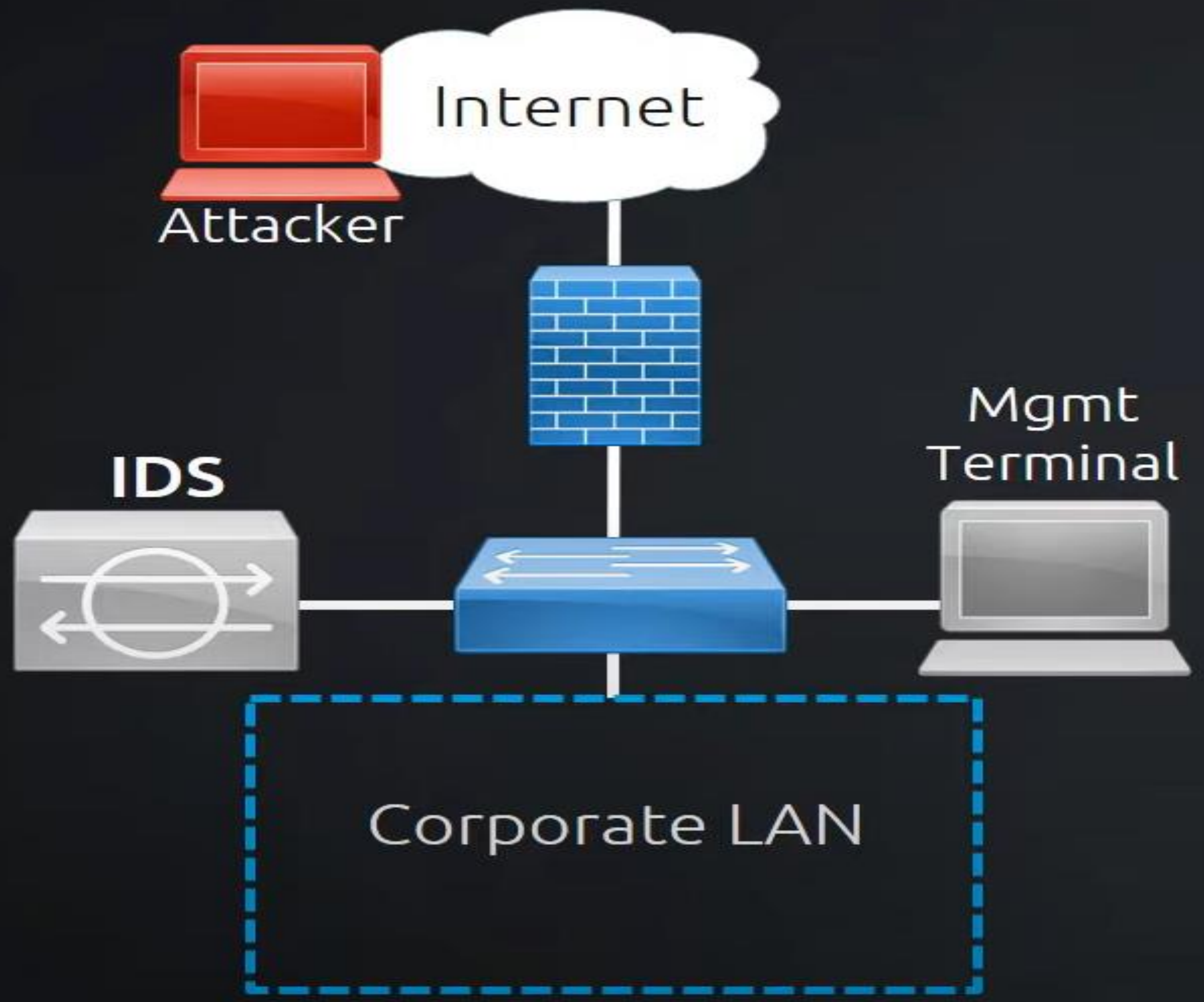
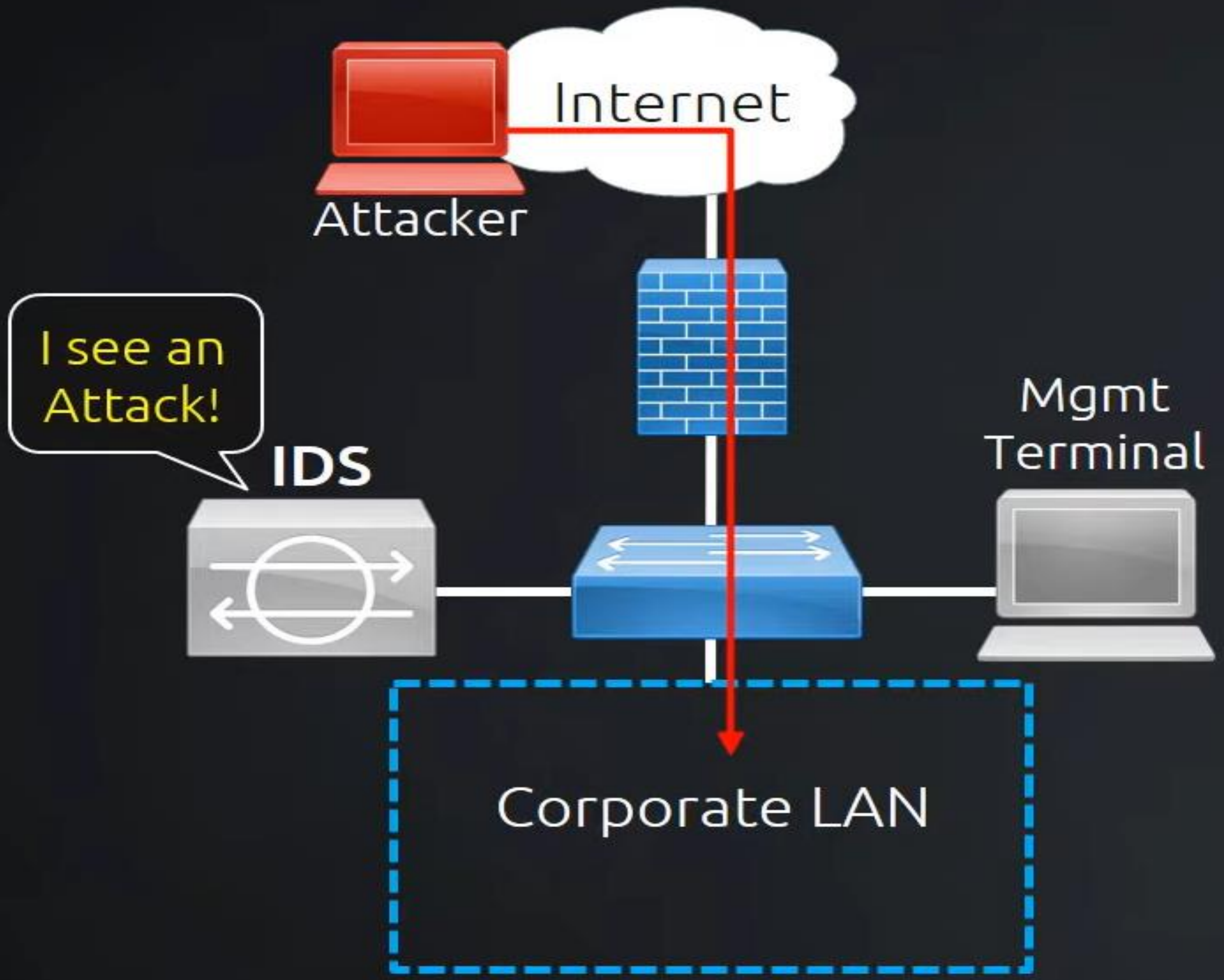
NETWORK INTRUSION DETECTION SYSTEM

Using UNSW-NB15 Dataset

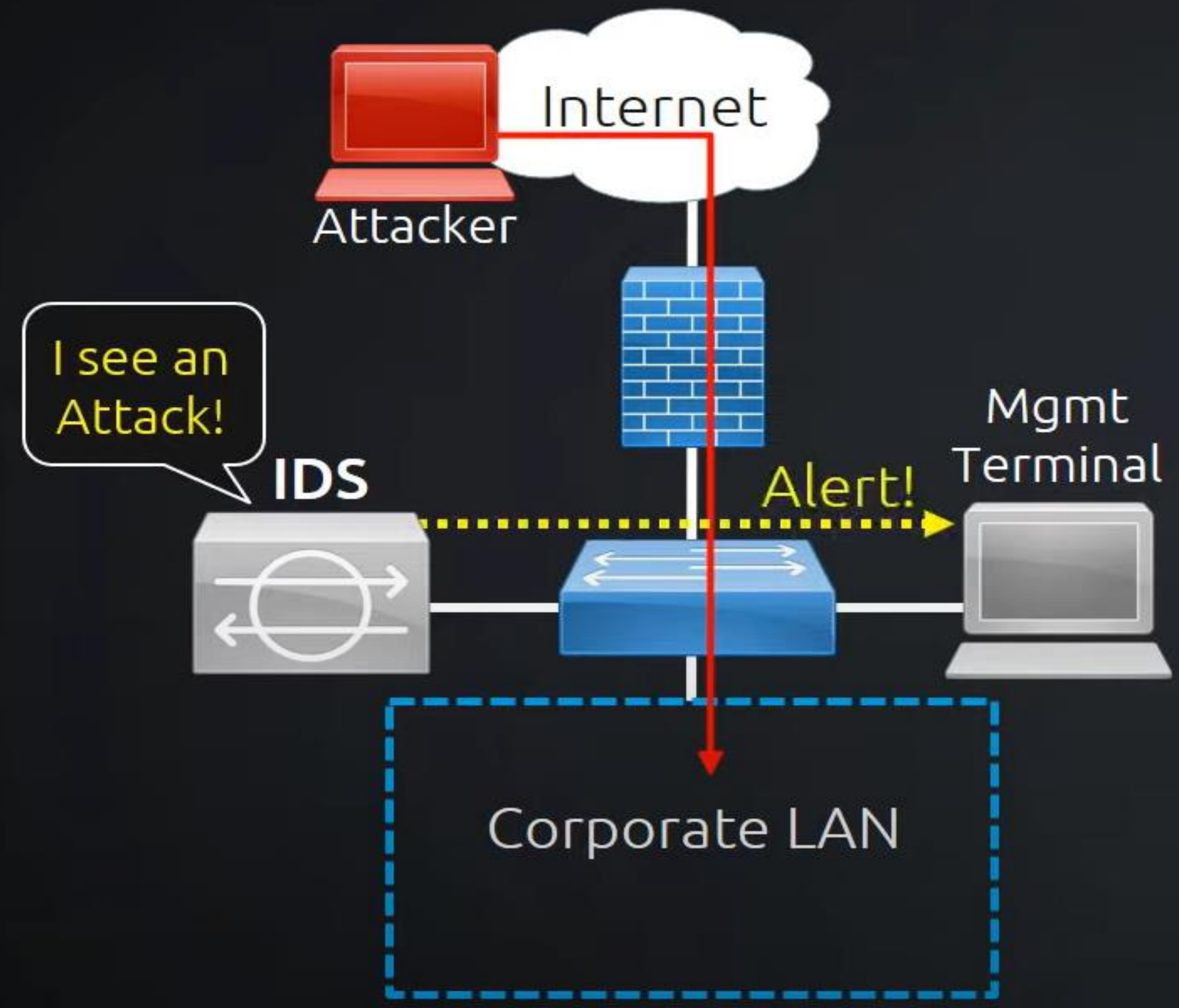
Intrusion Detection System



Intrusion Detection System



Intrusion Detection System



Why IDS?

STRAIGHT FORWARD REASON

- To Protect Data and System Integrity

FACT

- Cannot be done with ordinary password and file security

MISCONCEPTION

- A network firewall will keep the bad guys off my network, right?
- My anti-virus will recognize and get rid of any virus I might catch, right?
- And my password-protected access control will stop the office cleaner trawling through my network after I've gone home, right?

So that’s it – “I am Fully Protected”

HERE IS THE REALITY

1 Anti-virus systems are only good at detecting viruses they already know about.

2 Passwords can be hacked or stolen or changed by other

3 Firewalls DO NOT recognize attacks and block them

4 Simply a fence around your network

- No capacity to detect someone is trying to break-in(digging a hole underneath it)
- Can't determine whether somebody coming through gate is allowed to enter or not.
- Roughly 80% of financial losses occur hacking from inside the network

"I In April 1999, many sites were hacked via a bug in ColdFusion. All had firewalls to block other access except port 80. But it was the Web Server that was hacked. "

UNSW-NB15 Dataset

We operate on the UNSW-NB15 dataset that is currently one of the best representatives of modern attacks.

FEATURE	DESCRIPTION
Scrip, Dstip, Proto, State, Service,….	Corresponds to categorical variables which determines details on the form of the source to destination traversal of packets.
Sport, Dsport, ct_srv_src, sttl, dttl, ….	Correspond to discrete integer values representing information on sub-details of protocol implementation
Dur, Sload, Dload, Tcprtt,…	Represent continuous values representing temporal details surrounding packet flows/connections.
Stime, Ltime, Label,…	Binary and Timestamped Features

TABLE : SOME EXAMPLES OF VARIOUS FEATURES OF DATASET

APPROACH

- A reasonably good Network Intrusion Detection System generally requires a high detection rate and a low false alarm rate in order to predict anomalies more accurately.
- Older datasets are unable to capture the schema of a set of modern attacks and therefore modelling based on these datasets lacked sufficient generalizability.
- We operate on the UNSW-NB15 dataset that is currently one of the best representatives of modern attacks and thereby suggest various kinds of models.
- We discuss various models and conclude our discussion with the model that performs the best using various kinds of evaluation metrics.
- Alongside modelling, a comprehensive data analysis on the features of the dataset itself using our understanding of correlation, variance and similar such factors for a wider picture is done for better modelling.

EXTREME GRADIENT BOOSTED TREE

An Extreme Gradient Boosted Tree is used with 350 estimators

DECISION TREE

A Decision Tree with a maximum depth of 9 levels was used.

RANDOM FOREST CLASSIFIER

A Random Forest Classifier with no maximum depth was implemented along 300 estimators

ADABOOST CLASSIFIER

An AdaBoost Classifier was used with its base estimator as a Random Forest of 3 estimators

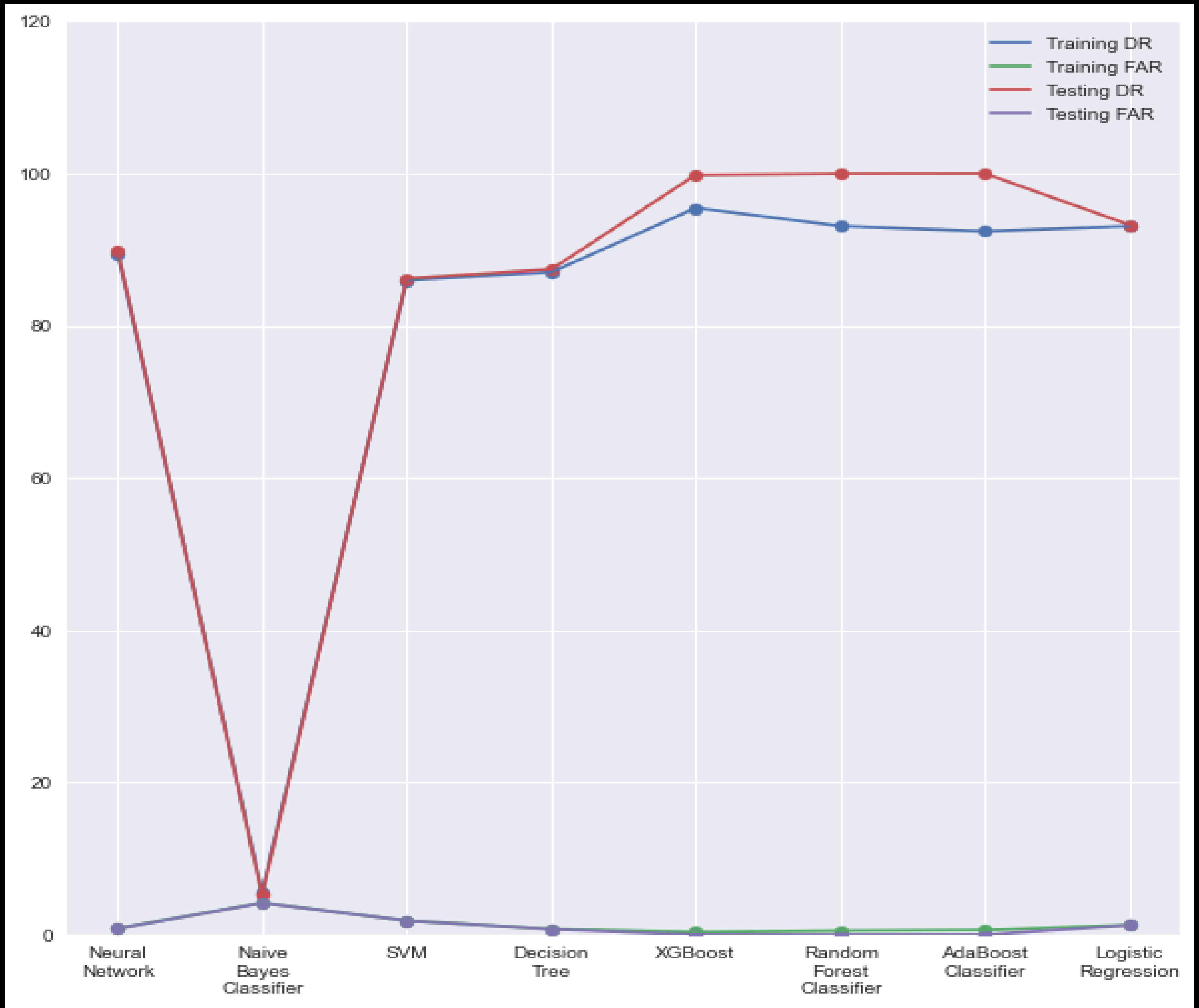
LOGISTIC REGRESSION

A logistic regression model was fitted using the non-linear conjugate gradient method

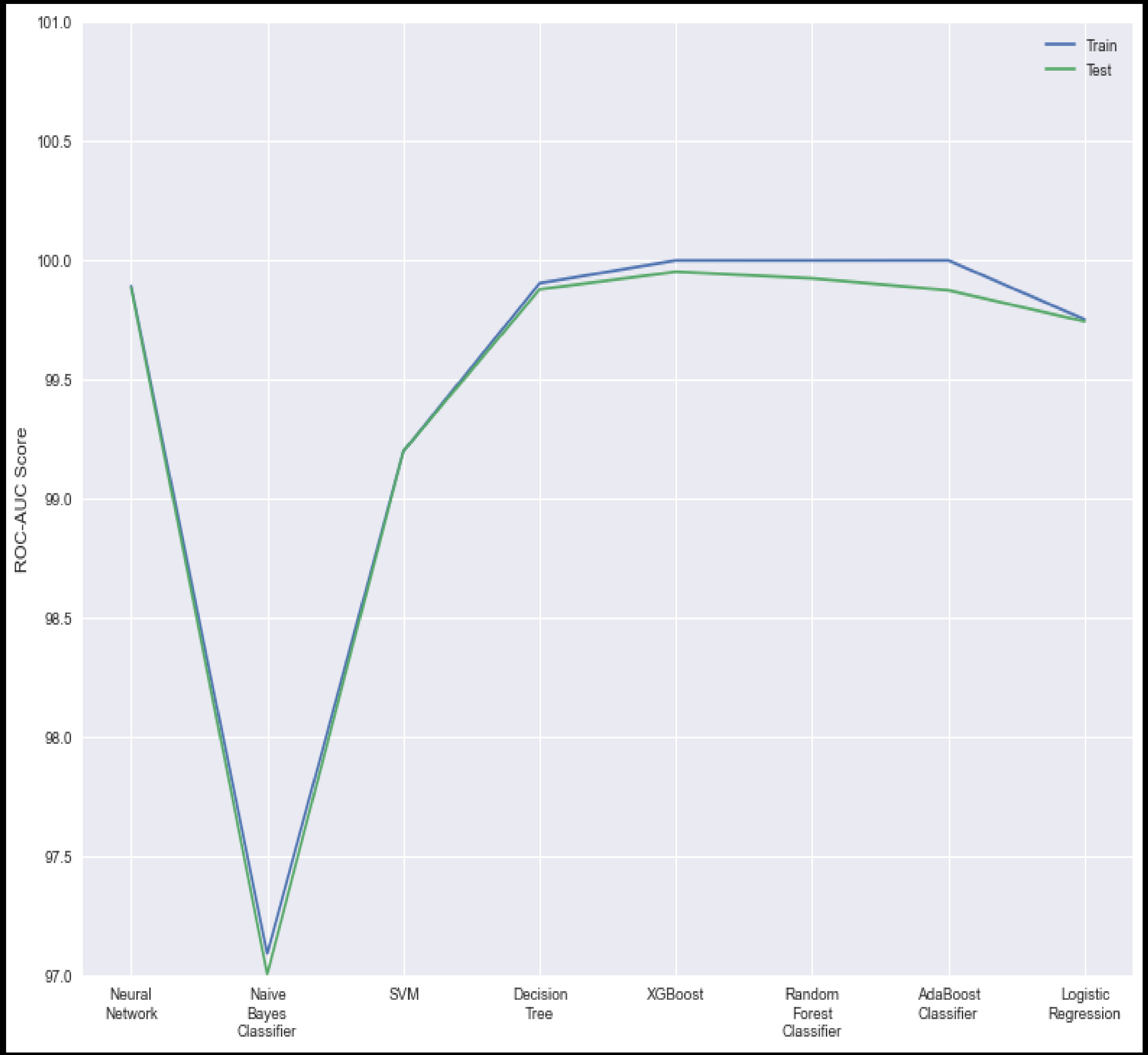
RESULT

MODELS	TRAINING ACCURACY	TESTING ACCURACY
Neural Network	99.189 ± 1.111e-6	99.153 ± 1.111e-6
Naive Bayes Classifier	95.838 ± 3.154	95.804 ± 2.154
SVM	98.149 ± 0.266	98.143 ± 0.155
Decision Tree	99.274 ± 0.113e-16	99.235 ± 0.113e-16
XGBoost	**99.987 ± 0.212e-15**	**99.646 ± 0.232e-15**
Random Forest Classifier	99.999 ± 0.321e-14	99.473 ± 0.351e-14
AdaBoost Classifier	100.0 ± 0.778e-6	99.381 ± 0.797e-6
Logistic Regression	98.745 ± 0.567	98.719 ± 0.667

RESULT

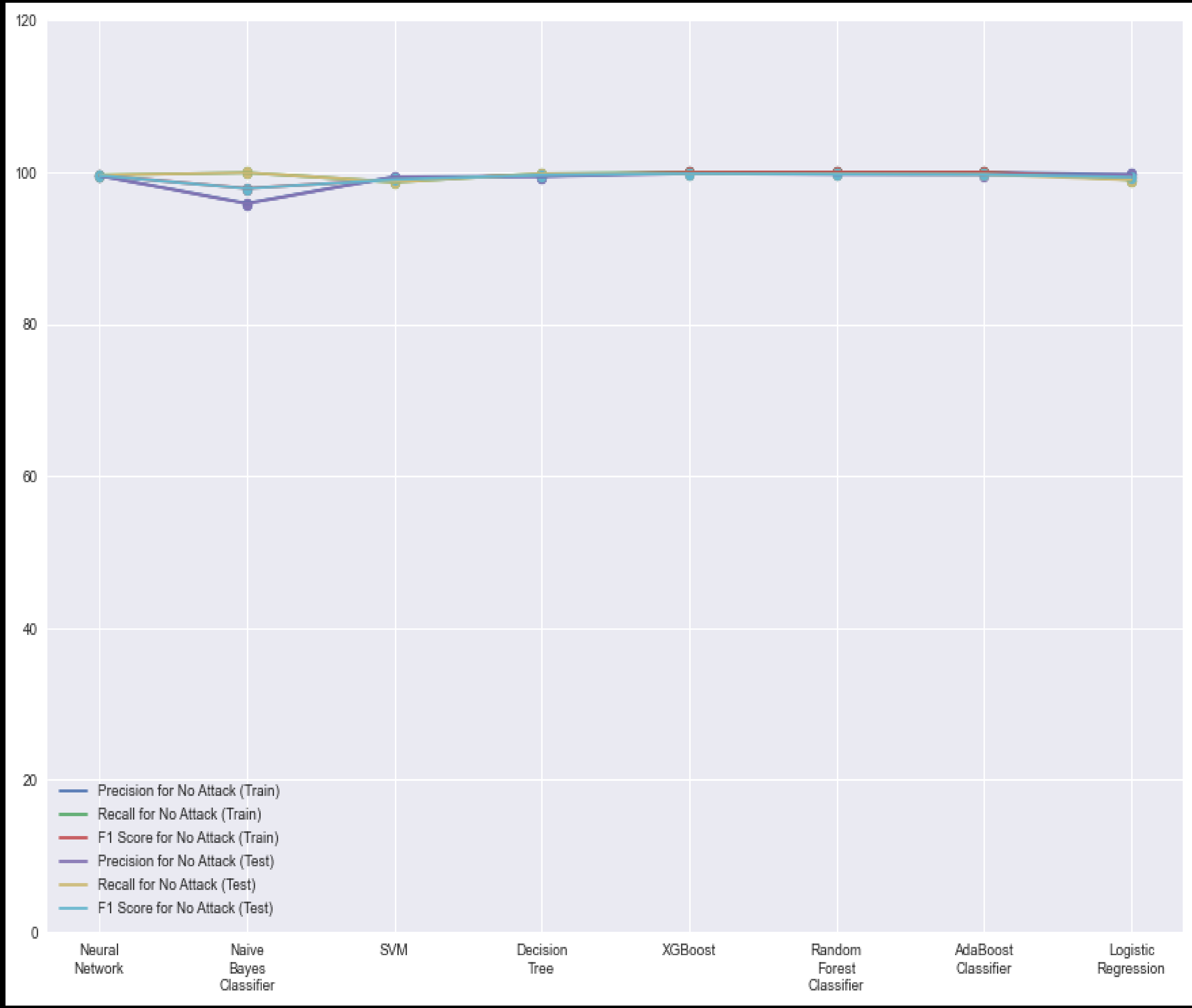


DRs and FARs of all Models



ROC-AUC SCORES OF ALL MODELS

RESULT

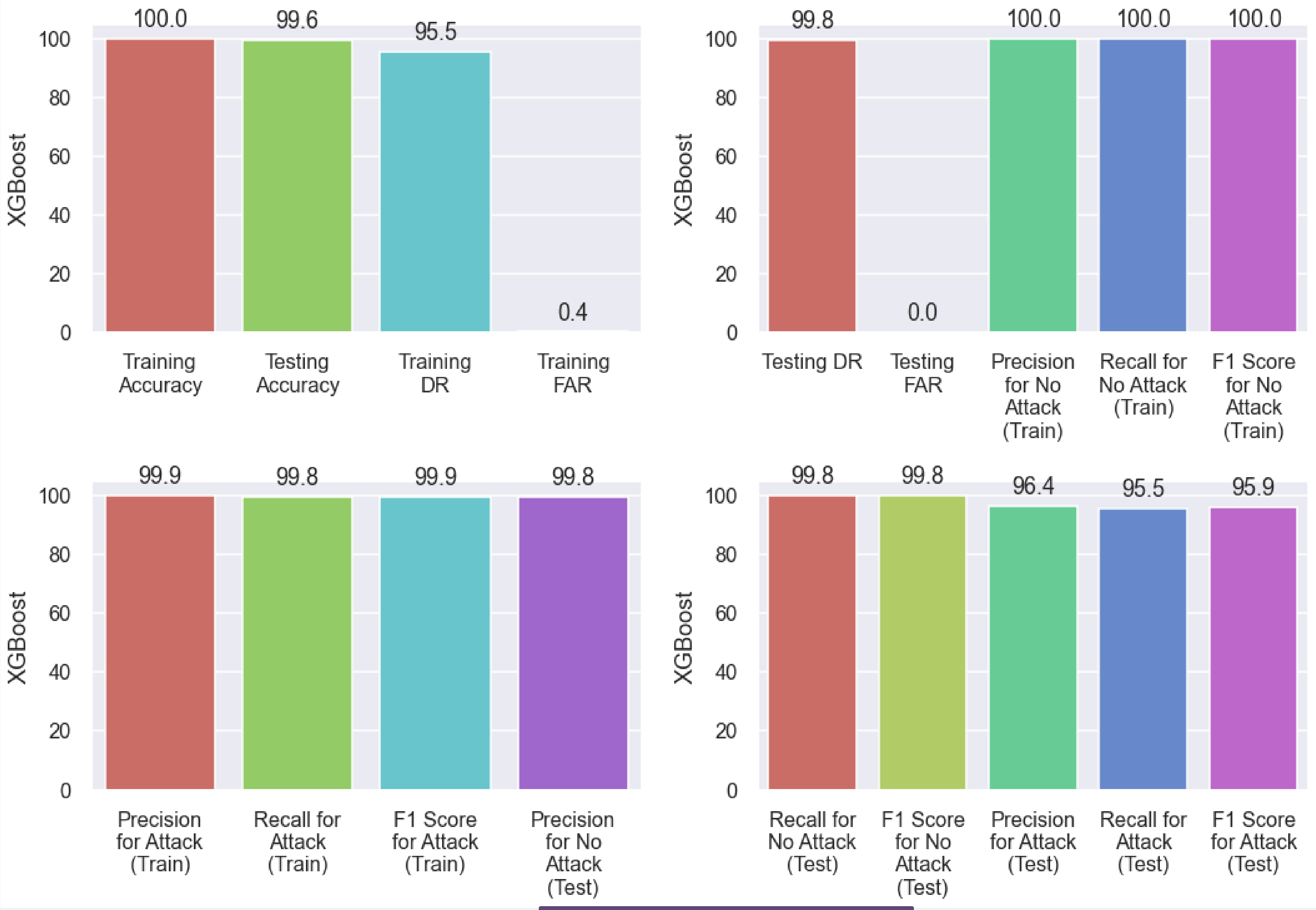


Classification Report for 'NO Attack' Label

Classification Report for 'Attack' Label

CONCLUSION

On the evaluation of various models, the Extreme Gradient Boosted Tree stood out of them all and has performed reasonably well as compared to other models in the literature. Our model gave us an excellent performance as expected from an ideal NIDS.



XGBoost Scores

Raghav Verma
AI/ML Teaching Assistant at Bennett University | Certified Ethical Hacker | ...
5m • Edited •

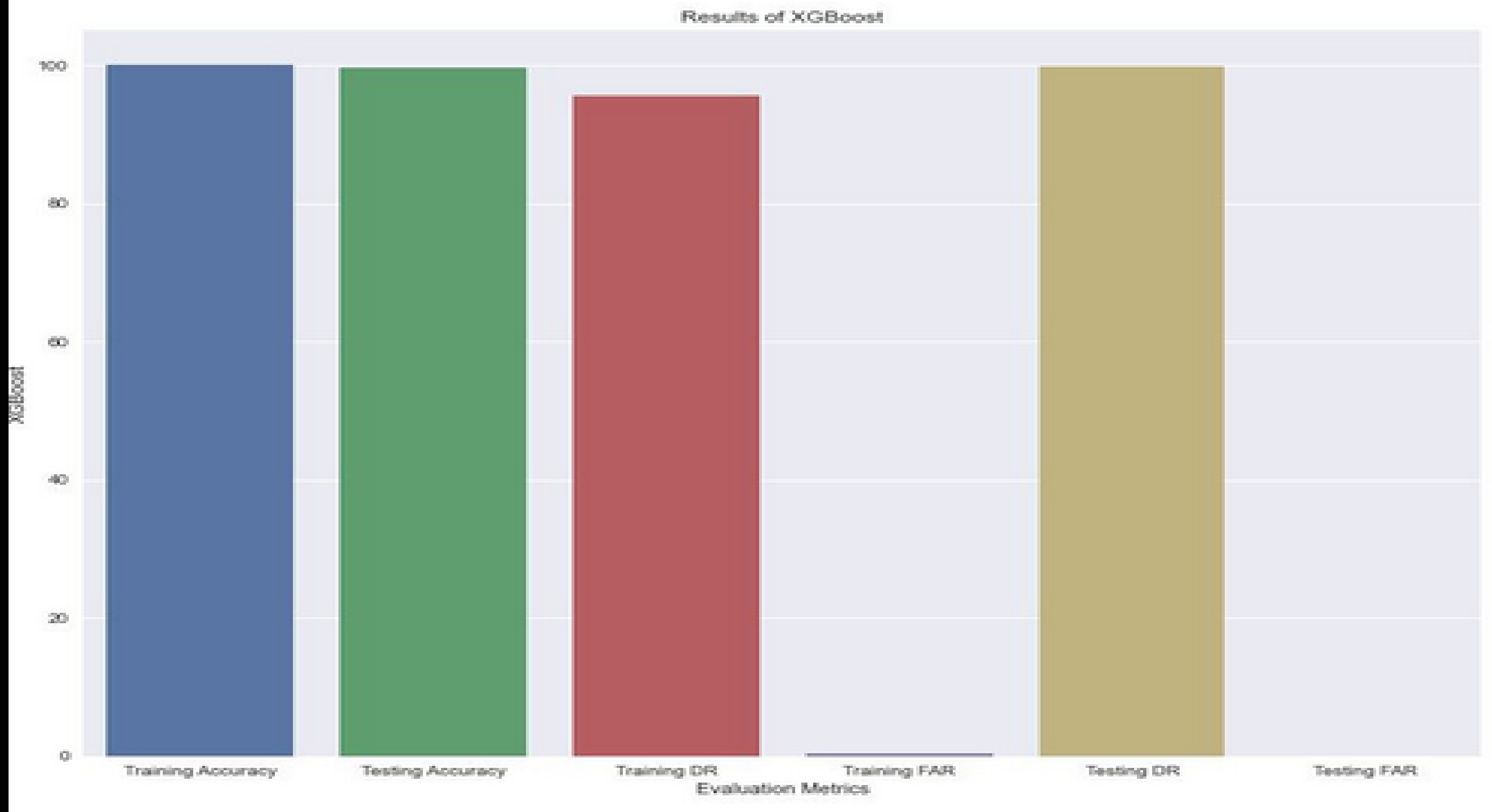
In this digital age, security has become more of a concern with time and critical infrastructure is hugely affected by malicious agents. We are in a dire need of an automated system for detecting malicious activities so that we can lessen losses for everyone involved in this mesh of the internet.

Today I would like to present our project on modelling for a NIDS (Network Intrusion Detection Systems). UNSW-NB15 is a comprehensive dataset for modern-day attacks. In this project, Priyanshu Prasad, Prahalad V Rao and I have tried various kinds of models on this very dataset such as Decision Tree, Multi-Layered Artificial Neural Network, AdaBoost, Logistic Regression, Gradient Boosted Tree (XGBoost), SVM, Naive Bayes and Random Forest. We finalized the XGBoost model as it gave the best results with a very high Detection Rate and a very low False Alarm Rate.

We would like to thank Vipul Kumar Mishra, Dr. Apeksha Aggarwal, and Dr. Dilbag Singh for their constant guidance while carrying out this project.

Details relevant to the project and the associated research paper can be found here - https://lnkd.in/eY4RatE

#deeplearning #datascience #artificialintelligence #ai #security #bigdata #machinelearning #cybersecurity #project #network #iot



THANK YOU