Major Project

INTRUSION DETECTION SYSTEM

For Web-Based Attack

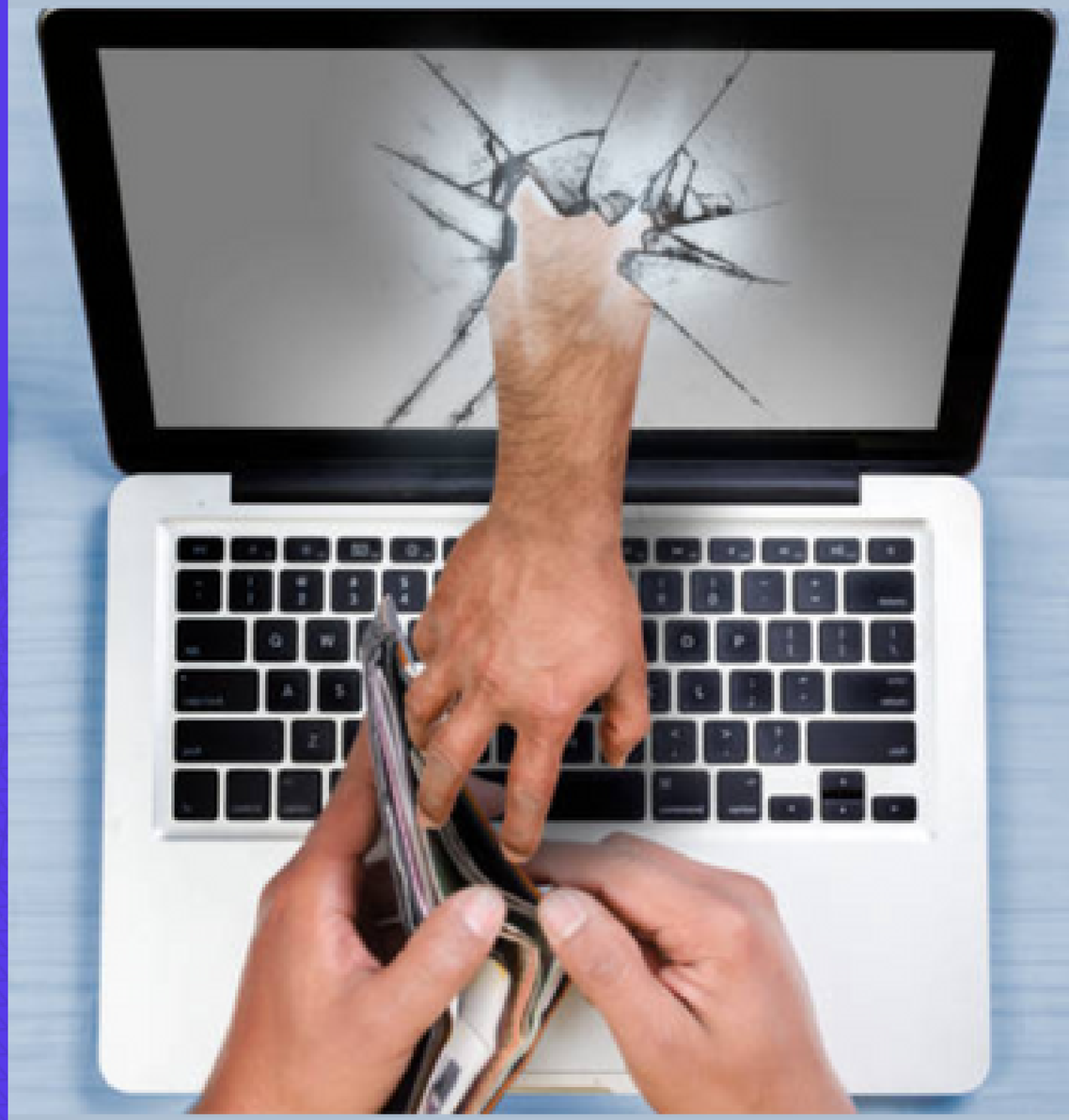
Under the Guidance of
Dr. Malay Kumar

Table Of Content

Major Project

INTRODUCTION

Intrusion Detection Systems (IDS) are a crucial component in protecting computer networks and systems against malicious attacks. IDS are software or hardware-based systems that monitor and analyze network traffic or system activity to identify and alert users of any suspicious or unauthorized behavior.



PROBLEM STATEMENT

We will build an IDS system based on best performing ML & DL algorithms and also embed the concept of EFFST framework before we go to the classification task

WHY IDS IS ESSENTIAL?

- Early Detection of Intrusions
- Protection Against Known Threats
- Protection Against Unknown Threats
- Compliance with Regulations
- Enhanced Network Security

CASE STUDY

Equifax data breach– 2017

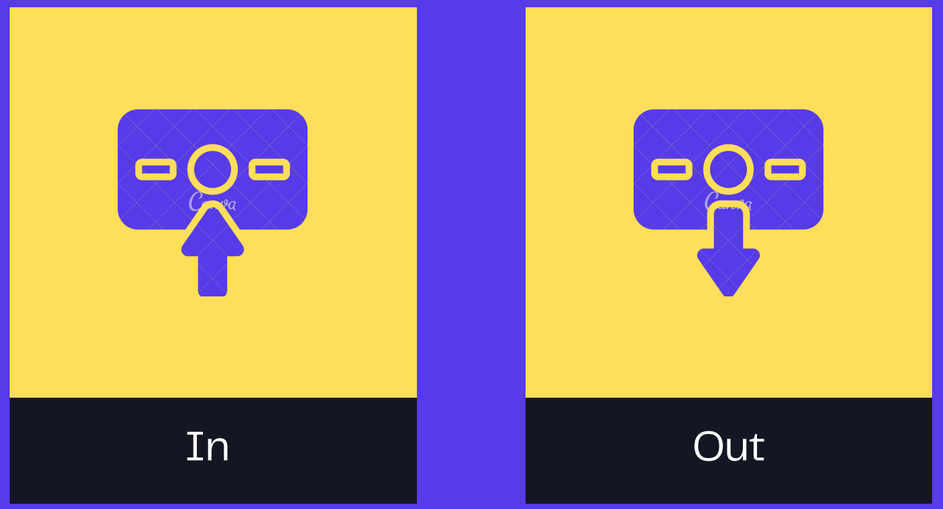
- Equifax, one of the largest credit reporting agencies in the US, suffered a data breach in 2017.
- The breach exposed sensitive information of over 143 million people, including Social Security numbers, birth dates, and addresses.
- An IDS system could have detected the attack and alerted the Equifax security team before the breach occurred, potentially preventing reputational and financial damage.

LITERATURE REVIEW

PAPER TITLE	YEAR	APPROACH
Deep Learning-Based Intrusion Detection Systems for Web Applications	2019	CNN
Web-based Intrusion Detection System Using DL with Attention Mechanism	2020	LSTM with attention
A ML based Web Intrusion Detection System using Hybrid Feature Selection Technique	2018	Hybrid Feature Selection
AN IDS system for web based attacks using IBM Watson	2022	SVM, NB, and RF
Towards an intrusion detection system for detecting web attacks based on an ensemble of filter feature selection techniques	2022	EFFST
Intrusion Detection System using ML Algorithms	2021	Various ML Algorithms
ML based IDS for web based attacks	2020	RF, SVM, and KNN
Web-Based Intrusion Detection System using Machine Learning and Feature Selection Techniques	2021	Feature Selection

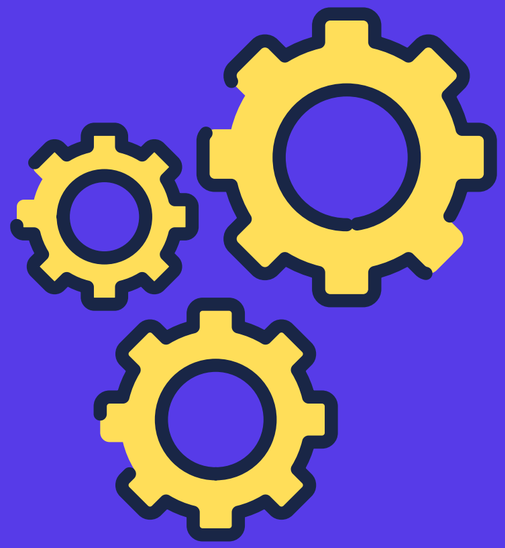
DATA SET

- CICIDS2017 dataset contains benign and the most up-to-date common attacks, which resembles the true real-world data (PCAPs).
- It also includes the results of the network traffic analysis using CICFlowMeter with labeled flows based on the time stamp, source, and destination IPs, source and destination ports, protocols and attack
- For this dataset, they built the abstract behavior of 25 users based on the HTTP, HTTPS, FTP, SSH, and email protocols.
- Created a new multi-class dataset by merging all the data we acquired, which is around 5 lakh data points having various attcks such as DDos, Web attacks, Cross site scripting, port scanner, bruteforce attacks etc.



DATA PREPROCESSING

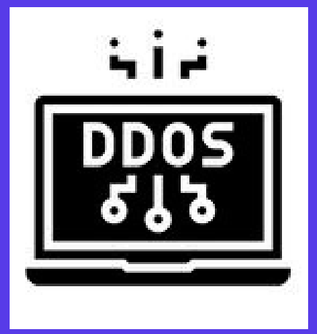
- A smaller portion of the CICIDS dataset was initially tested for binary classification to determine suitable algorithms for multi-class classification.
- Random Forest and XGBoost classifiers were selected for multi-class classification based on the outcomes.
- To handle the large dataset size, a new dataset was created using a 1:1 ratio of benign and respective attacks.
- Data cleaning was performed to remove any duplicate or missing values from the dataset.
- Normalization was applied to standardize the dataset by centering it to a mean of zero and scaling it to a standard deviation of one.
- Normalization ensures that machine learning algorithms are not biased towards features with larger scales, which is crucial for accurate analysis.
- Finally obtained dataset is around 5lakhs, which will be sent into our EFFST framework



RANGE OF ATTACKS

SQL INJECTION

Malicious SQL queries are injected into web applications' input fields to manipulate databases, and IDS can detect suspicious SQL syntax or abnormal query patterns.



DDOS ATTACK

Multiple systems flood a target system or network with overwhelming traffic, causing service disruptions, and IDS can monitor for unusual traffic patterns and raise alerts.

CROSS-SITE SCRIPTING

Malicious scripts are injected into web pages viewed by other users, and IDS can analyze web content for suspicious scripts or HTML patterns.

BRUTE FORCE

Repeatedly attempting to gain unauthorized access to a system or account by systematically trying different combinations of usernames and passwords, and IDS can detect patterns of repetitive login attempts.

BUFFER OVERFLOW

Exploiting a vulnerability to overflow a buffer in a system's memory, potentially allowing an attacker to execute arbitrary code, and IDS can detect abnormal memory manipulation patterns.

IMPORTANT TERMS

Correlation – CR

the statistical measure of the relationship between two variables

Information Gain – IG

the reduction in entropy

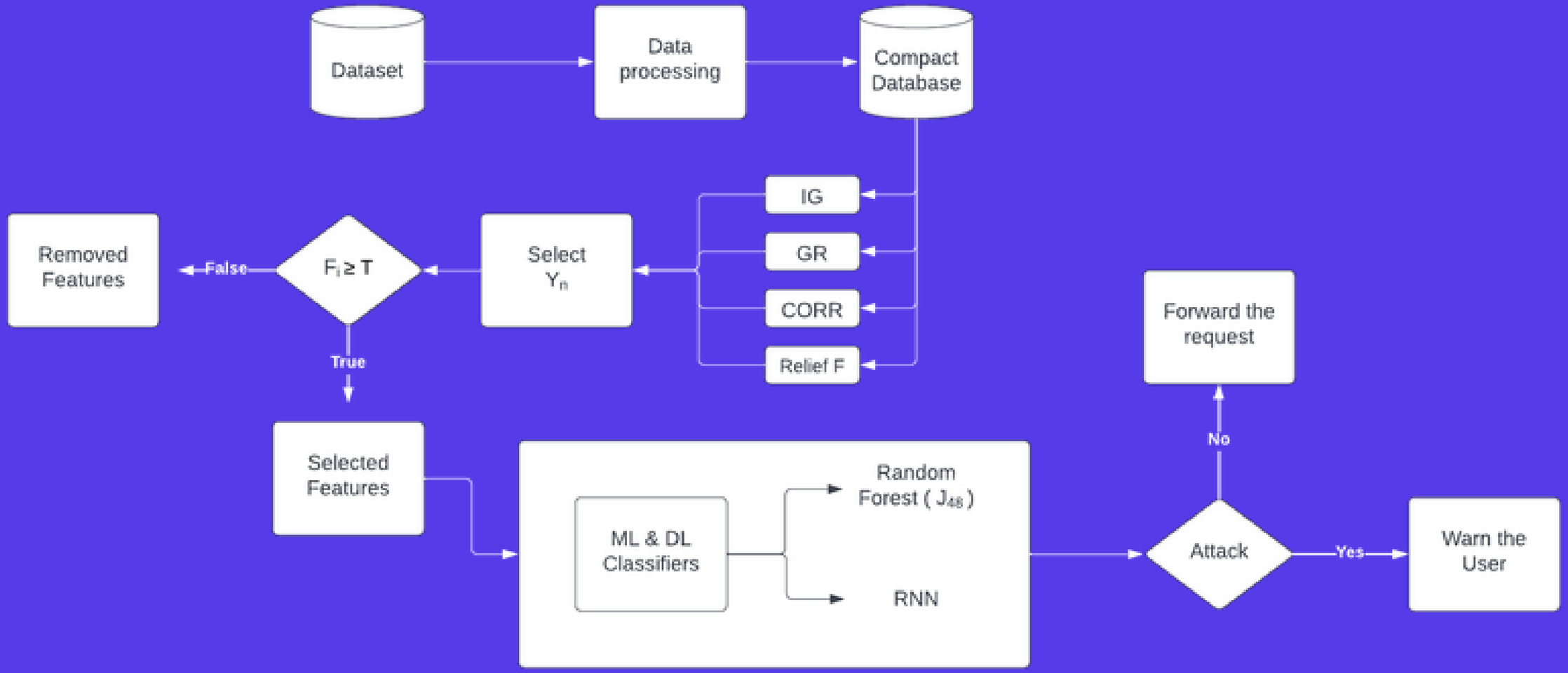
Gain Ratio – GR

modification of information gain that reduces its bias

ReliefF

to estimate the quality of attributes on the basis of how well the attribute can distinguish between instances that are near to each other

PROCESS FLOW DIAGRAM



ROLE OF AI IN IDS

Machine Learning (ML) and Deep Learning (DL) can play a crucial role in building efficient Intrusion Detection Systems (IDS) for web-based attacks. Here are some ways in which ML and DL can be used

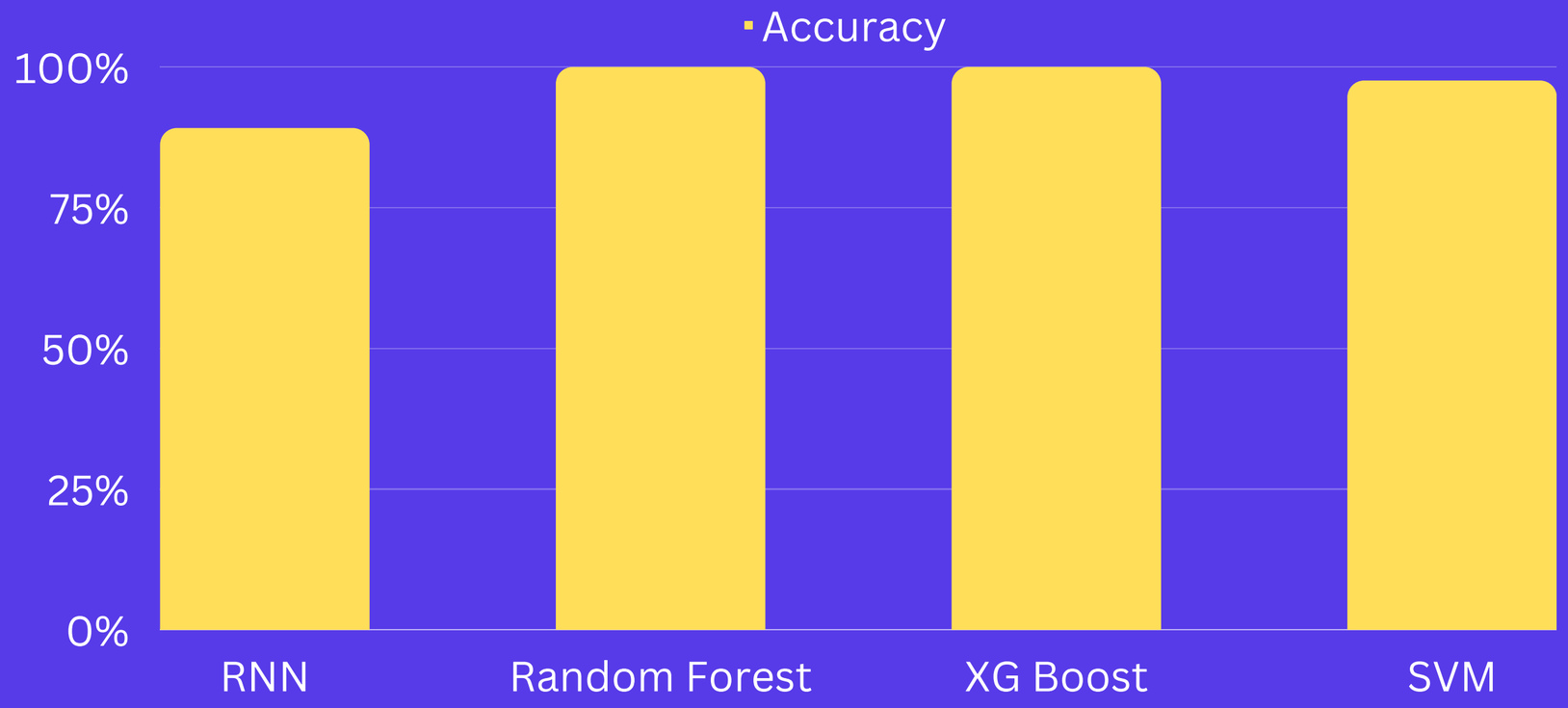
01 Anomaly Detection

02 Multi – Modal Learning

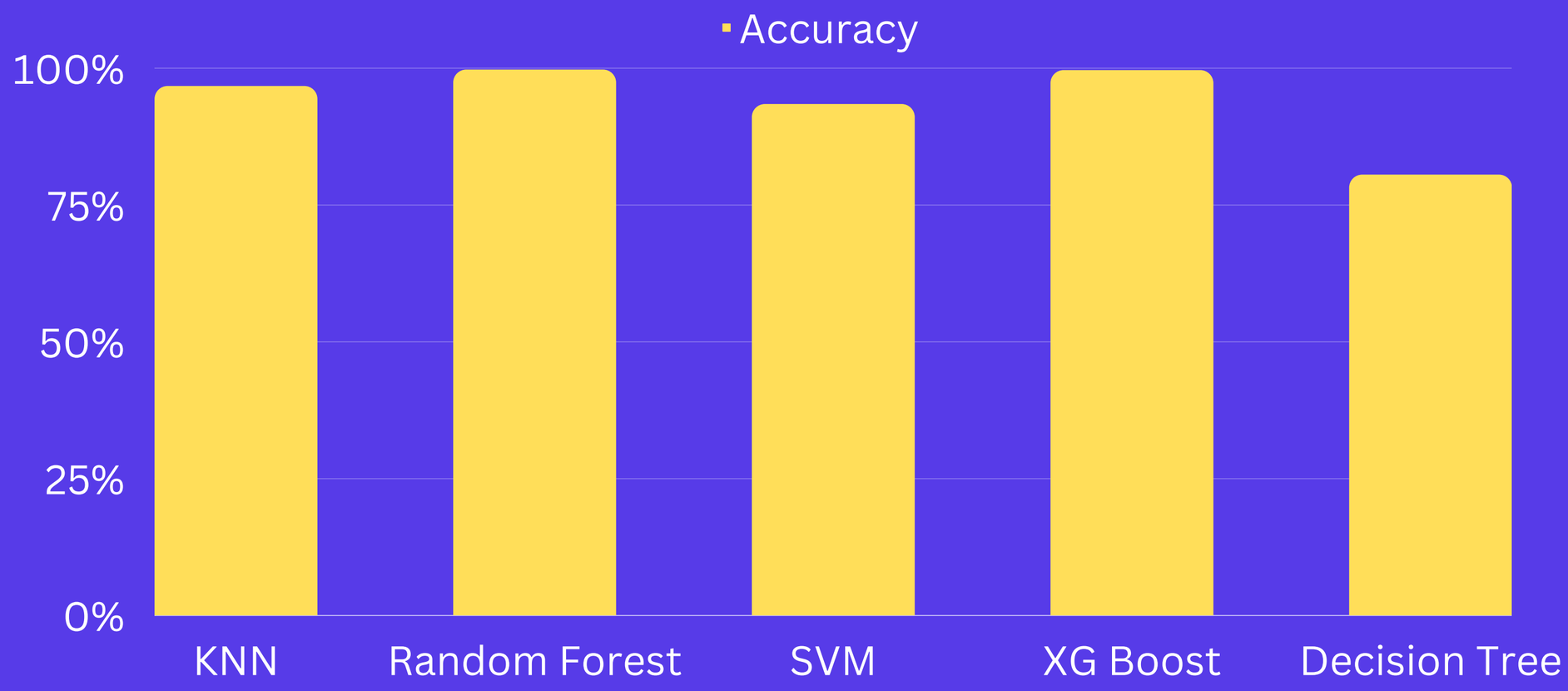
03 Speed & Efficiency

04 Continuos Learning

RESULTS & ANALYSIS



Binary Classification – DDos Attack (Binary class dataset)



Mutli-Class Classification – New Data set (CICIDS – 2017)

EVALUATION METRICS

Training Sample : 396087

Testing Sample : 99022

TYPE	VALUE (%)
Accuracy	99.74
F1 Score	99.75
Precision	99.78

Random Forest Model Evalutaion Metrics

CONTRIBUTION

Stage 1

Development with small dataset

Stage 2

Achieving results with prepared Multi-class Dataset

Stage 3

Implementation (Optional)

IMPLEMENTATION

Attacker Machine
(H3ping, Nmap)

User Machine

WIRE SHARK :
Capture traffic and
analyzation tool

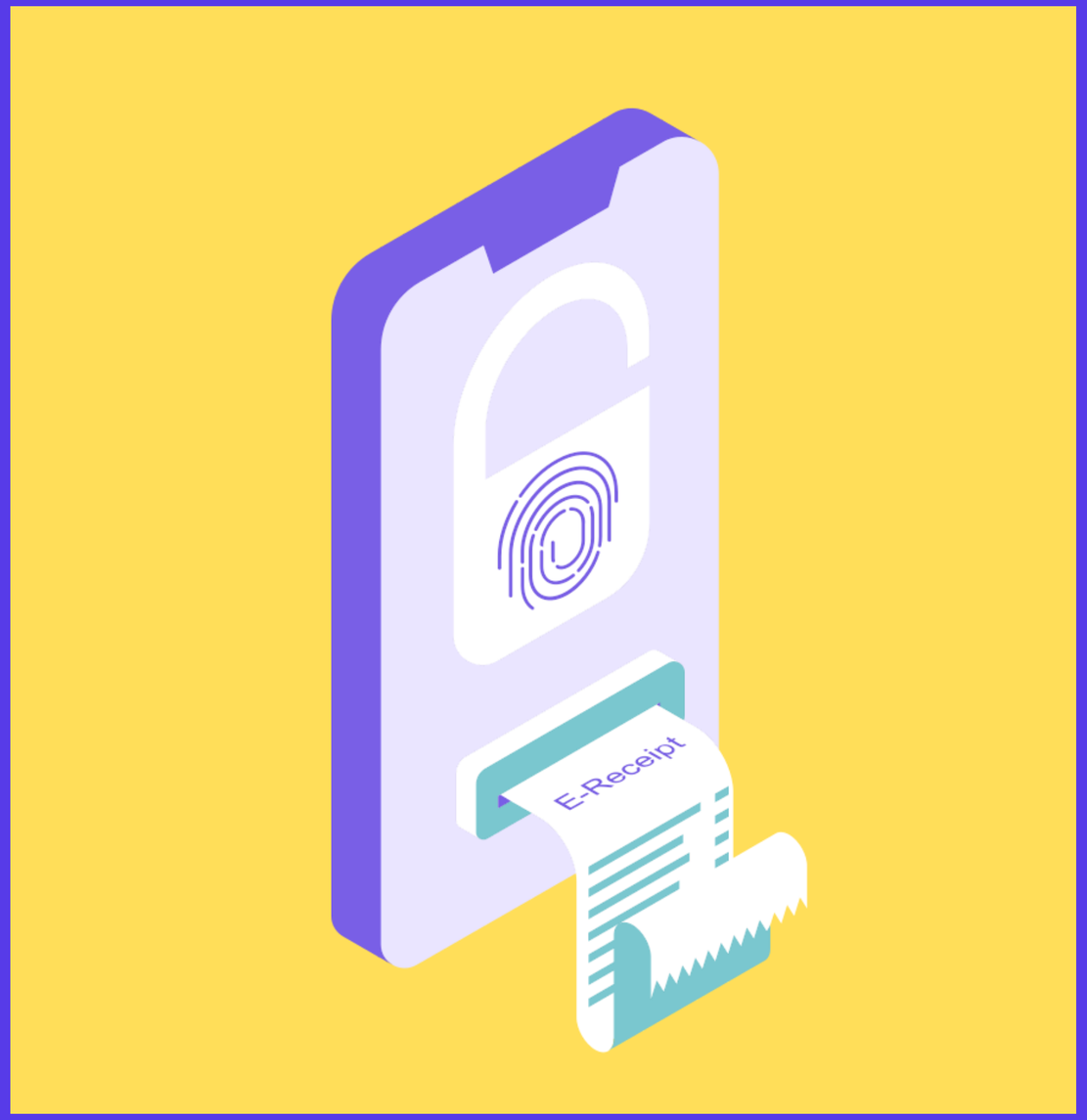
SCAPY : Data
manipulation tool &
extract features

Predict using our
Model

Process Flow of our implementation

Conclusion

- IDS systems can use ML & DL algorithms to detect and prevent web-based attacks by analyzing web traffic and system activity.
- Anomaly-based IDS systems use ML & DL to identify abnormal behavior that may be indicative of a web-based attack.
- Continuous learning with ML & DL allows IDS systems to improve over time and stay up-to-date with new and evolving attack patterns.
- IDS systems using ML & DL can provide a more comprehensive and effective defense against web-based attacks, helping to prevent data breaches and system failures.



Thank You

Our Team

Ch . Bhawan kumar – 19BCS031

D. Tarun Varma – 19BCS036

K. Sai Ganesh – 19BCS055

K. Chandra Sekhar Sharma – 19BCS059