Au niveau des métrics d'analyse :
Il peut étre intéressant de calculer le evasion rate : nb_attaque malclasse/nb_attaque
Indicateurs pour évaluer la robustesse : taux d'évasion , taux de détection , taux de fausse alerte.

Au sujet du CLI:
Il permettra une clarté pour la démo, peut être remplacé par un joli notebook. Il risque de prendre trop de temps pour ce dernier sprint.
Le pré-entrainement sera autorisé. Prévoir des dossiers test à part.
Prévoir un environnement exprès pour la démo.

Le substitut est plus efficace, car il permet de mapper les résultats aux prédictions voulus. On se rapproche beaucoup plus du même taux de prédiction du RF comme ça. (Oui je l'ai constaté aujourd'hui).

Le taux d'évasion semble binaire ? À vérifier mais cela semble dépendre du dataset considéré (vérifier le code)

Pour les MLP, pour ceux moins bien entrainés la perturbation est plus longue à prendre pas. Notre piste est la sensibilité de ces MLP mais ce n'est pas forcément une bonne chose de conseiller des MLP pas trop entrainé.

Proposer une piste de défense :
complexifier les architectures, en général c'est RF ou knn qui sont utilisé dans IDS.
Faire la comparaison, mettre en avant les différences entre architectures. Voir le substitut pour le RF (fait le 25/11).
Proposer juste des pistes d'amélioration, c'est ok, on risque de pas avoir le temps de modifier les architectures des réseaux. (mais le client veut quand même des pistes d'amélioration de robustesse donc il faut en permettre).
Faire une ouverture serait intéressant notamment sur la robustesse.

Comparaison des différentes architectures : statistiques , métrics etc
à voir avec les autres intervenants, mais on risque de manquer de temps pour refaire des tests.
Ouverture sur les meilleurs résultats qu'on a par exemple.

Au sujet du SMOTE :
Si il y a création d'échantillon synthétique incohérent, on risque d'entrainer les modèles sur de mauvaises données. Cela donnera un risque d'apprentissage de mauvaises relations logiques.
Dans le cas de CICID on crée moins de 15% de données par SMOTE (donc pas trop problématique).
Le fonctionnement biclasse est ok pour le dataset UNSW et permet de comparer multiclasse/biclasse avec deux datasets différents.
Envoyer un mail au PO sur ce qu'il attend en terme de résultats de robustesse.

Entraînement des modèles adverserials pour améliorer la robustesse des modèles -> piste pour l'ouverture/résultats
Augmenter les performances des modèles.
Scale sur les KNN? Apparemment oui on scale pour permettre la comparaison entre features de différent type.

Passons le moins de temps possible sur knn