Dr DIAKO	**Projet** **Détection avancée des botnets à l'aide des algorithmes du Machine Learning**	

Sujet : Détection des botnets par les algorithmes du Machine Learning : une approche comparative entre K-Modèle des plus proches voisins, Modèle d'arbre de décision et Modèle gaussien de Naive Bayes

Introduction :

Les botnets sont des réseaux de machines infectées par des logiciels malveillants, contrôlés par des cybercriminels pour effectuer diverses actions illicites. La détection précoce et précise de ces botnets est cruciale pour assurer la sécurité des réseaux informatiques et prévenir les cyberattaques. En tant qu'expert en cybersécurité et Intelligence Artificielle, nous vous proposons d'explorer la détection des botnets à l'aide des algorithmes du Machine Learning.

Dataset et variable cible :

Pour cette étude, nous utiliserons le dataset "network-logs.csv", qui contient des logs réseau avec les caractéristiques suivantes : REMOTE_PORT, LATENCY, THROUGHPUT. Notre objectif sera de prédire la variable cible ANOMALY, qui indique la présence d'une activité anormale pouvant être associée à un botnet.

Caractéristiques du dataset :

REMOTE_PORT : Le numéro de port distant (REMOTE_PORT) est un entier compris entre 0 et 65535 qui identifie un point de terminaison de communication dans un réseau informatique. Les ports sont utilisés pour acheminer les données vers des processus ou des services spécifiques sur un hôte distant.

LATENCY : La latence (LATENCY) est une mesure du temps de retard entre l'envoi d'un paquet de données et la réception de la réponse correspondante. Elle est généralement exprimée en millisecondes (ms).

THROUGHPUT : Le débit (THROUGHPUT) est une mesure de la quantité de données transmises entre deux hôtes en une unité de temps donnée, généralement exprimée en bits par seconde (bps) ou en octets par seconde (Bps).

Algorithmes de Machine Learning :

Nous mettrons en œuvre et comparerons trois algorithmes de Machine Learning :

K-Modèle des plus proches voisins (K-Nearest Neighbors, KNN) : Cet algorithme classe les instances en fonction de leur similarité avec les instances voisines dans l'espace des caractéristiques. Nous évaluerons les performances de KNN en ajustant le nombre de voisins K et en utilisant différentes métriques de distance.

Modèle d'arbre de décision : Les arbres de décision sont des algorithmes de classification qui apprennent à partir des données en construisant un arbre de décision, où chaque nœud représente une décision basée sur une caractéristique et chaque feuille représente une classe. Nous évaluerons les performances de l'arbre de décision en ajustant les paramètres tels que la profondeur de l'arbre et le critère de division.

Modèle gaussien de Naive Bayes : Cet algorithme est basé sur le théorème de Bayes et l'hypothèse d'indépendance conditionnelle des caractéristiques. Nous évaluerons les performances du modèle gaussien de Naive Bayes en supposant que les caractéristiques suivent une distribution gaussienne.

Comparaison et recommandations :

En comparant les performances de ces trois algorithmes sur le dataset "network-logs.csv", nous pourrons déterminer lequel est le plus adapté à la détection des botnets et fournir des recommandations pour améliorer la sécurité des réseaux informatiques.