CAHIER DES CHARGES TECHNIQUE

- PROJET DE PIPELINE DE CYBERSECURITE -

SOMMAIRE

Infrastructure et environnement

Le projet sera exécuté sur une infrastructure cloud pour garantir la flexibilité et la scalabilité. Les ressources nécessaires incluent :

- Instance de calcul haute performance pour l'entraînement des modèles.
- Base de données pour stocker les données prétraitées.
- Serveur web pour le déploiement de l'API.
- Système de gestion de version (Git) pour le suivi du code source.

Les versions des logiciels à utiliser sont Python 3.11 pour la programmation, TensorFlow 2.5 pour la construction des modèles et Flask et Docker pour le déploiement de l'API.

Traitement des données

Les données sont liées aux activités de sécurité, à la surveillance des systèmes et à la détection des menaces, voici les sources :

- DataSet
- Journaux d'événements du réseau.
- Alertes de sécurité provenant de systèmes de détection d'intrusion.
- Données sur les vulnérabilités connues.
- Informations sur les activités anormales des utilisateurs.

Les traitements appliqués seront :

- Nettoyage : Suppression des doublons, traitement des valeurs manquantes par imputation ou suppression, vérification de la cohérence des données.
- Transformation : Codage one-hot pour les caractéristiques catégorielles, normalisation des données numériques.
- Intégration : Fusion des données de sources multiples et élimination des incohérences.

Développement des modèles

Le développement des modèles de prédiction se fera en plusieurs étapes :

- Exploration des données pour comprendre les distributions et les tendances.
- Division des données en ensembles d'entraînement, de validation et de test.
- Sélection de l'algorithme le plus pertinent pour la détection des attaques (après exploration des différents algorithmes : RNN,LSTM,régression linéaire multiple).
- Optimisation des hyperparamètres par recherche par grille ou validation croisée pour évaluer la performance des modèles.

Déploiement

Une fois le modèle entraîné et évalué, il sera déployé avec Docker en tant qu'API RESTful à l'aide du framework Flask.

- L'image Docker contiendra le modèle entraîné, l'API et toutes les dépendances.
- L'API recevra des requêtes HTTP pour les données d'entrée et renverra les prédictions.
- Le modèle sera hébergé sur un serveur web accessible par une adresse URL dédiée.

Sécurité et confidentialité

Les données seront stockées dans une base de données sécurisée et les communications avec l'API seront chiffrées à l'aide du protocole HTTPS.

L'accès à l'API sera contrôlé par des clés d'API et des mécanismes d'authentification.

Les données sensibles seront anonymisées avant d'être utilisées pour l'entraînement du modèle.

Plan de test

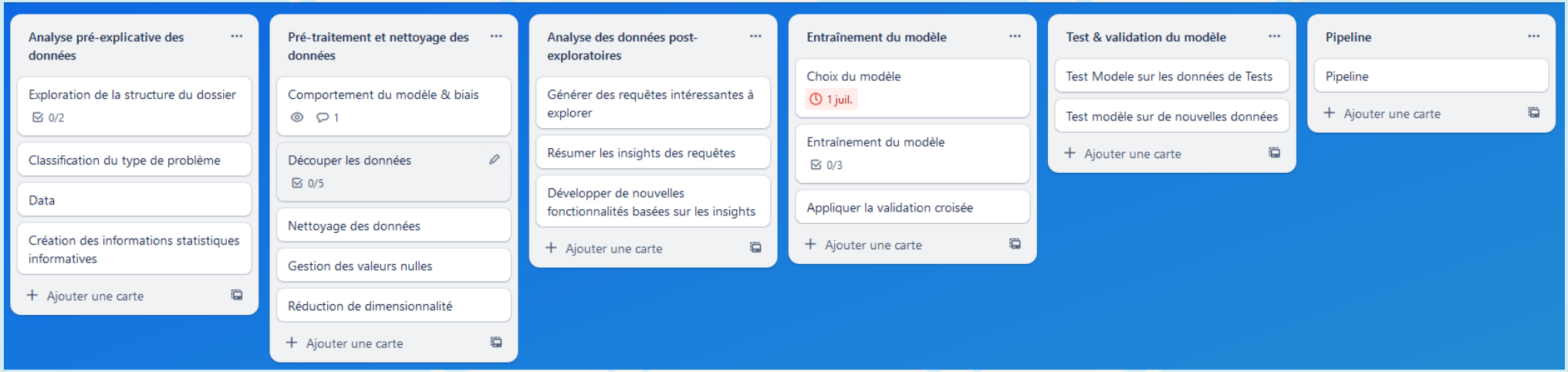
Voici les différents tests qui seront effectués :

- Tests unitaires : Vérification de chaque composant du système.
- Tests d'intégration : Vérification des interactions entre les composants.
- Tests de charge : Évaluation de la réactivité du modèle et de l'API sous charge.
- Tests de sécurité : Vérification des vulnérabilités potentielles.

Plan de Gestion de Projet

Les échéances clés sont les suivantes :

- Identification des tâches clés : Prétraitement des données, développement de modèles, déploiement, tests.
- Planification : Assignation des tâches, définition des échéances et des jalons.
- Suivi : Suivi continu de l'avancement, révision des échéances si nécessaire.
- Communication : Mise en place de réunions régulières pour partager les mises à jour et résoudre les problèmes.



Organigramme

La Maitrise d'œuvre (MOE)
L'équipe du projet est composée des membres suivants avec leurs rôles respectifs :

Jérôme FREGEFON
Data Scientist
Responsable du développement des modèles et de l'évaluation des performances.

Serge LACOMBE
Data Protection Officier
Responsable d'assurer la conformité aux RGPD et de superviser les mesures de sécurité des données

Pauline PERCEROU
Data Scientist & Développeur Full Stack

En charge de la collecte, du nettoyage et de l'intégration des données et responsable du déploiement et de la gestion de l'infrastructure cloud.

Conformité au RGPD

La conformité au Règlement général sur la protection des données (RGPD) revêt une importance cruciale dans le cadre de ce projet de pipeline de cybersécurité. En tant que réglementation majeure sur la protection des données personnelles, le RGPD impose des obligations strictes quant à la collecte, au stockage et au traitement des données des individus. Pour ce faire, des mesures de sécurité rigoureuses doivent être mises en place, incluant le cryptage des données sensibles, la gestion des consentements et le respect des droits des personnes concernées. Le pipeline doit garantir la traçabilité des données, faciliter la réponse aux demandes d'accès et assurer la notification en cas de violation de données. En veillant à la conformité au RGPD tout au long du processus de développement, notre entreprise CyberSentinel s'engage à maintenir les normes de protection des données requises, renforçant ainsi la confiance des utilisateurs et minimisant les risques juridiques potentiels.