GROUPE: Bérangère BAZIRE, Jean-Charles MONCEAU, Vincent MOUTON-DUBOSC

Sujet: Challenge Cybersécurité de la Sorbonne. Création d'un algorithme d'analyse des comportements de logiciels et applications Windows, potentiellement malveillants, afin de fournir aux experts cyber des rapports automatisées permettant de focaliser leurs efforts.

1. Identifier les risques et incertitudes liés au contexte opérationnel, aux données et au manque de connaissances

Discussion d'experts (graphes, cybersécurité) : Organiser des ateliers avec les parties prenantes (experts métier, data scientists, ingénieurs ML) pour identifier collectivement les risques potentiels du type (cas rares, graphes trop grands, graphes avec représentation erronée du logiciel original, …)

Analyse de la documentation existante : Examiner les analyses de données(graphes), les modèles ML existants (l'état de l'art : cyber et graphes) et la documentation recherche sur les modèles de graphes et interprétation de logiciels désassemblés pour déterminer leur comportement

Audit des données : Réaliser un audit approfondi des données pour évaluer leur qualité, leur représentativité et leur pertinence. Identifier les biais potentiels, les valeurs manquantes et les anomalies.

- Filtrage des graphes vides, et fichier en erreur
- Répartition des labels par graphes et par rapport au dataset d'entrainement (class unbalanced, multi-label, ..)
- Cas rare: label present dans un seul graph d'entraînement => pas les moyens(pas ls connaissances) de faire de la data augmentation pour accroître la représentativité de ces cas
- Biais: commencer avec des features de graphs classiques , qui peuvent présenter des biais (spurious, confounding) sur les features réellement nécessaires

ML modele:

- Risques sur le volume de données d'entrée (100e de gigaoctets)
- Risques sur le temps / RAM pour exécution d'un modèle (GPU , RAM, …)
- Coût de mise en oeuvre vs bénéfice

Gestion de projet:

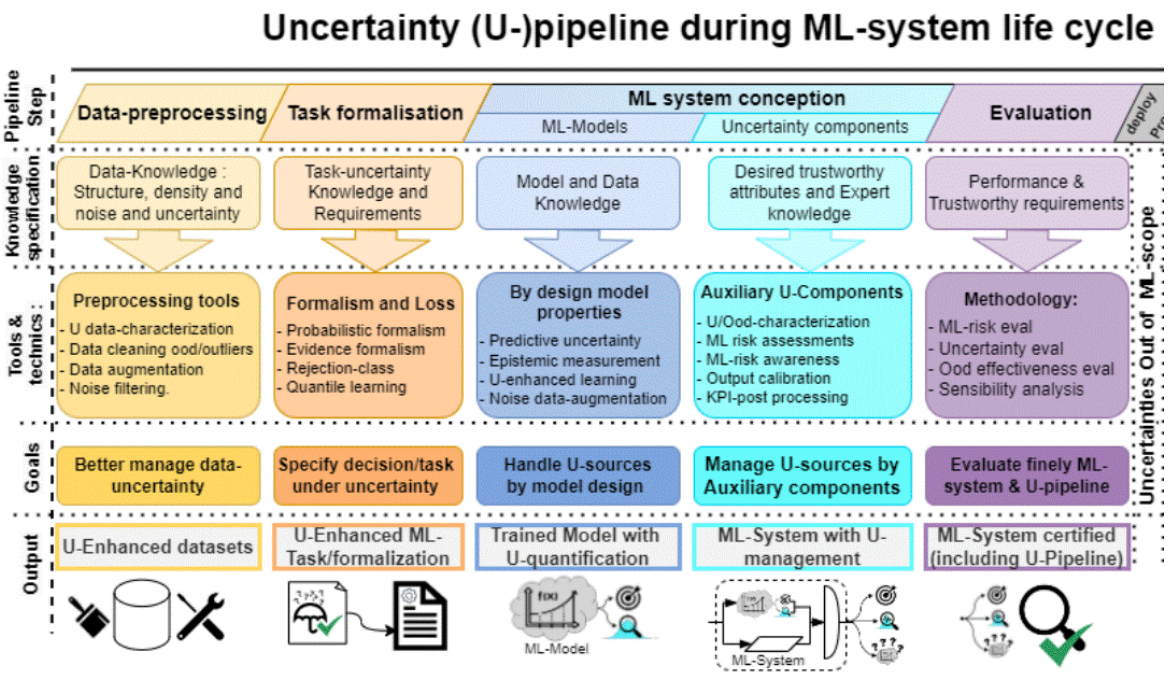
- Combiner le temps de conception par rapport au temps imparti pour le projet
- La répartition des tâches et collaboration en équipe (rôles, tâche, parallélisation, interface, documentation)

Matrice des risques : Créer une matrice des risques pour classer les risques par probabilité et impact, afin de prioriser les actions.

2. Analyse de leur impacts

Risks	Impacts	Mitigation / remediation / Activity
Délai / Planning projet	Pas fournir de livrable	Definir des lots/MVP, RACI
(manque d') expertise	Système pas performant	S'entourer d'experts (cyber, graphs), veille documentaire (recherche)
faisabilité/matériel	Perte de temps, limitation/incapacité à évaluer le système	- Sampling - Cloud computing - Commencer par Modèle + simple et augmenter progressivement la complexité vs performance
Qualité et complétude des Données	Performance du modèle (grande incertitude sur les résultats pour certains labels)	Conseils d'experts nécessaires (feature déterminantes) Gouvernance de la gestion des données
Applicabilité du modèle pour les utilisateurs finaux	Utilisation du modèle	Définir des KPI pertinents pour l'utilisateurs,

3. Spécification des activitées d'AI-Conception et U-mitigation dans le ML life cycle:



AI conception: Suivre et implémenter le Confiance BOK

- Data pre-processing: parcours des graphs (focus aux graphscomple, gestion de la taille des graphs,..), synthèse des infos sous forme de features
- Spécifications des métriques et KPI d'incertitudes pendant la creation du modele:
    - Confidence interval (par labels)
    - Robustesse : par groupe de labels et globale pour tout le model
    - Sensitivity: par groupe labels et par features

- AI-Design module:
    - Modèle Histogram-based Gradient Boosting Classification Tree utilisant un ensemble de métadonnées générées à partir des Control Flow Graph et des instructions en assembleur récupérés dans les fichiers (par exemple: nombre de noeuds et arêtes d'un graphe, taille du fichier, nombre d'instructions JMP présentes, etc.). Ce modèle simplifie grandement le volume de données à traiter, par la création d'un dataset de métadonnées qui s'abstrait des problèmes de variabilité due notamment à la taille des fichiers et du nombre d'instructions qu'ils contiennent.

- U-Mitigation module:
    - Human-in-the-loop : étant donné que notre système se base sur des métadonnées pour ses décisions de classification, il ne permet pas de fournir une explication "fine" du comportement logiciel (à la ligne de commande ou bloque d'instruction prêt). Par conséquent l'avis d'un expert est nécessaire pour trancher sur une décision, ceci peut être mis en place via un formulaire intéractif que l'expert peut compléter afin d'entériner la classification.

5. Protocole d'évaluation et enjeux de Monitoring.

- F1 score multi label (macro)
- Exactitude par étiquette
- Matrice de confusion multi label
- ROC multi label

Enjeu du monitoring:

- vérifier que le modèle reste pertinent lorsque de nouveaux binaires/graphs sont soumis au modèle.
- Le modèle reste fiable pour détecter des risques de cybersécurité