# 📄 Project Summary Report: IndustrialShield - Modbus Intrusion Detection System

## 🔍 Problem Statement:

Industrial networks using Modbus protocol are increasingly targeted by cyber-attacks. The goal of this project is to detect intrusion attempts in Modbus communication using supervised machine learning models.

## 📊 Dataset Overview:

- Source: Industrial IoT Modbus communication logs
- Shape: 31,106 rows × 8 columns
- Target Variable: Attack (Binary: Normal or Attack)
- Class Balance: Includes both normal and intrusion activities with realistic variance.

## 🛠 EDA & Preprocessing:

- Conducted distribution plots and boxplots to understand patterns and outliers.
- No outlier removal performed due to the domain-specific nature (cyberattacks may appear as outliers).
- Applied label encoding for categorical features (e.g., function_code, Address) where applicable.
- Feature scaling applied using StandardScaler.

## 📊 Feature Selection & Modeling:

- Feature importance visualized using Random Forest.
- Trained and compared three models:
- Logistic Regression (Tuned)
- Random Forest (Tuned)
- XGBoost was excluded due to compatibility issues during training.

## ✅ Best Model:

- Random Forest (Tuned) achieved the highest accuracy.
- Evaluated all models using:
  i. Accuracy
  ii. Confusion Matrix
  iii. ROC & AUC Curve
  iv. Learning Curve for overfitting/underfitting detection

## 🔬 Hyperparameter Tuning:

- Performed GridSearchCV for Logistic Regression and Random Forest.
- Improved model generalization through cross-validation and parameter optimization.

## 📉 Final Takeaway:

The tuned Random Forest model is well-suited for detecting intrusion patterns in Modbus-based industrial communication. Outlier handling and model comparisons were thoughtfully executed based on real-world cybersecurity constraints.