# ML Project

## Intrusion Detection System

**

*Préparé par :*

*Mohamed Wahchi*

*Malek Hmem*

*Kalthoum Dridi*

*Selim Sahli*

*Nour Hassine*

*Khouloud Zograni*

Introduction

À l'ère de la connectivité numérique, où les réseaux informatiques sont devenus les fondations essentielles de notre monde interconnecté, la sécurité des données et des systèmes revêt une importance capitale. Les progrès technologiques ont créé un paysage dynamique, mais ont également ouvert la porte à une diversité croissante de menaces cybernétiques. Dans ce contexte, la préservation de l'intégrité des réseaux devient une priorité absolue.

Notre projet s'inscrit dans cet environnement complexe et en constante évolution, cherchant à renforcer la sécurité des réseaux à travers des approches innovantes. Alors que les défenses traditionnelles peuvent parfois se révéler insuffisantes face à des attaques sophistiquées et émergentes, notre démarche se concentre sur l'intégration de solutions avancées.

Nous explorons spécifiquement les possibilités offertes par les techniques d'apprentissage automatique pour anticiper, détecter, et contrer les menaces émergentes, tout en minimisant les risques de faux positifs et de faux négatifs.

En adoptant une perspective holistique, notre projet vise à contribuer à l'évolution des défenses numériques, offrant une réponse adaptative aux défis de sécurité contemporains. En mettant l'accent sur l'innovation, nous aspirons à renforcer la résilience des réseaux, assurant ainsi la stabilité et la sécurité des infrastructures informatiques au cœur de notre société interconnectée.

C'est dans ce paysage complexe de la cybersécurité que les Systèmes de Détection d'Intrusion (IDS) émergent comme des acteurs essentiels, dédiés à la protection des réseaux contre des menaces en constante évolution. Ces systèmes, conçus pour anticiper et contrer les activités malveillantes, jouent un rôle crucial dans le maintien de la sécurité des réseaux. Leur importance découle de leur capacité à identifier non seulement les attaques connues, mais aussi à détecter des anomalies et des schémas inédits grâce à des méthodes avancées telles que l'apprentissage automatique.

1. Etude du sujet :
2. **Définition de l’IDS :**

Les Systèmes de Détection d'Intrusion (IDS) sont des piliers essentiels de la cybersécurité contemporaine, jouant un rôle crucial dans la protection des réseaux informatiques contre une diversité de menaces en constante évolution. Ces systèmes sont conçus pour surveiller en temps réel les activités du réseau, identifiant les comportements malveillants et les anomalies qui pourraient indiquer une tentative d'intrusion.

L'importance cruciale des IDS dans la cybersécurité moderne réside dans leur capacité à anticiper, détecter et répondre rapidement aux menaces émergentes.

En offrant une protection proactive, les IDS renforcent la résilience des réseaux, assurant ainsi la confidentialité, l'intégrité et la disponibilité des données dans un paysage numérique en constante évolution.

1. **Types de l’IDS :**

La construction d'un IDS repose sur deux approches fondamentales : les IDS basés sur des signatures et les IDS basés sur des anomalies.

**IDS basés sur des Signatures :**

* Identification des attaques par la comparaison avec des signatures préalablement identifiées.
* Utilisation de bases de données de signatures pour repérer des schémas de données malveillants.
* Efficaces contre des attaques bien connues et documentées.
* Limitations face à des attaques inédites qui ne correspondent pas à des signatures connues.
* Fonctionnent principalement sur la reconnaissance de modèles spécifiques associés à des attaques connues.

**IDS basés sur des Anomalies :**

* Exploitation de techniques d'apprentissage automatique pour analyser le trafic en temps réel.
* Détection de comportements anormaux en comparant le flux de données actuel avec des modèles historiques.
* Adaptatifs aux attaques nouvelles et émergentes.
* Risque de générer des faux positifs en identifiant parfois des activités normales comme anormales.
* Recherche de déviations par rapport aux schémas normaux plutôt que de se baser sur des signatures spécifiques.

Ces deux approches complémentaires sont souvent intégrées dans des systèmes d'IDS hybrides pour tirer parti de leurs forces respectives et atténuer leurs faiblesses respectives :

**IDS Hybrides :**

* Intègrent à la fois des signatures et des techniques d'analyse comportementale pour une détection plus robuste.
* Combinent les avantages des approches basées sur des signatures et des anomalies pour une meilleure couverture de sécurité.

1. **Les formes de l’IDS:**

Les systèmes de détection d'intrusions (IDS) sont des gardiens essentiels de la cybersécurité, adoptant diverses formes pour prévenir et identifier les menaces numériques. Que ce soit à travers des approches basées sur des signatures ou des anomalies, des déploiements réseau, hôtes, sans fil, physiques, virtuels ou même dans le cloud, chaque forme d'IDS offre des avantages spécifiques dans la protection des infrastructures contre les attaques informatiques.

1. **L’IDS Réseau :**

* Surveillent le trafic sur le réseau pour identifier les activités suspectes.
* Peuvent être déployés à différents niveaux du réseau, du périmètre externe aux segments internes.

1. **L’IDS Hôte :**

* Se concentrent sur l'activité au niveau des hôtes individuels, tels que les serveurs ou les postes de travail.
* Détectent les tentatives d'intrusion en surveillant les fichiers système, les journaux d'événements, etc.

1. **IDS Sans Fil :**

* Spécifiquement conçus pour surveiller et détecter les menaces sur les réseaux sans fil.
* Identifient les comportements suspects dans le trafic Wi-Fi et les communications sans fil.

1. **L’IDS Physiques et Virtuels :**

* Les IDS physiques sont des appareils matériels dédiés, tandis que les IDS virtuels sont des logiciels exécutés sur des machines virtuelles.
* Offrent une flexibilité de déploiement en fonction des besoins de l'infrastructure.

1. **Le Cloud-based IDS :**

* S'adaptent aux environnements cloud en surveillant le trafic et les activités au sein des services cloud.
* Fournissent une visibilité dans des environnements cloud dynamiques et évolutifs.

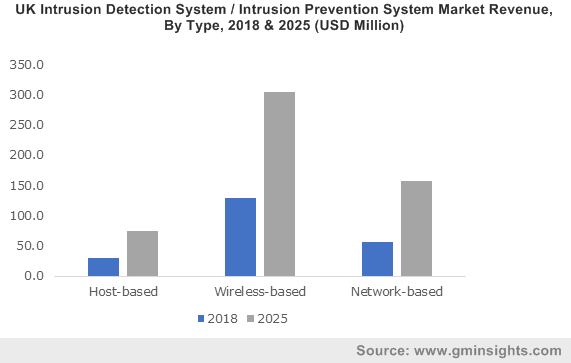
### **Statistiques mondiales sur l’IDS:**

La menace croissante de violations de données et la croissante de la numérisation ont été responsables de la croissance de la part de marché des systèmes de détection d'intrusions au cours des dernières années.

La mise en œuvre de systèmes de détection d'intrusions (IDS) dans le système réseau permet incontestablement aux administrateurs réseau d'évaluer l'emplacement précis des menaces et de les traiter. Ces systèmes contribuent également à prévenir de nouvelles menaces en ces endroits, accélérant ainsi la croissance de l'industrie des IDS.

Le graphique ci-dessus illustre la répartition des revenus du marché britannique des IDS selon trois types : basé sur l'hôte, basé sur le sans fil et basé sur le réseau, en 2018 et les projections pour 2025.

On observe une croissance significative des revenus, mettant en évidence la demande croissante pour ces types spécifiques de systèmes de détection d'intrusions au Royaume-Uni. Ces chiffres reflètent la reconnaissance croissante de l'importance des IDS dans le contexte de la cybersécurité au Royaume-Uni.

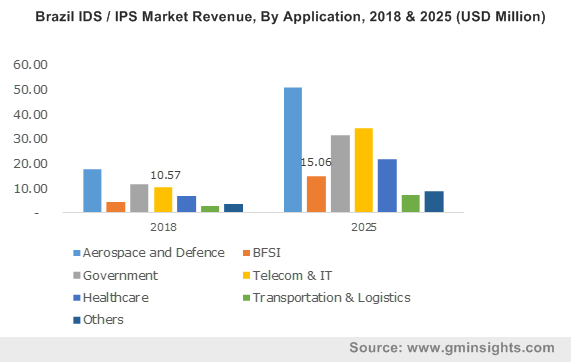


On peut aussi citer que selon des statistiques récentes, la demande d'IDS a connu une croissance significative dans divers secteurs tels que la finance, la santé, les télécommunications et l'énergie.

Une analyse des parts du marché au Brésil, de 2019 à 2025, révèle une tendance marquée vers une adoption accrue de l'IDS.

Le secteur de la défense aérospatiale et le domaine de la santé, confrontés à des défis croissants liés à la confidentialité des dossiers médicaux, ont intensifié leur utilisation d'IDS pour garantir la sécurité des informations médicales.

Cette tendance est corroborée par un graphique illustrant la répartition des parts de marché par revenu, soulignant l'importance croissante de l'IDS dans le paysage de la cybersécurité au Brésil. Cette évolution démontre clairement que les entreprises reconnaissent l'IDS comme une stratégie indispensable pour faire face aux menaces cybernétiques de plus en plus sophistiquées et souligne la nécessité continue d'investir dans des solutions de détection des intrusions pour assurer la résilience des infrastructures numériques.



1. Business Problem :

Dans le paysage en constante évolution de la cybersécurité, la sophistication croissante et la diversité des menaces cybernétiques posent un défi significatif aux mesures de sécurité traditionnelles telles que les pare-feu.

L'objectif principal du projet de Système de Détection d'Intrusion (IDS) est de remédier à l'insuffisance des défenses conventionnelles et de renforcer le réseau contre les attaques cybernétiques nouvelles et inconnues. Conscient que la détection basée sur des signatures est efficace contre les attaques connues mais limitée face aux menaces émergentes, l'accent est mis sur les IDS basés sur des anomalies utilisant des techniques d'apprentissage automatique.

Le lancement de ce projet découle du besoin critique de renforcer les défenses de cybersécurité face à l'évolution des menaces cybernétiques. En explorant l'application d'algorithmes d'apprentissage automatique, en particulier Decision Tree et KNN, sur des ensembles de données bien établis tels que notre dataset NSL KDD, le projet vise à apporter des connaissances précieuses sur l'efficacité de ces algorithmes pour la détection d'intrusions.

En fin de compte, ce projet vise à faire progresser la compréhension des applications pratiques de l'apprentissage automatique dans la détection d'intrusions et à contribuer aux efforts continus pour renforcer la sécurité des réseaux dans un monde de plus en plus numérisé.

1. Business Understanding :
2. Data Understanding :
3. **Les données sur l’utilisateur :**

|  |  |
| --- | --- |
| Nom de la colonne : | Description : |
| Gender | **Le sexe de l’utilisateur :**  Male / Femelle (Donnée de type string) |
| Age | **L’age moyen de l’utilisateur :**  21/46/26/31/41/50plus/36/below21 (Donnée de type string) |
| Marital Status | **L’état civil de l’utilisateur:**   Unmarried partner, Single, Married partner, Divorced, Widowed (Donnée de type string) |
| Has\_Children | **Variable indiquant si l’utilisateur a des enfants :**  1:has children/0:no children (donnée de type int) |
| Education | **Le niveau d’éducation de l’utilisateur:**  Some college — no degree, Bachelors degree, Associates degree, High School Graduate, Graduate degree (Masters or Doctorate), Some High School (Donnée de type string) |
| Occupation | **L’occupation que l’utilisateur occupe:**  Unemployed, Architecture & Engineering, Student,Education&Training&Library, Healthcare Support,Healthcare Practitioners & Technical, Sales & Related, Management,Arts Design Entertainment Sports & Media, Computer & Mathematical,Life Physical Social Science, Personal Care & Service, Community & Social Services, Office & Administrative Support, Construction & Extraction, Legal, Retired, Installation Maintenance & Repair, Transportation & Material Moving, Business & Financial, Protective Service, Food Preparation & Serving Related, Production Occupations, Building & Grounds Cleaning & Maintenance, Farming Fishing & Forestry (Donnée de type string) |
| Income | **Le salaire annuel de l’utilisateur :**  Less than $12500,$12500 — $24999,$25000 — $37499,$37500 — $49999,$50000 — $62499,$62500 — $74999,$75000 — $87499,$87500 — $99999,$100000 or More (Donnée de type string) |
| Car | **Le véhicule que l’utilisateur conduit :**  Scooter and motorcycle, crossover, Mazda5 (Donnée de type string) |
| Bar | **Combien de fois l’utilisateur fréquente le bar chaque mois :** never, less1, 1~3, 4~8, gt8, nan (Donnée de type string) |
| CoffeeHouse | **Combien de fois l’utilisateur fréquente un salon de thé chaque mois :**  never, less1, 1~3, 4~8, gt8, nan (Donnée de type string) |
| CarryAway | **Combien de fois l’utilisateur prend sa commande à emporter chaque mois :**  never, less1, 1~3, 4~8, gt8, nan (Donnée de type string) |
| RestaurantLessThan20 | **Combien de fois l’utilisateur fréquente un restaurant avec un montant inférieur à 20$ chaque mois :**  never, less1, 1~3, 4~8, gt8, nan (Donnée de type string) |
| Restaurant20To50 | **Combien de fois l’utilisateur fréquente un restaurant avec un montant moyen de 20$ à 50$ chaque mois :**  never, less1, 1~3, 4~8, gt8, nan (Donnée de type string) |

1. **L’attribut cible (output) :**

Conclusion