trees. It builds the model in a stage-wise fashion like other boosting methods do, and it generalizes them by allowing optimization of an arbitrary differentiable loss function.

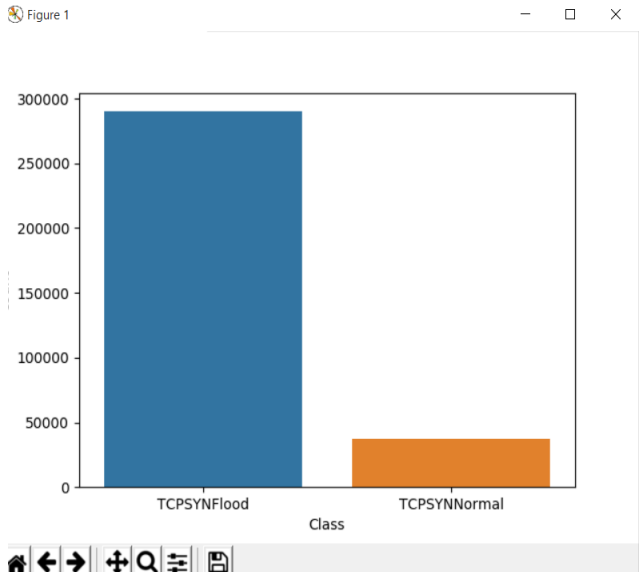
VISUALIZATION OF DATASET

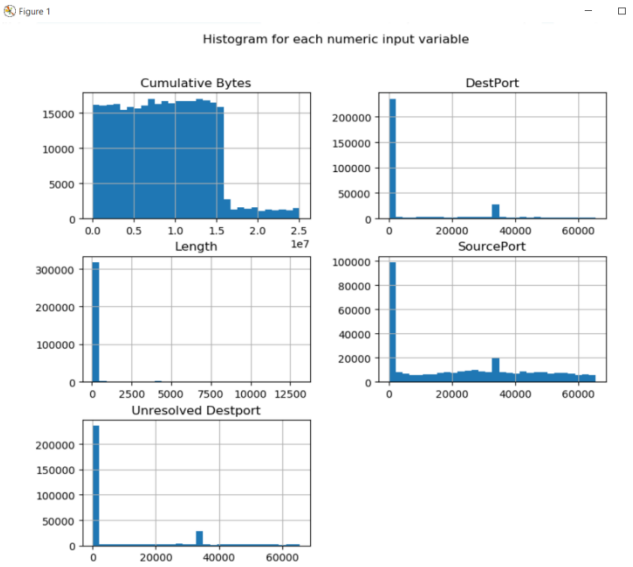
1.) UDP





2.) TCPSYN





3.) ICMP

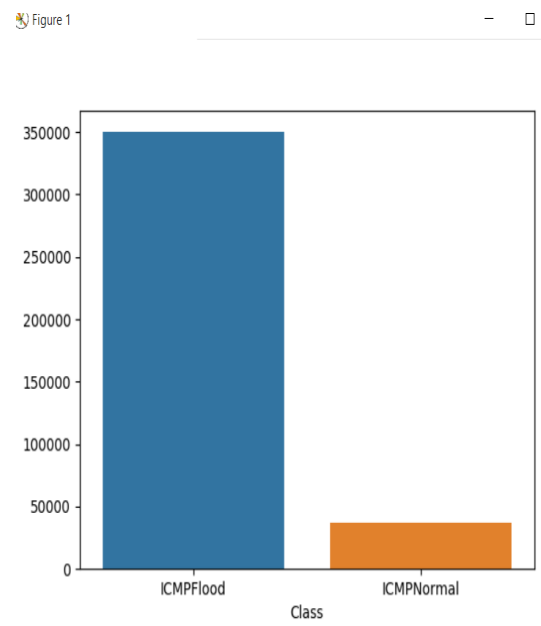


Figure 1
350000
300000
250000
200000
150000
100000
50000
0
ICMPFlood
ICMPNormal
Class

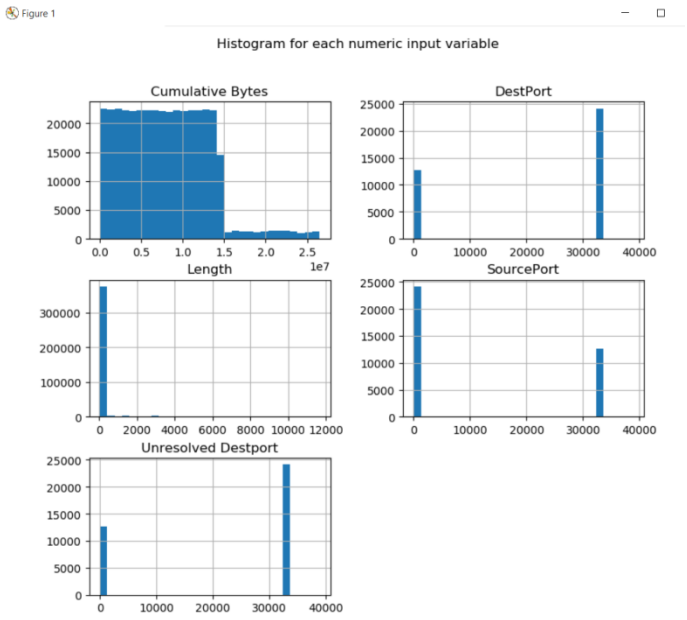


Figure 1
Histogram for each numeric input variable
Cumulative Bytes
DestPort
Length
SourcePort
Unresolved Destport
1e7

4.) LAND

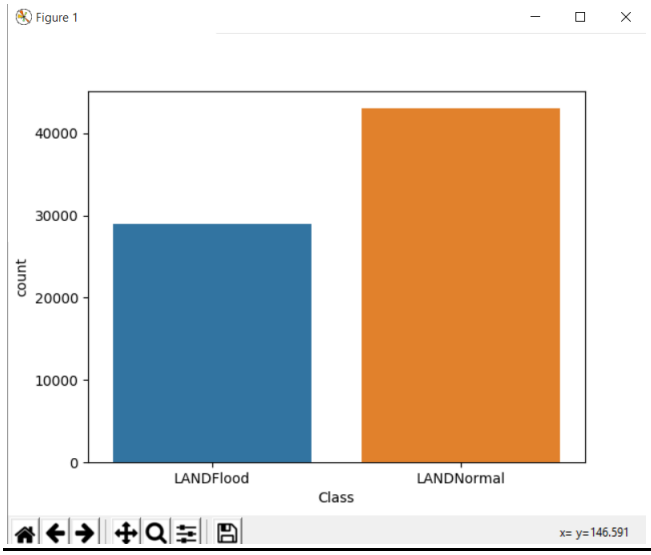
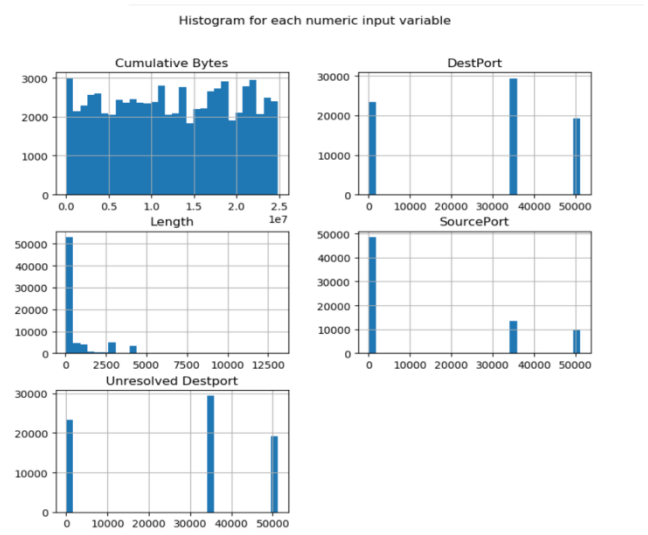
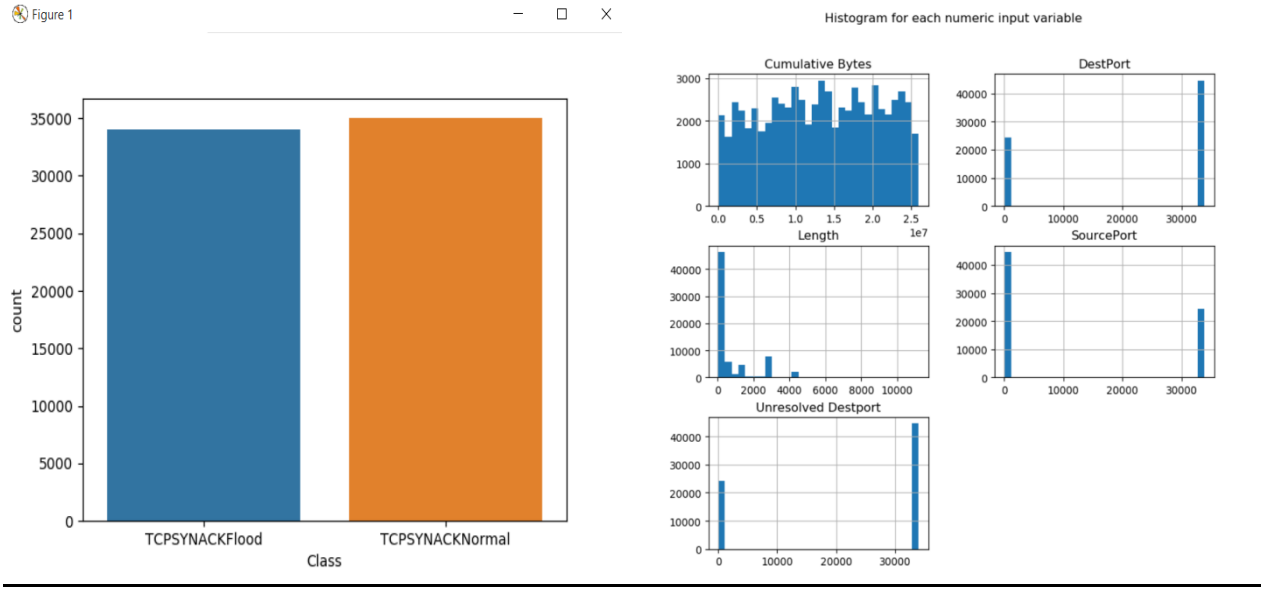


Figure 1
40000
30000
20000
10000
0
count
LANDFlood
LANDNormal
Class
x= y=146.591


Histogram for each numeric input variable
Cumulative Bytes
DestPort
Length
SourcePort
Unresolved Destport
1e7

5.) TCPSYNACK



By Varying Different Parameters for Classifiers

1.)Decision Tree

Attack	Accuracy
UDP	0.9952
TCPSYN	0.999311
ICMP	0.999418
LAND	0.834027
TCPSYNACK	0.772717

2.)Gradient Boosting

Varying Parameter: Learning Rate

Attack	Learning Rate	Accuracy
UDP	0.2	0.9997266666666667
	0.5	0.9997266666666667
	1.0	0.9997266666666667
	1.5	0.97086
	2.0	0.9707666666666667

Attack	Learning Rate	Accuracy
TCPSYN	0.2	0.9996253822629969
	0.5	0.9996253822629969
	1.0	0.9996253822629969
	1.5	0.9820336391437309
	2.0	0.8861238532110092

Attack	Learning Rate	Accuracy
ICMP	0.2	0.9996640826873385
	0.5	0.999657622739018
	1.0	0.9996770025839793
	1.5	0.9995025839793281
	2.0	0.9993023255813953

Attack	Learning Rate	Accuracy
LAND	0.2	0.6538541666666666
	0.5	0.6625694444444444
	1.0	0.6758333333333333
	1.5	0.6765277777777777
	2.0	0.5698958333333334

Attack	Learning Rate	Accuracy
TCPSYNACK	0.2	0.5839855072463768
	0.5	0.5942391304347826
	1.0	0.6068115942028985
	1.5	0.608804347826087
	2.0	0.5204710144927536

3.)Gaussian Process

Varying Parameter: Kernel

Attack	Kernel	Accuracy
UDP	0.5	0.9158466666666667
	1.0	0.9158466666666667
	1.5	0.9158466666666667
	2.0	0.9158466666666667

Attack	Kernel	Accuracy
TCPSYN	0.5	0.970565749235474
	1.0	0.9501376146788991
	1.5	0.9501376146788991
	2.0	0.9501376146788991

Attack	Kernel	Accuracy
ICMP	0.5	0.9246511627906977
	1.0	0.9436692506459948
	1.5	0.9516020671834625
	2.0	0.9516020671834625

Attack	Kernel	Accuracy
LAND	0.5	0.553125
	1.0	0.553125
	1.5	0.553125
	2.0	0.553125

Attack	Kernel	Accuracy
TCPSYNACK	0.5	0.5244565217391305
	1.0	0.5115217391304347
	1.5	0.4914855072463768
	2.0	0.4914855072463768

VALIDATION METRICES-:

1.) UDP

For Decision Tree:

TP:149928
FN:72
FP:37
TN:135779
Recall :0.99952
Precision :0.999753275764345
F-measure :0.9996366242728318

For Gradient Boosting:

TP:295543
FN:4457
FP:39
TN:271593
Recall :0.9851433333333334
Precision :0.9998680569182156
F-measure :0.9924510814631738

For Gaussian:

TP:432920
FN:17080
FP:39
TN:407409
Recall :0.9620444444444445
Precision :0.9999099221866273
F-measure :0.9806117837861101

2.) TCPSYN

For Decision Tree:

TP:130710
FN:90
FP:50
TN:115829
Recall :0.9993119266055046
Precision :0.9996176200672988
F-measure :0.9994647499617678

For Gradient Boosting:

TP:246615
FN:14985
FP:50
TN:231708
Recall :0.9427178899082569
Precision :0.9997972959276752
F-measure :0.9704189743539294

For Gaussian:

TP:370893
FN:21507
FP:51
TN:347586
Recall :0.9451911314984709
Precision :0.9998625129399586
F-measure :0.9717584732440421

3.) ICMP

For Decision Tree:

TP:154710
FN:90
FP:43
TN:139898
Recall :0.9994186046511628
Precision :0.9997221378583937
F-measure :0.9995703482117763

For Gradient Boosting:

TP:309402
FN:198
FP:104
TN:279778
Recall :0.9993604651162791
Precision :0.9996639806659645
F-measure :0.9995121998494604

For Guassian:

TP:456710

FN:7690
FP:104
TN:419719
Recall :0.9834409991386736
Precision :0.9997723362243714
F-measure :0.9915394251498567

4.) LAND

For Decision Tree:

TP:24015
FN:4785
FP:2244
TN:9256
Recall :0.8338541666666667
Precision :0.9145435850565521
F-measure :0.8723369476379884

For Gradient Boosting:

TP:40428
FN:17172
FP:8992
TN:14008
Recall :0.701875
Precision :0.8180493727235937
F-measure :0.7555223322743412

For Gaussian:

TP:56358
FN:30042
FP:16416
TN:18084
Recall :0.6522916666666667
Precision :0.7744249319812021
F-measure :0.7081307248671266

5.) TCPSYNACK

For Decision Tree:

TP:21310
FN:6290
FP:2978
TN:10587
Recall :0.7721014492753623
Precision :0.8773880105401844
F-measure :0.8213845205057045

For Gradient Boosting:

TP:35675
FN:19525
FP:4839
TN:22291
Recall :0.646286231884058

Precision :0.8805598064866466
F-measure :0.7454499864178699

For Gaussian:

TP:49240
FN:33560
FP:4839
TN:35856
Recall :0.5946859903381643
Precision :0.9105197951145546
F-measure :0.7194675589389168