Project Overview

Problem Statement:

In today's digital age, the volume and sophistication of cyber threats are increasing rapidly. Our goal is to develop an IDS that can detect different types of network attacks with high accuracy using machine learning models trained on real traffic data.

Key Tasks: Data cleaning, EDA, model training, evaluation, and deployment (optional).

Your Tasks for Week 1

Gitbook Link:

https://technocollabs.gitbook.io/cybersecurity-internship-1/week-1-tasks-for-interns

Here's what you're expected to accomplish this week:

1. Understand the Domain
    - Learn how IDS systems work (signature vs anomaly-based).
    - Study common attack types in the dataset.
    - Submit a 1-page summary: IDS_Domain_Overview.
2. Download & Explore the Dataset
    - Download at least 1–2 daily CSV files (e.g., Friday-02-03-2018.csv).
    - Use pandas to explore the data (shape, columns, label distribution).
    - Create a Jupyter Notebook: Data_Exploration.ipynb.
3. Clean & Preprocess the Data
    - Handle missing/null/infinite values.
    - Drop irrelevant columns (Flow ID, Source IP, etc.).
    - Encode labels and create a small sample dataset for testing.
    - Submit: clean_data.py and cleaned_data.csv.
4. Perform EDA
    - Generate visualizations (label distribution, correlation heatmap, etc.).
    - Save visuals for later use in reporting.

- Submit: EDA_Report.ipynb.

5. Organize Your Work
    - Use a proper folder structure (shared GitHub/GDrive):
    - CyberIDS-Project/
    - ├── data/
    - │   ├── raw/
    - │   └── cleaned/
    - ├── notebooks/
    - ├── scripts/
    - ├── reports/
    - └── README.md