

Machine Learning



- Dispensé par Dr. Msc. Ir. **MWAMBA KASONGO Dahouda**
- Docteur en génie logiciel et systèmes d'information
- Machine and Deep Learning Engineer
- E-mail : dahouda37@gmail.com

- Assisté Ass. **MWAMBA KASIKO Evariste**
E-mail : mwambakasiko@aol.com

Heure : 08H00 – 12H00

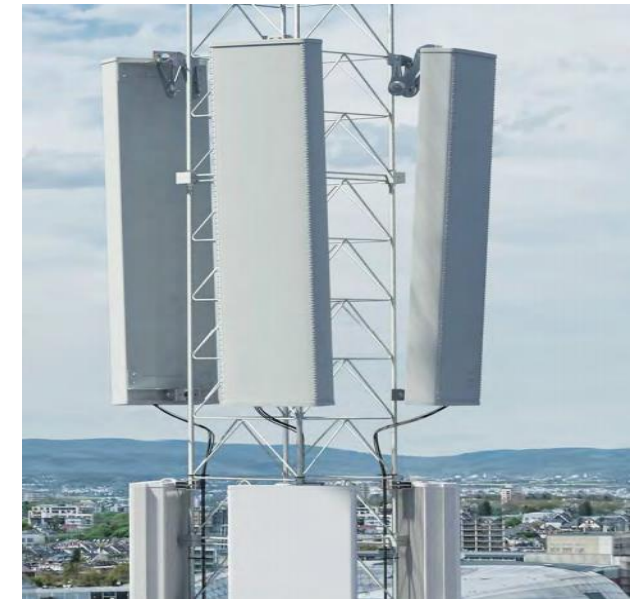
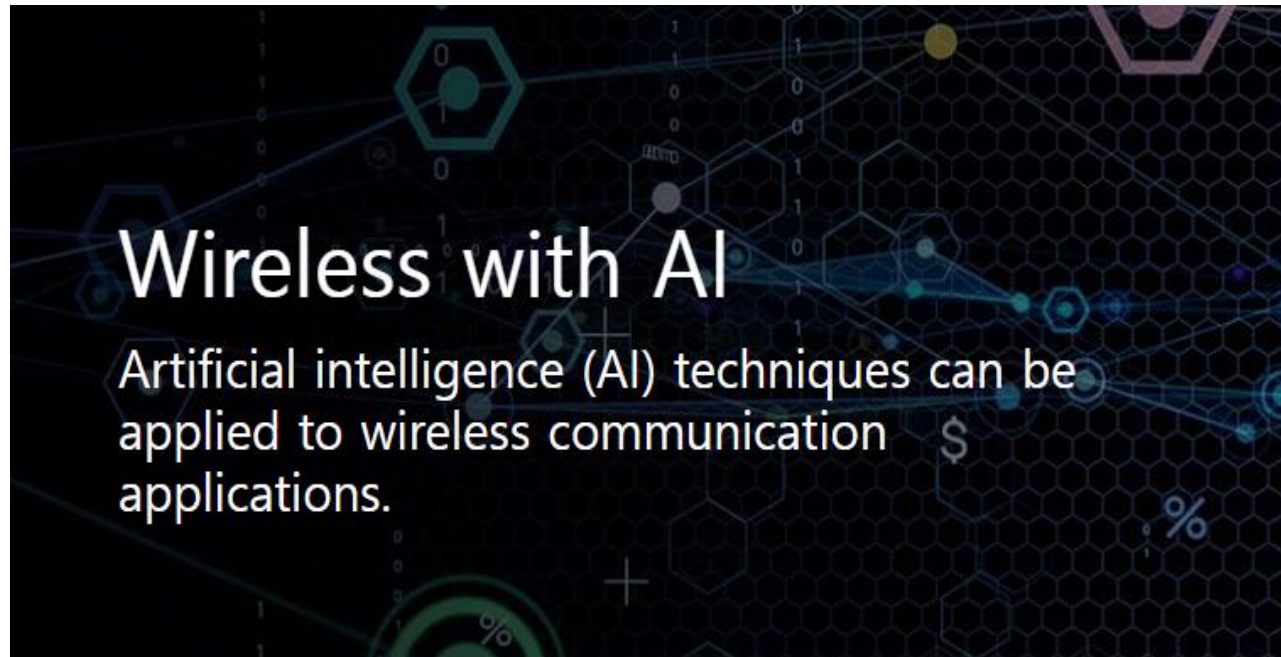


CHAPITRE 1 L'IA dans les Réseaux



Ce chapitre vise à fournir une compréhension approfondie de la manière dont l'IA peut être intégrée aux technologies de réseau pour optimiser les performances, améliorer la sécurité et automatiser les tâches de gestion.

Nous allons comprendre les principales techniques d'IA et leurs applications dans les infrastructures de réseau modernes.



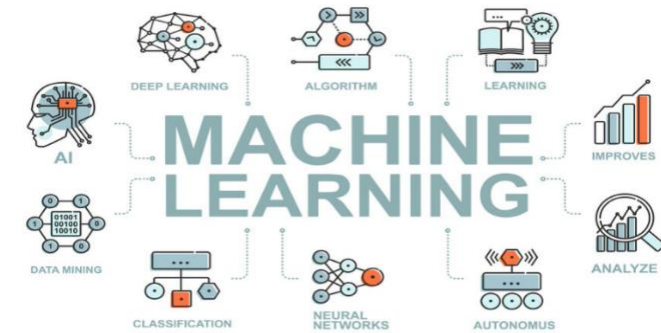
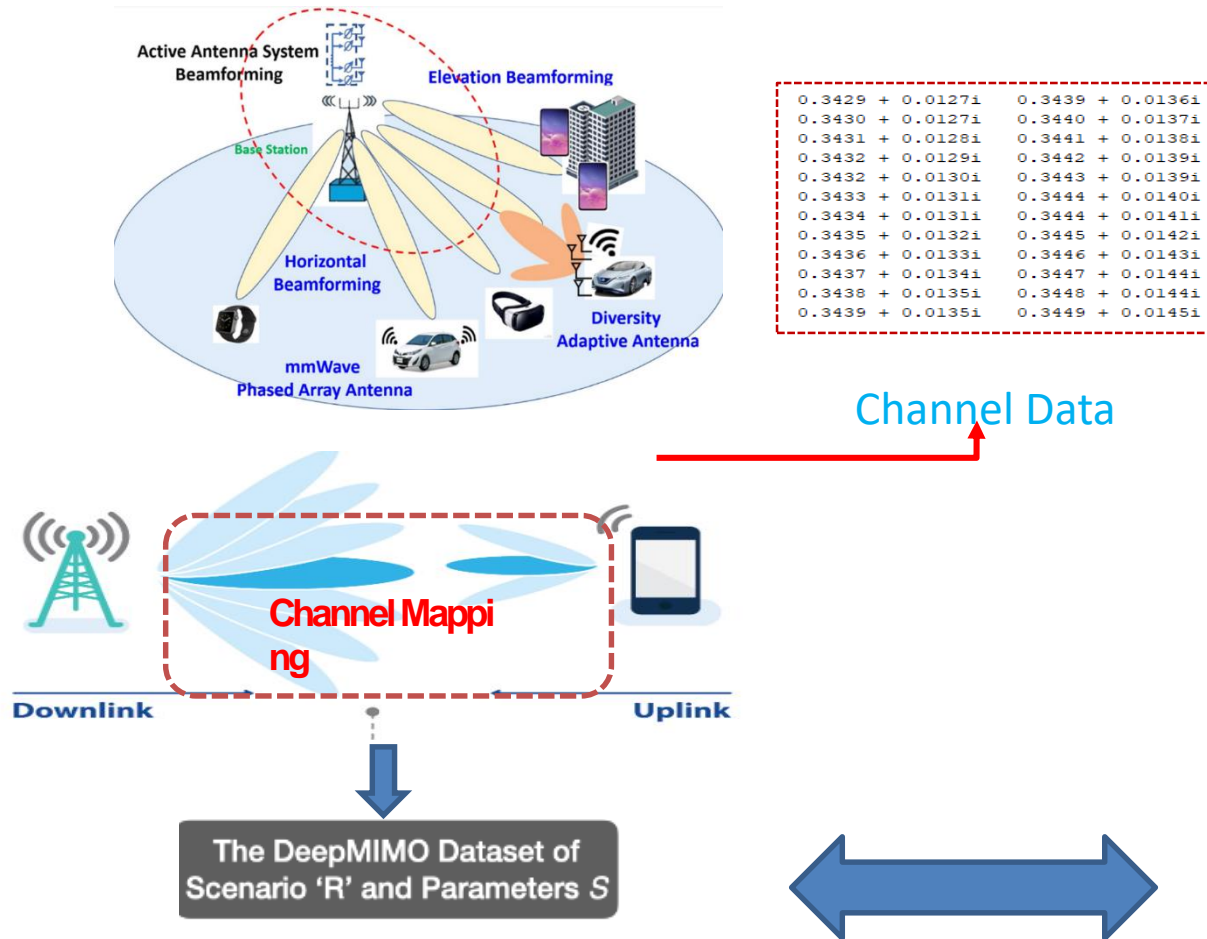
L'intelligence artificielle (IA) pour le sans fil est déjà là, avec des applications dans des domaines tels que la gestion de la mobilité, la détection et la localisation, la signalisation intelligente et la gestion des interférences ainsi que la sécurité de Traffic réseaux.



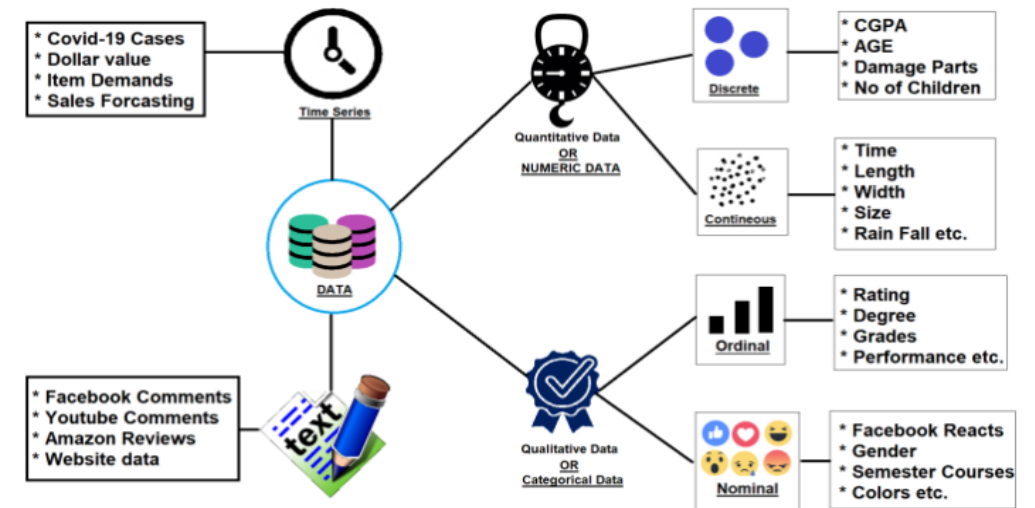
Machine Learning

CHAPITRE 1 L'IA dans les Réseaux

1.1. Applications de machine Learning les réseaux 5G



Comprendre les données pour l'apprentissage automatique



DeepMIMO est un ensemble de données générique pour les canaux mmWave/massive MIMO.

➤ En tant que données de canal

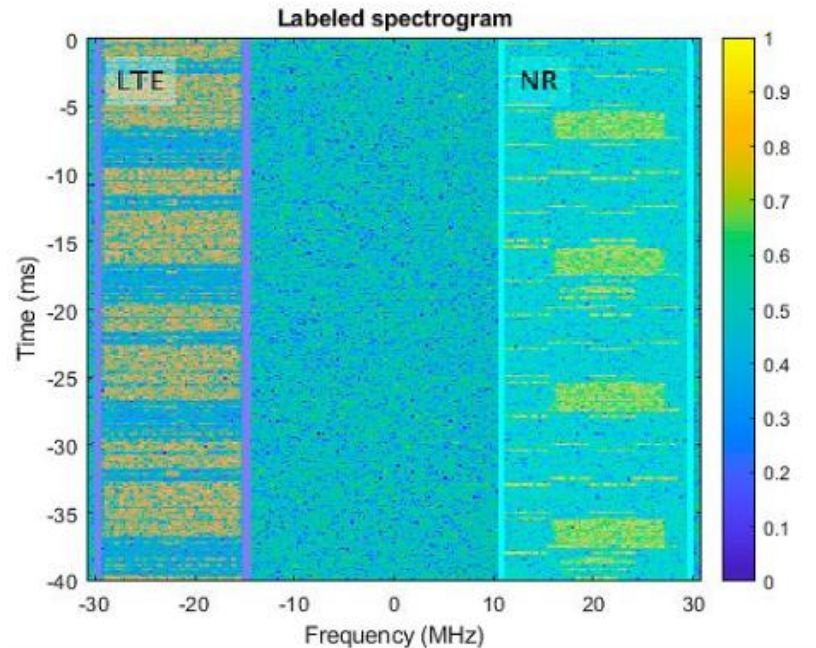
IA (Reseau et Telecom)-Dahouda, Ph.D.



1.1. Applications de machine Learning les réseaux 5G

❖ Pourquoi utiliser l'IA pour le sans fil ?

1. Détection spectrale et classification du signal



Application de Deep Learning à la communication sans fil :

- ✓ Identification des signaux 5G et LTE avec détection de spectre à l'aide de l'apprentissage profond.

La 5G est le réseau de nouvelle génération conçu pour répondre aux limites de la LTE et permettre l'utilisation de nouvelles technologies. Elle offre des vitesses plus rapides, une latence plus faible et la possibilité de prendre en charge davantage d'appareils.

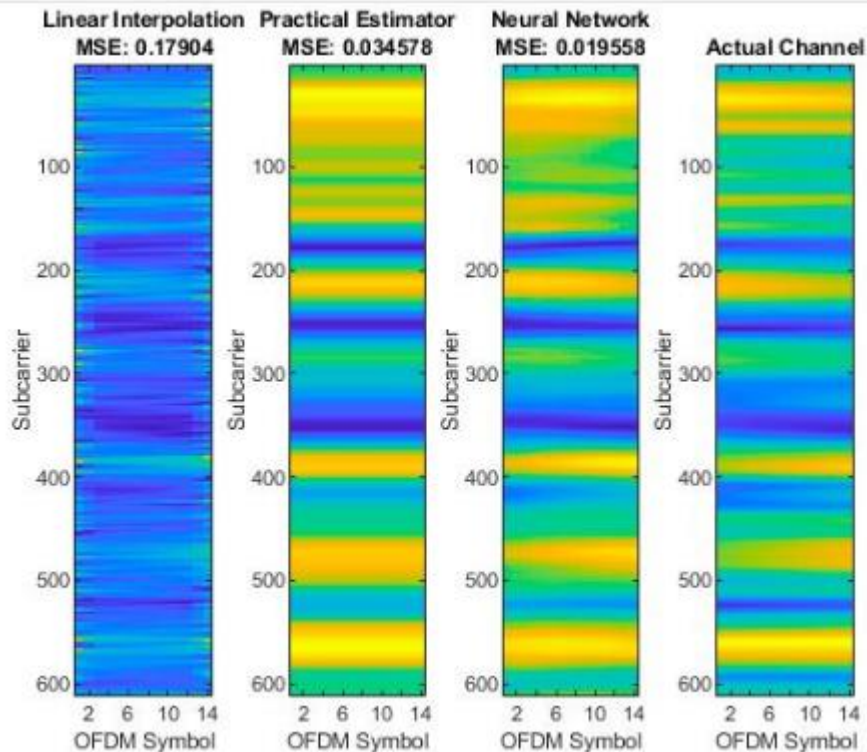
La LTE reste un réseau essentiel et hautement performant, mais la 5G est son successeur, conçu pour répondre à la demande croissante de connectivité, d'appareils intelligents et d'applications gourmandes en données.



1.1. Applications de machine Learning les réseaux 5G

❖ Pourquoi utiliser l'IA pour le sans fil ?

2. Gestion des faisceaux et estimation des canaux



Application de Deep Learning à la communication sans fil :

- ✓ Réseau neuronal pour la sélection de faisceaux
- ✓ Synthèse de données pour l'estimation des canaux 5G avec Deep learning
- ✓ Les réseaux neuronaux peuvent être utilisés pour réduire la complexité de calcul dans les tâches de sélection de faisceaux **5G NR (New Radio)**.
- ✓ Un CNN peut être formé pour l'estimation des canaux 5G NR



1.1. Applications de machine Learning les réseaux 5G

❖ 5G NR (New Radio)

- ❑ 5G NR (New Radio) est la norme mondiale pour l'interface aérienne des réseaux 5G.
- ❑ Développé par 3GPP (3rd Generation Partnership Project).
- ❑ La 5G NR utilise la technologie MIMO massive, qui consiste à déployer de grands réseaux d'antennes sur une seule station de base pour augmenter la capacité et améliorer la couverture.
Cela permet une meilleure gestion de plusieurs utilisateurs et appareils dans les zones encombrées.

❖ Formation de faisceaux [Beamforming]

- ❑ La 5G NR utilise la formation de faisceaux, une technique qui concentre les signaux radio dans une direction spécifique vers un appareil, plutôt que de diffuser des signaux dans toutes les directions.
Cela améliore à la fois la puissance du signal et la couverture, en particulier dans les zones présentant de nombreux obstacles.
- ❑ La formation de faisceaux et le Machine Learning (ML) jouent un rôle synergique dans l'optimisation des communications sans fil, en particulier dans les réseaux 5G



1.1. Applications de machine Learning les réseaux 5G

❖ 5G NR (New Radio)

Applications de machine Learning (ML) dans la formation de faisceaux :

- ✓ **Gestion des faisceaux en temps réel** : Dans les environnements comportant de nombreux obstacles (par exemple, les zones urbaines avec des bâtiments), les signaux peuvent être réfléchis ou bloqués. Les modèles ML peuvent prédire la direction optimale pour la formation de faisceaux en analysant les données passées sur les réflexions de signaux et les modèles d'interférence.
- ✓ **Prédiction des mouvements de l'utilisateur** : Les algorithmes de ML peuvent prédire les mouvements de l'utilisateur (par exemple, en voiture ou à pied), permettant au réseau d'ajuster de manière proactive la direction du faisceau. Cela garantit une connectivité continue et solide, même lorsque l'utilisateur se déplace à grande vitesse.
- ✓ **Estimation des canaux** : le ML peut être utilisé pour prédire les meilleurs canaux de transmission en analysant le bruit, les interférences et la charge de trafic. Cela permet d'optimiser la sélection de fréquence pour la formation de faisceaux.
- ✓ **Allocation dynamique du spectre** : ML peut aider à allouer de manière dynamique des ressources de spectre à différents utilisateurs en fonction de la demande du réseau en temps réel et des modèles d'utilisation. Cela optimise encore davantage la manière dont la formation de faisceaux dirige les signaux.



CHAPITRE 1 L'IA dans les Réseaux



1.1. Applications de machine Learning les réseaux 5G

❖ Technique de formation de faisceaux)

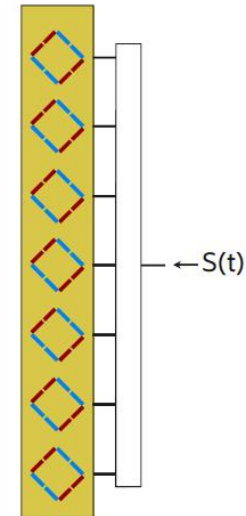
■ Qu'est-ce que la formation de faisceaux ?

La formation de faisceaux est une technique utilisée pour améliorer le rapport signal/bruit des signaux reçus (SNR), éliminer les sources d'interférences indésirables et concentrer les signaux transmis vers des emplacements spécifiques.

La formation de faisceaux est essentielle aux systèmes dotés de réseaux de capteurs, notamment aux systèmes de communication sans fil MIMO tels que 5G, LTE et WLAN.

■ Comment ça marche ?

- Les antennes sont fournies avec des diagrammes d'antenne (faisceaux) :
- ✓ Chaque antenne a son propre diagramme de faisceau ;
- ✓ Les antennes à inclinaison électrique sont fournies avec plusieurs diagrammes de faisceau, un pour chaque inclinaison.
- Mais qu'est-ce qui se cache derrière la forme et la direction du faisceau ?



CHAPITRE 1 L'IA dans les Réseaux

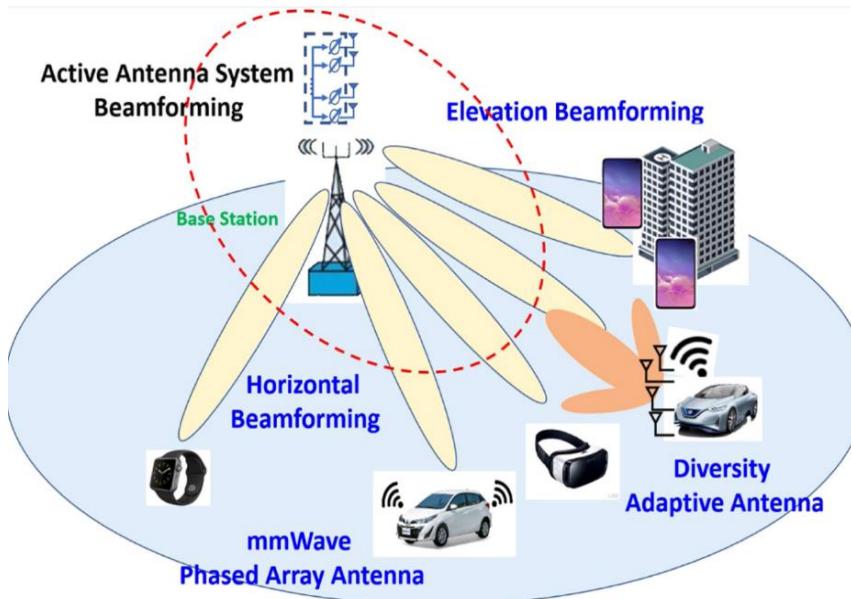
1.1. Applications de machine Learning les réseaux 5G

❖ Technique de formation de faisceaux)

- Concept de réseau d'antennes dans la formation de faisceaux faisceaux ?

La formation de faisceaux est une technique qui focalise un signal sans fil vers un appareil de réception spécifique, plutôt que de le diffuser dans toutes les directions, comme avec une antenne de diffusion.

La connexion directe qui en résulte est **plus rapide** et **plus fiable** que ce qu'elle serait sans formation de faisceaux.



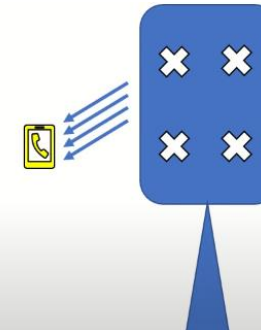
2T2R

This is a 2T2R antenna which can do 2x2 MIMO (low capacity)



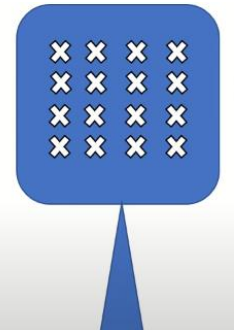
4T4R

This is a 4T4R antenna which can do 4x4 MIMO (Medium capacity)



16T16R

This is a 16T16R antenna which can do 16x16 MIMO (High capacity)



CHAPITRE 1 L'IA dans les Réseaux

1.1. Applications de machine Learning les réseaux 5G

❖ Technique de formation de faisceaux)

- Mécanisme de formation de faisceau : Derrière la forme du faisceau

Taille de Tableau petite

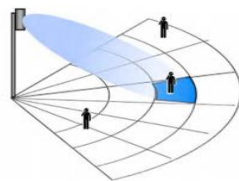
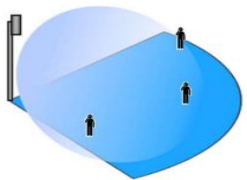
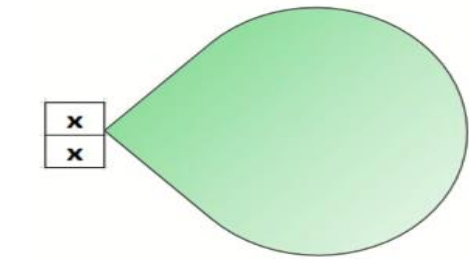
Un petit nombre d'éléments Tx peut générer des faisceaux avec **une largeur de faisceau plus grande**.

Par conséquent, ils sont utiles dans les cas où nous souhaitons couvrir un espace plus large avec un coût minimum.

Taille de Tableau plus grande

Cependant, à mesure que nous ajoutons plus d'éléments Tx, le faisceau devient **plus étroit**.

Mais le faisceau devient également **plus directionnel**.



Massive MIMO (Low Tx/Rx)



Massive MIMO (High Tx/Rx)



CHAPITRE 1 L'IA dans les Réseaux

1.1. Applications de machine Learning les réseaux 5G

❖ Technique de formation de faisceaux)

▪ Gains de formation de faisceau

- Plus le nombre d'éléments est élevé :
- ✓ Plus le faisceau est étroit
- ✓ Plus le gain d'antenne est élevé
- Si nous avons plus de faisceaux directionnels, **la couverture sera plus forte.**
- Cela signifie que si des données sont envoyées sur ces faisceaux, les utilisateurs pourront obtenir un meilleur **SINR**, et par conséquent un meilleur débit.
- Cela augmentera également la limite de la cellule de trafic, car les utilisateurs obtiendront de **meilleurs débits** de données même dans des conditions de couverture plus faibles.



Multiple antenna increase coverage and capacity.



CHAPITRE 1 L'IA dans les Réseaux



1.1. Applications de machine Learning les réseaux 5G

❖ Technique de formation de faisceaux)

- Gains de formation de faisceau: Bon débit

Zone à forte interférence

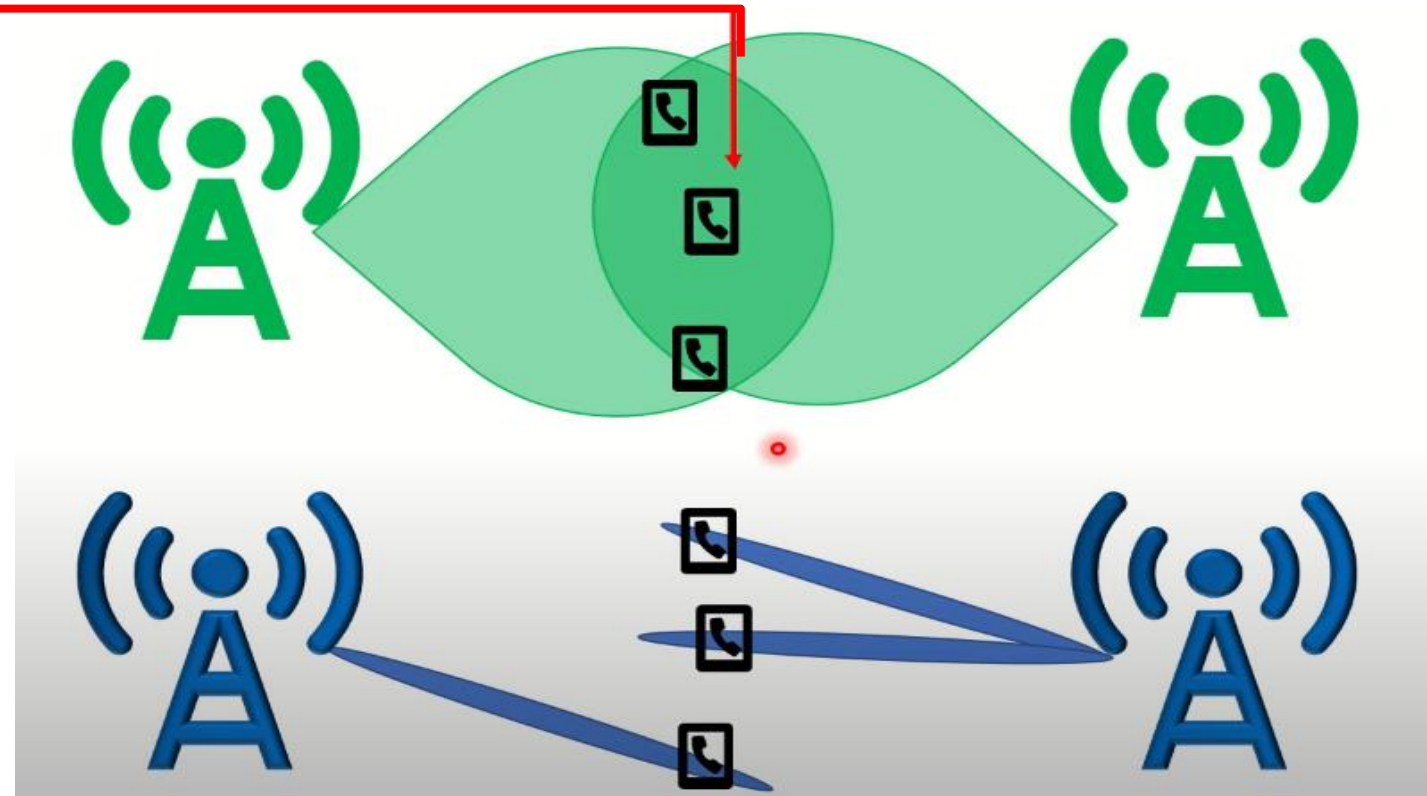
Comme les faisceaux sont beaucoup plus larges, ils auront donc plus d'interférences et les utilisateurs obtiendront un **SNIR plus faible** et un **débit plus faible**.

Zone à faible interférence

Comme les faisceaux sont plus directionnels et qu'il y a un chevauchement minimal des faisceaux entre les différentes cellules, l'interférence sera donc **plus faible** et le **SNIR/débit sera plus élevé**.

Massive MIMO Radios

Normal Radios



1.2. Les défis de l'IA dans les télécommunications

L'intelligence artificielle (IA) dans les réseaux transforme la manière dont les réseaux sont conçus, gérés et optimisés, les rendant plus efficaces, fiables et évolutifs.

❖ Mauvaise gestion du réseau

Le trafic mondial et le besoin d'équipements réseau supplémentaires augmentent considérablement, ce qui entraîne une gestion du réseau plus complexe et plus coûteuse.

❖ Manque d'analyse des données

Les entreprises de télécommunications ont du mal à exploiter les vastes quantités de données collectées auprès de leurs bases de clients massives au fil des ans. Les données peuvent être fragmentées ou stockées dans différents systèmes, non structurées et non catégorisées, ou simplement incomplètes et peu utiles.

❖ Coûts élevés

Après des investissements massifs dans les infrastructures et la numérisation, les analystes du secteur s'attendent à ce que les dépenses d'exploitation mondiales des télécoms augmentent de plusieurs milliards de dollars.

De nombreuses entreprises de télécoms sont confrontées à une crise financière et doivent trouver des moyens d'améliorer leurs résultats.

1.2. Les défis de l'IA dans les télécommunications

❖ IA pour la gestion et l'automatisation des réseaux

- ✓ **IA pour la gestion et l'automatisation des réseaux auto-optimisés [Self-Optimizing Networks] (SON)** : l'IA permet aux réseaux de s'optimiser automatiquement en ajustant des paramètres tels que la puissance du signal, l'allocation.
- ✓ **Maintenance prédictive** : les modèles d'IA analysent les données des périphériques réseau pour prédire les pannes potentielles avant qu'elles ne se produisent. Cette maintenance proactive réduit les temps d'arrêt et améliore la fiabilité, ce qui permet de planifier les réparations ou les mises à niveau en dehors des heures de pointe.

❖ L'IA au service de la gestion et de l'optimisation du trafic

- ✓ **Prévision du trafic** : l'IA peut analyser les données historiques pour prédire les pics de trafic et les goulots d'étranglement, permettant aux réseaux d'ajuster les ressources (comme la bande passante) à l'avance pour maintenir la qualité de service (QoS). Cela est essentiel pour des services tels que le streaming, les jeux et les événements à grande échelle.

1.2. Les défis de l'IA dans les télécommunications

❖ L'IA au service de la sécurité des réseaux

- ✓ **Détection et prévention des intrusions** : l'IA joue un rôle essentiel dans la reconnaissance des schémas d'activité malveillante (par exemple, les attaques DDoS, le phishing et les logiciels malveillants). Elle apprend en permanence des nouvelles menaces, identifie les anomalies et signale les failles de sécurité potentielles plus rapidement que les systèmes traditionnels basés sur des règles.
- ✓ **Détection des anomalies** : l'IA peut identifier les écarts par rapport au comportement normal du réseau, tels que les pics de trafic inhabituels ou les accès non autorisés, en analysant de vastes quantités de données en temps réel. Cela est essentiel pour prévenir les menaces internes et externes.
- ✓ **Réponse automatisée aux menaces** : en conjonction avec la détection des anomalies, l'IA peut automatiser les réponses aux menaces de sécurité, telles que l'isolement des appareils affectés ou le réacheminement du trafic loin des nœuds compromis, minimisant ainsi l'impact des attaques.

1.3. Machine Learning pour les ingénieurs réseau

Cette sous-section présente les concepts fondamentaux de l'apprentissage automatique (ML) dans un format adapté aux ingénieurs réseau et autres personnes qui sont nouvelles dans le domaine de l'intelligence artificielle et du ML et qui souhaitent expérimenter et apprendre comment appliquer les techniques de ML pour améliorer leurs performances.

1. Le rôle de l'apprentissage automatique (ML) dans l'ingénierie réseau :

- L'apprentissage automatique peut automatiser la prise de décision dans les tâches d'ingénierie réseau.
- Distinguer les différents types de problèmes traités par le ML, tels que **la classification, la prédiction, le clustering, l'optimisation et la génération.**

2. Identification des principales applications de ML pour le réseau :

- Examiner les cas d'utilisation spécifiques tels que **la détection d'anomalies, la prévision du trafic, l'optimisation de la qualité de service (QoS), la maintenance prédictive et la configuration du réseau.**
- Associer ces applications à des types de problèmes de Machine Learning généraux, en mettant l'accent sur leur pertinence et leur application aux scénarios de mise en réseau.

1.3. Machine Learning pour les ingénieurs réseau

3. Mettre en œuvre un algorithme ML simple (régression linéaire) :

- Comprendre les principes de base de la régression linéaire.
- Utiliser la régression linéaire pour modéliser la relation entre les données d'entrée et les prévisions de sortie.
- Générer des données synthétiques et adapter un modèle de régression linéaire à ces données.
- Ajuster les paramètres d'entrée pour obtenir la précision de prédiction souhaitée.

4. Apprendre différentes techniques d'apprentissage ML :

- Comprendre l'apprentissage supervisé, non supervisé et par renforcement.
- Comprendre comment ces techniques sont appliquées efficacement pour résoudre divers défis d'ingénierie.

5. Effectuer l'évaluation et la validation du modèle :

- Connaître certaines des différentes mesures permettant d'évaluer les modèles ML (par exemple : exactitude, précision, rappel, score F1, RMSE).
- Évaluez les performances du modèle de régression d'arbre de décision et discutez de la manière d'interpréter les résultats.



1.3. Machine Learning pour les ingénieurs réseau

6. Automatisation, apprentissage automatique et principaux types de problèmes

- Les domaines des réseaux, des logiciels, de l'informatique et de l'ingénierie connaissent aujourd'hui une innovation et un changement presque constants. Au cours des 10 dernières années, l'automatisation et la programmabilité sont devenues essentielles pour l'ingénieur moderne.

Une nouvelle ère : l'apprentissage automatique, un aspect de l'intelligence artificielle où nous formons des modèles pour Classifier les étiquètes. Les modèles, ainsi que leurs prédictions, sont le produit final du processus de Machine Learning, mais l'apprentissage automatique commence par les techniques et les algorithmes utilisés pour préparer et traiter les données.

- La formation d'un modèle consiste à alimenter les données via ces algorithmes et à ajuster les paramètres.
- Ce processus repose souvent sur la force brute et les méthodes d'essais et d'erreurs avant d'arriver à un modèle suffisamment précis.



1.3. Machine Learning pour les ingénieurs réseau

6. Automatisation, apprentissage automatique et principaux types de problèmes

- Pour les ingénieurs réseau, l'un des principaux avantages du ML est l'automatisation des « **décisions** ».
- La délégation des décisions de base libère les ingénieurs pour superviser et améliorer les modèles et leur permet de se concentrer sur des décisions, des problèmes et des défis de conception plus complexes.
- Le Machine Learning vise à imiter les processus de prise de décision humaine et, en tant que telle, elle doit être respectée, mais pas redoutée. Il s'agit simplement d'un autre outil permettant d'améliorer les performances.
- Chaque décision, qu'elle soit de planification, de conception, de configuration ou de provisionnement, peut potentiellement bénéficier de l'application du ML.



1.3. Machine Learning pour les ingénieurs réseau

7. Types de problèmes généraux que le ML résout

❑ Les algorithmes de ML utilisés dans différentes disciplines d'ingénierie sont souvent très similaires.

Même si les types ou classes de problèmes sont essentiellement les mêmes. Voici quelques types de problèmes clés que le ML peut résoudre pour l'ingénierie en général, avec des détails sur la manière dont ils s'appliquent spécifiquement à l'ingénierie réseau :

1. Classification

- **La classification consiste à trier les données en catégories.**

Dans le domaine des réseaux, la classification peut signifier l'identification de différents types de trafic réseau (par exemple : faire la distinction entre le streaming vidéo et la navigation sur le Web) ou la détection d'anomalies telles que les menaces de sécurité et les intrusions.

Exemple : Utiliser le ML pour classer les paquets entrants comme trafic normal ou menaces potentielles afin d'améliorer la sécurité du réseau.



1.3. Machine Learning pour les ingénieurs réseau

7. Types de problèmes généraux que le ML résout

2. Prédiction ou régression.

- **La prédiction ou la régression impliquent la prévision de valeurs futures en fonction de données passées.**

Pour les ingénieurs réseau, cela peut inclure la prévision des charges de trafic réseau, l'estimation des besoins en bande passante ou la prévision des points de défaillance potentiels avant qu'ils ne se produisent.

Exemple : Prédire la congestion du réseau et ajuster de manière proactive l'allocation de bande passante pour garantir des performances fluides.

3. Clustering

- **Le clustering est la tâche consistant à regrouper des points de données similaires.**

Dans le contexte de la mise en réseau, le clustering peut aider à segmenter les utilisateurs du réseau en fonction de leurs modèles d'utilisation, à identifier des périphériques réseau similaires ou même à regrouper des événements pouvant indiquer une cause commune.

Exemple : Regrouper des périphériques réseau ayant des modèles d'utilisation similaires pour améliorer l'allocation et la gestion des ressources.



1.3. Machine Learning pour les ingénieurs réseau

7. Types de problèmes généraux que le ML résout

4. Optimisation

- **L'optimisation consiste à trouver la meilleure solution parmi un ensemble d'options possibles.**

Les ingénieurs réseau peuvent utiliser des algorithmes d'optimisation pour améliorer les performances du réseau.

Les utilisations pratiques incluent l'optimisation des chemins de routage, l'équilibrage de charge ou la minimisation de la latence.

Exemple : Utilisation du ML pour optimiser les chemins de routage dans un environnement réseau dynamique, réduisant ainsi la latence et améliorant l'efficacité globale.

5. Génération

- **La génération fait référence à la création de nouvelles instances de données.**

Dans le domaine des réseaux, la génération peut impliquer la génération de trafic réseau synthétique pour tester la résilience du réseau, la création de nouvelles configurations pour les périphériques réseau ou la simulation de scénarios réseau à des fins de formation.

Exemple : Générer du trafic réseau synthétique pour tester la robustesse et la sécurité d'un réseau dans diverses conditions.



1.3. Machine Learning pour les ingénieurs réseau

8. Mappage des types de problèmes ML aux applications d'ingénierie réseau

Sur les page précédentes, Automatisation, apprentissage automatique et principaux types de problèmes, vous avez découvert les types de problèmes généraux que ML aborde et leurs applications spécifiques dans l'ingénierie réseau.

Vous pouvez maintenant observer les relations entre les types de problèmes généraux ML et les façons spécifiques dont ML est appliqué pour résoudre les problèmes d'ingénierie réseau.

Pour votre premier exercice, vous essayez de mapper 5 applications spécifiques de ML pour les problèmes de réseau aux 5 types de problèmes généraux ML abordés sur la page précédente.

1. Détection d'anomalies

- Les algorithmes ML peuvent apprendre à quoi ressemble le trafic réseau normal et détecter des modèles inhabituels qui peuvent indiquer des failles de sécurité ou des pannes de réseau.
- ❖ **Choix** : (1) Clustering, (2) Génération, (3) Classification, (4) Optimisation, (5) Prédiction ou régression



1.3. Machine Learning pour les ingénieurs réseau

8. Mappage des types de problèmes ML aux applications d'ingénierie réseau

2. Prévion du trafic [Traffic Forecasting]

- Le ML peut analyser les données historiques pour prévoir les modèles de trafic réseau futurs, aidant ainsi à l'allocation de bande passante et à la planification du réseau.
- ❖ **Choix :** (1) Clustering, (2) Génération, (3) Classification, (4) Optimisation, (5) Prédiction ou régression

3. Qualité de service (QoS)

- Le ML peut choisir les meilleurs chemins réseau pour différents types de trafic, garantissant ainsi que les services prioritaires maintiennent la qualité.
- ❖ **Choix :** (1) Clustering, (2) Génération, (3) Classification, (4) Optimisation, (5) Prédiction ou régression



1.3. Machine Learning pour les ingénieurs réseau

8. Mappage des types de problèmes ML aux applications d'ingénierie réseau

4. Segmentation du trafic réseau

- Le ML peut analyser et regrouper le trafic réseau en fonction de caractéristiques similaires, ce qui permet d'identifier des modèles de trafic distincts et d'optimiser les performances du réseau.
- ❖ **Choix :** (1) Clustering, (2) Génération, (3) Classification, (4) Optimisation, (5) Prédiction ou régression

5. Configuration réseau

- Les algorithmes ML peuvent recommander ou appliquer automatiquement des modifications de configuration réseau pour améliorer les performances ou la sécurité.
- ❖ **Choix :** (1) Clustering, (2) Génération, (3) Classification, (4) Optimisation, (5) Prédiction ou régression



1.3. Introduction aux techniques d'apprentissage pour l'apprentissage

- ❑ L'apprentissage automatique englobe diverses techniques qui permettent aux ordinateurs d'apprendre à partir de données et de décider sans programmation explicite.
- ❑ Ces techniques jouent un rôle crucial dans la résolution de problèmes complexes dans différents domaines, y compris la mise en réseau.
- ❑ La compréhension des différentes techniques de ML est essentielle pour les appliquer efficacement aux tâches d'ingénierie réseau.
- ❑ Les ingénieurs guident l'apprentissage automatique en sélectionnant des algorithmes appropriés, en optimisant les paramètres du modèle et en s'assurant que les prédictions du modèle correspondent aux données du monde réel dans le contexte des tâches d'ingénierie réseau.

Apprentissage supervisé

4. Segmentation du trafic réseau

- Le ML peut analyser et regrouper le trafic réseau en fonction de caractéristiques similaires, ce qui permet d'identifier des modèles de trafic distincts et d'optimiser les performances du réseau.
- ❖ **Choix** : (1) Clustering, (2) Génération, (3) Classification, (4) Optimisation, (5) Prédiction ou régression



1.3. Introduction aux techniques d'apprentissage pour l'apprentissage

Apprentissage supervisé

- L'apprentissage supervisé implique la formation de modèles sur des données étiquetées, où l'algorithme apprend à mapper les données d'entrée à la sortie correcte en fonction de ces exemples étiquetés.
- Cette technique vise à généraliser les prédictions à de nouvelles données invisibles avec précision.

Exemple : Classification du trafic réseau

- ✓ **Tâche** : Classification du trafic réseau comme normal ou malveillant.
- ✓ **Données** : Données historiques étiquetées où chaque échantillon est étiqueté comme trafic normal ou malveillant.
- ✓ **Méthode** : Entraîner un modèle d'apprentissage supervisé (par exemple : k-Nearest Neighbors, arbres de décision) pour classer le nouveau trafic réseau en fonction des modèles appris à partir des données étiquetées.



1.3. Introduction aux techniques d'apprentissage pour l'apprentissage

Apprentissage Non supervisé

- L'apprentissage non supervisé entraîne les modèles sur des données non étiquetées, leur permettant de découvrir de manière autonome des modèles et des structures au sein des données.
- Cette technique est essentielle pour identifier les relations et les informations cachées.

Exemple : Regroupement de périphériques réseau

- ✓ **Tâche** : Regroupement de périphériques réseau en fonction de leurs modèles de trafic.
- ✓ **Données** : Données de trafic réseau non étiquetées collectées à partir de divers périphériques.
- ✓ **Méthode** : Application d'algorithmes de clustering (par exemple : K-Means, DBSCAN) pour identifier des groupes de périphériques présentant des comportements de trafic similaires sans connaissance préalable d'étiquettes spécifiques.



1.4. Description de donnees

The DDoS attack dataset is a SDN specific dataset that is generated by making use of the mininet emulator and is mainly used for the classification of traffic by numerous deep learning and machine learning algorithms.

The process involved for the creation of the dataset includes the creation of ten topologies in mininet where the switches were connected to a single Ryu controller.

The network simulation runs for the both the benign UDP, ICMP and TCP traffic as well as the collection of malicious traffic for TCP Syn attack, ICMP attack and UDP flood attack.

The dataset includes 23 features in total where some of the data is extracted from the switches and others were calculated. Extracted features which are present in the dataset include: -



1.4. Description de donnees

The dataset includes 23 features in total where some of the data is extracted from the switches and others were calculated. Extracted features which are present in the dataset include: -

Packet_count – refers to the count of packets

byte_count – refers to the count of bytes in the packet

Switch-id – ID of the switch

duration_sec – packet transmission (in seconds)

duration_nsec – packet transmission (in nanoseconds)

Source IP – IP address of the source machine

Destination IP – IP address of the destination machine

Port Number – Port number of the application

tx_bytes – number of bytes transferred from the switch port

rx_bytes – number of bytes received on the switch port

dt field – shows the date and time which has been converted into a number and the flow is monitored at a monitoring interval of 30 seconds



1.4. Description de donnees

The calculated features present in the dataset include:

Byte Per Flow – byte count during a single flow

Packet Per Flow – packet count during a single flow

Packet Rate – number of packets transmitted per second and calculated by dividing the packet per flow by monitoring interval

number of Packet_ins messages – messages that are generated by the switch and is sent to the controller

Flow entries of switch – entries in the flow table of a switch which is used to match and process packets

tx_kbps – Speed of packet transmission (in kbps)

rx_kbps - Speed of packet reception (in kbps)

Port Bandwidth – Sum of tx_kbps and rx_kbps

The output feature is the last column of the dataset i.e. class label which classifies the traffic type to be benign or malicious. The malicious traffic is labelled as 1 and the benign traffic is labelled as 0.

The simulation of the network was run for approximately 250 minutes and 1,04,345 instances of data were collected and recorded. Further, the simulation was run for a given interval to collect more instances of data.