CPT_S 570 Machine Learning
Project Proposal

Team: SHARATH KUMAR KARNATI, 011852253.

Project Aim: Intrusion Detection Using Machine Learning Algorithms

With the dramatic growth in the use of computer networks and the vast increase in the number of applications running on them, network security has become increasingly critical. Many systems are prone to security vulnerabilities, leading to potential attacks with serious economic impacts. Detecting these vulnerabilities in real time, with high accuracy, is essential.

This project aims to develop models for intrusion detection using multiple machine learning algorithms, including Naive Bayes, Support Vector Machines (SVM), Decision Trees, and Random Forest. These models will be trained on network packet data to distinguish between normal and attack traffic. A comparative analysis of the models will be performed by plotting test vs. train performance curves for each algorithm.

Methodology

- **Data Collection**: Use publicly available datasets like KDDCup99 or NSL-KDD for network packet data containing normal and attack traffic.
- **Data Preprocessing**: Clean and preprocess data (scaling, encoding, feature extraction), then split into training and testing sets.
- **Model Development**: Build Naive Bayes, SVM, Decision Tree, and Random Forest models.
- **Model Training**: Train models on the dataset, applying hyperparameter tuning using cross-validation.
- **Model Evaluation**: Evaluate models using accuracy, precision, recall, F1 score, and ROC-AUC. Plot test vs. train performance for each.

Final Product

- Trained models (Naive Bayes, SVM, Decision Tree, Random Forest) optimized for intrusion detection.
- Detailed performance reports and test vs. train graphs.
- Graphical comparisons (confusion matrices, ROC curves) for model effectiveness.
- Best-performing model for real-time or batch intrusion detection.