MICROSOFT: CLASSIFYING CYBERSECURITY INCIDENTS

Agenda:

- Introduction
- Data Cleaning & Preprocessing
- Exploratory Data Analysis
- Prediction Modelling
- Conclusion

INTRODUCTION

In modern cybersecurity landscapes, Security Operation Centers (SOCs) play a crucial role in detecting, analyzing, and responding to security incidents.

However, due to the increasing number of alerts generated by security systems, SOC analysts often face alert fatigue, making it difficult to prioritize true threats.

To alleviate this, our goal is to develop a machine learning-based classification model capable of predicting the triage grade of cybersecurity incidents.

Categorizing them as True Positive (TP), Benign Positive (BP), or False Positive (FP).

DATA CLEANING

- **Handling Missing Values:** Missing values in features such as Mitre Techniques, Action Grouped, and other optional fields were handled by either imputing or dropping rows based on their importance in the model.
- **Duplicate Removal:** Any duplicate records that might skew the training were identified and removed.
- **Outlier Detection:** Certain features such as time and incident severities were checked for outliers. Outliers were either removed or appropriately capped to prevent skewing the model.

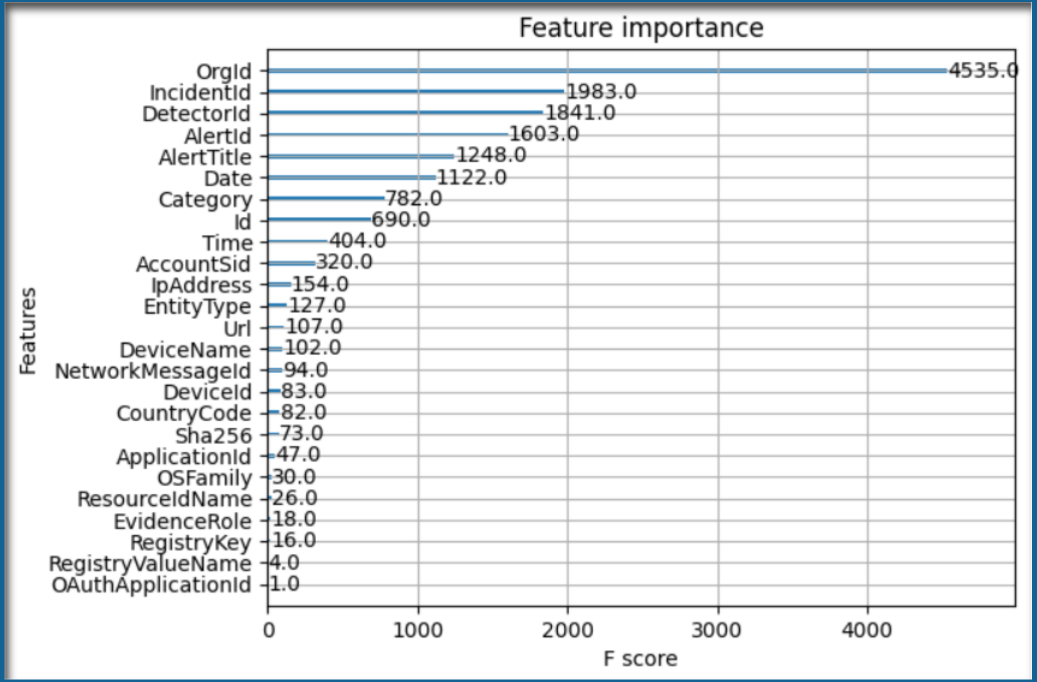
DATA PREPROCESSING

- **Label Encoding:** The target variable Incident Grade was **label-encoded**, converting the triage grades (TP, BP, FP) into numeric form for the model.
  - Benign Positive → 0
  - False Positive → 1
  - True Positive → 2

- **Feature Scaling:** **Min Max Scaler** was applied to normalize the features and ensure that the models treat all variables equally.

- **Train-Test Split:** The training dataset was split into a **training and validation** set to assess the model's performance during development.

EXPLORATORY DATA ANALYSIS (EDA)



- **Correlation Analysis:** A correlation matrix was generated to assess relationships between numerical features.
- **Feature Importance:** Using preliminary model XGBoost, feature importance scores were computed. Features such as OrgId, IncidentID, DetectorID, and AlertID were found to be most predictive for determining the triage grade.

PREDICTION MODELLING

This project leverages the GUIDE dataset, which contains historical cybersecurity incidents, to train the model. The dataset includes various features, By utilizing this data, we aim to build a classification model that not only demonstrates high accuracy but also generalizes effectively to **unseen data**, making it practical for real-world use.

The machine learning models used in this project include

- Logistic Regression
- Decision Tree Classifier
- Random Forest Classifier
- XGBoost Classifier (XGBClassifier)

After thorough evaluation, the **XGBoost Classifier** emerged as the **best model** with a prediction accuracy of **91%** on unseen data.

CONCLUSION

In this project, we successfully developed a machine learning model to predict the triage grade of cybersecurity incidents with high accuracy. The **XGBoost Classifier** emerged as the best model, with an accuracy of **91%**, along with strong accuracy score, precision, recall score and f1_score. This model has the potential to significantly enhance the efficiency of Security Operation Centers by reducing false positives and providing SOC analysts with actionable insights for triaging incidents.

Reference : https://github.com/Syed-Abuthahir-M/Microsoft-Classifying-Cybersecurity-Incidents-with-Machine-Learning