Anomaly-based IDS with ML - Project Guide

Project Guide: Anomaly-based Intrusion Detection System with Machine Learning

1 What is an IDS? Why is it important?

An Intrusion Detection System (IDS) is a security tool that monitors a computer network for suspicious or malicious activity. Anomaly-based IDS detects attacks by identifying unusual patterns in network traffic, even if the attack is unknown.

2 What does this project do?

This project builds an anomaly-based IDS that:

- Collects network traffic data
- Extracts features from the data
- Trains an ML model on normal traffic
- Flags unusual traffic as suspicious
- Displays results on a dashboard

3 System Architecture Diagram

Packet/Data Capture Feature Extraction ML Model Inference Alert & Dashboard

4 Tools & Technologies Required

- Python 3.x
- Libraries: pandas, scikit-learn, matplotlib, seaborn
- Streamlit/Dash (for dashboard)
- Dataset: NSL-KDD / CICIDS2017
- Wireshark/tcpdump (optional, for real traffic)

5 Dataset(s) to use

NSL-KDD (recommended for beginners, small & cleaned)

CICIDS2017 (larger, realistic, modern)

6 Step-by-Step Implementation Plan

Week 1: Research IDS & datasets, install tools

Week 2: Load & explore dataset, clean & extract features

Week 3: Train Isolation Forest model, test on dataset

Week 4: Build dashboard & add real-time components

Week 5: Write README, test, record demo

7 Optional Enhancements

- Real-time packet sniffing
- Autoencoder-based deep learning
- Email/SMS alerts
- Docker deployment

8 Expected Deliverables

- Python code
- Trained ML model
- Dashboard app
- Dataset & cleaned CSVs
- README with usage instructions

9 How to showcase it on Resume & GitHub

Resume line:

> Designed and built an anomaly-based Intrusion Detection System using Isolation Forest, achieving 92% detection accuracy on NSL-KDD dataset with real-time dashboard.

GitHub:

- Clean repo structure: /src, /data, /notebooks, README.md
- Include sample screenshots & demo video link

References

- NSL-KDD: http://www.unb.ca/cic/datasets/nsl.html
- CICIDS2017: https://www.unb.ca/cic/datasets/ids-2017.html
- Isolation Forest: https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.IsolationForest.html