

## Group 4 Project Proposal

### Topic Name:

Cybersecurity Intrusion Detection using Supervised Binary Classification

### Goal and motivation for the project:

Create a robust machine learning model to detect intrusion and intrusive behaviors through analysis of network data. We want to be able to accurately detect and classify network traffic as either **benign** or an **attack**. Practically, intrusion detection is critical in cybersecurity, as it helps organizations detect malicious activity early, minimize damage, and ultimately improve trust in network systems. So, given features of a network session, we want to predict whether it's benign (0) or an attack (1).

### Dataset Used:

<https://www.kaggle.com/datasets/dnkumars/cybersecurity-intrusion-detection-dataset/data>

- Network traffic records
- The dataset contains over 9000 records with features such as: protocol, session duration, whether encryption was used, and attack detected.

### Members (roles and responsibilities):

1. Saba Siddiqi (Project Manager) – model tuning, evaluation, documentation
2. Jaswanth Singamsetty (Modeling Lead) – model development and training
3. Matt Christiansen (Data Analyst) – data cleaning, feature engineering, analysis, and documentation
4. ThaiHa Dang (Validation/Visualization Lead) - data visualization, and model evaluation

### Schedule with milestones and expected outcomes:

Date	Milestone	Expected Outcome
18 Sep 25	Project Proposal	Finished project proposal with details on schedule, roles, responsibilities, dataset, and overall goal
2 Oct 25	Data Preprocess (Phase 1)	Data should be fully preprocessed for goal, identify class distributions and missing values
16 Oct 25	Data Preprocess (Phase 2)	Feature encoding, normalization, train/test split. Train baseline models
30 Oct 25	Training	Train advanced models (e.g., Random Forest)
13 Nov 25	Finalize results	Prepare report and findings <ol style="list-style-type: none"><li>1. A trained intrusion detection classifier with accuracy</li></ol>

		2. Insights into which features are most predictive of intrusions
TBD	TBD	TBD