**Project Report: Cybersecurity Incident Classification**

**1. Introduction**

**1.1 Project Background**

With the rise in cybersecurity threats, organizations are increasingly exposed to incidents that demand rapid response and precise classification. Efficiently categorizing incidents based on severity, threat type, and related features can streamline incident management workflows and support proactive threat mitigation.

**1.2 Objective**

The primary goal of this project is to classify cybersecurity incidents using machine learning models. By analyzing incident features such as alert titles, severity grades, categories, timestamps, and user information, the project aims to predict and prioritize incidents, enabling effective and timely response.

**2. Data Overview**

**2.1 Data Description**

The dataset contains key information on cybersecurity incidents, with fields such as:

* **Alert Title**: Descriptive label of the security alert.
* **Incident Grade**: Severity level (e.g., benign, true positive, false positive).
* **Category**: Type of cybersecurity threat (e.g., credential access, malware).
* **Timestamp**: Date and time of the incident.
* **User Account**: Identifier for the user involved in the incident.

**2.2 Data Preprocessing**

Data preprocessing involved:

* **Handling Missing Values**: Imputed missing values to ensure completeness.
* **Encoding Categorical Features**: Transformed categorical variables (e.g., Incident Grade, Category) using encoding techniques.
* **Feature Engineering**: Extracted relevant temporal features (e.g., year, month, hour) from timestamps and converted textual features to numeric representations.

**3. Methodology**

**3.1 Model Selection**

Multiple classification models were explored to identify the best-performing model for incident classification:

* **Random Forest Classifier**: A robust ensemble model based on multiple decision trees.
* **Decision Tree Classifier**: A simple yet interpretable model that divides data based on feature splits.
* **Logistic Regression**: Used as a baseline due to its simplicity and interpretability.
* **Gradient Boosting**
* **XG Boost**

**3.2 Model Training and Hyperparameter Tuning**

* Models were trained on a balanced dataset using **SMOTE (Synthetic Minority Oversampling Technique)** to address any class imbalance.
* **Hyperparameter tuning** was performed using GridSearchCV or RandomizedSearchCV to optimize model performance, adjusting parameters like n\_estimators (for Random Forest) and max\_depth (for Decision Tree).

**3.3 Model Evaluation Metrics**

The models were evaluated using the following metrics:

* **Accuracy**: Proportion of correctly classified incidents.
* **Precision**: Measure of true positives among predicted positives.
* **Recall**: Measure of true positives among actual positives.
* **F1 Score**: Harmonic mean of precision and recall, balancing both metrics.
* **Confusion Matrix**: Displayed model performance across different incident classes.

**4. Results and Analysis**

**4.1 Model Performance**

The Random Forest model emerged as the best-performing model, achieving the highest F1 score and accuracy across test data.

**4.2 Feature Importance**

Using feature importance analysis on the Random Forest model, the following features were identified as having the greatest influence on predictions:

* **Incident Grade**: Severity level had the highest impact on classification.
* **Alert Title**: Certain alert titles correlated strongly with specific incident categories.
* **User Account and IP Address**: Helped distinguish between benign and critical incidents

**5. Conclusion**

**5.1 Summary**

This project demonstrated that machine learning models, particularly Random Forest, can effectively classify cybersecurity incidents by analyzing key incident features. The model’s interpretability, enhanced through feature importance and SHAP values, provided actionable insights for cybersecurity operations.