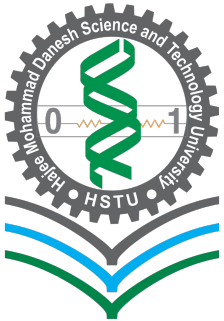
CSE 470

Machine Learning and Pattern Recognition Sessional



Presented By

Mahfujur Rahman
Student ID: 1902006

Md Nur Alam
Student ID: 1902067

Chiranjib Chakraborty
Student ID: 1902059

HAJEE MOHAMMAD DANESH SCIENCE AND TECHNOLOGY UNIVERSITY, DINAJPUR

DDoS Attack Classification By Machine Learning

OUTLINE

- Introduction
- Project Pipeline
- About Datasets
- Data Preprocessing
- Data Exploring
- Data Splitting
- Model Training
- Evaluation and Result
- Model Comparison
- Conclusion

INTRODUCTION

A **Distributed Denial-of-Service** (DDoS) attack is a malicious attempt to disrupt normal traffic of a targeted server, service or network by overwhelming the target or its surrounding infrastructure with a flood of Internet traffic.

DDoS attacks achieved by -

- Utilizing multiple **compromised computer systems (BOTNET)** as source of attack traffic.
- Machines can include computers and other networked resources such as IoT devices.

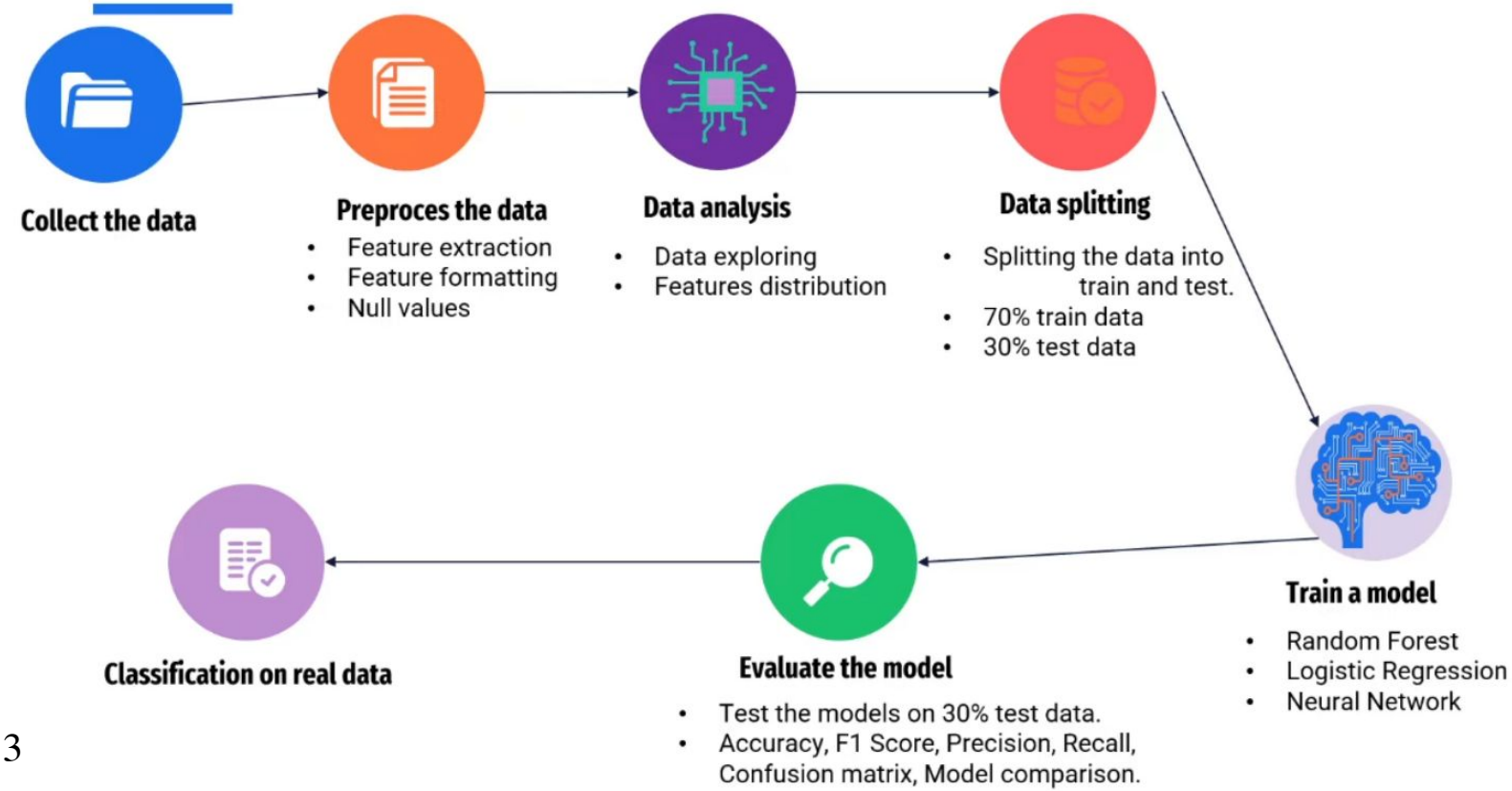
INTRODUCTION cont.

Distributed Denial of Service (DDoS) attacks are significant threats to the stability and reliability of online services.

For detecting this we focuses to made a machine learning project which classify the DDoS attack. We employ 3 ML algorithms -

- Random Forest
- Logistic Regression
- Neural Networks

PROJECT PIPELINE



ABOUT DATASET

We use the public dataset “Intrusion detection evaluation dataset (CIC-IDS2017)” available at internet . Link.

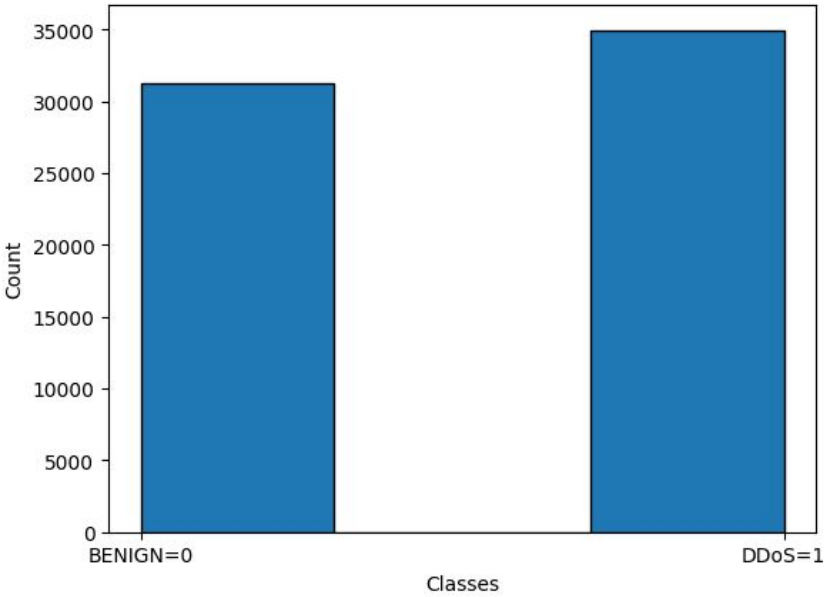
The data capturing period started at 9am Monday, July 3, 2017 and ended at 5pm. Friday July 7, 2017 for a total 5 days.

Table 2: Daily Label of Dataset.

Days	Labels
Monday	Benign
Tuesday	BForce,SFTP and SSH
Wednes.	DoS and Hearbleed Attacks slowloris, Slowhttptest, Hulk and GoldenEye
Thurs.	Web and Infiltration Attacks Web BForce, XSS and Sql Inject. Infiltration Dropbox Download and Cool disk
Friday	DDoS LOIT, Botnet ARES, PortScans (sS,sT,sF,sX,sN,sP,sV,sU, sO,sA,sW,sR,sL and B)

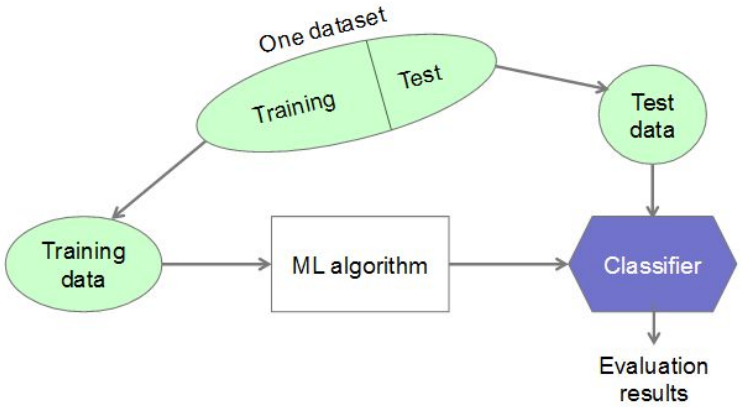
DATA PREPROCESSING

- Remove spaces before the column names.
- Take unique values in the targeted columns.
- Take the 3 target labels .
  - BENIGN
  - DDos
  - NaN
- Removing the null values in the targeted features data.
- Convert the levels into classes . Such that BENIGN = 0 and DDoS=1
- Take this in a separate data frame.



DATA ANALYSIS

- Plot the distributions of every features and understand the relationships.
- Splitting the total dataset into 2 sets.
  - Training Sets (70%)
  - Testing Sets (30%)
- We use those sets to following this stateriges .



TRAIN MODELS

Random Forest :

```
# Random Forest
rf_model = RandomForestClassifier(n_estimators=50, random_state=42)
rf_model.fit(X_train, y_train)
rf_pred = rf_model.predict(X_test)
```

Logistic Regression :

```
lr_model = LogisticRegression(random_state=42)
lr_model.fit(X_train, y_train)
lr_pred = lr_model.predict(X_test)
```

Neural Networks:

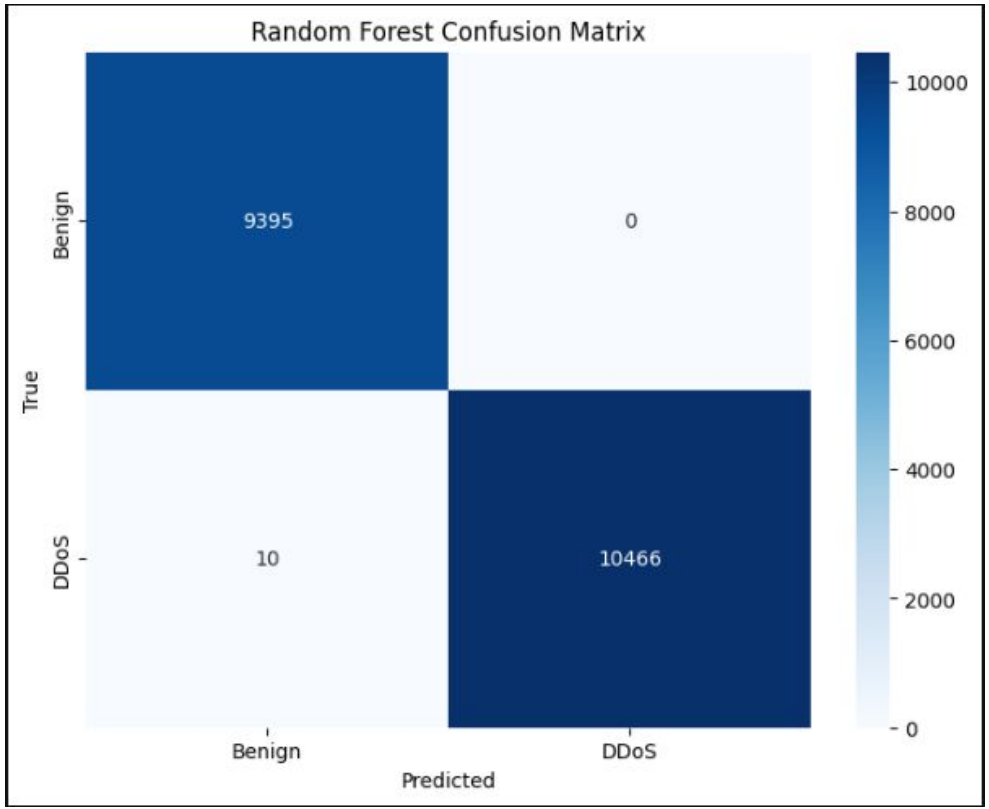
```
nn_model = MLPClassifier(hidden_layer_sizes=(10,), max_iter=1000, random_state=42)
nn_model.fit(X_train, y_train)
nn_pred = nn_model.predict(X_test)
```

EVALUATIONS AND RESULTS

Evaluation of Random Forest :

Random Forest Matrices:

- Accuracy: 0.9995
- F1 Score: 0.9995
- Precision: 1.0000
- Recall : 0.9990

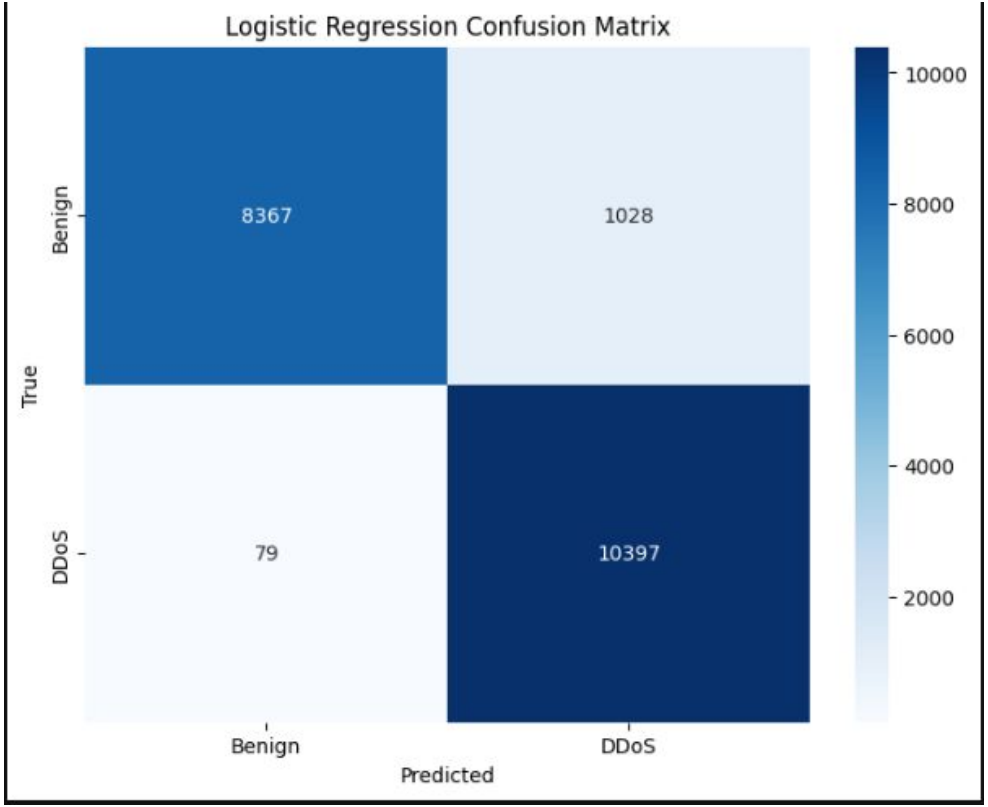


EVALUATIONS AND RESULTS cont.

Evaluation of Logistic Regression :

Logistic Regression Matrices:

- Accuracy: 0.9443
- F1 Score: 0.9495
- Precision: 0.9100
- Recall : 0.9925

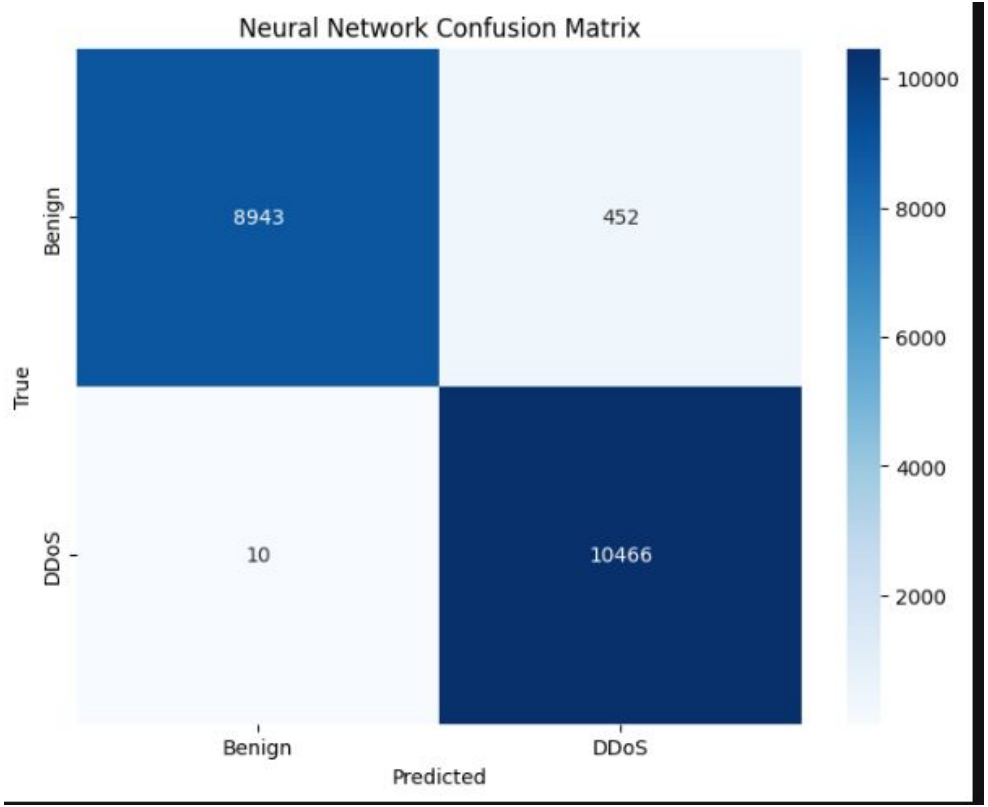


EVALUATIONS AND RESULTS cont.

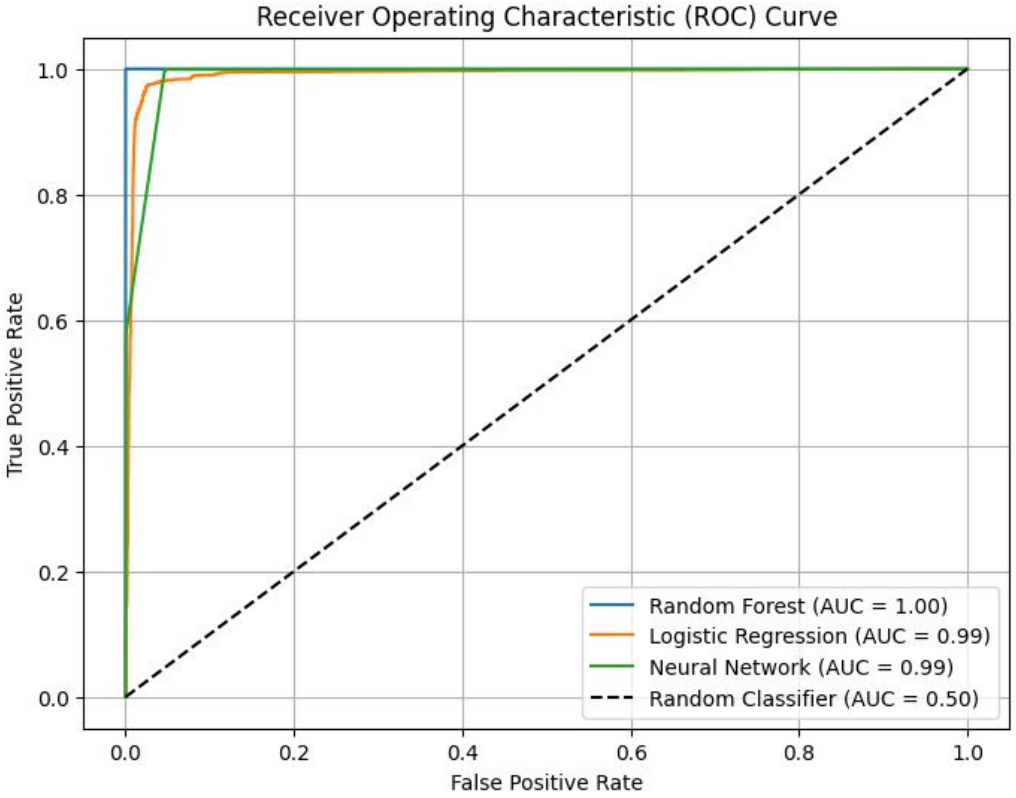
Evaluation of Neural Network :

Neural Network Matrices:

- Accuracy: 0.9768
- F1 Score: 0.9784
- Precision: 0.9586
- Recall : 0.9990



MODEL COMPARISON



CONCLUSION

- Random Forest achieved perfect classification with an AUC of 1.00, indicating no misclassifications between attack and non-attack instances.
- Logistic Regression and Neural Network both showed nearly perfect performance with AUC (Area Under Curve) of 0.99
- Handling class imbalance was a challenge .We mitigated this by using techniques like oversampling
- This model can be integrated into real time network monitoring system to detect and mitigate DDoS attacks proactively
- We aim to test the model on larger datasets and advanced deep learning techniques to further improve detection accuracy.