# 

# ML-based Network Intrusion Detection System

Name: Haider Nadeem  
Enrollment: 03-134221-012  
Course: Information Security  
Instructor: Dr. Nadeem Sarwar

## Summary

In today's interconnected digital world, network intrusions have become increasingly sophisticated. Traditional security systems like firewalls and antivirus programs are no longer sufficient to detect and mitigate complex attacks. This project presents a Machine Learning-based Intrusion Detection System (IDS) designed to classify network traffic as either benign or malicious. Using the CIC-IDS2017 dataset, which includes real-world traffic patterns such as DDoS and brute-force attacks, a Random Forest classifier was developed and evaluated. The model achieved high accuracy and recall, demonstrating the potential of ML to enhance traditional NIDS solutions.

## Introduction

Organizations face constant threats from cyber attackers who exploit vulnerabilities in network infrastructure. The role of an Intrusion Detection System (IDS) is to identify unauthorized access or malicious activities in real-time. As a security analyst at SecureNet Corp, the goal is to enhance existing detection systems by integrating an intelligent, data-driven approach. This proof-of-concept uses machine learning to improve detection accuracy and reduce false positives.

## Methodology

The workflow followed these stages:

1. Dataset Selection: The CIC-IDS2017 dataset was chosen for its realistic network behavior and diverse attack types such as DDoS, PortScan, and Brute Force.  
2. Data Preprocessing: Using Python libraries like pandas and numpy, missing values were handled, categorical features (e.g., protocol type) were encoded, and numerical features were normalized.  
3. Model Development: A Random Forest Classifier was selected for its robustness and ability to handle complex relationships. The model was trained on 80% of the data and tested on the remaining 20%.  
4. Evaluation Metrics: Accuracy, Precision, Recall, F1-Score, and Confusion Matrix were used to evaluate performance.

## Results and Security Analysis

The Random Forest model achieved the following sample performance metrics:  
• Accuracy: 97.6%  
• Precision: 95.8%  
• Recall: 96.9%  
• F1-Score: 96.3%  
  
The high recall value indicates the model’s effectiveness in identifying most attack instances, reducing the risk of undetected intrusions. A strong precision score means fewer false alarms, which helps security analysts focus on real threats.

The confusion matrix showed minimal false negatives, which is critical for security applications. Feature importance analysis revealed that flow-based metrics such as packet rate, byte count, and connection duration contributed most to attack detection.

## Conclusion and Future Enhancements

This project demonstrates that machine learning can serve as an intelligent extension to traditional IDS solutions. By automating traffic classification and continuously learning from new patterns, such systems can enhance threat visibility and incident response. Future improvements could include testing deep learning models, implementing SMOTE to balance classes, and integrating the model into a live monitoring system for real-time intrusion alerts.

## References

1. CIC-IDS2017 Dataset – Canadian Institute for Cybersecurity  
2. Scikit-learn Documentation – https://scikit-learn.org/  
3. Python Pandas and NumPy Libraries  
4. SecureNet Corp. Simulation Scenario

## GitHub Repository

https://github.com/YourUsername/CLO4-IDS-ML-Solution