Master 1, Data Science, IDSI – INPHB

Matière : **Forensic**

Enseignant : **Jean Marc Kouassi-Zessia**

Projet Tutoré

Classification et priorisation des évènements de sécurité par les méthodes de Machine Learning

1. Présentation du Projet

Le SOC (Security Operations Center) d'Orange CI est en charge de la détection et du traitement de tous les évènements sécurité. Les membres de cette équipe de supervision doivent analyser par jour des milliers d'évènements sécurité et déceler les plus pertinents en vue de répondre aux incidents sécurité menaçant les infrastructures de la structure.

En raison du volume de trafic croissant traité par les équipements d'Orange CI, ce centre reçoit de plus en plus d'évènements sécurité avec un fort taux de "**faux positif**", ce qui conduit les équipes à passer beaucoup plus de temps dans la phase "**Triage Alerte**" au lieu d'investiguer sur des cas réels d'incident sécurité. Cela crée d'énormes retards et ralentit considérablement la résolution des événements de sécurité critiques.

Pour résoudre ce problème, vous êtes contactés en tant que « **Expert Sécurité & IA** » pour mettre en place une **plateforme de classification et priorisation** de ces évènements à l'aide **de techniques de Machine Learning** dans le but de permettre aux analystes sécurité de déceler rapidement les cas d'incident réels.

Le thème de ce projet tutoré est donc : **Classification et priorisation des évènements de sécurité par les méthodes de Machine Learning.**

2. Objectif du projet

Comme susmentionné, le travail sera d'utiliser les techniques de Machine Learning adéquates pour réduire le nombre de fausses alertes dans la gestion des évènements de sécurité ; ce qui permettra de vite déceler les attaques en cours ou à venir.

Pour y parvenir nous procèderons comme suit :

- Classifier les évènements sécurité des données de l'EDR
- Prioriser ces évènements sécurité par degré de criticité
- Réduire le nombre de faux positifs dans les alertes fournies aux analystes sécurité

3. Livrable

A terme, on obtiendra une application **Web** fournissant un système de **visualisation dynamique riche** et interactif des évènements sécurité qui permettra de charger à la volé des extractions (format CSV) de la plateforme EDR et présenter le résultat des modèles de machine learning développés.

Un notebook Juyter vous est aussi demandé pour présenter le processus de traitement des « données labélisées » et de création de vos modèles de Machine Learning.

4. Démarche

- Features Engineering sur la base des données labélisées fournies
- Choix du model de machine Learning adéquat
- Développement de la plateforme de visualisation
- Test avec de nouvelles extractions de l'EDR