

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche
Université des Sciences et de la Technologie Houari Boumediene



Faculté de Génie Electrique
Département de télécommunications

Mémoire de MASTER
Domaine : Sciences et Technologies
Filière : Télécommunications
Spécialité : Télécommunication Radio Mobile

THEME

Optimisation des réseaux 4G LTE en utilisant des algorithmes de machine learning

Réalisé par : ZITOUCHE Dounia

KASDI Sarah

Devant le jury composé de :

Président : Pr. DEBYECHE Mohamed

Examineur : Pr. MEKSEN Thouraya

Promotrice : Pr. FERGANI Lamya

Co-promoteur : Mr. CHEKIREB Hamza

Co-promoteur : M. DERAMCHI Mohamed Rafik

Remerciements

« Ils t'interrogeront sur l'esprit ; dis : l'esprit est du domaine exclusif de mon seigneur et vous n'avez reçu de la science que fort peu de choses » Coran - le voyage nocturne 84 (El Israa)

Nous tenons à remercier Le Grand Dieu de nous avoir donné la force et le courage pour poursuivre nos études, veuille t'il guider nos pas dans le droit chemin.

Ensuite, nous remercions infiniment et au sens propre du mot nos chers parents qui ont tout donné pour nous, et ont fait de nous les personnes que nous sommes, sans eux nous ne serons jamais arrivés là, merci infiniment pour tout ce que vous avez fait pour nous.

Nous remercions *M. Mohamed Rafik DERAMCHI*, notre maître de stage, de nous avoir guidé durant notre travail, pour tout l'aide qu'il nous a apporté et les connaissances qu'il a su partager avec nous.

On le remercie aussi pour sa disponibilité, sa patience et ses précieux conseils, ses orientations et surtout ses encouragements. Ça a été un honneur de travailler avec vous.

Nous adressons nos plus sincères remerciements à *Mme Lamya FERGANI*, notre promotrice, qui nous a orienté vers ce stage et nous a fourni toutes les informations pour son bon déroulement.

On adresse nos remerciements également à notre Co-promoteur *Mr. CHEKIREB Hamza*, pour tout le temps qu'il nous a consacré et pour la qualité du suivi durant toute la période de notre stage.

Nos remerciements les plus sincères aux membres du jury qui ont bien voulu examiner ce modeste travail.

Nous tenons à remercier, enfin, toutes les personnes qui ont contribué de loin ou de près au succès de notre projet de fin d'études et qui nous ont aidé lors de la rédaction de ce mémoire.

Dédicaces

Je dédie ce mémoire à :

Ma mère, ABAOUB Souad, qui a œuvré pour ma réussite, par son amour, son soutien, tous les sacrifices consentis et ses précieux conseils, pour toute son assistance et sa présence dans ma vie, reçois à travers ce travail aussi modeste soit-il, l'expression de mes sentiments et de mon éternelle gratitude.

Mon père, Abdelmalek, qui peut être fier et trouver ici le résultat de longues années de sacrifices et de privations pour m'aider à avancer dans la vie. Puisse Dieu faire en sorte que ce travail porte son fruit.

Merci pour les valeurs nobles, l'éducation et le soutien permanent venu de toi.

Vous êtes ma force dans cette vie.

A mes frères et ma sœur, Wahid, Rayan et Lyna, à qui je souhaite un parcours réussi, je vous aime.

A mes cher(e)s ami(e)s pour leur soutien moral, leur appui et leur encouragement.

A mon ami, Nadir, en particulier, qui a été d'une grande aide pour moi.

A ma binôme Sarah Kasdi, qui a fourni énormément d'efforts, je suis très ravie d'avoir fait ce projet avec toi.

À tous ceux qui ont contribué de près ou de loin pour que ce projet soit possible,

Je vous remercie et je vous aime.

ZITOUCHE Dounia

Je dédie ce modeste travail

A mes très chers parents Aboulkacem et Assia, et aussi à ma très chère tante Khadija et à ma grand-mère qui m'ont toujours soutenu et encouragé à aller plus loin et à me surpasser, j'aimerais vous remercier pour tous vos sacrifices, votre amour et votre bienveillance, c'est grâce à vous que je suis là aujourd'hui, je ne pourrai décrire toute l'admiration et le respect que j'ai envers vous, je vous dois tout, je vous aime.

A mon frère Sofiane.

A ma binôme, ZITOUCHE Dounia je tiens à te remercier pour ton soutien ta patience et ton courage et C'est un honneur d'avoir réalisé ce travail avec toi.

A mes ami(e)s qui m'ont soutenue et aidé et encourager a réalisé ce mémoire et de continuer jusqu'à la fin

Kenza, Ichrak, Sabrina, Kamilia, Imad, Anis, Oussama, Ishak je tiens à vous remercier car c'est grâce à vous.

A toute ma famille et ami(e)s.

Et Merci à tous ceux qui ont contribué de près ou de loin à ma réussite.

KASDI Sarah

Résumé

A cause des conditions opérationnelles et environnementales, le processus de production dans les entreprises et équipementiers de télécommunications est très sensible à la détérioration et aux défaillances. Donc, il est indispensable de définir la stratégie de maintenance correcte afin de réduire au minimum les temps d'arrêt causés par les pannes inattendues et de minimiser par conséquent les coûts très élevés de maintenance. Récemment, la maintenance prédictive a profité du développement dans les domaines de la science de données et de machine learning pour surpasser plusieurs limitations de la maintenance conventionnelle.

Le but de ce projet est d'explorer les outils de l'apprentissage automatique dans la prédiction des KPIs, grâce à cette technologie émergente qui est l'IA pour faciliter les tâches des ingénieurs, en particulier, ça les aidera à anticiper l'emplacement et le jour de la dégradation avant son apparition.

Pour résoudre cette problématique de régression, nous allons d'abord vers une stratégie de modélisation, qui est formée en exécutant à l'aide de Python des modèles d'intelligence artificielle (IA), plus précisément, les modèles de régression de la machine learning qui font partie de l'apprentissage supervisé. Ensuite, nous allons étudier leur efficacité et leur performance grâce à une étude comparative entre eux à l'aide des métriques d'évaluation des performances adéquates.

Abstract

Due to operational and environmental conditions, the production process in telecommunications companies and equipment manufacturers is very susceptible to deterioration and failure. Therefore, it is essential to define the correct maintenance strategy in order to minimize the Call Drops caused by unexpected failures and consequently minimize the very high maintenance costs. Recently, predictive maintenance has taken advantage of developments in the fields of data science and machine learning to overcome several limitations of conventional maintenance.

The purpose of this project is to explore the tools of machine learning in the prediction of KPIs, thanks to this emerging technology which is AI to facilitate the tasks of engineers, in particular, it will help them to anticipate the location and the day of the degradation before its appearance.

To solve this regression problem, we first go to a modeling strategy, which is trained by running artificial intelligence (AI) models using Python, more specifically, machine learning regression models. that are part of supervised learning. Next, we will study their effectiveness and performance through a comparative study between them using the appropriate performance evaluation metrics.

ملخص

نظرًا للظروف التشغيلية والبيئية، فإن عملية الإنتاج في شركات الاتصالات ومصنعي المعدات معرضة بشدة للتدهور والفشل لذلك، من الضروري تحديد استراتيجية الصيانة الصحيحة لتقليل وقت التوقف عن العمل الناجم عن الأعطال غير المتوقعة وبالتالي تقليل تكاليف الصيانة المرتفعة للغاية. في الآونة الأخيرة، استفادت الصيانة التنبؤية من التطورات في مجالات علم البيانات والتعلم الآلي للتغلب على العديد من قيود الصيانة التقليدية.

الغرض من هذا المشروع هو استكشاف أدوات التعلم الآلي في التنبؤ بمؤشرات الأداء الرئيسية، وذلك بفضل هذه التكنولوجيا الناشئة وهي الذكاء الاصطناعي لتسهيل مهام المهندسين، على وجه الخصوص، ستساعدهم على توقع موقع ويوم الحدث. تدهور قبل ظهوره لحل مشكلة الانحدار هذه.

بعد ذلك، سوف ندرس فعاليتها وأدائها من خلال دراسة مقارنة بينها باستخدام مقاييس تقييم الأداء المناسبة.

Table des matières

<i>Dédicaces</i>	III
Liste des Tableaux.....	IX
Liste des Figures.....	X
Introduction Générale	1
<i>CHAPITRE 1 : Généralités sur les réseaux mobiles</i>	3
1.1 Introduction.....	4
1.2 L'évolution des réseaux cellulaires.....	4
1.2.1 La première génération 1G.....	5
1.2.2 La deuxième génération 2G	5
1.2.3 La troisième génération 3G	5
1.2.4 La 4eme génération (LTE)	6
1.3 Définition de la LTE.....	6
1.3.1 Les principaux objectifs visés par la 4ème génération LTE	6
1.3.2 Les caractéristiques de l'accès LTE	7
1.3.3 Architecture de la LTE.....	7
1.3.4 User Equipment.....	8
1.3.5 Réseau d'accès radio (E-UTRAN)	8
1.3.6 EPC (Evolved Packet Core).....	9
Les principaux composants de l'EPC sont :	10
MME (Mobility Management Entity)	10
HSS (Home Subscriber Server)	10
PDN GW (Packet Data Network Gateway)	10
Policy and Charging Rules Function (PCRF)	11
Serving Gateway (SGW).....	11
1.3.4 Les interfaces d'un réseau LTE	11
1.3.5 Modes de Duplexage en LTE	12
1.3.6 Technique de modulation en LTE	13
1.3.7 Techniques d'Accès en LTE :	14
1.3.8 L'OFDMA/SC-FDMA.....	14
1.3.9 Techniques d'antennes multiples (MIMO).....	15
1.4 Le réseau LTE-Advanced.....	16
1.5 La cinquième génération 5G	17
1.5.1 Les objectifs de la cinquième génération :	17
1.5.2 La 5G SA et 5G NSA.....	18

1.5.3	Les avantages	18
1.5.4	Les différentes générations de communications mobiles.....	18
1.5.5	Intégration des IoT dans les réseaux 5G.....	19
1.5.5.1	Définition de l'IoT (Internet of Things)	19
1.5.5.2	Intelligence artificielle des choses (AIoT)	20
1.6	Conclusion	20
<i>CHAPITRE 2 : Optimisation d'un Réseau LTE</i>		21
2.1	Introduction.....	22
2.2	La qualité du service dans le réseau LTE	22
2.2.1	Le but de la QoS.....	22
2.2.2	Les paramètres de la QoS.....	23
2.3	L'optimisation du réseau LTE	23
2.3.1	Définition de l'optimisation du réseau radio	23
2.3.2	L'objectif de l'optimisation du réseau LTE	24
2.3.3	Processus d'optimisation radio	24
2.3.3.1	Pré optimisation	25
2.3.3.2	Optimisation continue.....	26
2.3.4	Le Drive Test	26
2.4	Les compteurs et les indicateurs clés de performance	26
2.4.1	Les compteurs	26
2.4.2	Les PI (Performance Indicator)	26
2.4.2.1	Accessibilité (Accessibility)	28
2.4.2.2	Mobilité (Mobility).....	29
2.4.2.3	Intégrité (Integrity)	29
2.4.2.4	Maintien de Service (Retainability)	30
2.5	Perte des appels (Call drop)	30
2.5.1	Les causes d'un call drop	30
2.5.2	Les solutions au call drop	31
2.5.3	Optimisation de Maintenabilité	31
2.6	Conclusion	32
<i>CHAPITRE 3 : L'Apprentissage Automatique</i>		33
3.1	Introduction.....	34
3.2	L'intelligence Artificielle IA.....	34
3.3	L'apprentissage automatique.....	35
3.3.1	Les types d'apprentissage automatique.....	35
3.3.1.1	L'apprentissage supervisé	35

3.3.1.2	Apprentissage non supervisé	36
3.3.1.3	Apprentissage semi-supervisé.....	36
3.3.1.4	Apprentissage par renforcement	36
3.3.2	Algorithmes du Machine Learning	37
3.3.2.1	Régression linéaire (SLR)	37
3.3.3	Techniques de régularisation	38
3.3.3.1	Technique de régularisation de crête (Ridge)	38
3.3.3.2	Technique de régularisation Lasso	39
3.3.3.3	Régression Elastic Net	39
3.3.3.4	Régression polynomiale	40
3.3.3.5	Support Vector Regression SVR.....	41
3.3.3.6	Arbre de décision :.....	42
3.3.3.7	Réseaux de neurones :	43
3.3.3.8	K-plus proche voisin :.....	45
3.3.3.9	Les forêts aléatoires pour la régression (Random Forest Regression) :.....	45
3.3.4	L'apprentissage ensembliste :	47
3.3.4.1	Méthodes de l'apprentissage ensembliste	47
3.3.4.2	Bagging	48
3.3.4.3	Le boosting	49
3.3.4.4	Le stacking	49
3.3.4.5	Conclusion sur les méthodes d'ensemble.....	49
3.4	Conclusion	50
<i>CHAPITRE 4 : Expérimentations et Analyse des Résultats</i>		51
4.1	Introduction.....	52
4.2	Python pour l'IA.....	52
4.2.1	Présentation du langage Python	52
4.2.2	Principales caractéristiques du langage Python	53
4.2.3	Environnement de développement pour Python	53
4.2.4	Framework de programmation	53
4.2.5	Principaux modules et bibliothèques utilisés dans nos programmes	54
4.2.5.1	Scikit –learn	54
4.2.5.2	Pandas	54
4.2.5.3	NumPy	55
4.2.5.4	Matplotlib	55
4.3	La présentation des données utilisées	55
4.4	Approche proposée	56

4.4.1	Importation et nettoyage des données.....	57
4.4.2	Préparation des données (Split)	59
4.4.3	Le choix du bon modèle	60
4.5	Evaluation et sélection du modèle	61
4.5.1	Résultats	62
4.5.2	Comparaison des modèles	68
4.6	Conclusion	69
	Conclusion générale	70
	Bibliographie.....	72
	Annexes	73

Liste des Tableaux

Tableau 1 : les interfaces es d'un réseau LTE [14]	11
Tableau 2 : Différence entre LTE et LTE-Advanced.....	17
Tableau 3 : Résumé des générations de communications mobiles	18
Tableau 4 : KPIs d'Accessibilité de la quatrième génération	28
Tableau 5 : KPIs de mobilité de la quatrième génération	29
Tableau 6 : KPIs de mobilité de la quatrième génération	29
Tableau 7 : KPIs de continuité de la quatrième génération.....	30
Tableau 8 : Capture écran du fichier Excel de données reçu	55
Tableau 9 : Descriptif des colonnes utiles du fichier Excel reçu.	56
Tableau 10 : Evaluation du modèle Lasso	62
Tableau 11 : Evaluation du modèle Ridge	62
<i>Tableau 12 : Evaluation du modèle Decision Tree</i>	62
Tableau 13 : Evaluation du modèle SLR	63
Tableau 14 : Evaluation du modèle Random Forest	63
Tableau 15 : Evaluation du modèle Elastic_Net	63
Tableau 16 : Evaluation du modèle SVR (noyau=linéaire)	63
<i>Tableau 17 : Evaluation du modèle SVR (noyau=rbf)</i>	63
Tableau 18 : Comparaison des performances des modèles	68

Liste des Figures

Figure 1 : Evolution des réseaux mobiles	4
Figure 2 Architecture de l'UMTS	6
Figure 3 : Architecture de la LTE [9].....	8
Figure 4 : Architecture de l'E-UTRAN.....	9
Figure 5 : Architecture de l'EPC.....	10
Figure 6 : La représentation des porteuses FDD en DL/UL.....	12
Figure 7 : La représentation de la porteuse TDD	13
Figure 8 : Les différents types de modulation LTE	13
Figure 9 : OFDM/OFDMA	14
Figure 10 : L'OFDMA et la SC-FDMA [18].....	15
Figure 11 : Différents types d'antennes	16
Figure 12 : Processus d'optimisation radio LTE.....	25
Figure 13 : Les différents paramètres d'un KPI.....	27
Figure 14 : Types des KPI en LTE.....	28
Figure 15 : le positionnement des notions d'IA, machine learning et deep learning imbriquées les unes aux autres.....	34
Figure 16 : Les types d'apprentissage automatique	35
Figure 17 : Représentation graphique d'un exemple sur la régression linéaire	37
Figure 18 : Représentation de la régression polynomiale	40
Figure 19 : Représentation des hyperplans en utilisant l'algorithme SVM	41
Figure 20 : Représentation de deux classes en utilisant l'algorithme SVM.....	41
Figure 21 : SVR linéaire à une dimension	42
Figure 22 : Principe de fonctionnement de l'algorithme arbre de décision	43
Figure 23 : Représentation d'un réseau de neurone [48]	44
Figure 24 : Exemple de classification par l'algorithme KNN [49]	45
Figure 25 : Description de l'algorithme de la forêt aléatoire	46
Figure 26 : Bagging [50]	46
Figure 27 : Le type de méthodes ensemblistes [53]	48
Figure 28 : Les étapes du Bagging	48
Figure 29 : Description du fonctionnement du Stacking [57].....	49
Figure 30 : Logo de Python.....	53
Figure 31 : Logo d'anaconda.....	54
Figure 32 : Logo de l'environnement de développement Spyder	54
Figure 33 : Vue d'ensemble de l'algorithme utilisé dans notre travail	56
Figure 34 : Visualisation de la Base de Donnée avant le filtrage dans l'environnement Spyder	57
Figure 35 : Visualisation de la Base de donnée sous forme de nuage de points.....	57
Figure 36 : Les valeurs aberrantes avant la suppression	58
Figure 37 : Les valeurs aberrantes après la suppression	58
Figure 38 : Les valeurs manquantes avant la suppression.....	58
Figure 39 : Les valeurs manquantes après la suppression.....	59
Figure 40 : Visualisation de la nouvelle Base de Donnée après le filtrage dans l'environnement Spyder	59
Figure 41 : Comparatif de la taille des deux Base de donnée, Data et df.....	59
Figure 42 : Dimension de l'ensemble de données.....	60

Figure 43 : Préparation et entraînement des données par Python	60
Figure 44 : La carte de Scikit-Learn du Site officiel [64]	61
Figure 45 : Graphe représentatif des valeurs réels et prédites Test des PI.....	65
Figure 46 : Graphe représentatif des valeurs réels et prédites Train des PI.....	65
Figure 47 : Graphe représentatif des valeurs réelles et prédites Test des PI.....	66
Figure 48 : Graphe représentatif des valeurs réels et prédites Train des PI.....	66
Figure 49 : Graphe représentatif des valeurs réels et prédites Test des PI.....	67
Figure 50 : Graphe représentatif des valeurs réels et prédites Train des PI.....	67

Introduction Générale

Le réseau radio mobile est aujourd'hui un domaine en pleine effervescence. Pendant la dernière décennie, les évolutions de télécommunications ont explosé une nouvelle gamme de service qui a écarté les services classiques afin de satisfaire l'augmentation du nombre des utilisateurs et les exigences de taux de données élevées. Cette motivation, laisse les générations mobiles se succéder et se développent de la technologie GSM vers un système de paquets tout IP optimisé dénommé LTE (Longue Term Evolution).

Les utilisateurs attendent des services plus diversifiés et une vitesse de chargement et de téléchargement plus rapide. Les opérateurs donc à leur tour essayent de satisfaire le besoin de leurs abonnés et leurs assurer la maintenabilité des services de haute qualité. Pour ce faire, les réseaux LTE doivent être optimisés et surveillés pendant la phase de leur déploiement et après leur mise en œuvre.

L'optimisation vise à améliorer les performances du réseau tout en réduisant les coûts d'exploitation et d'investissement dans les brefs délais. Une bonne optimisation du réseau peut réduire significativement le niveau d'interférence global du réseau, améliorer le taux de réussite de l'établissement de connexions, réduire les interruptions de service (Call Drop), améliorer le débit des données, augmenter le taux de réussite des transferts (handover) entre cellules dans le réseau entier et améliorer la capacité du réseau.

C'est dans ce contexte que s'inscrit le sujet de notre projet de fin d'études, nous avons choisis d'utiliser une branche de l'Intelligence Artificielle qui est le Machine Learning et ses méthodes pour prédire le KPI de maintenabilité plus précisément le Call Drop dans le but de pouvoir préparer le réseau à d'éventuelles coupures radio.

L'objectif principal de ce travail est d'étudier le comportement aléatoire des PI du réseau LTE en leur appliquant les algorithmes de machine learning essayant de trouver le bon estimateur prédictif.

Ce mémoire est donc organisé en quatre chapitres. Dans le premier chapitre, nous décrivons l'évolution de la téléphonie vers la LTE, les principaux éléments de la technologie 4G LTE, en particulier, l'architecture du réseau, les modes de duplexage, les techniques de modulation et d'accès au réseau, ainsi que la notion MIMO. On le clôture avec la cinquième génération et l'Intégration des IoT dans ce dernier.

Dans le second chapitre, nous présenterons le concept de l'optimisation d'un réseau radio LTE et de la QoS. Nous présenterons également les principaux indicateurs clés de performance, en suite, nous nous focaliserons sur les KPI de maintenabilité.

Le troisième chapitre sera consacré à la théorie des différents algorithmes du Machine Learning. Nous expliquerons dans un premier lieu l'IA et la notion du ML, l'apprentissage des machines et l'utilisation de quelques algorithmes. Nous détaillerons ensuite le principe de fonctionnement de chacun. Enfin, les méthodes d'apprentissage ensemblistes.

Le dernier chapitre décrit la partie application dans laquelle nous appliquerons ces algorithmes sur notre cas d'étude, les données de taux de drop ou la maintenabilité.

Nous effectuerons par la suite une étude comparative des performances des différents algorithmes de régression. Grâce à ça, nous pourrons tirer l'algorithme qui fonctionnera le mieux pour prédire les valeurs du drop les plus proches possible aux valeurs réelles. Nous finissons par une conclusion et des perspectives.

CHAPITRE 1 :

Généralités sur les réseaux mobiles

1.1 Introduction

Au cours des années, les réseaux mobiles ont connu d'énormes évolutions quittant des réseaux de première génération (1G) aux réseaux de cinquième génération (5G). Ceci, grâce à diverses techniques qui ont permis l'amélioration de la capacité apportant une évolution remarquable surtout dans les domaines du débit et de la bande passante en autorisant la transmission de tout type de données numériques et procurant ainsi une grande amélioration de la qualité de service. Ce chapitre présente, dans un premier temps, les étapes majeures du développement des réseaux mobiles, de leurs premiers pas dans les années 1970 à leurs dernières évolutions.

Ces rappels permettront de mieux appréhender le contexte d'émergence du LTE (Long Term Evolution), qui constitue un système dit de quatrième génération. Ce chapitre a été basé essentiellement sur la connaissance globale de cette dernière, de son architecture et de sa manière de fonctionnement. Puis, les motivations pour la définition d'une nouvelle génération de systèmes mobiles sont examinées et l'intégration des IoT dans les réseaux 5G.

1.2 L'évolution des réseaux cellulaires

Le réseau cellulaire est un réseau de communications spécialement destiné aux équipements mobiles. Il permet la communication entre ces unités mobiles ainsi qu'avec l'ensemble des abonnés au téléphone mobile. L'onde radio dans le cas d'un réseau cellulaire est le lien entre l'abonné et l'infrastructure de l'opérateur chaque antenne couvre un territoire défini et lors des déplacements de l'utilisateur le téléphone mobile change de cellule.

La figure suivante illustre l'évolution de la téléphonie mobile à travers les années.[1]



Figure 1 : Evolution des réseaux mobiles

1.2.1 La première génération 1G

Les réseaux mobiles de première génération appelés RTC sont à l'origine de la téléphonie mobile. Parut dans les années 80, elle permettait juste le transport de la voix en mode analogique à l'échelle national avec de très faible débit. La 1G avait beaucoup de défauts, comme les normes incompatibles d'une région à une autre, une transmission analogique non sécurisée (écouter les appels), pas de roaming vers l'international (roaming est la possibilité de conserver son numéro sur un réseau d'un autre opérateur). Cette dernière sera abandonnée au détriment de la deuxième offrant un débit amélioré. [1]

1.2.2 La deuxième génération 2G

Le GSM est apparu dans les années 90. Il s'agit de la norme 2G. Cette génération marque le passage de l'analogique vers le numérique dans les techniques de modulation et de codage. Elle a mis en œuvre le concept de TDMA. Son principe, est de passer des appels téléphoniques, s'appuyant sur les transmissions numériques permettant une sécurisation des données (avec cryptage), il a connu un succès et a permis de susciter le besoin de téléphoner en tout lieu avec la possibilité d'émettre des minimessages (SMS, limités à 80 caractères). Ainsi qu'il autorise le roaming entre pays exploitant le réseau GSM.

Devant le succès, il a fallu proposer de nouvelles fréquences aux opérateurs pour acheminer toutes les communications, et de nouveaux services sont aussi apparus, comme le MMS. Le débit de 9.6 kbps proposé par le GSM est insuffisant, dans ce concept, ils ont pensé à développer de nouvelles techniques de modulations et de codages qui ont permis d'accroître le débit pour la nouvelle génération.[2]

1.2.3 La troisième génération 3G

La troisième génération de réseau mobile connu sous l'appellation d'UMTS (Universal Mobile Telecommunication System), conduit à une amélioration significative des vitesses de transmission pouvant atteindre 2 Mbps. Cette technologie est basée sur la technique d'accès multiple W-CDMA (WideBand Code Division Multiple Access) qui consiste à exploiter une large bande de fréquence pour envoyer un grand nombre de données par paquet. [1]

L'UMTS comprend deux grandes parties qui sont le réseau d'accès appelé UTRAN (Universal Terrestrial Radio Access Network) et le réseau cœur appelé CN (Core Network) comme le montre la figure ci-dessus : [3]

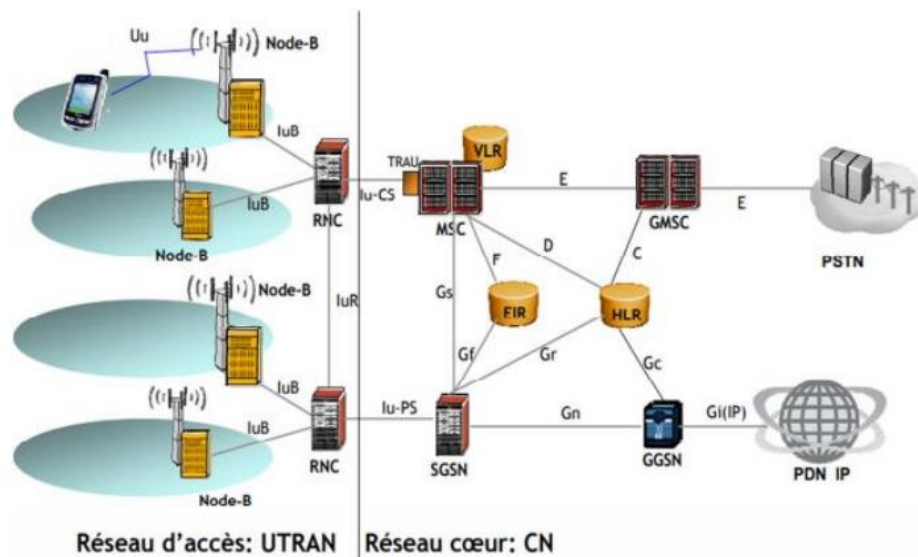


Figure 2 Architecture de l'UMTS

1.2.4 La 4eme génération (LTE)

1.3 Définition de la LTE

LTE (Long Term Evolution) est le nom d'un projet au sein du 3GPP (Organisme de standardisation) qui vise à produire les spécifications techniques de la norme de réseau mobile de quatrième génération (4G).

Contrairement aux anciennes générations, la LTE définit une architecture d'un système basé sur la commutation purement paquets (réseau tout IP) qui est dite architecture SAE (System Architecture Evolution). [4]

La LTE s'inscrit dans la continuité des réseaux 3G actuels et exploite principalement deux techniques :

- L'OFDM (Orthogonal frequency division multiplexing) procédé de codage de signaux numériques déjà utilisé notamment pour l'ADSL et le Wi-Fi.
- Le MIMO (Multiple inputs multiple outputs) une technologie qui utilise plusieurs antennes au niveau de l'émetteur et du récepteur alors que le Wi-Fi n'en utilise qu'une. [5]

En théorie, cette technologie permet d'atteindre des débits de l'ordre de 50Mbps en liaison montante et de 100 Mbps en liaison descendante avec une portée plus importante, un nombre d'appel par cellule supérieure et une faible latence.

1.3.1 Les principaux objectifs visés par la 4ème génération LTE

- Assurer la continuité de la session en cours.

- Réduire les délais et le trafic de signalisation.
- Fournir une meilleure qualité de service.
- Optimiser l'utilisation des ressources.
- Réduire le délai de relèvement, le délai de bout-en-bout, la gigue et la perte de paquets.
- Minimiser le coût de signalisation. [6]

1.3.2 Les caractéristiques de l'accès LTE

- Débit sur l'interface radio : 100 Mbit/s descendant et 50 Mbit/s montant. L'interface radio E-UTRAN doit pouvoir supporter un débit maximum descendant instantané (du réseau au terminal) de 100 Mbit/s en considérant une allocation de bande de fréquence de 20 MHz pour le sens descendant et un débit maximal montant instantané (du terminal au réseau) de 50 Mbit/s en considérant aussi une allocation de bande de fréquence de 20 MHz.
- Connexion permanente : principe des accès haut débit où la connectivité est permanente pour l'accès à Internet. Même si la connexion est permanente au niveau du réseau, il est nécessaire pour le terminal de passer de l'état IDLE à l'état ACTIF lorsqu'il s'agira d'envoyer ou de recevoir un trafic. Ce changement d'état s'opère en moins de 100 ms. Le réseau pourra recevoir le trafic de tout terminal rattaché puisque ce dernier dispose d'une adresse IP, mettre en mémoire ce trafic, réaliser l'opération de paging afin de localiser le terminal et lui demander de réserver des ressources afin de pouvoir lui relayer son trafic.
- Mobilité : assurée à des vitesses comprises entre 120 et 350 km/h. Le handover pourra s'effectuer dans des conditions où l'utilisateur se déplace à grande vitesse. La LTE ne permet que le hard handover et non pas le soft handover.
- Latence : il existe principalement deux (02) types de latence :
 - La Latence du plan de contrôle qui est le temps d'établissement d'une connexion avec un service (l'émission et la réception de la signalisation).
 - La latence du plan usager qui est le temps d'émission d'un paquet de données appartenant à l'utilisateur du réseau, et ce, après l'établissement de la connexion. [7]

1.3.3 Architecture de la LTE

Pour les opérateurs, la LTE implique de modifier le cœur du réseau et les émetteurs radio, le réseau s'appelle EPS (Evolved Packet System) constitué de trois principales parties :

- UE (User Equipment).
- E-UTRAN (Evolved UMTS Terrestrial Access Network).
- EPC (Evolved Packet Core).

L'EPS, Evolved Packet System, représente l'ensemble du réseau à savoir LTE et SAE.

Il a les caractéristiques suivantes :

Il possède une architecture plate et simplifiée comparée à celle hiérarchique des systèmes 2G/3G. Puisque l'entité RNC est enlevée, ses fonctions sont réparties entre l'eNodeB et le réseau cœur, la grande partie étant assignée à l'eNodeB. Il s'agit d'une architecture uniquement paquet comparée à l'architecture 2G/3G circuit et paquet. Son interface radio est totalement partagée entre tous les usagers en mode connecté, comparée à des ressources dédiées et partagées dans l'architecture 2G/3G. Il permet des Handover vers les autres réseaux 2G/3G afin d'assurer des communications sans coupure en environnement hétérogène. [8]

La figure 3 représente l'architecture globale du réseau LTE : [9]

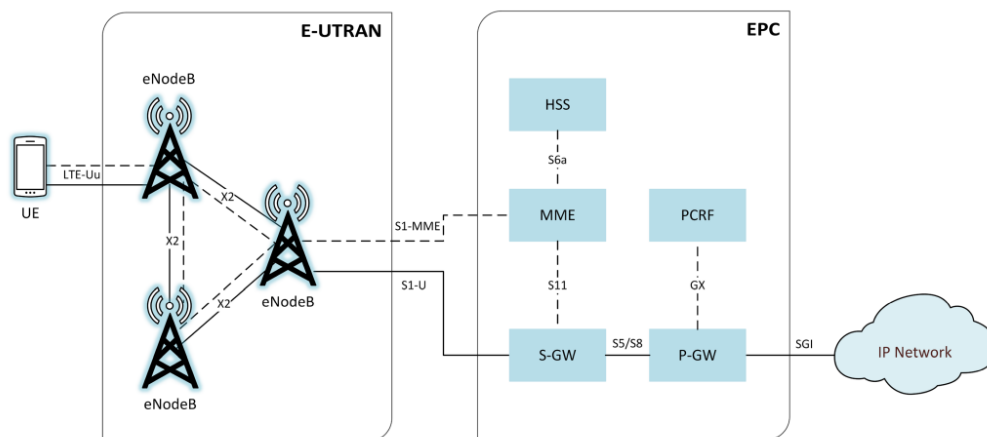


Figure 3 : Architecture de la LTE [9]

1.3.4 User Equipment

L'UE se connecte par une liaison radio aux stations de base eNode B (Evolved Node B). Il peut s'agir d'un téléphone mobile, smartphone ou d'un ordinateur portable. Ce terminal est authentifié et autorisé à accéder au réseau mobile RAN de l'opérateur. Il se compose de deux éléments distincts :

- ME (Mobile Equipement) qui assure le flux de données.
- UICC (Universal Integrated Circuit Card) : connu comme la carte SIM (Subscriber Identity Module) pour les équipements LTE. [10]

1.3.5 Réseau d'accès radio (E-UTRAN)

Le réseau d'accès consiste en plusieurs eNodeBs, qui sont interconnectées à travers l'interface X2. Le principal rôle de cette dernière est de minimiser les pertes de paquets lors de la mobilité

de l'utilisateur durant le Handover. L'eNodeB est responsable de la transmission et de la réception radio avec l'UE et dispose d'une interface S1 avec le réseau cœur. L'interface S1 consiste en S1-C (S1- Contrôle) entre l'eNodeB et le MME et S1-U (S1-Usager) entre l'eNodeB et le Serving GW. [11]

L'architecture du réseau d'accès radio terrestre UMTS évolué (E-UTRAN) est illustrée ci-dessous dans la Figure 4.[11]

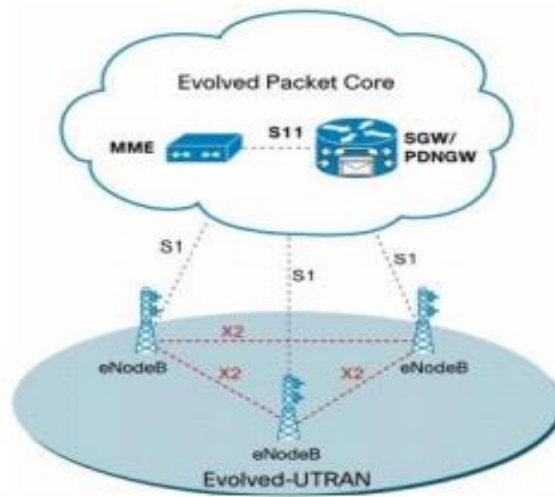


Figure 4 : Architecture de l'E-UTRAN

1.3.6 EPC (Evolved Packet Core)

Le cœur de réseau appelé « PC » (Evolved Packet Core) utilise des technologies « full IP », c'est-à-dire basées sur les protocoles Internet pour la signalisation qui permet des temps de latence réduits, le transport de la voix et des données. Ce cœur de réseau permet l'interconnexion via des routeurs avec les autres eNodeB distants, les réseaux des autres opérateurs mobiles, les réseaux de téléphonie fixe et le réseau Internet.

EPC est l'élément principal de l'architecture SAE. Il a un rôle équivalent du NSS dans le réseau GPRS.

Le réseau cœur EPC, quant à lui, est constitué principalement du MME (Mobility Management Entity) qui est le plan contrôle, du SGW (Serving Gateway) qui constitue le plan utilisateur, et du PGW (Packet Data Network Gateway). L'architecture en détails est présentée dans la figure suivante : [11]

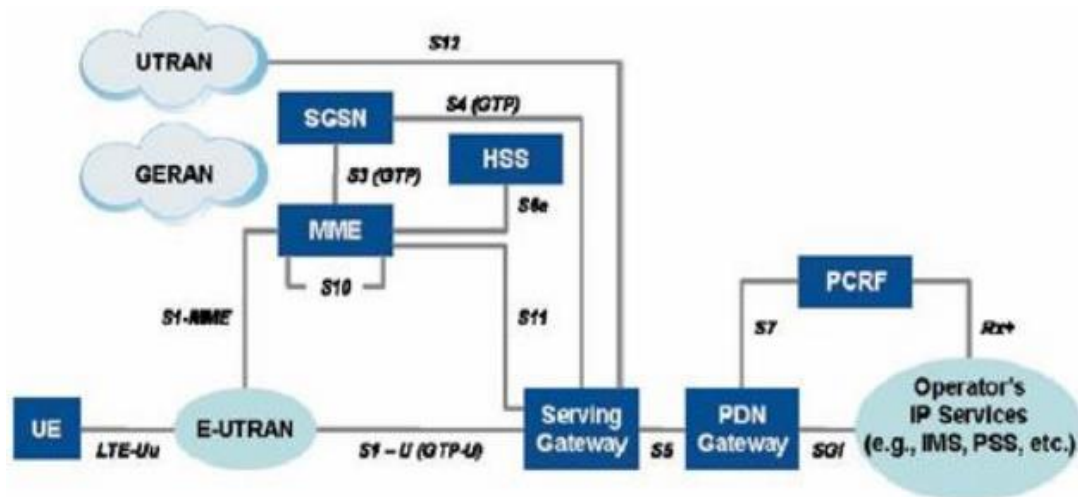


Figure 5 : Architecture de l'EPC

Les principaux composants de l'EPC sont :

MME (Mobility Management Entity)

Cette partie est responsable de la localisation et la poursuite du terminal mobile (UE) entre les appels et la sélection d'une bonne S-GW (Serving-Gateway) à travers une connexion. Comme elle réalise le dernier point de la protection par codage, donc par conséquent c'est le point qui valide l'interception de signalisation. Ainsi, qu'elle contrôle le signal entre le UE (Utilisateur Equipment) et le réseau cœur, et assure l'établissement, la maintenance, et l'élargissement de la porteur radio services. [12]

HSS (Home Subscriber Server)

Avec la technologie LTE, le HLR est réutilisé et renommé Home Subscriber Server (HSS). Le HSS est donc un HUR évolué qui contient l'information de souscription pour les réseaux GSM, GPRS, 3G, LTE et IMS. Le HSS est une base de données qui est utilisée simultanément par les réseaux 2G, 3G, LTE/SAE et IMS appartenant au même opérateur. Il supporte donc les protocoles MAP et DIAMETER.

Il est connecté au MME par l'intermédiaire de l'interface S6a qui s'appuie sur le protocole DIAMETER (protocole du monde IP). [12]

PDN GW (Packet Data Network Gateway)

C'est l'interface entre le réseau et les réseaux externes (Internet et Intranets). Le PDN GW est l'entité qui termine le réseau mobile EPS et assure l'interface aux réseaux externes IPv4 ou IPv6. Il sa charge également de l'allocation de l'adresse IP de l'UE.

Policy and Charging Rules Function (PCRF)

Le PCRF est un nœud optionnel au sein de l'architecture EPC. Toutefois, il inclut les points suivants:

- La gestion du trafic et de la facturation.
- L'indication des caractéristiques de la QoS à appliquer par le PCF sur les flux de données.
- S'assure que le traitement appliqué est en accord avec le profil de souscription de l'utilisateur. [3]

Serving Gateway (SGW)

La S-GW est un élément du plan de données au sein de LTE. Son objectif principal est de gérer la mobilité dans le plan utilisateur. Elle agit également comme une frontière principale entre le RAN et le CN. [10]

La S-GW maintient également les chemins de données entre les eNodeBs et les passerelles PDN-GW à partir de l'interface S5. Aussi quand les UEs se déplacent dans les régions desservies par des eNodeB différentes la S-GW veille à ce que le chemin de données soit maintenu. [13]

1.3.4 Les interfaces d'un réseau LTE

Tableau 1 : les interfaces es d'un réseau LTE [14]

Interfaces	Description
Uu	Est l'interface radio qui relie les UE aux eNodeBs. Il gère tous les messages de signalisation entre l'eNodeB et le MME ainsi que le trafic de données entre l'UE et le S-GW.
X2	Est le point de référence de contrôle et de plan utilisateur entre deux nœuds E-UTRAN.
S1	Elle relie l'E-UTRAN et l'EPC à la fois pour l'utilisateur et les plans de contrôle. Il comporte deux parties : le S1-U, appartenant au plan utilisateur et le S1-MME, appartenant au plan de contrôle et basé sur la transmission IP (Il transmet les messages de signalisation de la couche réseau radio de l'E-UTRAN)
S1-U (eNB et S-GW)	Elle permet le tunneling par porteur de plan utilisateur et la commutation de chemin entre eNB pendant le transfert.
S1-MME (eNB et MME)	Elle fournit un chemin de plan de contrôle entre eNB et MME.
S6a (MME et HSS)	Elle permet le transfert des données d'abonnement et d'authentification pour authentifier/autoriser l'accès des utilisateurs aux services réseau.
S10 (MME et MME)	Elle est utilisée pour le déplacement de MME et le transfert d'informations MME à MME.
S11 (MME et S-GW)	Elle fournit le point de référence entre MME et S-GW.

1.3.5 Modes de Duplexage en LTE

Le duplexage définit la manière dont sont séparées les transmissions sur la voie descendante et sur la voie montante. La voie descendante aussi appelée DownLink (DL) correspond à la transmission de la station de base vers l'UE. La voie montante, aussi appelée UpLink (UL), correspond à la transmission de l'UE vers la station de base.

Il existe deux principaux modes de duplexage, gérés par l'interface radio du LTE :

- Le duplexage en fréquence ou Frequency Division Duplex (FDD)
- Le duplexage en temps ou Time Division Duplex (TDD).

En mode FDD, les voies montante et descendante opèrent sur deux fréquences porteuses séparées par une bande de garde. [15]

En mode TDD, les voies montante et descendante utilisent la même fréquence porteuse, le partage s'effectuant dans le domaine temporel, comme illustré sur la figure suivante :[15]

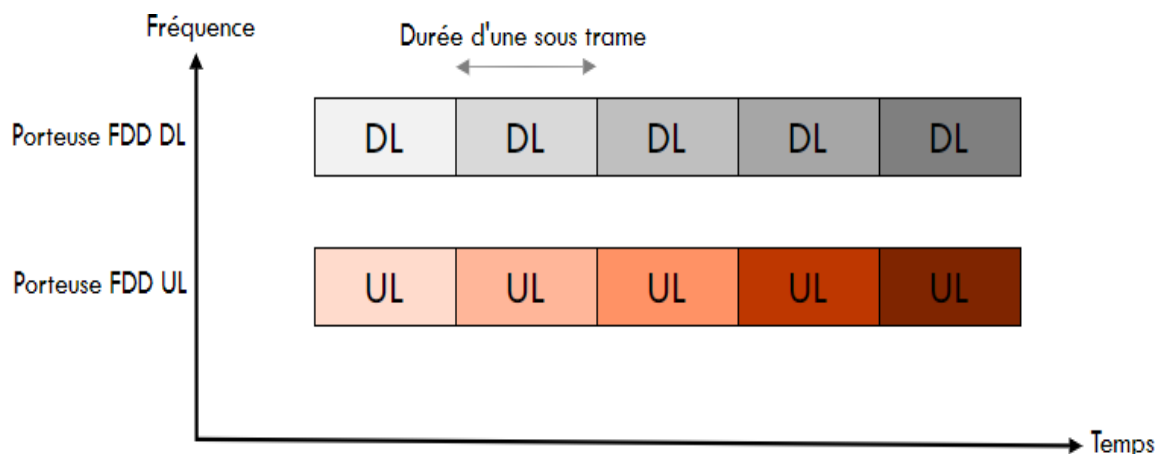


Figure 6 : La représentation des porteuses FDD en DL/UL.

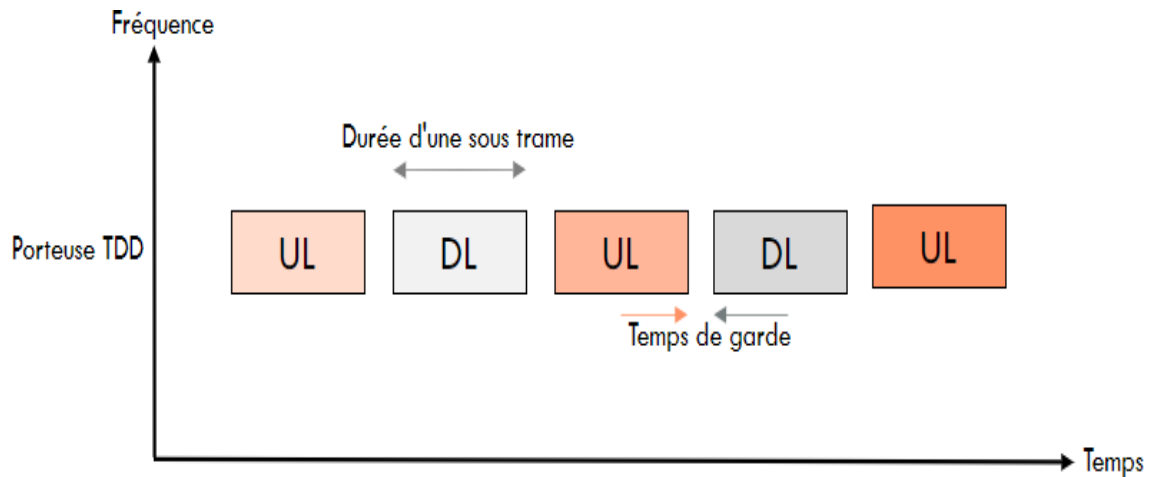


Figure 7 : La représentation de la porteuse TDD

1.3.6 Technique de modulation en LTE

La modulation utilisée en LTE est une modulation adaptative qui varie en fonction de la distance séparant l'abonné de l'eNodeB. Chaque sous-porteuse est modulée à l'aide de différents niveaux de modulation :

- QPSK (Quadrature Phase Shift Keying).
- 4-QAM, 16-QAM et 64-QAM (Quadrature Amplitude Modulation).

Par exemple, si les modulations disponibles sont le QPSK et le 16-QAM, dans le cas où le canal est marqué comme bon, on utilisera la modulation 16-QAM, qui offre un meilleur débit mais une plus faible robustesse. Par contre, si le canal est marqué comme dégradé, on utilisera la modulation QPSK, permettant un débit plus faible, mais plus robuste (moins sensible aux interférences). [16]

La figure suivante représente les techniques de modulation en LTE :

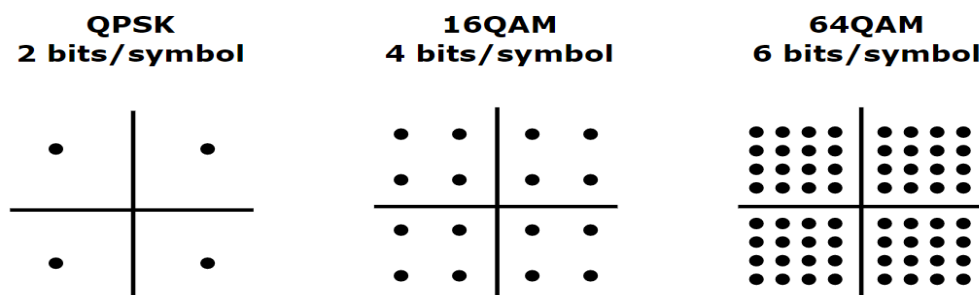


Figure 8 : Les différents types de modulation LTE

1.3.7 Techniques d'Accès en LTE :

La modulation du LTE est basée essentiellement sur l'utilisation de la technologie OFDM. L'OFDM consiste à répartir les symboles modulés sur plusieurs sous-porteuses. On distingue deux types : l'OFDMA pour la voie descendante et la SC-FDMA pour la voie montante.

1.3.8 L'OFDMA/SC-FDMA

L'OFDM (Orthogonal Frequency Division Multiplexing) est une modulation à porteuses multiples dont la principale caractéristique est l'orthogonalité des canaux supportés par chaque porteuse (appelés aussi sous-porteuses). Les modulations à porteuses multiples ont été introduites pour combattre l'interférence entre symboles tout en maintenant un débit élevé.

La méthode d'accès multiple correspondant à l'OFDM est l'OFDMA, Orthogonal Frequency Division Multiple Access, qui est introduite pour assurer une meilleure gestion du spectre. En effet, la bande de fréquences est divisée en plusieurs sous-porteuses de 15kHz, et les ressources sont allouées aux utilisateurs d'une manière variée, dépendant des requis des services demandés par chaque utilisateur. [17]

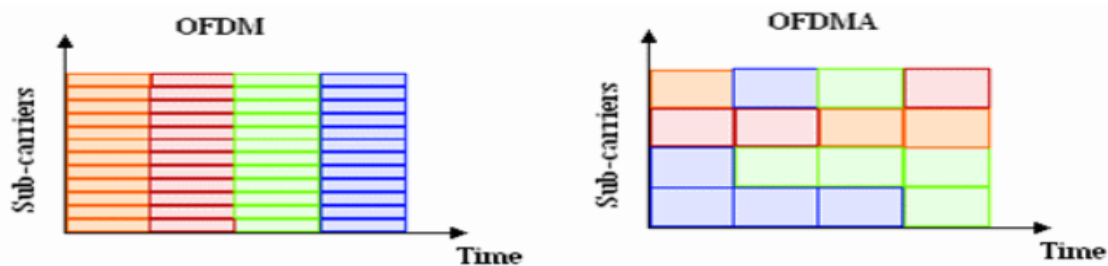


Figure 9 : OFDM/OFDMA

Cette technique a apporté plusieurs avantages, tels qu'une efficacité spectrale plus élevée et plus de protection contre l'effet des interférences. Cependant, elle présente un inconvénient majeur qui est un facteur de crête (PAPR : Peak-to-Average Power Ratio) élevé, du fait que la puissance reçue est la somme des puissances de toutes les sous-porteuses. Plus le PAPR est élevé, plus élevé est le niveau de puissance au-dessus duquel l'amplificateur doit fonctionner pour décoder le signal. [17]

Afin de remédier à ce problème, un amplificateur puissant est nécessaire. Cependant, celui-ci a des requis en termes de coût et de taille, ce qui le rend difficile à implémenter au niveau des équipements utilisateurs, mais il est possible à implémenter dans l'eNodeB. Pour cela, on n'utilise l'OFDMA que pour la voie descendante.

Pour la voie montante, on utilise une technique un peu similaire, qui utilise l'accès multiple monoporteuse (SC-FDMA), à la différence que celle-ci n'est pas sensible au PAPR (Peak to

Average Power Ratio) en raison de sa structure simple. Elle utilise une modulation simple avec une égalisation dans le domaine fréquentiel pour adapter l'accès multi-utilisateur. La figure suivante fournit une comparaison entre l'OFDMA et la SC-FDMA. [17]

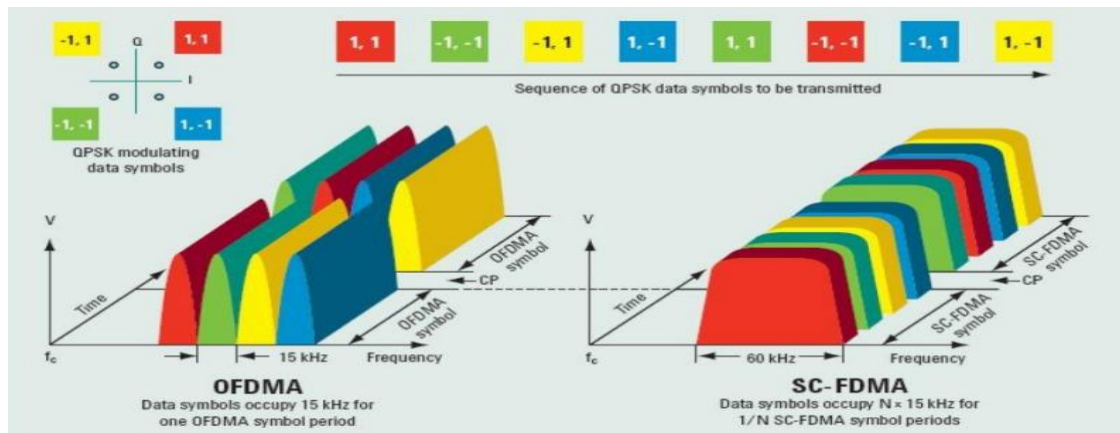


Figure 10 : L'OFDMA et la SC-FDMA [18]

1.3.9 Techniques d'antennes multiples (MIMO)

Les antennes MIMO jouent un rôle important dans le réseau LTE. Elles se basent sur l'utilisation de plusieurs antennes en émission et en réception. Les antennes MIMO font partie intégrante de du LTE. Cette technologie permet la transmission simultanée de données sur la même fréquence avec plusieurs antennes au niveau de l'émetteur et du récepteur.

En utilisant plusieurs antennes à l'émission et à la réception on obtient la diversité spatiale dont l'objectif est de combattre l'évanouissement causé par de multi-trajet.

Cette technologie MIMO est utilisée dans le but d'augmenter le débit et la portée des réseaux sans fil.

Lorsqu'un système comprend une antenne en émission et plusieurs antennes à la réception, on parle de SIMO, s'il comporte plusieurs antennes à l'émission et une antenne à la réception, il est nommé MISO et si les deux cotés (émission-réception) comportent une antenne chacune on parle de SISO. [19]

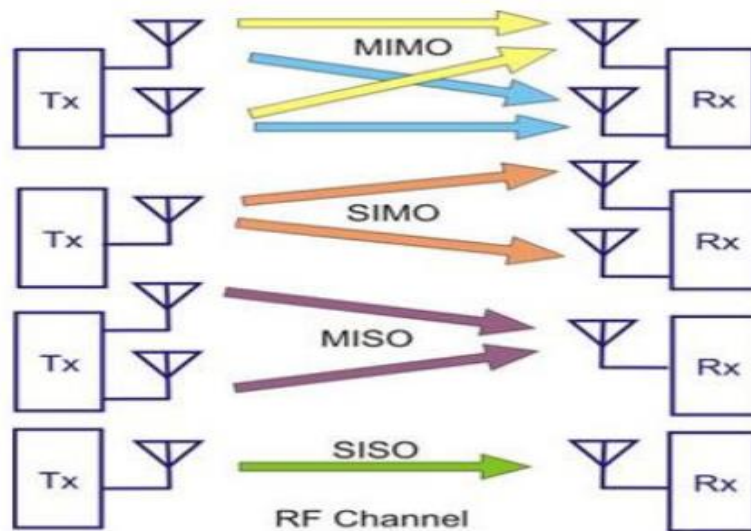


Figure 11 : Différents types d'antennes

Avec une telle structure mise en place (antenne MIMO) dans le réseau LTE cela permet d'améliorer le rapport signal sur bruit (SNR) et donc le taux d'erreur binaire.

1.4 Le réseau LTE-Advanced

Le LTE Advanced est une (version) évolution de la norme LTE lui permettant d'atteindre le statut de la véritable norme 4G tout en gardant une compatibilité avec le LTE.

Grâce aux fréquences identiques et aux techniques de codage telles que la technique OFDMA (Orthogonal Frequency Division Multiple Access) pour le sens descendant, la technique SCFDMA (Single Carrier Frequency Division Multiple Access) pour le sens montant et la technique MIMO (Multiple Input Multiple Output) pour l'émission-réception qui sont déjà utilisées dans le réseau LTE.

Le LTE-Advanced permet de fournir des débits descendants (pour le téléchargement) supérieurs à 1Gbit/s et 500 Mbit/s pour la liaison montante. Le LTE-Advanced est caractérisé par :

- Un débit maximal descendant de 1Gbit/s
- Latence : moins 5 ms
- Une bande passante de 100 MHZ
- Mobilité : elle a la même mobilité que le LTE
- Compatibilité : elle est compatible avec le LTE et les Systèmes existants 3GPP (2G/3G) [19]

Tableau 2 : Différence entre LTE et LTE-Advanced

	LTE	LTE ADVANCED
Max downlink speed (bps)	150 M	1G
Max uplink speed (bps)	50 M	500 M
Temps de latence	≈10 ms	Less than 5 ms
3GPP releases	Rel 8/9	Rel 10
Année de déploiement	2009/10	2011
Méthode d'accès	OFDMA/SC-FDMA	OFDMA/SC-FDMA
Spectre hertzien (MHZ)	20	100

1.5 La cinquième génération 5G

La 5G (cinquième génération) est une norme de réseau de téléphonie mobile. Elle succède à la quatrième génération, appelé 4G, en proposant des débits plus importants et une latence fortement réduite, tout en évitant le risque de saturation des réseaux lié à l'augmentation des usages numériques (smartphones, tablettes, objets connectés). [20]

1.5.1 Les objectifs de la cinquième génération :

- 10 fois plus de débit expérimenté : l'ère des débits de pointe plus uniformes et multi-Gbps.
- 10 fois moins de temps de latence : les niveaux de latence devraient être aussi bas que 1 ms.
- 10 fois plus de densité de connexion : activation de la connectivité Internet d'objets (IOT) avec peu de complexité et de surcharge de signalisation.
- Augmentation de 3 fois dans l'efficacité spectrale: une utilisation plus efficace de la bande passante.
- 100 fois plus de capacité de trafic : réseaux très densifiés avec plus de points d'accès partout.

- 100 fois plus d'efficacité du réseau : réseau énergétiquement efficace avec traitement de signal et matériel efficaces. [21]

1.5.2 La 5G SA et 5G NSA

Pour les principaux opérateurs de réseaux mobiles (MNOs), la 5G est principalement une question d'offre de connectivité haute vitesse aux consommateurs, sur des appareils qui supporte les services de réseau de cinquième génération. Pour passer en douceur du noyau existant à la 5G, les MNOs ont deux voies : l'architecture non autonome (NSA) ou autonome (SA). Et bien qu'ils soient tous deux des moyens à la même fin, NSA et SA sont structurellement et fonctionnellement différents. [22]

La 5G NSA (non-standalone), Il s'agit de continuer d'utiliser le cœur de réseau 4G LTE de l'opérateur tout en ajoutant petit à petit des antennes 5G, et permettre notamment l'utilisation de hautes fréquences en 5G NR.

Par opposition, la 5G SA, ou 5G Standalone, représente l'idéal du déploiement de la 5G, où un appareil peut utiliser les technologies 5G aussi bien sur les basses et les hautes fréquences, avec un cœur de réseau entièrement migré vers la 5G NR (new radio). Dans cette situation, l'appareil ne se repose plus sur les technologies de la 4G LTE. Cela demande des investissements bien plus conséquents, et ne sera donc disponible qu'à long terme. [22]

1.5.3 Les avantages

- NSA est extrêmement faible en coût par rapport à SA.
- La NSA facilite les déploiements de réseaux 5G en réutilisant les installations 4G existantes ce qui laisse un délai rapide pour commercialiser le haut débit mobile 5G.
- Avec la NSA, le déploiement est plus rapide et le délai de mise sur le marché est plus faible, car les emplacements 4G peuvent être utilisés pour installer la radio5G.
- SA nécessite la construction de stations de base 5G et le réseau de base 5G backend pour réaliser pleinement les caractéristiques et les fonctions de la 5G.
- SA implique un noyau 5G avec SBA pour l'évolutivité et la flexibilité pour fournir un réseau ultra-rapide avec une latence ultra-faible pour les cas d'utilisation 5G avancés. [22]

1.5.4 Les différentes générations de communications mobiles

Tableau 3 : Résumé des générations de communications mobiles

Génération	1G	2G	3G	4G	5G
1^{ère} année de déploiement	1981	1992	2001	2010	2020
Débit de données	2 Kb/s	64 Kb/s	2 Mb/s	100 Mb/s	10 Gb/s
Fréquences	900 MHz	900 MHz et 1,8 GHz	800/900MHz 1,7 à 1,9GHz 2100 MHz	800MHz 900MHz 1800MHz 2100MHz 2600MHz	28GHz 37 GHz 39 GHz 64 –71 GHz
Description fonctionnelle générale	Téléphones cellulaires analogiques	Téléphones cellulaires numériques (GSM/CDMA)	Première bande passante mobile utilisant des protocoles IP (WCDMA 2000)	Le haut débit mobile sur un standard unifié (LTE)	Internet tactile - Améliorer le réseau de communication M2M

1.5.5 Intégration des IoT dans les réseaux 5G

Le grand développement des télécommunications a créé une compatibilité entre l'Internet des objets et la cinquième génération, où ils seront intégrés pour améliorer la qualité de la vie sociale et économique comme l'éducation, les soins de santé, la surveillance, la sécurité et les foyers intelligents. Cette intégration a également donné un élan aux mondes physique et numérique du hardware-to-software, permettant la mise en œuvre de logiciels sur des dispositifs reconfigurables.

1.5.5.1 Définition de l'IoT (Internet of Things)

L'Internet des Objets, communément appelé en anglais Internet of Things (IoT) désigne une technologie d'avant-garde, où les objets traditionnellement non connectés qui nous entourent (comme des lampes, machines, vêtements, etc.), qu'ils soient physiques ou virtuels, ont désormais la capacité de communiquer entre eux en temps réel. Ce réseau d'objets connectés permet le partage de leurs données par l'intermédiaire d'une plateforme *Cloud* et ce, sans intervention humaine. Grâce à l'optimisation des interactions entre les humains et les machines et à la multiplication des flux de données, que les objets connectés offrent la possibilité de définir les besoins précis d'un individu, de sorte à lui offrir un bien ou un service unique. [24]

1.5.5.2 Intelligence artificielle des choses (AIoT)

L'intelligence artificielle des choses (AIoT) est la combinaison des technologies de l'intelligence artificielle (IA) avec l'infrastructure de l'internet des choses (IoT) pour réaliser des opérations plus efficaces de l'IoT, améliorer les interactions homme-machine et améliorer la gestion et l'analyse des données.

L'IA peut être utilisée pour transformer les données de l'IoT en informations utiles pour une meilleure prise de décision, créant ainsi une base pour une technologie plus récente telle que l'IoT Data as a Service (IoTDaaS).

L'AIoT est transformationnel et mutuellement bénéfique pour les deux types de technologie, car l'IA apporte une valeur ajoutée à l'AIoT grâce aux capacités d'apprentissage machine (Machine learning) et l'AIoT apporte une valeur ajoutée à l'IA grâce à la connectivité, la signalisation et l'échange de données. À mesure que les réseaux d'IoT se répandront dans les grandes industries, il y aura de plus en plus de données non structurées orientées vers l'homme et générées par des machines. L'IoT peut fournir un support pour les solutions d'analyse de données qui peuvent créer de la valeur à partir des données générées par l'IoT. [24]

1.6 Conclusion

Une bonne connaissance de l'architecture de la technologie et ses caractéristiques permet aux planificateurs de mieux gérer les ressources, de faciliter l'évolution du réseau en intégrant des technologies plus performantes et de fournir à la fois des services de bonne qualité.

Dans ce chapitre introductif, nous avons décrit d'une façon globale l'évolution des réseaux mobiles (1G, 2G, 3G, 4G et la 5G) en mettant l'accent sur l'interface radio des réseaux 4G LTE puisqu'elle constitue l'objet de notre travail. On a enchaîné avec une introduction détaillée de l'architecture des systèmes LTE, nous avons présenté ses objectifs, les caractéristiques de l'accès LTE, son architecture, les différentes techniques d'accès et la technologie Multi-Antenne (MIMO).

Alors que la LTE n'a pas fini de nous montrer ses performances, à l'heure actuelle, nous avons la 5G ou l'ultra haut débit mobile. Cette technologie abordera l'évolution au-delà de l'internet mobile vers l'IOT (Internet des Objets). Ce dont la dernière partie de ce chapitre parle.

Le chapitre suivant se focalisera sur l'optimisation d'un réseau 4G LTE.

CHAPITRE 2 :
Optimisation d'un
Réseau LTE

2.1 Introduction

L'objectif de tout réseau de télécommunication installé est d'offrir les meilleures performances aux abonnés. Cependant, la qualité de service (Quality of Service QoS) est un enjeu majeur dans le domaine des télécommunications. Elle consiste à satisfaire au minimum les exigences des utilisateurs et au maximum, atteindre les valeurs théoriques des paramètres fixées par le 3GPP.

Pour des raisons commerciales, chaque opérateur mobile fixe comme objectif le regroupement du plus grand nombre d'abonnés possible sans altérer le bon fonctionnement de son réseau, pour ce faire, l'opérateur doit se préoccuper d'offrir une qualité de service convenable au nombre d'utilisateurs servi en essayant d'optimiser les paramètres de la QoS ainsi que d'assurer une couverture satisfaisante aux clients, ceci se fera à travers des indicateurs clés de performance (KPI : Key Performance Indicator).

Dans ce chapitre, nous allons nous intéresser aux différents paramètres de la QoS ainsi qu'aux performances des réseaux LTE puis nous décrirons les différents KPI généralement utilisés par les équipes d'optimisation, tout en indiquant la manière dont ils sont calculés en se basant sur le KPI de maintien de service (call drop) que nous utiliserons pour la suite.

2.2 La qualité du service dans le réseau LTE

La qualité de service (QDS) ou Quality of Service (QoS) est définie comme un ensemble de technologies mises en œuvre afin d'assurer les meilleures performances pour la transmission des données dans de bonnes conditions en respectant certaines valeurs minimales des attributs du réseau tels que la disponibilité, le débit, la latence, la gigue, le taux de perte en paquets...etc. La qualité de service (QoS) correspond à la manipulation du trafic de sorte qu'un équipement réseau, tel qu'un routeur ou un commutateur, puisse transférer ce trafic conformément aux comportements requis de la part des applications à l'origine de ce trafic. En d'autres termes, la QoS permet à un équipement réseau de différencier le trafic et de lui appliquer différents comportements.

2.2.1 Le but de la QoS

Son but est d'optimiser les ressources du réseau et de garantir de bonnes performances aux applications. Cela permet d'offrir aux utilisateurs des débits et des temps de réponse différenciés par application suivant les protocoles mis en œuvre au niveau de la couche réseau.] [25]

Elle permet ainsi aux fournisseurs de services de s'engager formellement auprès de leurs clients sur les caractéristiques de transport des données applicatives sur leurs infrastructures IP. [26]

Selon les types d'un service envisagé, la qualité pourra résider :

- Le débit (téléchargement ou diffusion vidéo).
- Le délai (pour les applications ou la téléphonie).
- La disponibilité (accès à un service partagé).
- Le taux de pertes de paquets.

2.2.2 Les paramètres de la QoS

La QoS peut se mesurer de manière quantitative, à l'aide notamment des paramètres suivants :

- La perte de paquets qui se produit lorsque les liaisons réseau se congestionnent et que les routeurs et les commutateurs commencent à ne plus tenir compte des paquets. Quand ce phénomène a lieu lors d'une communication en temps réel, par exemple au cours d'un appel vocal ou vidéo, des instabilités ou des trous peuvent apparaître lors de la transmission.
- La gigue peut apparaître en cas d'encombrement du réseau, de dérive temporelle ou de changement d'acheminement. Si elle est trop importante, la qualité de la communication vocale ou vidéo se dégrade.
- La latence est la durée du trajet d'un paquet de sa source jusqu'à sa destination. Elle doit s'approcher autant que possible de zéro. Si un appel vocal sur IP (VoIP) a une latence élevée, on entend de l'écho et un mélange de sons.
- La bande passante est la capacité maximale de transmission, d'un point à un autre, d'un volume de données sur une ligne de communication en un temps donné. La QoS optimise le réseau en en gérant la bande passante et en hiérarchisant les applications selon les ressources dont elles ont besoin.
- La note d'opinion moyenne (MOS, Mean Opinion Score) mesure la qualité de la voix sur une échelle d'un à cinq, la plus haute note reflétant la meilleure qualité.
- La mobilité : La gestion du changement de cellule lors du mouvement des terminaux est nécessaire afin de garantir un réseau mobile optimal. L'objectif de l'optimisation de la mobilité est d'ajuster les paramètres de transfert et de permettre aux zones de transfert d'être plus raisonnables. [27]

2.3 L'optimisation du réseau LTE

2.3.1 Définition de l'optimisation du réseau radio

L'optimisation du réseau radio désigne l'amélioration de la performance du réseau d'accès en utilisant les ressources existantes. [28]

C'est un processus qui se fait aux étapes suivantes :

- Collecte et vérification des données : via les Drive Test, les statistiques de trafic et les informations d'alarmes de l'OMC, ainsi que les plaintes des clients.
- Analyse des données à l'aide de logiciel spécifique et en comparant les KPIs (Key Performance Indicator) avec les seuils fixés par l'opérateur.
- Ajustement des paramètres du matériel.
- Confirmation des résultats de l'optimisation et préparation des rapports.

2.3.2 L'objectif de l'optimisation du réseau LTE

Pour satisfaire au mieux leurs clients, les opérateurs œuvrent pour assurer la continuité de la délivrance des services avec une qualité optimale. Une fois le réseau actif, l'opérateur doit veiller sur son bon fonctionnement. Il est nécessaire de réaliser un suivi de la qualité de service et d'adapter le réseau aux différentes fluctuations en vue de son amélioration et de son expansion. Ainsi l'optimisation d'un réseau cellulaire est motivée par deux objectifs principaux : améliorer la qualité de service offerte aux utilisateurs et augmenter le volume de trafic écoulé par le réseau avec les équipements existants. [28] [29]

2.3.3 Processus d'optimisation radio

L'optimisation du réseau d'accès LTE fait référence à deux phases primordiales : la phase de pré-optimisation qui s'effectue pendant le déploiement progressif du réseau et après sa mise en service lorsqu'il y a besoin d'installer un nouveau site et la phase d'optimisation continue dite aussi optimisation basée sur les compteurs 1 (counter based optimization) qui s'effectue de façon régulière, dès la mise en service du réseau.

Le schéma suivant illustre les différents types d'optimisation des réseaux LTE :

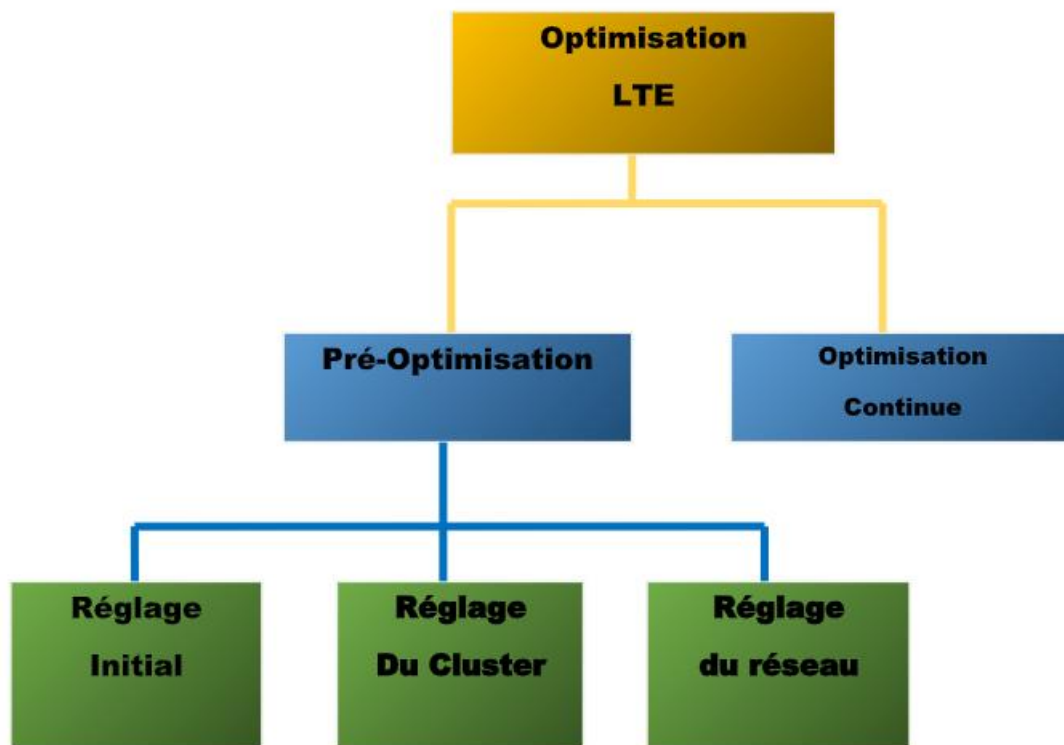


Figure 12 : Processus d'optimisation radio LTE

2.3.3.1 Pré optimisation

Cette phase est exécutée juste après le déploiement du réseau, les modifications opérées et les actions exécutées sont de nature physique liées au matériel (tilts, azimuts des antennes, puissance et activation de certaines fonctionnalités de l'eNodeB comme le MIMO, les hauteurs des sites, etc), les étapes de la pré-optimisation sont :

- Le SSV (Single Site Verification) : La vérification par site constitue la première étape à effectuer lors du déploiement du réseau pour la vérification du fonctionnement des nouveaux sites. Il a pour but de veiller à ce que chaque site soit correctement installé et que les paramètres soient correctement configurés. Autrement dit, le SSV implique l'exécution d'une procédure de vérification basée sur des mesures collectées via un drive test réalisé sur un nouveau site. Le SSV se fait selon deux méthodes différentes, la première consiste à faire le test dans le cas statique en choisissant un point particulier, dans la majorité des cas on le fait en face de la cellule. Tandis que dans le second est réalisé dynamiquement tout en se déplaçant autour des cellules. [30]
- CV (Cluster Verification) : L'étape de vérification par cluster peut être effectuée lorsque plus de 80% des sites ont été déployés dans le cluster en question. Les résultats de ce drive test permettent de vérifier la configuration physique des sites tels que le tilt et l'azimut de

l'antenne, la puissance de transmission, les paramètres du handover, et la liste des cellules voisines, afin de minimiser les interférences et améliorer l'exécution du handover.[30]

2.3.3.2 Optimisation continue

L'optimisation continue est essentielle pour conserver le niveau de performance défini par les KPI dont les valeurs nominales sont exigées par l'opérateur dès la phase de déploiement des réseaux. Durant cette phase, des compteurs établis au niveau du plan de contrôle de l'UE et de l'e-NodeB sont calculés régulièrement et collectés pour avoir une vue centralisée du fonctionnement global du réseau et donc permettent un suivi régulier de sa QoS en vue de son amélioration.

2.3.4 Le Drive Test

Les Drive test sont une phase d'optimisation indispensable avant le lancement du réseau, pour atteindre certains KPI sur le terrain. Il est effectué pour s'assurer de la possibilité de maintenir une session dans une zone définie. Les données collectées lors de ce test sont utilisées pour analyser et localiser tout problème lié à la conception du réseau. [31]

Cette méthode permet d'évaluer la couverture, la capacité et l'interférence. Aussi, elle donne des informations sur les Handover, les coupures de lien radio, les cellules voisines.

2.4 Les compteurs et les indicateurs clés de performance

2.4.1 Les compteurs

Un compteur peut être défini comme une valeur incrémentale d'un événement spécifique répétitif. Dans la norme UMTS, un événement peut correspondre à un message de signalisation. Par exemple, lors d'un appel téléphonique des milliers de messages de signalisations sont échangés entre le UE et le MSC. La mise à jour des compteurs à un certain point se fait via les messages échangés. Ce point est appelé le « point de déclenchement ». Et comme des milliers de messages de signalisation existent, de nombreux compteurs peuvent également exister. Cependant, le nombre de compteurs utilisables dépend de la stratégie des fournisseurs des équipements.

2.4.2 Les PI (Performance Indicator)

De nombreux PI peuvent être basés sur des compteurs. Le PI peut également se présenter sous la forme de métriques qui montrent les performances de certaines parties d'un système. Ces PI n'ont pas nécessairement un impact sur les KPI. La majorité des données sont

fondamentalement mesurées au niveau de la cellule et avec une granularité ROP de 15 minutes. L'opérateur peut calculer le PI sur plusieurs cellules ou des intervalles de temps plus longs. [32]
Formule de calcul : Une formule signifie une combinaison mathématique des compteurs qui a comme conséquence un indicateur significatif.

Il existe plusieurs statistiques indiquant le comportement du réseau (nombre de tentatives d'appel, échecs et succès d'appel et de Handover, etc.). Afin de faciliter l'interprétation du comportement du réseau, un ensemble de données rapportant la performance, dits KPIs, sont définis à partir des formules qui utilisent ces compteurs. [32]

Un KPI est le résultat d'une formule qui est appliquée aux compteurs (appelés les indicateurs de performance). Au niveau de l'entreprise, les KPIs sont extraits dans des rapports de format prédéfinis. Ces rapports sont adressés ensuite aux ingénieurs d'optimisation pour les employer pour différents buts. Le mécanisme est illustré sur la figure suivante :

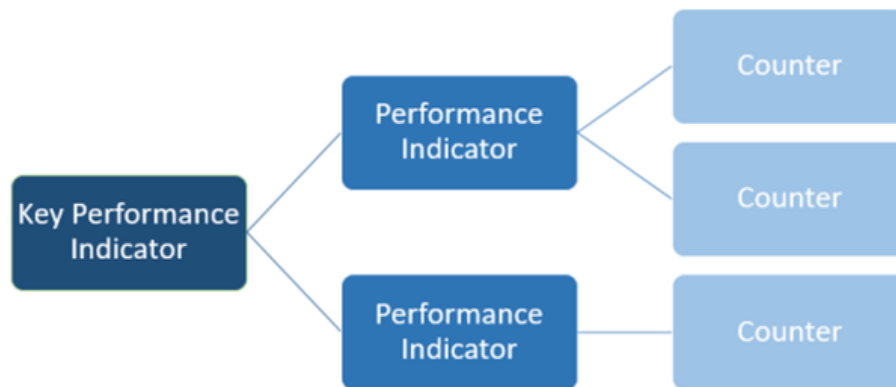


Figure 13 : Les différents paramètres d'un KPI

Les KPIs sont formulés pour mesurer les performances du réseau en termes d'accessibilité, d'intégrité, de mobilité, de conservation et de qualité perçue par l'abonné. Ils peuvent être définis comme ensemble de résultats qui mesurent les performances durant les heures chargées ou les heures normales sur le réseau entier. Le KPI est le résultat d'une formule qui est appliquée aux indicateurs de performance (Performance Indicators, PIs). Le PI peut être extrait d'un secteur, une eNodeB (dans le réseau 4g), un TRX ou à un niveau d'une cellule adjacente. Des centaines de KPIs existent et Ils emploient des compteurs d'une ou plusieurs mesures et peuvent être calculés à partir d'un compteur ou d'une formule de plusieurs compteurs. La période de l'observation se rapporte à la durée des échantillons rassemblés : heure, jour, semaine, mois, etc. [32] [33]

La qualité de la performance du réseau est principalement évaluée par la mesure des KPI. En effet, on distingue deux méthodes de mesure des KPI : mesure statistiques et mesure de test sur site.

Les indicateurs clé de performance LTE sont divisées en quatre familles comme représenté dans la figure ci-dessous :

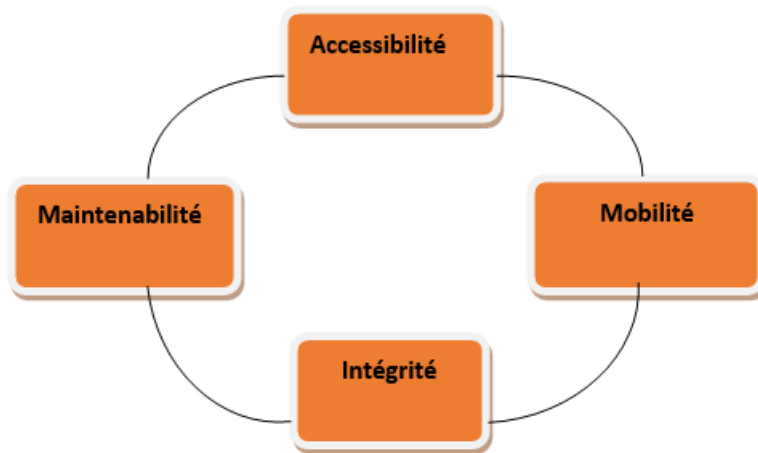


Figure 14 : Types des KPI en LTE

2.4.2.1 Accessibilité (Accessibility)

Ces KPI mesurent la possibilité qu'un service soit obtenu à des tolérances près et selon d'autres conditions, quand l'utilisateur le demande. En d'autres termes, c'est la capacité de joindre le réseau et l'opérateur doit surveiller le taux de succès d'établissement d'appel, le taux de succès de procédure de paging et la probabilité de blocage. [34]

Nous regroupons les KPIs d'Accessibilité de la quatrième génération dans le tableau ci-dessous où les différents compteurs ont été représentés par des lettres expliqué au tableau de l'annexe 2. [35]

Tableau 4 : KPIs d'Accessibilité de la quatrième génération

<i>La famille des KPIs</i>	KPIs	Les formules des KPIs
<i>Accessibilité</i>	RRC CSSR : Taux de succès d'établissement d'appel (RRC CSSR)	$100 * \frac{[A1]}{[A2] - [A3]}$
	E RAB SR : Taux de succès d'établissement d'un E RAB	$100 * \frac{[B1] + [B2]}{[B3] + [B4]}$
	S1 CSSR : Taux de succès d'établissement d'une connexion S1	$100 * \frac{[C1]}{[C2]}$

2.4.2.2 Mobilité (Mobility)

Il s'agit de résoudre les problèmes concernant la gestion de la mobilité ce qui constitue un important défi technique à relever, afin d'empêcher la terminaison forcée de l'appel et permettre l'exécution des applications d'une manière transparente à la mobilité.

Les KPIs de mobilité sont utilisés pour évaluer les performances de la mobilité EUTRAN/UTRAN, qui est essentielle à l'expérience. Plusieurs catégories de KPI de mobilité sont définies en fonction des types de Handover intra-fréquence, inter-fréquence et inter-radio accès technologie (RAT). [34]

Nous regroupons les KPIs de mobilité de la quatrième génération dans le tableau ci-dessous où les différents compteurs ont été représentés par des lettres expliquées au tableau à l'annexe 2. [35]

Tableau 5 : KPIs de mobilité de la quatrième génération

<i>La famille des KPIs</i>	KPIs	Les formules des KPIs
<i>Mobilité</i>	HO ESR : Taux de succès sur l'exécution du passage de cellule	$100 * \frac{[D1]+[D2]}{[D3]+[D4]}$

2.4.2.3 Intégrité (Integrity)

Les KPI d'intégrité de service indiquent les impacts d'E-UTRAN sur la qualité du service fourni à l'utilisateur final : le débit fourni, la latence. Il représente la qualité éprouvée par l'utilisateur pendant l'appel ou la session. Ces KPIs peuvent être calculés par cellule ou par cluster. Plusieurs catégories de KPI de mobilité sont définies en fonction des types de Handover intra-fréquence, inter-fréquence et inter-radio accès technologie (RAT).

Nous regroupons les KPIs d'intégrité de la quatrième génération dans le tableau ci-dessous où les différents compteurs ont été représentés par des lettres expliquées au tableau à l'annexe 2.

Tableau 2-3 KPIs d'intégrité de la quatrième génération. [35]

Tableau 6 : KPIs de mobilité de la quatrième génération

<i>La famille des KPIs</i>	KPIs	Les formules des KPIs
<i>Intégrité</i>	DL Cell TP : Débit du canal descendant	$\frac{[E1]}{[E2]}$
	UL Cell TP : Débit du canal montant	$\frac{[E3]}{[E4]}$

2.4.2.4 Maintien de Service (Retainability)

Les KPI de maintien de service sont utilisés pour évaluer la capacité du réseau à retenir les services demandés par un utilisateur pour une durée souhaitée une fois que l'utilisateur est connecté aux services. Ces compteurs peuvent être calculés par eNodeB (en regroupant tous les compteurs des cellules) ou par cellule.

Ces KPIs sont importants pour évaluer si notre système réseau cœur et radio peut maintenir la qualité de service. Ce KPI est un indicateur de la quantité du handover ainsi que le call drop dont la visée de ce travail est plus concentrée.

Nous regroupons les KPIs de maintien de service de la quatrième génération dans le tableau ci-dessous où les différents compteurs ont été représentés par des lettres expliqué au tableau à l'annexe 2. [35]

Tableau 7 : KPIs de continuité de la quatrième génération

<i>La famille des KPIs</i>	KPI	Les formules des KPIs
<i>Maintien de Service</i>	E-RAB Drop : Taux de coupure E-RAB	$100 * \frac{[F1] + [F2]}{[F3] + [F4] + [F5]}$

2.5 Perte des appels (Call drop)

Une interruption d'appel en LTE signifie que la session en cours d'un utilisateur est interrompue nécessitant l'initiation d'une nouvelle connexion pour reprendre les services. Les appels interrompus entraînent non seulement une mauvaise expérience pour l'utilisateur mais également une diminution du nombre d'abonnés, ce qui affecte les activités des opérateurs cellulaires. D'autre part les interruptions d'appels provoquent une libération anormale des ERAB (Radio Acces Bearer Abnormal Release).

Dans un système à commutation de paquets tel que LTE, il est naturel d'établir et de libérer des ERAB. Les E-RAB n'ont pas besoin d'être libérés simplement parce qu'ils sont inactifs, ils peuvent être conservés pour avoir un accès rapide une fois que de nouvelles données arrivent. La définition d'une libération anormale est que la libération de l'E-RAB a eu un impact négatif sur l'utilisateur final. [36]

2.5.1 Les causes d'un call drop

La capacité de maintien de service est définie comme la capacité d'un utilisateur à conserver le service demandé une fois connecté pendant la durée souhaitée. Elle peut être mesurée par le

nombre d'appels interrompus par seconde ou en pourcentage. Il peut y avoir de nombreuses raisons pour lesquelles un appel est interrompu sur LTE.

- Des mauvaises conditions radio qui peuvent engendrer un nombre de retransmission atteignant le maximum défini ou/et dés synchronisation avec l'eNodeB.
- Hand Over mal réglé.
- Relations de voisinage non établies
- Rejet d'admission, en raison du manque de licences.
- Un problème de transport au niveau de la liaison de l'interface S1 comme une congestion ou une interruption.

2.5.2 Les solutions au call drop

- Un réglage avec une précision de la liste des voisins serait la solution.
- Augmentation du rayon de la cellule cible afin qu'il ne dépasse pas son rayon prévu.
- Le bilan de fonctionnement de l'eNodeB ou l'optimisation physique peuvent aider.
- Ajustement fin des paramètres HO pour effectuer le handover de manière optimisée, c'est-à-dire éviter les HO retardés, mais aussi les HO trop précoces ou les ping-pongs fréquents.
- Mise à niveau de la capacité des licences ou déchargement de l'eNodeB.

2.5.3 Optimisation de Maintenabilité

Il est important de vérifier les messages de la couche 3 afin de comprendre la cause de drop. L'emplacement de ce drop doit également être vu par rapport au RSRP (Reference signal Receive Power) et au SINR (Signal Interference Noise Ratio) et l'emplacement des sites en tenant compte de la géographie de l'endroit. [36]

Lorsque le problème est défini, on passera à l'analyse de ce dernier en établissant les indicateurs clés de performance basée sur des paramètres de maintenabilité qui sont composés de plusieurs sous-paramètres calculés selon des procédures bien définies :

- La connexion RRC (Radio Resource Control).
- La configuration de l'architecture du système d'évolution E-RAB.

L'optimisation et l'amélioration de ces KPI se fait manuellement par des recommandations proposées par des ingénieurs en analysant les dégradations de ces indicateurs et c'est l'étape qui va prendre du temps alors que le service est non disponible.

De ce fait, notre travail consiste à réduire cet abs de temps en implémentant la notion de l'Intelligence Artificielle et essayer de détecter les dégradations avant qu'elle ne se produise en se basant sur des algorithmes d'apprentissage automatique (Machine Learning).

Le machine Learning (ML) est un champ d'étude de l'intelligence artificielle, qui consiste à programmer une machine pour pouvoir effectuer des tâches sans instructions explicites, similaires au processus d'apprentissage d'un humain par des exemples. Le chapitre prochain justement détaillera le concept du machine Learning, son fonctionnement ainsi que ces algorithmes.

2.6 Conclusion

L'optimisation radio est l'une des principales étapes d'amélioration des performances des réseaux cellulaires, l'optimisation consiste en plusieurs types d'analyse et d'actions à entreprendre afin de maintenir et améliorer la qualité et la capacité du réseau.

Il est important de noter que les KPIs dépendent les uns des autres, en essayant d'en améliorer un, il est fort probable d'en altérer un autre. En prenant en compte la complexité de ce réglage et l'énorme coût engendré par la sous-traitance d'une entité tierce, les entreprises de télécoms visent de nos jours l'automatisation de ce processus en utilisant l'Intelligence Artificielle.

Dans le prochain chapitre, nous allons présenter l'Intelligence Artificielle ainsi que ses méthodes de prédiction.

CHAPITRE 3 :
L'Apprentissage
Automatique

3.1 Introduction

Le Machine Learning, de son nom apprentissage automatique en français, est une forme d'intelligence artificielle permettant aux ordinateurs d'apprendre sans avoir été programmés de façon explicite. Cette technologie permet de développer des programmes informatiques pouvant changer en cas d'exposition à de nouvelles données.

Dans ce chapitre, nous exposons tout d'abord un bref aperçu sur l'intelligence artificielle, l'apprentissage automatique ainsi que ses types. Ensuite, on va se concentrer sur l'apprentissage supervisé et présenter les principaux types d'algorithmes de régression en apprentissage automatique pour envisager leurs applicabilités à notre jeu de données.

3.2 L'intelligence Artificielle IA

L'intelligence artificielle (IA) est un ensemble de théories et de techniques permettant de simuler l'intelligence humaine à l'aide de machines utilisant des programmes informatiques complexes. Les différentes méthodes d'apprentissage automatique (ou *Machine Learning*, ML) ne sont qu'un sous-groupe de l'IA. Elles émanent de la science informatique. Elles utilisent des algorithmes établis à partir de l'analyse de bases de données pour pouvoir accomplir des tâches. Parmi ces méthodes figurent les arbres décisionnels ou forêts aléatoires, les machines à vecteurs de support ainsi que les réseaux de neurones artificiels. [37]

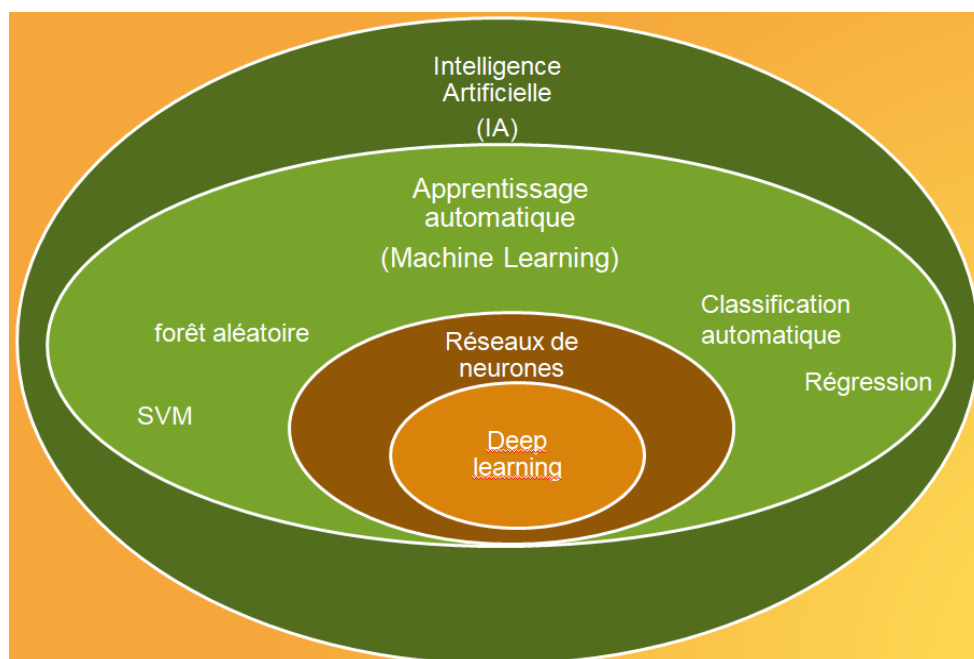


Figure 15 : le positionnement des notions d'IA, machine learning et deep learning imbriquées les unes aux autres

Comme le montre la figure ci-dessus, l'IA c'est la grande famille qui englobe toutes ces méthodes « de nouvelles formes d'intelligence ». Ainsi le machine learning fait partie de la famille des intelligences artificielles. Le deep learning quant à lui est un sous domaine du machine Learning (sa particularité étant l'utilisation de réseaux de neurones profond). [37]

3.3 L'apprentissage automatique

L'apprentissage automatique ou Machine Learning (ML) est une branche de l'intelligence artificielle qui consiste à programmer des algorithmes permettant d'apprendre automatiquement à partir des données et d'expériences passées ou par interaction avec l'environnement. Ce qui rend l'apprentissage machine vraiment utile est le fait que l'algorithme peut "apprendre" et adapter ses résultats en fonction de nouvelles données sans aucune programmation à priori.

Il existe plusieurs façons d'apprendre automatiquement à partir des données dépendamment des problèmes à résoudre et des données disponibles. [38]

3.3.1 Les types d'apprentissage automatique

La figure 3.1 ci dessous donne un sommaire des types d'apprentissage automatique les plus connus.

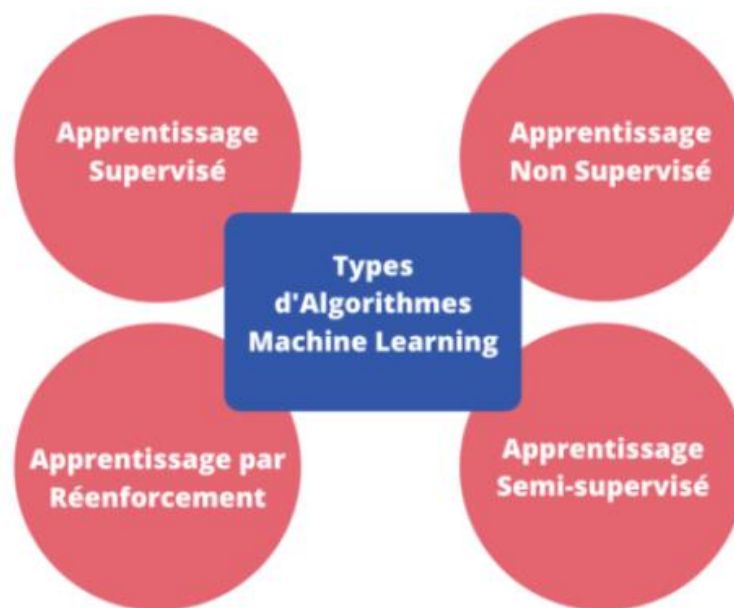


Figure 16 : Les types d'apprentissage automatique

3.3.1.1 L'apprentissage supervisé

Le terme apprentissage supervisé vient de l'idée qu'à partir d'un ensemble de données de formation (Training Data) on construit un modèle qui peut prédire le résultat des nouvelles données imprévues sur la base des exemples précédentes.

L'apprentissage automatique surveillé s'entraîne avec une base de données marquée. C'est-à-dire que les données sont étiquetées avec des informations que le modèle est destiné à déterminer pour l'apprentissage automatique, et qui peuvent même être classées de manière à ce que le modèle soit destiné à classer les données. Il nécessite moins de données d'apprentissage que les autres méthodes d'apprentissage automatique et facilite l'apprentissage car les résultats du modèle peuvent être comparés aux résultats réels marqués. Il existe deux sous-catégories au sein de l'apprentissage supervisé : la régression et la classification.

Dans les problèmes de régression, on cherche à prédire la valeur d'une variable continue, c'est à dire une variable qui peut prendre une infinité de valeurs. Dans un problème de classification, on cherche à classer un objet dans différentes classes, en d'autres termes, on cherche à prédire la valeur d'une variable discrète.

3.3.1.2 Apprentissage non supervisé

Dans le cadre de l'apprentissage non supervisé, les données sont non labellisées. L'algorithme trouve tout seul les cas de similarités parmi ses données d'entrée. Ainsi, le but de l'apprentissage non supervisé est de découvrir des modèles cachés, des corrélations, de faire des regroupements dans un ensemble de données, mais il est aussi utilisé pour la réduction de dimension. La réduction de dimension peut aider pour réduire la taille des données en sélectionnant les attributs qui se distinguent mieux pour la classification. [37]

3.3.1.3 Apprentissage semi-supervisé

Cette technique d'apprentissage est une alternative entre l'apprentissage supervisé et non supervisé. Il y a des cas où certaines données sont étiquetées mais pas toutes les données. On peut utiliser l'apprentissage non supervisé pour segmenter une grande quantité de données en deux clusters. On pourra donc labelliser ces données en fonction des caractéristiques de chaque cluster. Une fois les données labélisées, l'apprentissage supervisé servira pour la classification. [37]

3.3.1.4 Apprentissage par renforcement

Dans le cadre de l'apprentissage par renforcement, l'agent interagit avec son environnement en effectuant des actions pour recevoir des récompenses. La récompense peut être positive, négative ou nulle. Prenons par exemple, un robot qui joue au jeu de Tetris, quand il gagne il y a une récompense positive, et dans le cas contraire une récompense négative. [37]

3.3.2 Algorithmes du Machine Learning

En machine Learning, il existe plusieurs algorithmes, chacun est particulièrement pertinent pour exécuter une tâche donnée. De nombreux facteurs sont en jeu, tels que la taille et la structure de données. Par conséquent, il faut essayer de nombreux algorithmes différents pour résoudre le problème posé tout en utilisant un ensemble de données de tests pour évaluer les performances de chacun et sélectionner le meilleur. Bien sûr, les algorithmes a essayé doivent être adaptés au problème, c'est à dire choisir la bonne catégorie d'apprentissage. [39]

Dans le cas de notre étude, on s'intéresse aux algorithmes d'apprentissage supervisé qui permettent des modélisations prédictives. Il s'agit d'un problème de régression car la classification utilise l'apprentissage automatique pour prédire une valeur catégorielle, cette dernière doit être une valeur prédéfinie qui a déjà été vue dans l'ensemble de données. En revanche, le résultat d'un algorithme de régression est une valeur continue et non une valeur discrète comme dans la classification.

3.3.2.1 Régression linéaire (SLR)

Cet algorithme cherche à établir, sous forme d'une droite, une relation entre une variable expliquée et une variable explicative. Par exemple, prédire une note à un examen (variable expliquée) en fonction du nombre d'heures de révisions (variable explicative). En d'autres termes, les données d'une série d'observations sont représentées sous forme d'un nuage de points et l'on cherche à trouver une droite passant au plus près de ces points.

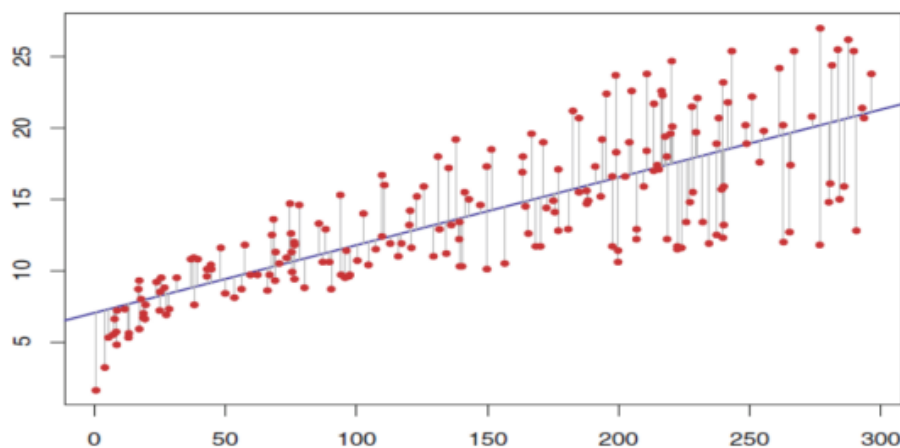


Figure 17 : Représentation graphique d'un exemple sur la régression linéaire

Sur l'image ci-dessus, il y a un exemple de dépendance entre la variable d'entrée x et la variable de sortie y . La ligne rouge dans le graphique ci-dessus est appelée la ligne droite la mieux ajustée. Sur la base des points de données donnés (exemples des notes d'examen), nous

essayons de tracer une ligne qui modélise le mieux les points. Dans le scénario du monde réel, nous avons normalement plus d'une variable d'entrée.

La régression linéaire est le modèle de régression le plus courant. [39]

3.3.3 Techniques de régularisation

Dans le contexte de la statistique et en particulier de l'apprentissage automatique, la régularisation est le nom donné à un processus visant principalement à réduire des problèmes liés au surapprentissage. Et ce, au travers d'une réduction de la variance ou d'une sélection du nombre de paramètres employés. Ce processus n'est pas neutre puisqu'il consiste à introduire une nouvelle information au problème, dans l'objectif de le "simplifier", que ce soit en introduisant une pénalité d'autant plus grande que la complexité du modèle l'est ou en "imposant" une distribution a priori des paramètres du modèle.

Le principe général de la régularisation consiste à pénaliser les valeurs extrêmes des paramètres (ce qui conduit souvent à la variance du surapprentissage). On cherche alors à minimiser à la fois le modèle par rapport à notre métrique de choix ainsi qu'une métrique sur la taille et le nombre de paramètres employés. [40]

Nous allons nous attarder sur une famille d'outils de régularisation liée aux modèles linéaires, nommément : Ridge, Lasso et Elastic net.

3.3.3.1 Technique de régularisation de crête (Ridge)

La régression Ridge est une méthode de réglage de modèle qui est utilisée pour analyser toutes les données qui souffrent de multi colinéarité. Cette méthode effectue la régularisation L2. Lorsque le problème de la multi colinéarité se produit, que les moindres carrés ne sont pas biaisés et que les variances sont importantes, les valeurs prédites sont donc éloignées des valeurs réelles.

La fonction de coût pour Ridge régression :

$$(\|y - x(\theta)\|_2^2 + \lambda \|\theta\|_2^2) \quad (1)$$

Lambda est le terme de pénalité. λ donné ici est noté par un paramètre alpha dans la fonction de Ridge. Ainsi, en changeant les valeurs d'alpha, nous contrôlons le terme de pénalité. Plus les valeurs d'alpha sont élevées, plus la pénalité est grande et donc l'amplitude des coefficients est réduite.

- Il réduit les paramètres. Par conséquent, il est utilisé pour empêcher la multi colinéarité.
- Il réduit la complexité du modèle par le retrait du coefficient.

Pour tout type de modèles d'apprentissage automatique de régression, l'équation de régression habituelle constitue la base qui s'écrit comme suit :

$$Y = X B + e \quad (2)$$

Où Y est la variable dépendante, X représente les variables indépendantes, B est les coefficients de régression à estimer et e représente les erreurs sont des résidus.

Une fois que nous ajoutons la fonction lambda à cette équation, la variance qui n'est pas évaluée par le modèle général est considérée. [41]

3.3.3.2 Technique de régularisation Lasso

Contrairement aux méthodes de régression classiques, la régression LASSO (Least Absolute Shrinkage and Selection Operator) introduit un terme de pénalisation (d'où l'appellation régression pénalisée) dans la fonction objectif, et ce afin de favoriser la parcimonie du modèle final. L'objectif principal de cette méthode est alors d'améliorer la robustesse du modèle sur un échantillon indépendant de la base d'apprentissage. Cette amélioration réside dans le fait que, bien que le modèle soit volontairement biaisé par le terme de pénalité, la variance du modèle est réduite.

$$E [Y/X] = g - 1 = (\beta_0 + \beta_1 * X_1 + \dots + \beta_p * X_p) \quad (3)$$

Où Y est la variable cible, $X = (X_1, \dots, X_p)$ est la matrice design contenant les variables explicatives, et $\beta = (\beta_0, \dots, \beta_p)$ sont les p+1 coefficient à estimer. Ainsi, qu'il s'agisse d'une minimisation de l'erreur quadratique dans le cadre d'une régression linéaire ou d'une maximisation de la vraisemblance, la régression LASSO ajoute un terme de pénalité correspondant à la norme L1 du vecteur β . [42]

3.3.3.3 Régression Elastic Net

La régularisation de type "Elastic net" consiste à combiner les deux régularisations précédentes (Ridge et Lasso) afin d'éviter la sélectivité trop forte que peut proposer Lasso tout en conservant possiblement des variables fortement corrélées, autrement dit cette technique de régression utilise les pénalités des techniques de lasso et de crête pour régulariser les modèles de régression.

La fonction de coût de la régression nette élastique :

$$cost\ function = \sum_{i=1}^n \left(Y_i - \sum_{j=1}^m \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^m |\beta_j| + \frac{1-\alpha}{2} \sum_{j=1}^m \beta_j^2$$

On peut donc choisir une valeur alpha entre 0 et 1 pour optimiser le filet. En effet, cela réduira certains coefficients et en définira certains sur 0 pour une sélection clairsemée.

La méthode d'Elastic Net améliore les limites du lasso, c'est-à-dire que le lasso prend quelques échantillons pour des données de grandes dimensions. La procédure d'Elastic net fournit l'inclusion d'un nombre "n" de variables jusqu'à saturation. Si les variables sont des groupes fortement corrélés, le lasso a tendance à choisir une variable de ces groupes et à ignorer complètement le reste. [43]

3.3.3.4 Régression polynomiale

Il est parfois difficile de trouver une droite pouvant passer parmi les points de la série d'observations de façon optimale ce qui est le cas en régression linéaire. Cependant, il est parfois possible de trouver un lien entre les variables à l'aide d'une courbe. C'est ce que permet la régression polynomiale en ajoutant des plis à la courbe à l'aide d'éléments appelés polynômes. [44]

Cet algorithme cherche à établir, sous forme d'un polynôme, une relation entre une variable expliquée et une variable explicative. Par exemple, prédire une note à un examen (variable expliquée) en fonction du nombre d'heures de révisions (variable explicative). En d'autres termes, les données d'une série d'observations sont représentées sous forme d'un nuage de points et l'on cherche à trouver une courbe passant au plus près de ces points. [44]

La représentation de la régression polynomiale est une équation qui décrit une courbe qui correspond au mieux à la relation entre les variables d'entrées x et les variables de sorties y tout en recherchant les coefficients β et en choisissant le bon degré m de chaque variable d'entrée x . Cependant, la sélection du degré exact de chaque variable nécessite naturellement une certaine connaissance du lien entre les données et la sortie. [44]

$$y = \beta_0 + \beta_1 \times x^1 + \dots + \beta_m \times x^m$$

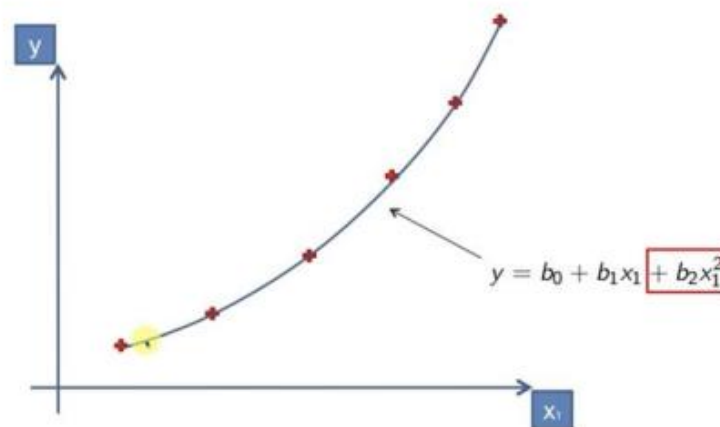


Figure 18 : Représentation de la régression polynomiale

3.3.3.5 Support Vector Regression SVR

Pour bien expliquer les SVR on doit passer par les machines à vecteurs de support qui sont un ensemble de méthodes d'apprentissage supervisé utilisées pour la classification, la régression. Toutes ces tâches sont courantes dans l'apprentissage automatique. Un simple classificateur SVM linéaire est l'un des algorithmes les plus populaires il fonctionne en traçant une ligne droite entre deux classes. Cela signifie que tous les points de données d'un côté de la ligne représenteront une catégorie et que les points de données de l'autre côté de la ligne seront placés dans une catégorie différente. Cela signifie qu'il peut y avoir un nombre infini de lignes parmi lesquelles choisir.

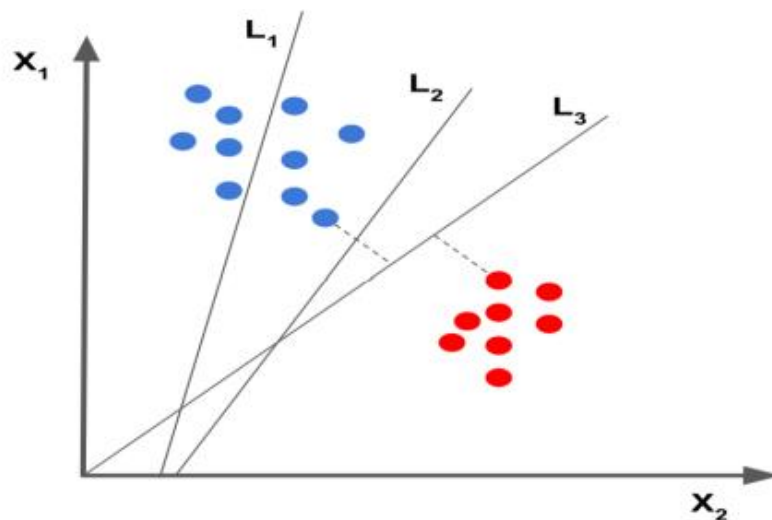


Figure 19 : Représentation des hyperplans en utilisant l'algorithme SVM

Ce qui rend l'algorithme SVM linéaire meilleur que certains autres algorithmes, c'est qu'il choisit la meilleure ligne pour classer les points de données. Il choisit la ligne qui sépare les données et qui est la plus éloignée possible des points de données proches. La Figure 3.4 Représente deux classes en utilisant l'algorithme SVM.

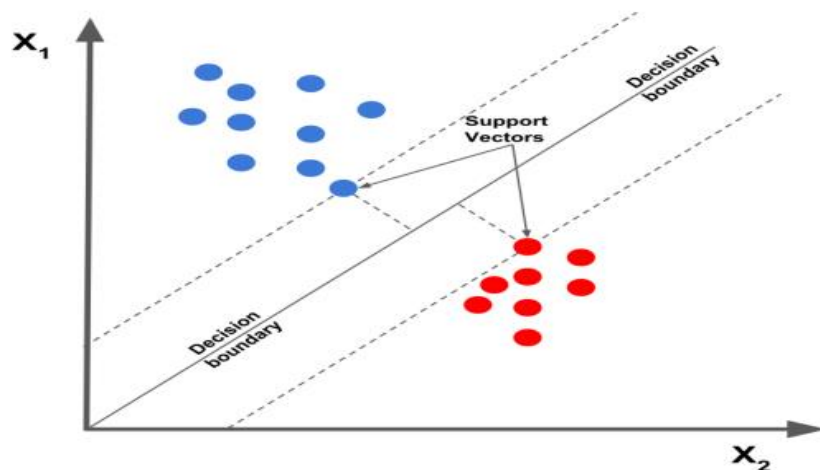


Figure 20 : Représentation de deux classes en utilisant l'algorithme SVM

Le problème de régression est une généralisation du problème de classification. Elle est accomplie en introduisant une région insensible à ε autour de la fonction, appelée ε -tube. Ce tube reformule le problème d'optimisation pour trouver le tube qui se rapproche le mieux de la fonction à valeur continue, tout en équilibrant la complexité du modèle et l'erreur de prédiction. Plus précisément, SVR est formulé comme un problème d'optimisation en définissant d'abord une perte convexe insensible à ε fonction, à minimiser et trouver le tube le plus plat qui contient la plupart des instances de formation. [45]

La fonction à valeur continue approchée peut être écrite comme dans l'équation suivante :

$$Y = f(x) = \langle w, x \rangle + b = \sum_j^M w_j x_j + b, y, b \in R, x, w \in R^M$$

Elle peut être représentée graphiquement comme suit :

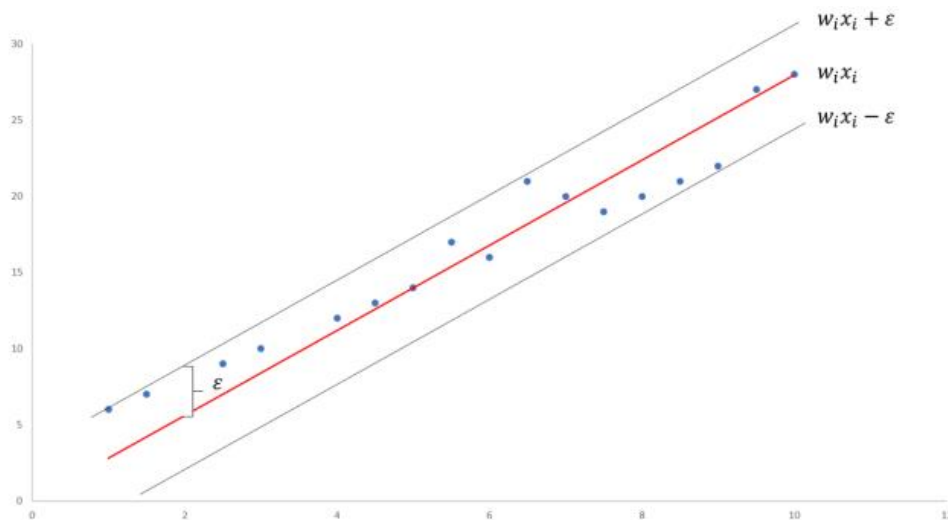


Figure 21 : SVR linéaire à une dimension

Sur ce graphe, la ligne noire représente l'hyperplan optimal. Les lignes représentent la ligne de marge, présent à une distance ε de l'hyperplan. Les SVR peuvent être utilisées dans des cas non linéaires et cela en jouant sur les paramètres d'optimisation du modèle. [46]

3.3.3.6 Arbre de décision :

Contrairement à d'autres algorithmes d'apprentissage supervisé, cet algorithme peut être utilisé pour résoudre des problèmes non seulement de classification mais aussi de régression. Le but de l'utilisation d'un arbre de décision est de créer un modèle qui peut être utilisé pour prédire la classe ou la valeur de la variable cible en apprenant des règles de décision simples déduites de données antérieures (données d'apprentissage).[47]

Dans l'algorithme d'arbre décision, pour prédire une étiquette de classe ou une valeur pour un enregistrement, nous partons de la racine de l'arbre. Nous comparons les valeurs de l'attribut racine avec l'attribut de l'enregistrement. Sur la base de la comparaison, nous suivons la branche correspondante à cette valeur et sautons au nœud suivant jusqu'à l'arriver à un nœud feuille et en sortie la valeur prédite à ce nœud feuille. [47]

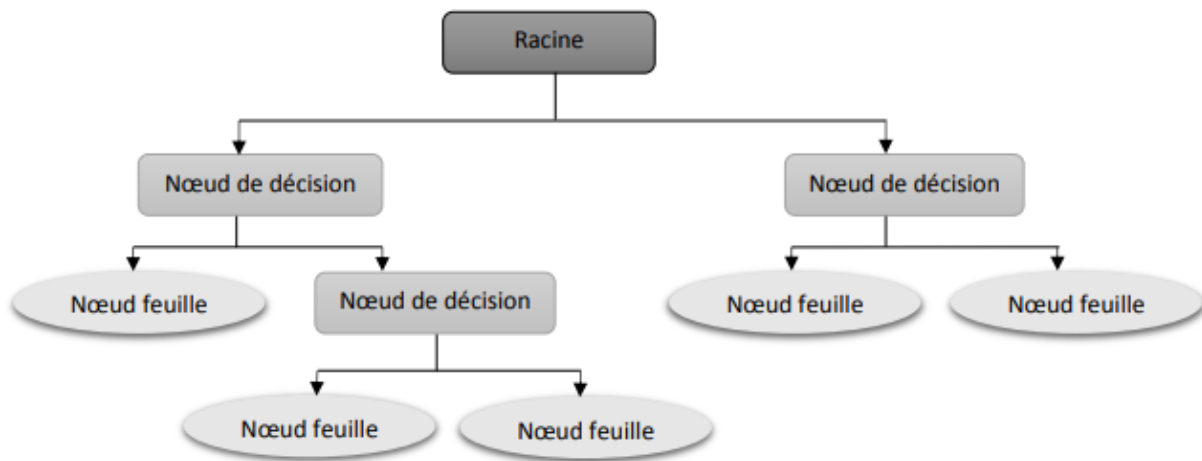


Figure 22 : Principe de fonctionnement de l'algorithme arbre de décision

3.3.3.7 Réseaux de neurones :

Un réseau de neurones artificiels, ou Artificial Neural Network en anglais, est un système informatique matériel et / ou logiciel dont le fonctionnement est calqué sur celui des neurones du cerveau humain.

Il s'agit là d'une variété de technologie Deep Learning (apprentissage profond), qui fait elle-même partie de la sous-catégorie d'intelligence artificielle du Machine Learning (apprentissage automatique). [48]

Les réseaux de neurones sont puissants pour faire des prédictions/hypothèses. Ils sont inspirés du cerveau humain. Les cercles ci-dessous représentent les nœuds et il y a plusieurs couches : couche d'entrée, couche de sortie et les couches cachées.

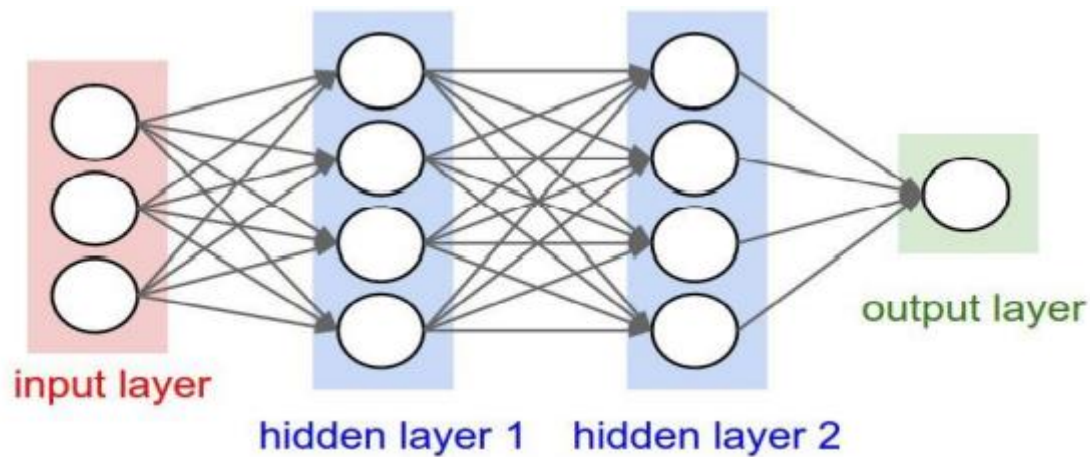


Figure 23 : Représentation d'un réseau de neurone [48]

En règle générale, un réseau de neurones repose sur un grand nombre de processeurs opérant en parallèle et organisés en tiers. Le premier tiers reçoit les entrées d'informations brutes, un peu comme les nerfs optiques de l'être humain lorsqu'il traite des signaux visuels.

Par la suite, chaque tiers reçoit les sorties d'informations du tiers précédent. On retrouve le même processus chez l'Homme, lorsque les neurones reçoivent des signaux en provenance des neurones proches du nerf optique. Le dernier tiers, quant à lui, produit les résultats du système.[48]

Par le biais d'un algorithme, le réseau de neurones artificiels permet à l'ordinateur d'apprendre à partir de nouvelles données. L'ordinateur doté du réseau de neurones apprend à effectuer une tâche en analysant des exemples pour s'entraîner. Ces exemples ont préalablement été étiquetés afin que le réseau puisse savoir ce dont il s'agit.

Contrairement à d'autres types d'algorithmes, les réseaux de neurones ne peuvent pas être programmés directement pour effectuer une tâche. A la manière du cerveau en développement d'un enfant, la seule instruction qu'ils ont est d'apprendre.

On distingue toutefois trois méthodes d'apprentissage distinctes. Dans le cas de **l'apprentissage supervisé**, l'algorithme s'entraîne sur un ensemble de données étiquetées et se modifie jusqu'à être capable de traiter la base de données pour obtenir le résultat souhaité. Dans le cas de **l'apprentissage non-supervisé**, les données ne sont pas étiquetées. Le réseau de neurones analyse l'ensemble de données, et une fonction-coût lui indique dans quelle mesure il est éloigné du résultat souhaité. Le réseau s'adapte alors pour augmenter la précision de l'algorithme.

Enfin, avec la méthode de **l'apprentissage renforcé**, le réseau de neurones est renforcé pour les résultats positifs et sanctionné pour les résultats négatifs. C'est ce qui lui permet d'apprendre au fil du temps, de la même manière qu'un humain apprend progressivement de ses erreurs.

3.3.3.8 K-plus proche voisin :

K-Nearest Neighbors (KNN), ou le k-plus proche voisin, est un algorithme simple d'apprentissage automatique supervisé qui peut être utilisé pour résoudre des problèmes de classification et de régression. Dans la classification, il est basé sur le vote majoritaire des voisins. Un objet est classé par un vote majoritaire de ses voisins, l'objet étant attribué à la classe la plus commune parmi ses k voisins les plus proches, où k est le numéro du voisin de l'objet. Le choix du paramètre k est très crucial dans cet algorithme dont le meilleur choix dépend des données. En général, des valeurs plus élevées de k réduisent l'influence du bruit sur la classification [49].

Dans la figure 2.7, si on choisit $k = 3$, l'algorithme cherche les trois plus proches voisins du cercle rouge pour pouvoir le classer soit dans la classe des cercles, soit dans la classe des carrés. Dans ce cas, les trois plus proches voisins du cercle rouge sont un carré et deux cercles. Par conséquent, l'algorithme classera le cercle rouge dans la classe des cercles.

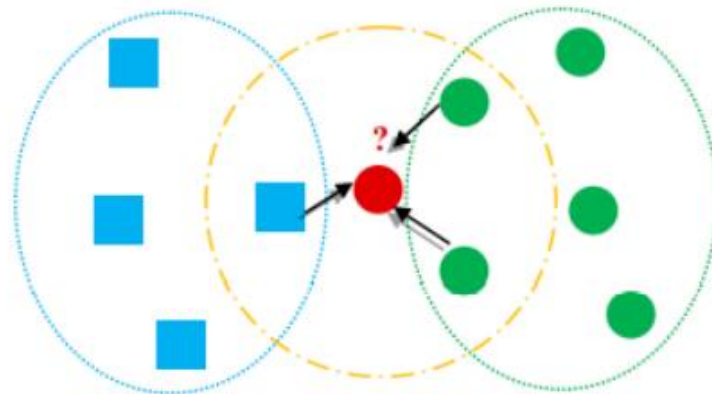


Figure 24 : Exemple de classification par l'algorithme KNN [49]

3.3.3.9 Les forêts aléatoires pour la régression (Random Forest Regression) :

Un des principaux inconvénients des arbres de décision est qu'ils ont tendance à se surajuster aux données de formation. Les forêts aléatoires sont un moyen de remédier à ce problème. Une forêt aléatoire est essentiellement une collection d'arbres de décision, où chaque arbre est légèrement différent des autres. L'idée derrière les forêts aléatoires est que chaque arbre peut faire un assez bon travail de prévision, mais qu'il va probablement sur-apprendre une partie des données. Si nous construisons plusieurs arbres, qui fonctionnent tous bien et se surajoutent de différentes façons, nous pouvons réduire le nombre de surajouts en faisant la moyenne de leurs résultats. Cette réduction du sur-ajustement, tout en conservant le pouvoir de prédiction des arbres, peut être démontrée par des calculs mathématiques rigoureux.

Pour mettre en œuvre cette stratégie, nous devons construire de nombreux arbres de décision. Chaque arbre doit permettre de prévoir l'objectif de manière acceptable et doit également être différent des autres arbres. Les forêts aléatoires tirent leur nom de l'injection de hasard dans la construction des arbres pour s'assurer que chaque arbre est différent. Les arbres d'une forêt aléatoire sont randomisés de deux manières : en sélectionnant les points de données utilisés pour construire un arbre et en sélectionnant les caractéristiques de chaque test de division. [52]

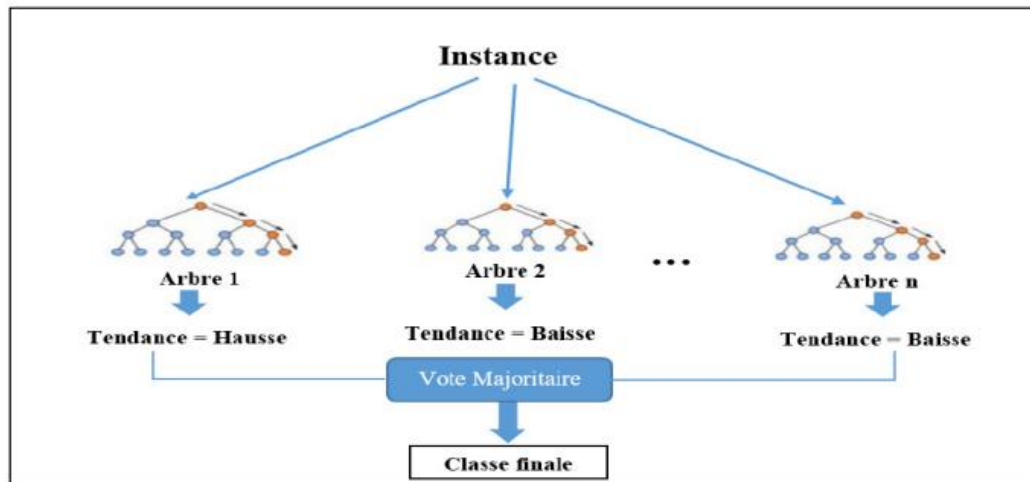


Figure 25 : Description de l'algorithme de la forêt aléatoire

Les forêts aléatoires est l'algorithme le plus connu de la technique d'apprentissage ensembliste « bagging » ou « bootstrap aggregating » qui consiste à créer plusieurs entités d'un même modèle (plusieurs arbres de décisions) et d'entraîner chacun de ces arbre sur une portion aléatoire d'une collection de données, après avoir entraîné chaque arbre, nous pouvons regrouper les résultats de chaque arbre afin d'effectuer la prédiction, la figure 3.10 montre la technique du « bagging ». [50]

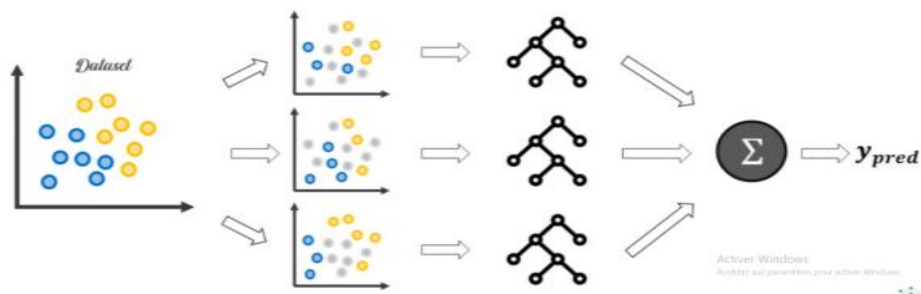


Figure 26 : Bagging [50]

3.3.4 L'apprentissage ensembliste :

L'apprentissage ensembliste (ou ensemble learning) est une technique qui repose sur la combinaison de multiples algorithmes de machine learning pour accroître les performances du modèle d'apprentissage, et parvenir à un niveau de précision supérieur à celui qui serait réalisé si on utilisait un de ces algorithmes pris séparément.

Selon les cas de figure et le résultat désiré, il est possible d'utiliser un seul et même type d'algorithme pour produire un résultat homogène, ou encore des algorithmes différents pour obtenir un résultat hétérogène. Le concept sur lequel repose la méthode ensembliste est que l'on peut améliorer les performances des modèles apprenants en les combinant.

3.3.4.1 Méthodes de l'apprentissage ensembliste

Il existe deux grands types de méthodes ensemblistes, la séquentielle et la parallèle. Avec la méthode d'ensemble learning séquentielle, les modèles sont entraînés à la suite, ce qui leur permet d'apprendre au fur et à mesure de leurs erreurs. Avec la méthode ensembliste parallèle, les modèles sont entraînés en simultané. Chaque modèle est entraîné sur un sous-échantillon de l'ensemble de données d'apprentissage. Pour déterminer le résultat final, on procède par un vote des résultats de chaque modèle pour une classification, ou par une moyenne pour une régression. [51]

La visée des méthodes d'ensemble est de combiner les prédictions de plusieurs estimateurs de base construits avec un algorithme d'apprentissage donné afin d'améliorer un modèle robuste et optimal. Les principales raisons d'utiliser cette méthode incluent les situations où il existe des incertitudes dans la représentation des données ou les techniques de modélisation. Les instances ou méthodes candidates sont appelées apprenants de base. Chaque apprenant de base fonctionne indépendamment comme une méthode de Machine Learning de base, et les résultats éventuels sont combinés pour produire un seul résultat robuste. La combinaison peut être effectuée à l'aide de l'une des méthodes de calcul de la moyenne (simple ou pondérée) et du vote (majoritaire ou pondéré) pour les méthodes de régression et de classification, respectivement.[52]

L'idée de l'ensemble learning vient d'un phénomène appelé the wisdom of the crowd. L'idée de ce dernier est qu'une foule d'individus a plus souvent raison qu'un individu tout seul. [55]

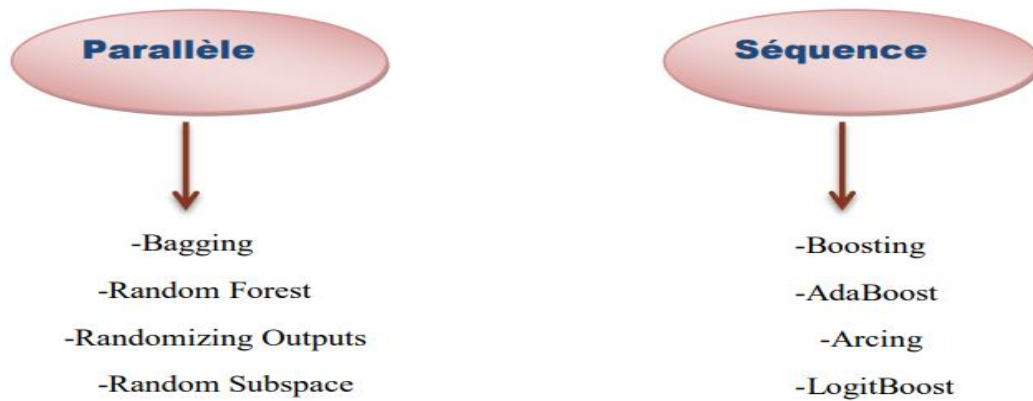


Figure 27 : Le type de méthodes ensemblistes [53]

Dans notre cas, on va s'intéresser à 3 méthodes d'ensembles :

3.3.4.2 Bagging

Le bagging (une méthode ensembliste parallèle), aussi appelé bootstrap aggregating, consiste à sous-échantillonner les données, en créant un [data set](#) pour chaque modèle. La motivation derrière cette méthode est que l'erreur de prédiction peut être réduite de manière significative en réduisant la variance.

Cette technique choisit un échantillon aléatoire dans la Base de donnée. Par la suite, chaque modèle est généré à partir des échantillons (échantillons bootstrap) fournis par les données d'origine avec remplacement connu sous le nom d'échantillonnage de ligne. Cette étape d'échantillonnage de lignes avec remise est appelée bootstrap. Désormais, chaque modèle est entraîné indépendamment, ce qui génère des résultats dissemblables. Le résultat final est basé sur la moyenne après avoir combiné les résultats de tous les modèles. On appelle cette dernière Agrégation.[54][56]

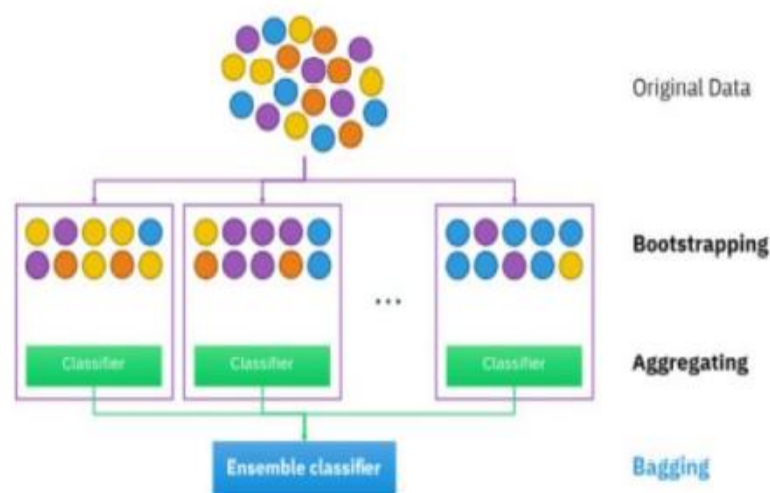


Figure 28 : Les étapes du Bagging

3.3.4.3 Le boosting

Pour cette méthode, les classifieurs de base sont générés de manière séquentielle (par exemple AdaBoost) et dépendante, contrairement aux méthodes parallèles. À chaque fois qu'un classifieur de base est entraîné, les instances mal classées précédemment sont pondérées avec un poids plus élevé, dans le but que lors des prochaines itérations, les nouveaux modèles corrigent les erreurs des modèles précédents, ce qui devra améliorer la performance globale.

3.3.4.4 Le stacking

Il s'agit d'un procédé qui consiste à combiner plusieurs modèles très différents dans le but d'améliorer la qualité de la prédiction finale. Le fonctionnement de cette technique est décrit dans le schéma en bas.

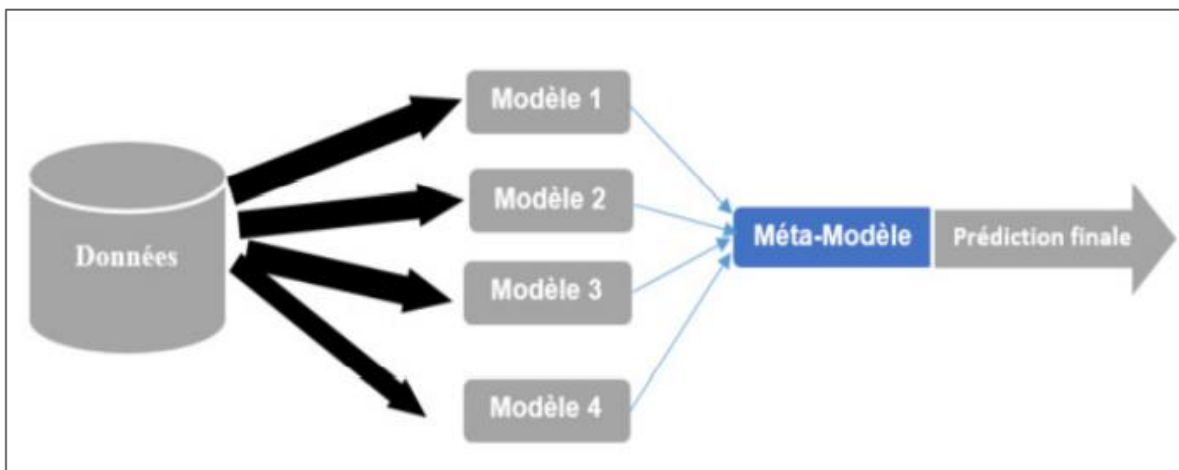


Figure 29 : Description du fonctionnement du Stacking [57]

On peut par exemple utiliser une combinaison de modèles, réseau de neurone artificiel, une régression logistique et un SVM pour construire un seul modèle. On attribuera un poids à chaque modèle pour aboutir à notre modèle agrégé.[58]

3.3.4.5 Conclusion sur les méthodes d'ensemble

Les méthodes d'ensemble ont connu du succès dans une variété de problèmes et elles ont démontré qu'ils sont capables de faire face à plusieurs défis d'apprentissage automatique comme le biais et la variance de prédiction. Toutefois, certains désavantages de ces méthodes méritent d'être soulignés. Cela inclut la mémoire supplémentaire et les coûts de calcul en combinant plusieurs classifieurs et en les gardant prêts à être traités. La perte d'interopérabilité peut également être une source de préoccupation, vu qu'on s'intéresse plus à la maximisation de la performance des prédictions.[59]

3.4 Conclusion

Dans ce chapitre, nous avons fait une description de l'intelligence artificielle et de toutes les ressources nécessaires en Machine Learning, plus précisément les algorithmes d'apprentissage automatique qui peuvent résoudre des problèmes de régression, comme c'est le cas pour nous. Pour cela, on a consacré ce chapitre pour présenter d'une façon générale quelques algorithmes indispensables à la conception d'un système intelligent qui répond à notre problématique.

Dans le prochain et le tout dernier chapitre, on va résoudre le problème de régression prédictif à partir d'une base de donnée fournis par Ericsson.

CHAPITRE 4 :
Expérimentations et
Analyse des Résultats

4.1 Introduction

Nous avons effectué un stage pratique chez l'entreprise Ericsson à Alger, nous avons été accueillis par les ingénieurs où nous avons eu une visite guidée. Ils nous ont expliqué le fonctionnement interne de l'entreprise et ce qu'ils font par rapport à la maintenance et l'installation des équipements LTE.

Ils nous ont ainsi expliqué les problèmes rencontrés pour maintenir une qualité de service satisfaisante et répondre aux besoins de leurs clients, plus précisément la problématique de la coupure d'appel Call Drop en optimisant les KPI de maintien de service qui impacte énormément la QoS de l'utilisateur.

Le but de ce travail est d'étudier le comportement aléatoire des PI du réseau LTE en leur appliquant les algorithmes de machine learning essayant de trouver le bon estimateur prédictif. Dans ce chapitre, nous démontrerons l'évaluation et la validité des différents algorithmes étudiés et présentés dans le chapitre précédent pour notre cas d'étude.

4.2 Python pour l'IA

Python est le langage le plus populaire dans le monde de l'intelligence artificielle. Il est orienté objet et se veut relativement facile d'accès. Il est très utilisé au sein de la communauté scientifique et particulièrement dans le domaine de l'intelligence artificielle. Il peut également être utilisé comme langage frontal. Les principaux frameworks de machine learning et deep learning sont effectivement disponibles en Python.

Nous avons utilisé le langage de programmation python car il consiste en une grande bibliothèque standard. Python est un langage de programmation (au même titre que le C, C++, fortran, java . . .), développé en 1989. [60] [61]

4.2.1 Présentation du langage Python

Le langage de programmation Python a été créé par Guido van Rossum en 1990 et est rendu disponible sous licence libre. Son développement est aujourd'hui assuré par la « Python Software Foundation ». Python est un langage de programmation puissant et facile à apprendre, il dispose de structures de données de haut niveau et permet une approche simple mais efficace de la programmation orientée objet. Parce que sa syntaxe est élégante, que son typage est dynamique et qu'il est interprété, Python est un langage idéal pour l'écriture de scripts et le développement rapide d'applications dans de nombreux domaines et sur la plupart des plateformes. [60]

4.2.2 Principales caractéristiques du langage Python

C'est un langage open source ; Python est gratuit même pour les usages commerciaux. De nombreux outils standards sont disponibles. Python est facile à apprendre et à utiliser. C'est un langage de script qui le permet aussi créer des applications fonctionnelles rapidement et avec peu d'effort. C'est un langage interprété; de nombreux interprètes interactifs sont disponibles. C'est orienté objet, mais la POO n'est pas obligatoire, il a aussi des outils pour le rendre possible programmation fonctionnelle ou impérative. C'est un langage très ouvert, il possède plusieurs interfaces avec C/C++/FORTRAN. Python a une excellente portabilité ; il fonctionne sur tous les systèmes d'exploitation et plateformes informatiques (les différentes variantes d'Unix, sur MacOS, BeOS, NeXTStep, MS-DOS et les différentes variantes de Windows). De nombreux modules sont disponibles dans les bibliothèques et librairies optimisées écrit en C, C++. En plus des modules intégrés dans la distribution Python standard, il existe des bibliothèques dans tous les domaines : scientifique, base de données, tests de fonctionnement et contrôle de qualité, graphiques 2D et 3D, etc ... [60]

4.2.3 Environnement de développement pour Python

Nous avons utilisé le langage de programmation Python la version 3.9. Python est un langage de programmation général, interprété, interactif, orienté objet et de haut niveau [61]. Il a été créé par Guido van Rossum entre 1985 et 1990. Comme Perl, le code source Python est également disponible sous licence GNU ou General Public License (GPL).

La figure 4.1 illustre le logo du langage Python :



Figure 30 : Logo de Python

4.2.4 Framework de programmation

Nous avons utilisé le framework Anaconda car c'est un distributeur libre et open source du langage de programmation Python appliqué au développement d'applications dédié à la science de données et à l'apprentissage automatique (traitement de données à grande échelle, analyse prédictive, calcul scientifique) [62] [63]

La figure 4.2 illustre le logo du framework de programmation Anaconda :



Figure 31 : Logo d'anaconda

Nous avons utilisé l'environnement de développement Spyder (nommé Pydee dans ses premières versions). C'est un environnement de développement pour Python. Libre (Licence MIT) et multiplateforme (Windows, Mac OS, GNU/Linux) [63] qui contient nombreuses bibliothèques d'usage scientifique : Matplotlib, NumPy, SciPy et Python.[63]



Figure 32 : Logo de l'environnement de développement Spyder

4.2.5 Principaux modules et bibliothèques utilisés dans nos programmes

4.2.5.1 Scikit –learn

C'est un outil simple utiliser pour l'analyse prédictive des données. Scikit-learn peut prendre en charge de nombreux algorithmes non supervisés et supervisés. Nous pouvons effectuer différentes tâches telles que la classification, la régression, le regroupement et d'autres tâches d'exploration de données avec cette bibliothèque Python. Cette bibliothèque comprend des algorithmes tels que les k-voisins, les forêts aléatoires, et prend en charge les machines vectorielles, etc. [63]

Le modèle ML est développé en utilisant la librairie Scikit-learn. Cette librairie utilise Scipy (Scientific Python) et contient les libraires Numpy, Matplotlib et Pandas.

4.2.5.2 Pandas

C'est une bibliothèque qui permet de manipuler facilement des données à analyser. Il entre en jeu avant que l'ensemble de données ne soit préparé pour l'entraînement. Les pandas facilitent le travail avec des séries chronologiques et des données multidimensionnelles structurées pour les programmeurs d'apprentissage automatique.

4.2.5.3 NumPy

Cette bibliothèque permet d'effectuer des calculs numériques avec Python. Elle introduit une gestion facilitée des tableaux de nombres.

4.2.5.4 Matplotlib

C'est une librairie compréhensive pour tracer et visualiser des diagrammes 2D et 3D, notamment des graphiques, histogrammes, diagrammes d'erreur et des nuages de points. [63]

4.3 La présentation des données utilisées

Nous avons reçu un fichier Excel de la part de l'entreprise Ericsson, montré dans le tableau 4.1, contenant les données recueillies à partir des différentes cellules d'un réseau 4G LTE.

Tableau 8 : Capture écran du fichier Excel de données reçu

Date	Site_ID	Sector	Band	Average RRC Connected Users	eNodeB_Drop [%]	pmErabRelAbnormalEnbAct	pmErabRelAbnormalEnb	pmErabRelNormalEnb	pmErabRelMme
01/04/2021	XX0300	Sector_1	L1800	25,32	0,22	163	323	72333	2829
01/04/2021	XX0300	Sector_2	L1800	48,51	0,37	440	669	114403	4703
01/04/2021	XX0300	Sector_3	L1800	32,42	0,23	196	281	81310	2226
01/04/2021	XX0301	Sector_1	L1800	30,43	0,22	171	330	76019	3009
01/04/2021	XX0301	Sector_2	L1800	57,16	0,29	450	820	150575	4101
01/04/2021	XX0301	Sector_3	L1800	12,82	0,44	163	239	34947	1466
01/04/2021	XX0302	Sector_1	L1800	31,08	0,34	287	493	80093	2679
01/04/2021	XX0302	Sector_2	L1800	58,69	0,30	448	789	145046	4774
01/04/2021	XX0302	Sector_3	L1800	29,51	0,22	172	279	76520	2520
01/04/2021	XX0303	Sector_1	L1800	9,97	0,24	73	211	29420	954
01/04/2021	XX0303	Sector_2	L1800	23,07	0,22	133	201	57061	1888
01/04/2021	XX0303	Sector_3	L1800	12,66	0,43	144	172	32207	1206
01/04/2021	XX0304	Sector_1	L1800	37,41	0,35	303	516	82561	2598
01/04/2021	XX0304	Sector_2	L1800	22,35	0,22	127	284	56527	1702
01/04/2021	XX0304	Sector_3	L1800	27,02	0,27	230	395	82969	2073
01/04/2021	XX0305	Sector_1	L1800	8,56	0,20	46	104	22139	520
01/04/2021	XX0305	Sector_2	L1800	6,52	0,17	30	53	17474	311
01/04/2021	XX0305	Sector_3	L1800	33,52	0,26	225	397	86019	1674
01/04/2021	XX0306	Sector_1	L1800	41,53	0,49	331	583	63401	3138
01/04/2021	XX0306	Sector_2	L1800	14,32	0,25	99	193	38427	1062
01/04/2021	XX0306	Sector_3	L1800	8,82	0,76	185	309	22347	1758
01/04/2021	XX0307	Sector_1	L1800	78,29	0,58	605	854	97414	5508
01/04/2021	XX0307	Sector_2	L1800	47,51	0,39	307	545	74622	4007
01/04/2021	XX0307	Sector_3	L1800	39,79	0,35	246	477	67683	2337

Cette base de donnée a été étendue sur 4 mois, du 01 avril 2021 jusqu'au 31 juillet 2021.

On a filtré notre base de données en prenant seulement les colonnes qui nous importent.

Tableau 9 : Descriptif des colonnes utiles du fichier Excel reçu.

Colonne	Explication
Date	La date de prise des données
Average RRC Connected Users	Nombre d'utilisateur moyen de la cellule pendant les 24hrs
pmErabRelAbnormalEnbAct	Le compteur de drop anormal de ERAB au niveau du EnodeB ou il y a des données dans le buffer.

Ces données ont été rassemblées grâce au logiciel de collecte des données de l'entreprise Ericsson.

4.4 Approche proposée

Afin de voir l'impact des modèles du Machine Learning sur nos données, on fera dérouler l'algorithme suivant, ses étapes sont représentées comme suit :

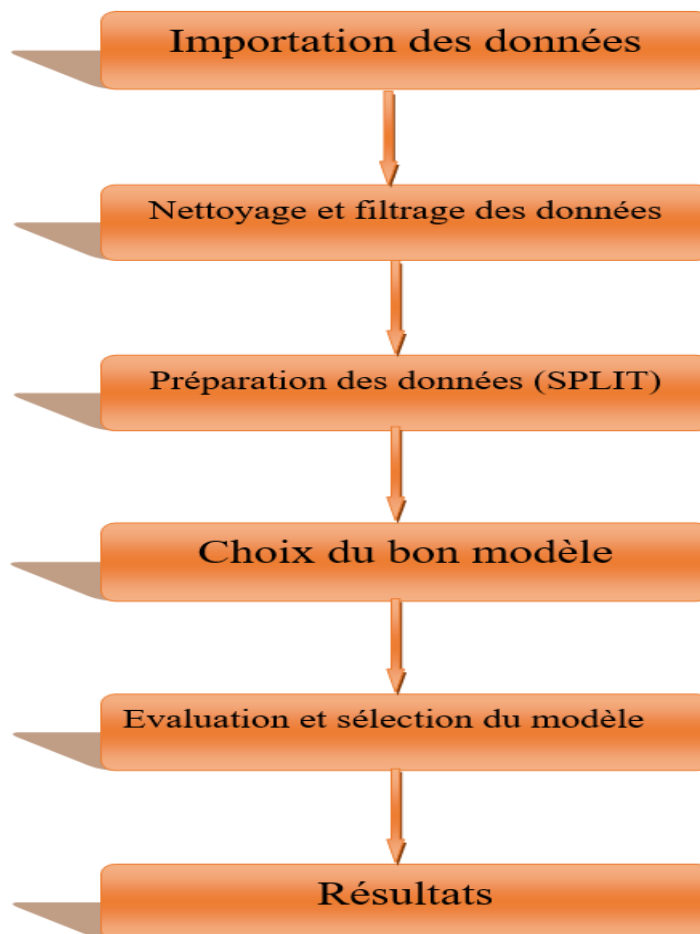


Figure 33 : Vue d'ensemble de l'algorithme utilisé dans notre travail

4.4.1 Importation et nettoyage des données

Cette étape est effectuée dans le cadre du prétraitement des données pour nettoyer les données. Après l'importation des données du fichier Excel avec la fonction `pd.read_excel` de la librairie Pandas, nous cherchons à détecter et traiter les valeurs aberrantes et les valeurs manquantes pour supprimer toutes les lignes qui contiennent une ou plusieurs cases vides.

Index	Date	Site ID	Sector	Band	Average RRC Connected Users	NodeB Drop Rate	pmErAbnAct	pmErAbnNb	pmErAbnNormalNb	pmErAbnTime
0	737881	XX0300	Sector_1	L1800	25.3189	0.215937	163	323	72333	21
1	737881	XX0300	Sector_2	L1800	48.5113	0.367355	440	669	114403	41
2	737881	XX0300	Sector_3	L1800	32.4226	0.233843	196	281	81310	21
3	737881	XX0301	Sector_1	L1800	30.4343	0.215479	171	330	76019	30
4	737881	XX0301	Sector_2	L1800	57.1628	0.289397	450	820	150575	41
5	737881	XX0301	Sector_3	L1800	12.8153	0.444723	163	239	34947	14
6	737881	XX0302	Sector_1	L1800	31.0806	0.344683	287	493	80093	20
7	737881	XX0302	Sector_2	L1800	58.6867	0.297459	448	789	145046	41
8	737881	XX0302	Sector_3	L1800	29.5084	0.216846	172	279	76520	21

Figure 34 : Visualisation de la Base de Donnée avant le filtrage dans l'environnement Spyder

On a pu visualiser la Base de donnée sous forme de nuage de points en utilisant cette ligne de code :

`sns.pairplot(data).`

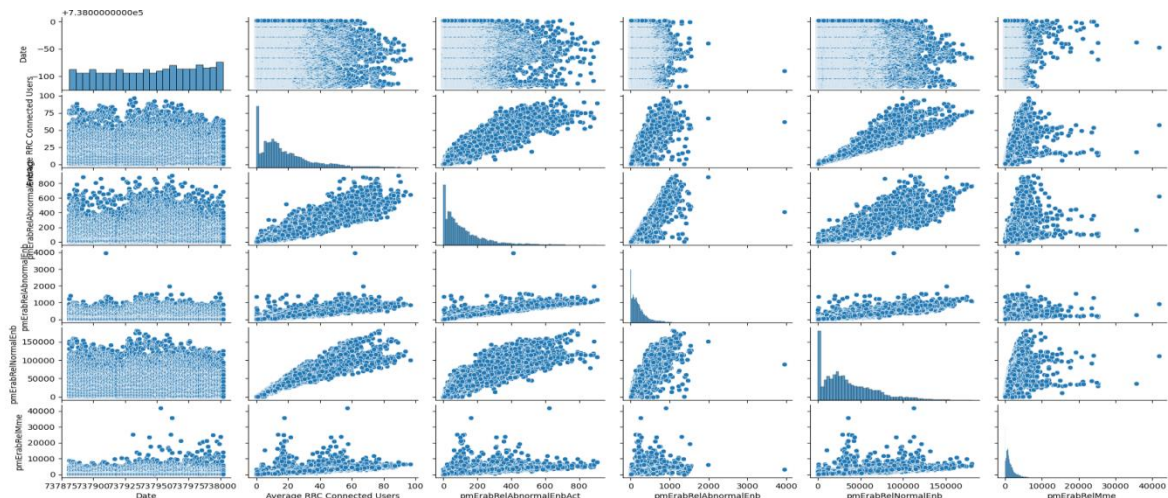


Figure 35 : Visualisation de la Base de donnée sous forme de nuage de points

Pour supprimer les valeurs aberrantes (voir annexe 4) de la Base de donnée, il faut d'abord les détecter. Pour cela, nous pouvons utiliser plusieurs méthodes, dans notre cas, on a utilisé la fonction `boxplot()` du diagramme en boîte connu sous le nom de boîte à moustache.

Une fois on les a détecté, on remplace ces valeurs par 'NaN' qui représente Not a Number en anglais pour les localiser afin de les supprimer en utilisant la fonction `df.dropna()` de la librairie Pandas.

```
Date      0
Site_ID    0
Sector     0
Band       0
Average RRC Connected Users  0
eNodeB_Drop [%]  0
pmErabRelAbnormalEnbAct      786
pmErabRelAbnormalEnb         0
pmErabRelNormalEnb          0
pmErabRelMme                 0
```

Figure 36 : Les valeurs aberrantes avant la suppression

```
Date      0
Site_ID    0
Sector     0
Band       0
Average RRC Connected Users  0
eNodeB_Drop [%]  0
pmErabRelAbnormalEnbAct      0
pmErabRelAbnormalEnb         0
pmErabRelNormalEnb          0
pmErabRelMme                 0
```

Figure 37 : Les valeurs aberrantes après la suppression

Pour supprimer les valeurs manquantes (voir annexe 4) qui représentent des cases vides, on utilise la fonction `df.dropna()`.

```
Date      0
Site_ID    0
Sector     0
Band       0
Average RRC Connected Users  21
eNodeB_Drop [%]  21
pmErabRelAbnormalEnbAct      0
pmErabRelAbnormalEnb        21
pmErabRelNormalEnb          21
pmErabRelMme                21
```

Figure 38 : Les valeurs manquantes avant la suppression

```

Date      0
Site_ID   0
Sector    0
Band      0
Average RRC Connected Users  0
eNodeB_Drop [%]  0
pmErabRelAbnormalEnbAct      0
pmErabRelAbnormalEnb         0
pmErabRelNormalEnb           0
pmErabRelMme                  0

```

Figure 39 : Les valeurs manquantes après la suppression

On a nommé la nouvelle Base de donnée df comme cela est montré dans la figure ci-dessous :

df - DataFrame

Index	Date	Average RRC Connected Users	pmErabRelAbnormalEnbAct
0	737881	25.3189	163
2	737881	32.4226	196
3	737881	30.4343	171
5	737881	12.8153	163
6	737881	31.0806	287
8	737881	29.5084	172
9	737881	9.96644	73
10	737881	23.0719	133
11	737881	12.659	144

Figure 40 : Visualisation de la nouvelle Base de Donnée après le filtrage dans l'environnement Spyder

Il est à noter que le filtrage des colonnes a été fait également avec la fonction `df.dropna()`.

Name	Type	Size	Value
data	DataFrame	(12843, 10)	Column names: Date, Site_ID, Sector, Band, Average RRC Connected Users ...
df	DataFrame	(12036, 3)	Column names: Date, Average RRC Connected Users, pmErabRelAbnormalEnbA ...

Figure 41 : Comparatif de la taille des deux Base de donnée, Data et df

4.4.2 Préparation des données (Split)

La préparation de données consiste à diviser aléatoirement la Base de donnée en deux sous-ensembles (train/test) afin de tester les performances de nos algorithmes.

- Train set : est la base de donnée utilisé pour entrainer le modèle.
- Test set: est la base de donnée utilisé pour évaluer le modèle.

Il y a un seul paramètre à régler qui est la taille de l'ensemble de données d'entraînement ou de test, souvent exprimé sous forme d'un pourcentage.

Nous avons divisé l'ensemble de données comme suit :

- 80% pour l'ensemble de données d'entraînement.
- 20% pour l'ensemble de données de test.

Name	Type	Size	Value
X_test	DataFrame	(2408, 2)	Column names: Date, Average RRC Connected Users
X_train	DataFrame	(9628, 2)	Column names: Date, Average RRC Connected Users
Y_test	Series	(2408,)	Series object of pandas.core.series module
Y_train	Series	(9628,)	Series object of pandas.core.series module

Figure 42 : Dimension de l'ensemble de données

Nous avons utilisé un ensemble de données composé de 12036 lignes et 3 colonnes, ce qui donne une valeur de 36108 échantillons. 9628 échantillons ont été utilisés comme données d'entraînement et 2408 ont été utilisés comme données de test.

Prenons en considération que les facteurs (features) dont le modèle va apprendre avec sont : Date et Average RRC Connected Users, ils représentent X_test et X_train. Et notre target variable Y est pmErabRelAbnormalEnbAct qui représente Y_test et Y_train comme cela est montré sur la figure ci-dessus.

```
def preprocess_inputs(df1):
    df1 = df1.copy()
    X = df1.drop('pmErabRelAbnormalEnbAct', axis = 1)
    Y = df1['pmErabRelAbnormalEnbAct']
    plt.show()
    X_train, X_test, Y_train, Y_test = train_test_split(X, Y, train_size = 0.8, shuffle = True, random_state = 1)
    return X_train, X_test, Y_train, Y_test
X_train, X_test, Y_train, Y_test = preprocess_inputs(df)
```

Figure 43 : Préparation et entraînement des données par Python

4.4.3 Le choix du bon modèle

Souvent, la partie la plus difficile de la résolution d'un problème d'apprentissage automatique peut être de trouver le bon estimateur pour le travail.

Y'a plusieurs modèles d'algorithmes d'apprentissage automatique mais pour ne pas se perdre, on a décidé de suivre la carte de Scikit-Learn suivante qui pose assez bien la problématique :

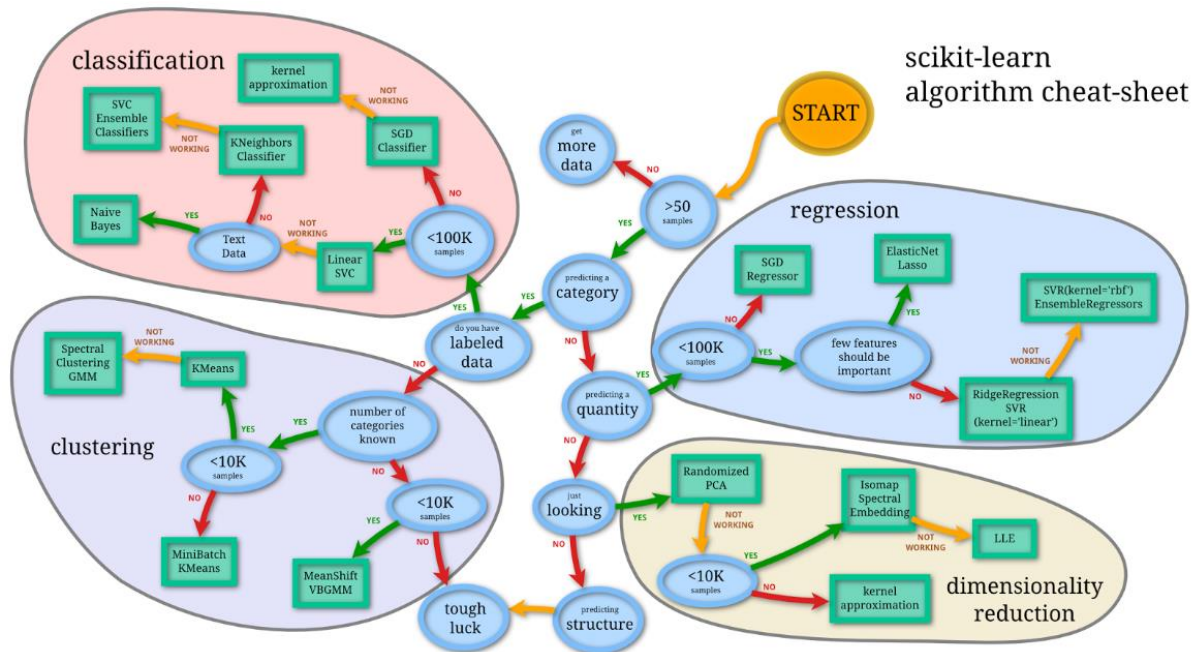


Figure 44 : La carte de Scikit-Learn du Site officiel [64]

La première condition sert à calculer le nombre d'échantillons de notre Base de donnée. Si ce dernier a plus de 100 000 échantillons, le régresseur SGD avec approximation du noyau peut être choisi car il a une complexité temporelle compétitivement moindre.

Le nombre d'échantillons de notre Base de donnée contient 36108 échantillons et comme ce dernier est inférieur à 100 000 échantillons, nous allons éliminer le SGD regressor de la liste prévue pour le meilleur estimateur.

Cela nous amène à une autre condition qui est « l'importance des facteurs ». Cette condition sert à calculer l'importance des facteurs de notre Base de donnée.

Comme nous n'avons pas su considérer cette condition, nous avons opté pour l'implémentation de tous les modèles de régression mentionnés dans le schéma qui sont : la régression Lasso, la régression Elastic Net, la régression Ridge, SVR avec un noyau linéaire, SVR avec un noyau RBF et Random Forest qui utilise la méthode d'ensemble.

Le processus de l'entraînement consiste à fournir des données à l'apprenant pour qu'il puisse construire un modèle. Nous formons nos modèles en utilisant un ensemble de données que nous avons préparé dans l'étape précédente.

4.5 Evaluation et sélection du modèle

Dans cette partie, nous avons évalué les performances de généralisation des modèles que nous avons cités précédemment à travers la carte Scikit-Learn en utilisant la technique de validation croisée (CV) afin de fiabiliser les résultats obtenues. Le score de validation nous donne une

idée sur les performances du modèle. En outre, cela nous permet d'identifier des problèmes tels que le sous-ajustement et le surajustement.

4.5.1 Résultats

Nous avons évalué les performances des modèles en comparant les métriques de la régression utilisant les paramètres par défaut.

On a commencé par l'implémentation des modèles dans le programme. Ensuite, nous avons évalué les modèles avec les paramètres par défaut.

Faut noter qu'on a eu le résultat du coefficient de détermination R^2 train et test, du RMSE train et test, du MAE train et test (voir annexe).

L'algorithme calcule ou prédit le $y_{predtrain}$ et $y_{predtest}$.

- L'évaluation du modèle de régression Lasso donne les résultats suivants :

Tableau 10 : Evaluation du modèle Lasso

	R² Train	R² Test	RMSE Train	RMSE Test	MAE Train	MAE Test
Lasso	0.8107	0.0830	39.1275	38.6003	27.5254	26.9711

- L'évaluation du modèle de régression Ridge donne les résultats suivants :

Tableau 11 : Evaluation du modèle Ridge

	R² Train	R² Test	RMSE Train	RMSE Test	MAE Train	MAE Test
Ridge	0.8107	0.8030	39.1275	38.6010	27.5219	26.9689

- L'évaluation du modèle de régression Decision Tree donne les résultats suivants :

Tableau 12 : Evaluation du modèle Decision Tree

Decision Tree	R² Train	R² Test	RMSE Train	RMSE Test	MAE Train	MAE Test
	0.9999	0.6775	0.0681	49.3923	0.0014	34.0067

- L'évaluation du modèle de régression SLR donne les résultats suivants :

Tableau 13 : Evaluation du modèle SLR

	R² Train	R² Test	RMSE Train	RMSE Test	MAE Train	MAE Test
SLR	0.8107	0.8030	39.1275	38.6010	27.5218	26.9689

- L'évaluation du modèle de régression Random Forest donne les résultats suivants :

Tableau 14 : Evaluation du modèle Random Forest

	R² Train	R² Test	RMSE Train	RMSE Test	MAE Train	MAE Test
Random Forest	0.9743	0.8101	14.3995	37.8977	9.8629	26.1404

- L'évaluation du modèle de régression Elastic_Net donne les résultats suivants :

Tableau 15 : Evaluation du modèle Elastic_Net

	R² Train	R² Test	RMSE Train	RMSE Test	MAE Train	MAE Test
Elastic_Net	0.8107	0.8030	39.1285	38.6002	27.5430	26.9862

- L'évaluation du modèle de régression SVR (noyau=linéaire) donne les résultats suivants :

Tableau 16 : Evaluation du modèle SVR (noyau=linéaire)

	R² Train	R² Test	RMSE Train	RMSE Test	MAE Train	MAE Test
SVR (noyau=Lineaire)	-0.0796	-0.0687	93.4795	89.9190	69.9903	67.6284

- L'évaluation du modèle de régression SVR (noyau=rbf) donne les résultats suivants :

Tableau 17 : Evaluation du modèle SVR (noyau=rbf)

	R² Train	R² Test	RMSE Train	RMSE Test	MAE Train	MAE Test
SVR (noyau=rbf)	-0.0796	-0.0687	93.4595	89.9190	69.9903	67.6284

La comparaison de différents modèles avec les deux métriques score (R^2) et erreur quadratique (RMSE) nous permet de qualifier à quel point un modèle de régression s'adapte à un ensemble de données.

le R^2 Train nous permet de savoir si le modèle est bien entraîné afin de déterminer, par la suite, si le modèle est performant grâce au R^2 Test, ce qui conduit à avoir une bonne prédiction. Il doit être supérieur à 0,6 et proche de 1 pour considérer que le modèle est bon. Dans le cas contraire, le modèle est considéré pas bon.

R^2 dépend de l'erreur et de la variance, pour que R^2 soit bon, il faut la variance soit supérieur à l'erreur sachant que de $0 < R^2 < 1$.

Partant de ces principes, on compare les valeurs des R^2 (les scores) des différents modèles de régression, on a observé que la valeur du R^2 Train du modèle Random Forest est égal à 0,9743 ce qui est très bon et la valeur du R^2 Test est égal à 0,8101 ce qui est également bon.

On observe aussi que les RMSE Train/Test de ce dernier sont très faibles par rapport aux autres, c'est-à-dire, le modèle contient un faible taux d'erreur. De ce fait, le modèle Random Forest est considéré comme meilleur modèle prédictif du Maching Learning comparé aux autres.

En revanche, on remarque que la valeur du R^2 Train du modèle de régression Decision Tree est égale à 0.999, une valeur très élevée contrairement à la valeur du R^2 Test qui est très faible. Aussi, la valeur du RMSE Train est très inférieure à la valeur du RMSE Test, ce qui provoque le phénomène de sur-ajustement (voir annexe).

Dans ce cas, le modèle Decision Tree ne nous mènera pas à une bonne prédiction.

Le RMSE Test et le RMSE Train du modèle SVR (noyau =rbf) sont très élevées ce qui signifie que le modèle contient beaucoup d'erreurs. De plus, sa valeur du R^2 est inférieure à 0,6.

On déduit que ce modèle est fortement biaisé, ses valeurs de prédiction moyennes sont loin des valeurs réelles. C'est ce qu'on appelle le phénomène de sous-ajustement (voir annexe).

Nous devons éviter les phénomènes de sur-ajustement et le sous-ajustement.

On conclut que le meilleur modèle de régression qui a obtenu un très bon score




Et un RMSE très faible par rapport aux autres modèles est Random Forest.

On va visualiser les graphes des valeurs réelles et prédites Test/Train des PI des 3 modèles de régression : Random Forest, Decision Tree et SVR (noyau=rbf).

Il est à noter que le calcul graphique des métriques se fait comme suit :

Le calcul de l'erreur quadratique RMSE, c'est la somme des carrés de la soustraction de chaque point de la courbe des valeurs prédites Test avec chaque point de la courbe des valeurs réelles sous la racine.

Le calcul du MAE, c'est la somme des carrés de la valeur absolue de chaque soustraction.

-  Le tracé bleu représente la distribution des valeurs prédites Test des PI (ypred test).
-  Le tracé jaune représente la distribution des valeurs réelles des PI (yreal).
-  Le tracé rouge représente la distribution des valeurs prédites Train des PI (ypred train).

- **Random Forest**

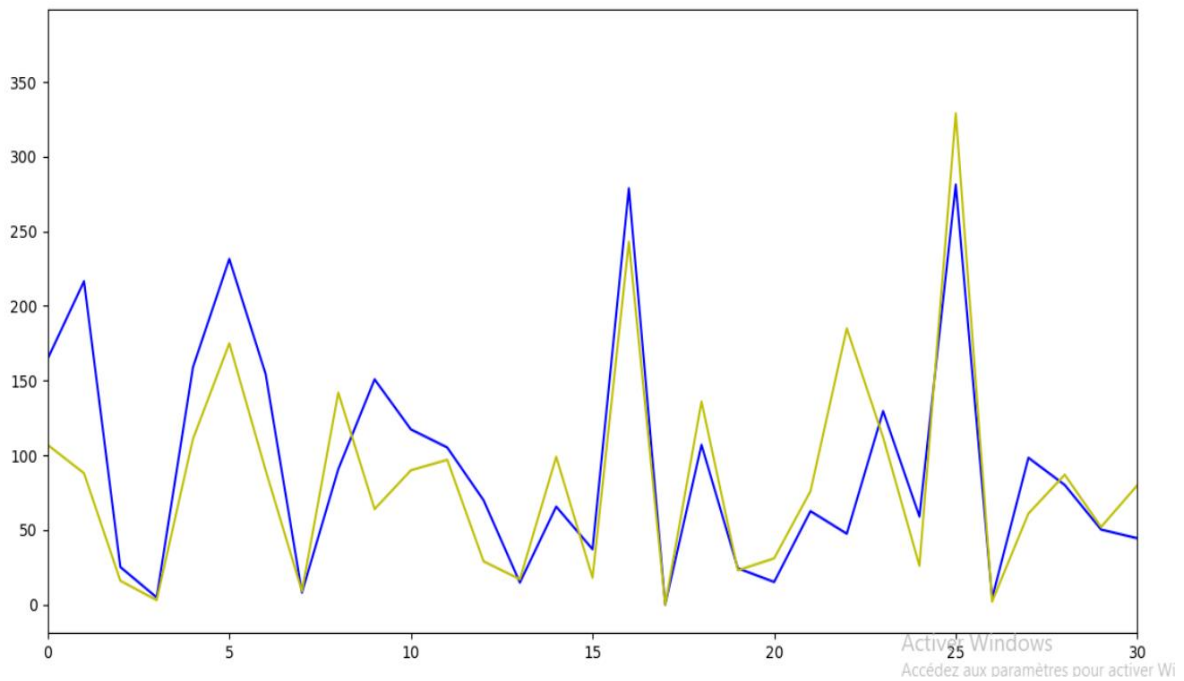


Figure 45 : Graphe représentatif des valeurs réelles et prédites Test des PI

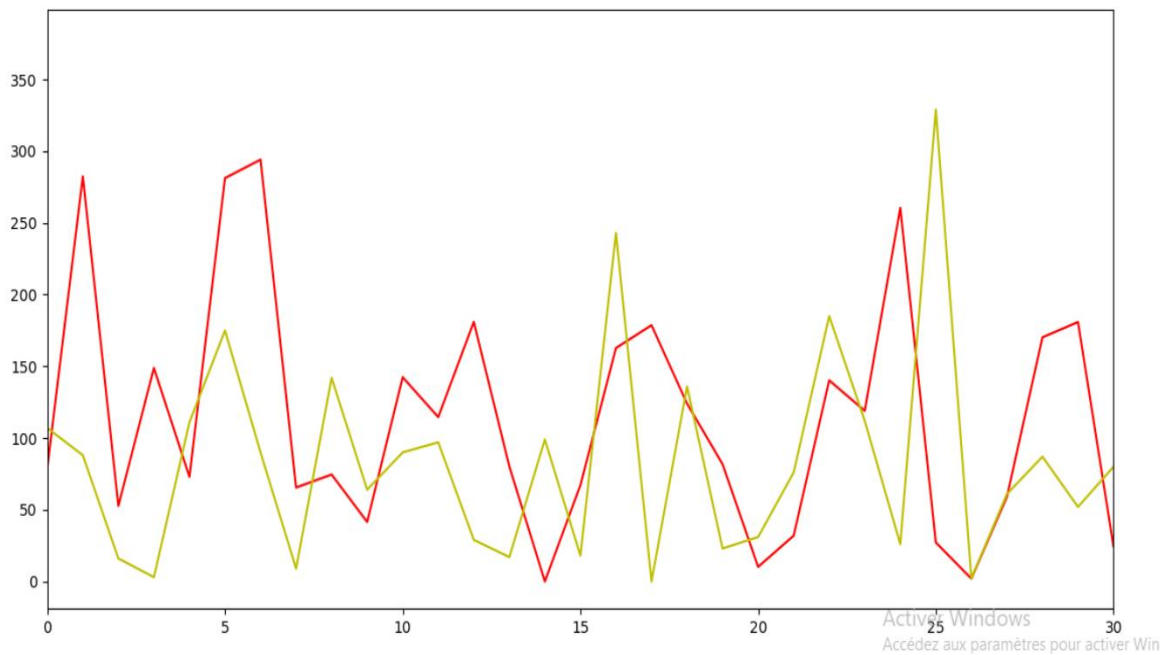


Figure 46 : Graphe représentatif des valeurs réelles et prédites Train des PI

On remarque bien que les deux courbes des valeurs réelles et prédites des deux figures sont quasi superposées, ça veut dire que le RMSE Test/Train et le MAE Test/Train sont très faibles. Ainsi que le $R^2=0,9679$. Donc, ce résultat est très bon.

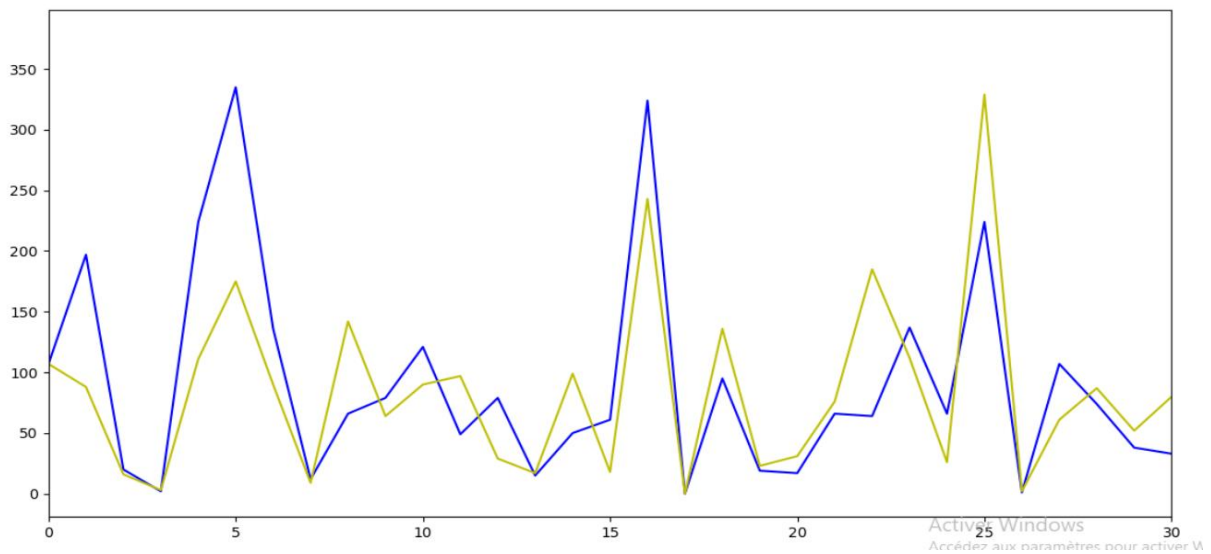
➤ **Decision Tree**

Figure 47 : Graphe représentatif des valeurs réelles et prédites Test des PI

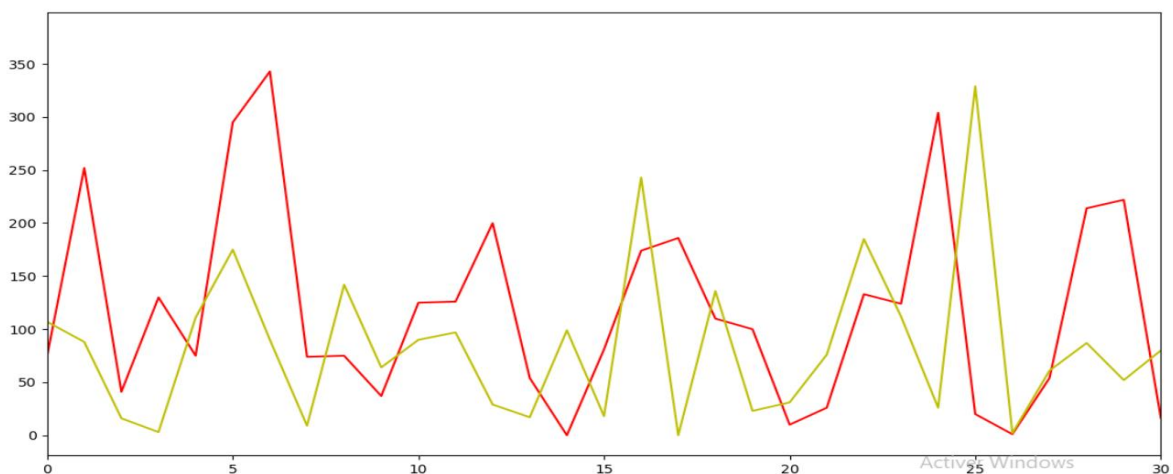


Figure 48 : Graphe représentatif des valeurs réels et prédites Train des PI

D'après les calculs graphiques du RMSE Test qui est très élevée (RMSE Test=49.3923) et le RMSE Train qui est plutôt très faible (RMSE Train=0.0014), avec la valeur du R^2 Train qui est trop parfaite et qui s'approche du 1 tandis que le R^2 train est égale à 0.6775. on est dans la catégorie du Overfitting ou le phénomène sur-ajustement.

- **SVR (noyau=rbf)**

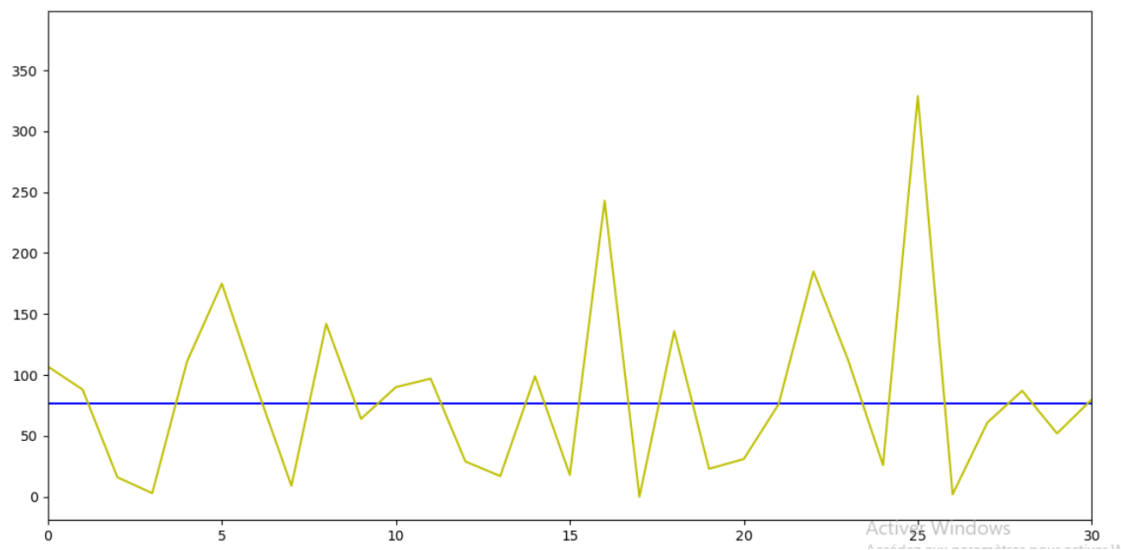


Figure 49 : Graphe représentatif des valeurs réelles et prédites Test des PI

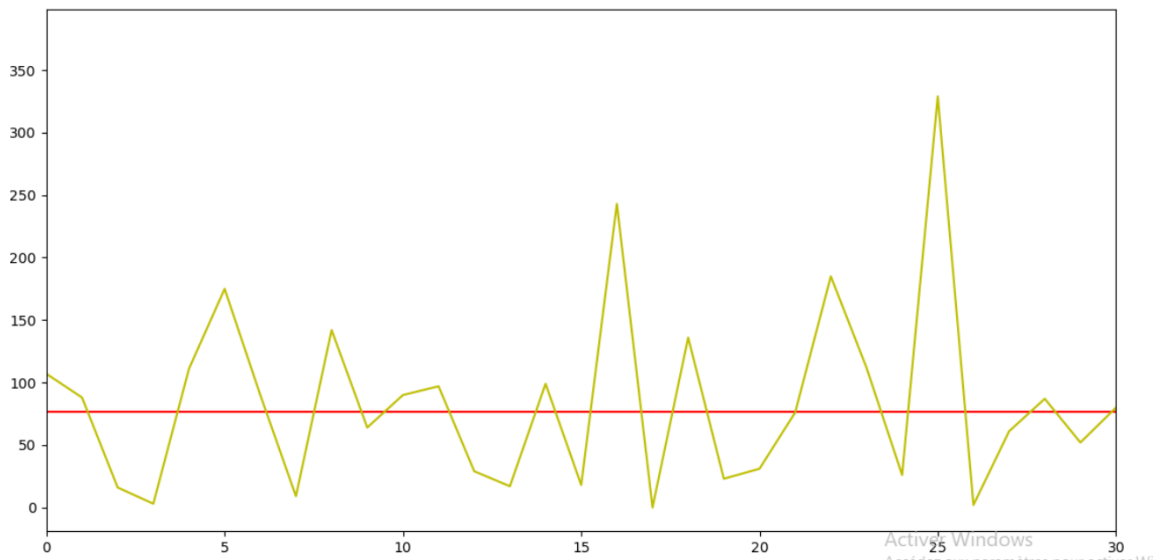


Figure 50 : Graphe représentatif des valeurs réelles et prédites Train des PI

Les deux tracés des valeurs prédites Test/Train des deux figures sont sous forme de droite, et à l'opposé du tracé des valeurs réelles. D'où la valeur du RMSE Test/Train et du MAE qui est très grande. Le modèle n'est pas du tout performant et il ne nous conduira pas à une bonne prédiction.

4.5.2 Comparaison des modèles

Pour confirmer les résultats précédents, nous allons utiliser une méthode nommée Cross Validation (voir annexe) qui nous permettra de tester les performances d'un modèle prédictif de Machine Learning.

Pour choisir le meilleur modèle, les moyennes des métriques pour chaque modèle sont illustrées dans le tableau ci-dessous :

Tableau 18 : Comparaison des performances des modèles

	Lasso	Ridge	Elastic_ Net	Random Forest	SVR(noyau= Linéaire)	SVR(noyau =rbf)
R² Moyenne (Test)	0.8101	0.8101	0.8100	0.8182	-278857518.6315	-0.08382
RMSE Moyenne (Test)	39.1442	39.1441	39.1451	38.3192	1498942.1143	93.5966
MAE Moyenne (Test)	27.5255	27.5391	27.5530	26.5188	1444654.4255	70.0114
R² Moyenne (Train)	0.81076	0.81076	0.81075	0.97448	-277876194.0483	-0.08261
RMSE Moyenne (Train)	39.1267	39.1266	39.1277	14.3858	1499415.1717	93.5866
MAE Moyenne (Train)	27.5251	27.5215	27.5425	9.8397	1443600.8998	60.9893

On remarque à travers le tableau ci-dessus, que le modèle Random Forest est le plus performant, avec des grandes valeurs de R² moyenne (Train) qui est égal à 0.97448, et de R² moyenne (Test) qui est égal à 0.8182, ainsi qu'un RMSE moyenne (Train/Test) très faible par rapport aux autres modèles. Ce qui nous confirme les résultats obtenus précédemment.

Cross Validation a rendu les résultats plus logiques car R^2 moyenne (Train) $>$ R^2 moyenne (Test), autrement dit, un modèle mieux entraîné donnera par la suite de meilleurs résultats dans la prédiction.

4.6 Conclusion

Nous avons présenté dans ce chapitre notre contribution à la recherche d'un modèle qui permettra la prédiction journalière du Drop d'une cellule avant qu'elles ne se produisent grâce aux algorithmes de machine Learning.

De ce fait, on a fait un nettoyage et filtrage de données pour la préparation de ce dernier. Par la suite, on a suivi la carte Scikit-Learn pour faire le choix du bon modèle. Nous avons entraîné et évalué nos modèles pour déterminer le meilleur modèle en les comparant aux différentes métriques.

Durant ce test, nous avons conclu que Random Forest est le modèle le plus performant face au problème de régression pour la base des données considérée.

Conclusion générale

L'un des soucis majeurs des opérateurs est de garantir la qualité de service à leurs clients, tout en minimisant les coûts. D'où la nécessité de l'optimisation du réseau, qui vise à améliorer les performances avec les ressources existantes.

Le premier objectif de ce projet de fin d'études était de maîtriser les techniques d'optimisation du réseau LTE. Pour ce faire, on était amené en premier lieu à étudier les principes fondamentaux de cette technologie afin de pouvoir assimiler la méthodologie de l'optimisation adoptée au sein d'Ericsson, et ce afin de pouvoir intégrer l'une des équipes et de participer au processus de l'optimisation.

Le deuxième objectif consiste à découvrir l'intelligence artificielle par l'utilisation de machine Learning pour la prédiction. Nous avons pris l'initiative d'appliquer le ML dans les KPIs de maintien de service et nous avons réussi à obtenir un bon estimateur prédictif qui sera d'une grande aide à l'équipe des ingénieurs de l'optimisation à anticiper et détecter des dégradations des KPI et d'agir de façon plus rapide.

Ce stage au sein de l'équipementier Ericsson était une expérience très enrichissante, autant sur le plan professionnel que sur le plan personnel. Il nous a permis d'approfondir nos connaissances théoriques et d'acquérir de nouvelles connaissances techniques relatives à l'ingénierie de l'optimisation et la science des données. De plus, On a eu l'occasion de découvrir l'ambiance de travail dans une multinationale, et de travailler avec des ingénieurs de différentes cultures et d'une grande expertise, ce qui a été d'un grand apport pour nous.

Pour conclure, nous estimons avoir satisfait les objectifs initialement fixés, et avoir acquis le savoir nécessaire pour une éventuelle carrière professionnelle dans ce domaine.

Bibliographie

- [1] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16, 321- 357, 2002.
- [2] A. F. M. Shahan Shah, "A Survey From 1G to 5G Including the Advent of 6G: Architectures, Multiple Access Techniques, and Emerging Technologies," *2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC)*, pp. 1117-1123, 2022.
- [3] Yannick bouguen, Eric hardouin et francois-xavier wolff . « LTE pour les réseaux 4G ».editions Eyrolles 2012.
- [4] Boucherit Yassamine et Madani Samir, « LTE RAN Design and Optimisation».Mémoire de Master, 2015, Institut National des Télécommunications et des Technologies de l'Information et de la Communication.
- [5] BOUCHENTOUF Hadjer et BOUDGHENE STAMBOULI Riyad, « ETUDE DES PERFORMANCES DES RESEAUX 4G (LTE) ».Mémoire de Master, 2013, Université Abu Bekr Belkaid de Tlemcen.
- [6] Melle BOUCHENTOUF Hadjer Mr BOUDGHENE STAMBOULI Riyad, « ETUDE DES PERFORMANCES DES RESEAUX 4G ».Mémoire de Master, Université Abou Bekr Belkaid Tlemcen, Algérie, 2012-2013
- [7] ERICSSON Proprietary and confidential, « LTE Overlay Networks », ERICSSON Design Guideline, Decenmbre 2011.
- [8] <https://developer.orange.com/od-uploads/Generalites-et-architecture-de-la-4G.pdf>
- [9] Jeffrey Cichonski, Joshua M. Franklin, Michael Bartock « Guide to LTE Security. » December,2017.
- [10] Amiar Abderaouf et Saidani Maroua, « Optimisation d'un réseau 4G LTE au sein de HUAWEI», 2016, institut national de la poste et des technologies de l'information et de la communication.
- [11] FROM GSM TO LTE, AN INTRODUCTION TO MOBILE NETWORKS AND MOBILE BROADBAND- MartinSauter- WirelessMoves, Germany.
- [12] 3GPP «Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN) », TS 36.300 V11.14.0 R11. 2015.
- [13] H.Holma et A.Toskala, «LTE for UMTS OFDMA and SC-FDMA Based Radio Access»,2009, USA: John Wiley & Sons Ltd.
- [14] S. Ahmadi, LTE-Advanced Practical Systems Approach to Understanding 3GPP LTE Releases 10 and 11 Radio Access Technologies, Amsterdam, Netherlands: Elsevier Inc, 2014.
- [15] KHELIFI , Salim et HADDAD , Aissa. 2018.Planification et dimensionnement d'un réseau mobile 4G/LTE. s.l. : Mémoire de Master universite Akli Mohand Oulhadj-Bouira, 2018.
- [16] <https://moniem-tech.com/2018/06/01/modulation-signal-quality-lte/>
- [17] Xavier Lagrange, « Réseaux mobiles 2G et 3G : Caractéristiques d'une interface radio ». Département RSM, ENST Bretagne, 12/2004.
- [18] ERICSSON Proprietary and Confidential, « LTE Overlay Networks (FDD) », ERICSSON Design Guideline, December 2011.

- [19] ZENATI Ilyas, BOUFERSAKHA Oussama « Dimensionnement et planification d'un réseau LTE ». Mém. de mast, Université Aboubakr Belkaid, Tlemcen, juin 2017
- [20] <https://fr.wikipedia.org/wiki/5G>.
- [21] Hajri, S. E. (2018). L'amélioration des performances des systèmes sans fil 5G par groupements adaptatifs des utilisateurs (Doctoral dissertation).
- [22] [5G SA vs 5G NSA deployment: What are the differences \(alepo.com\)](https://www.alepo.com/5G-SA-vs-5G-NSA-deployment-What-are-the-differences/)
- [23] M. H. Abu Saada « Design Of Efficient Millimeter Wave Planar Antennas For 5g Communication Systems » Mémoire de Master, Université Islamique de Gaza 2017.
- [24] <https://actualiteinformatique.fr/internet-of-things-iot/intelligence-artificielle-des-choses-aiot#:~:text=L'intelligence%20artificielle%20des%20choses,et%20l'analyse%20des%20donn%C3%A9es.>
- [25] D. Chahrazed et T. Chafia, «Contribution à l'optimisation d'un réseau LTE: Adpatation du Handover». Mém. de mast, Universty M'hamed Bougara, Boumerdes, Algeria, 2017.
- [26] <http://www.commentcamarche.net/contents/wireless/wlintro.php3>
- [27] <https://www.lemagit.fr/definition/QoS-qualite-de-service>
- [28] LAIHO, Jaana, WACKER, Achim, et NOVOSAD, Tomáš (ed.), "Radio network planning and optimization for UMTS". Mém. de mast, John Wiley & Sons, 2006.
- [29] Idrissi Kaitouni Hajar. Planification et optimisation des réseaux 3G/4G pour INWI. Mémoire de master. 2015
- [30] Dellys Billel et Ouamri Mohamed-Amine, « Aspects d'Optimisation de l'Interface Radio UTRAN d'un Réseau UMTS ». Mém. de mast, Université Abderrahmane Mira de Bejaïa, 2013-2014.
- [31] Seghier Oussama et Nabi Mohammed, « 4G Planning and initial tuning using Drive Test ». Mémoire de Master, 2016, Institut National des Télécommunications et des Technologies de l'Information et de communication.
- [32] Ericsson, «Key Performance Indicators,» Ericsson, Stockholm, 2013.
- [33] LAIHO, Jaana, WACKER, Achim, et NOVOSAD, Tomáš (ed.), "Radio network planning and optimization for UMTS", John Wiley & Sons, 2006.
- [34] HOCINE, Merah. Conception d'un MODEM de la quatrième génération (4G) des réseaux de mobiles à base de la technologie MC-CDMA. Mémoire de Magister En électronique, 2012.
- [35] Djami Schehrazed. Contribution à l'optimisation d'un réseau LTE : Adaptation du HANDOVER. Mémoire de master .2017.
- [36] Ericsson, «Key Performance Indicators,» Ericsson, Stockholm, 2013.
- [37] https://www.sciencedirect.com/science/article/abs/pii/S187705172200226X?fbclid=IwAR3a6dt_rkDG4-RJ3c15h2EA7Tf29Tq5EwNbbY3QCbPibDHwmFzIJTouD2g
- [38] <https://jafwin.com/2020/03/09/lessentiel-a-savoir-pour-debuter-en-machine-learning/>
- [39] Rachid MIFDAL, « Application des techniques d'apprentissage automatique pour la prédiction de la tendance des titres financiers ». Mémoire, Montréal, 12 Novembre 2019, École de Technologie Supérieure université du Québec.

- [40] G. James, D. Witten, T. Hastie, and R. Tibshirani, An introduction to statistical learning. Springer, vol. 112, 2013.
- [41] <https://zestedesavoir.com/billets/4104/regularisation-ridge-lasso-et-elastic-net-1/#2-regularisation>
- [42] Prashanth Ashok, «What is Ridge Regression», <https://www.mygreatlearning.com/blog/what-is-ridge-regression/>, 25/05/21.
- [43] Syphax Oould Rabah, «Modélisation de la prime pure de la garantie Vol/Incendie d'un contrat d'assurance automobile», thèse de master, sous la direction de Elodie Delle-Case, université paris dauphine, 133p
- [44] « Elastic Net », 1 septembre 2021
<https://corporatefinanceinstitute.com/resources/knowledge/other/elastic-net/>
- [45] https://complements.lavoisier.net/9782409020735_intelligence-artificielle-vulgarisee-le-machine-learning-et-le-deep-learning-par-la-pratique_Chapitre4.pdf
- [46] Selvi tamoul, « Why -Polynomial Regression and not Linear Regression ? », 9 février 2021, <https://www.numpyninja.com/post/why-polynomial-regression-and-not-linear-regression>
- [47] M. Awad et R. Khanna, «Support Vector Regression,» chez Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers, New York, New York, United States, Apress, pp. 67-80, 2015.
- [48] T. Sharp, «An Introduction to Support Vector Regression (SVR),» towardsdatascience, 3 03 2020. [En ligne]. Available:<https://towardsdatascience.com/an-introduction-to-support-vector-regression-svr-a3ebc1672c2>. [Accès le 05 08 2021].
- [49] Nagesh Singh Chauhan, « Algorithme d'arbre de décision »
- [50] <https://eldorhaan.wordpress.com/2019/04/06/reseau-de-neurones-artificiels-quest-ce-que-cest-et-a-quoi-ca-sert/>
- [51] Wendbenedo Arnaud, « Application des algorithmes d'apprentissage automatique pour la détection de défauts de roulements sur les machines tournantes dans le cadre de l'industrie 4.0 », thèse du grade de maître ès sciences appliquée en ingénierie, profil recherche, université du Québec à Chicoutimi, 2020, 69p.
- [52] Hacene BELLAHMER. «Implémentation et évaluation d'un modèle d'apprentissage automatique pour l'estimation de la valeur marchande de propriétés immobilières ». Mém. de mast. TIZI-OUZOU : UNIVERSITÉ MOULOUD MAMMERI, 2019/2020.
- [53] Le schemas Bühlmann, P. Bagging, boosting and ensemble methods. Dans Handbook of Computational Statistics. Springer. pp. 985-1022, 2012.
- [54] <https://www.journaldunet.fr/web-tech/guide-de-l-intelligence-artificielle/1501883-methode-ensembliste-ensemble-learning/?fbclid=IwAR3GP81PR35D5YYYuceI8SmcdpAHbdjbvWnWUg3niXN7WBuoLIIE2kUtmD>
- [55] https://en.wikipedia.org/wiki/Wisdom_of_the_crowd?fbclid=IwAR0s_OzIVeW6H2jhudDGDjT85c63aq-osiliKvSzT-W3BTfm0nalcBKCu7g
- [56] Zouari, H. K. (2004). Contribution à l'évaluation des méthodes de combinaison parallèle de classifieurs par simulation (Doctoral dissertation, Université de Rennes I).
- [57] Héli Khoufi Zouari, Contribution à l'évaluation des méthodes de combinaison parallèle de classifieurs [27-28-29-30-31-32-33-34]

- [58] https://fr.wikipedia.org/wiki/Fichier:Ensemble_Bagging.svg
- [59] Salzberg, S. L. (1994). C4. 5: Programs for machine learning by j. ross quinlan. morgan kaufmann publishers, inc., 1993. Machine Learning, 16(3), 235-240.
- [60] <https://python.developpez.com/tutoriels/apprendre-programmation> python/les bases/?page=le-langage-python
- [61] <https://enacit.epfl.ch/cours/python/introduction/outils-python.html#spyder>
- [62] Goudjil Ayyoub Hasni Bachir ,« Détection d’opinion à partir de Twitter ».Mémoire de fin d’étude, Ecole supérieure de commerce, 2018.
- [63] <https://docs.python.org/fr/3/tutorial>
- [64] <https://scikit-learn.org/stable/>

Annexes

A1. ERAB (Evolved Radio Acces Bearer):

Le concept de bearer est un concept virtuel et définit un ensemble de configurations de réseau qui fournissent la priorité à certains ensembles de trafic. Ainsi, différents supports prennent en charge différentes exigences de QoS. La combinaison d'un bearer radio et d'un bearer S1 est parfois appelé support d'accès radio évolué (E-RAB). Comme l'illustre la figure suivante :



Architecture de support de LTE

A2. Tableau Explicatif des compteurs de KPI

Famille De KPI	KPI	Compteurs	Symbole	Signification
Accessibilité	RRC CSSR	pmRrcConnEstabSucc	A1	Nombre total d'établissements de connexion RRC réussis
		pmRrcConnEstabAtt	A2	Nombre total des tentatives de demande de connexion RRC
		pmRrcConnEstabAttReatt	A3	Nombre total des tentatives de demande de connexion RRC qui sont considérées comme de nouvelles tentatives
	E RAB SR	pmErabEstabSuccInit	B1	Le nombre total d'établissements E-RAB initiaux réussis présents dans le message S1
		pmErabEstabSuccAdded	B2	Le nombre total des E-RAB ajoutés avec succès
		pmErabEstabAttInit	B3	Nombre total de tentatives d'établissement d'E-RAB initiales
		pmErabEstabAttAdded	B4	Nombre total de tentatives d'établissement d'E-RAB ajoutées
	S1 CSSR	pmS1SigConnEstabSucc	C1	Nombre total d'établissements de connexion de signalisation S1 réussis
		pmS1SigConnEstabAtt	C2	Nombre de tentatives d'établissement de connexion de signalisation S1

Tableau Explicatif des compteurs de KPI

Mobilité	HO ESR	pmHoExeSuccLteIntraF	D1	Nombre de transferts intra-fréquence intra LTE réussis
		pmHoExeSuccLteInterF	D2	Nombre de transferts inter-fréquences intra-LTE sortants réussis
		pmHoExeAttLteIntraF	D3	Nombre des tentatives d'exécution de transfert de intrafréquence intra LTE
		pmHoExeAttLteInterF	D4	Nombre de tentatives d'exécution de transfert inter-fréquences intra LTE sortantes
Intégrité	DL Cell TP	pmPdcPVolDlDrb	E1	Volume total sur les supports radio de données qui a été transféré (reconnu par l'UE) dans le sens descendant. Unité : kilobit (1000 bits)
		pmSchedActivityCellDl	E2	Nombre agrégé de milli secondes dans lequel les données devaient être planifiées dans la liaison descendante
	UL Cell TP	pmPdcPVolUlDrb	E3	Volume total sur les supports radio de données qui a été reçu dans le sens montant. Unité : kilobit (1 000 bits)
		pmSchedActivityCellUl	E4	Nombre agrégé de millisecondes dans lequel les données devaient être planifiées dans la liaison montante
Maintenabilité	E-RAB Drop	pmErabRelAbnormalEnbAct	F1	Nombre total de versions anormales d'E-RAB initiées par le RBS
		pmErabRelAbnormalMmeAct	F2	Nombre total de versions d'E-RAB initiées par le MME
		pmErabRelAbnormalEnb	F3	Nombre total de versions anormales d'E-RAB initiées par le RBS
		pmErabRelNormalEnb	F4	Nombre total de versions E-RAB normales déclenchées par le RBS
		pmErabRelMme	F5	Nombre total de versions d'E-RAB lancées par le MME

A3. Les métriques d'évaluation :

L'étape essentielle de tout modèle d'apprentissage automatique consiste à évaluer la précision du modèle. L'Erreur quadratique moyenne, l'Erreur absolue moyenne, l'Erreur quadratique moyenne racine et les métriques R-carré ou Coefficient de détermination sont utilisées pour évaluer les performances du modèle dans l'analyse de régression.

MSE

L'erreur quadratique moyenne représente la moyenne de la différence au carré entre les valeurs d'origine et prédites dans l'ensemble de données. Il mesure la variance des résidus.

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y})^2$$

L'erreur quadratique moyenne (MSE) indique à quel point une ligne de régression est proche d'un ensemble de points. Pour ce faire, on prend les distances entre les points et la droite de régression (ces distances sont les "erreurs") et on met au carré. La quadrature est nécessaire pour supprimer tout signe négatif. Cela donne également plus de poids aux différences plus importantes. C'est ce qu'on appelle l'erreur quadratique moyenne car elle donne la moyenne d'un ensemble d'erreurs. Plus l'EQM est faible, meilleures sont les prévisions.

MAE

L'erreur absolue moyenne représente la moyenne de la différence absolue entre les valeurs réelles et prédites dans l'ensemble de données. Elle mesure la moyenne des résidus dans l'ensemble de données.

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - y|^2$$

RMSE

L'erreur quadratique moyenne est la racine carrée de l'erreur quadratique moyenne. Elle mesure l'écart type des résidus.

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - y)^2}$$

R2

Le coefficient de détermination ou R2 représente la proportion de la variance de la variable dépendante qui est expliquée par le modèle de régression linéaire. Il s'agit d'un score sans échelle, c'est-à-dire que : que les valeurs soient petites ou grandes, la valeur de R au carré sera inférieure à un.

$$R^2 = 1 - \frac{\sum (y_i - y)^2}{\sum (y_i - y)^2}$$

La RMSE et le R2 quantifient à quel point un modèle de régression linéaire s'adapte à un ensemble de données. La RMSE indique dans quelle mesure un modèle de régression peut prédire la valeur d'une variable de réponse en termes absolus, tandis que le R2 indique dans quelle mesure les variables prédictives peuvent expliquer la variation de la variable de réponse.

A4. Les différences entre ces mesures d'évaluation :

L'erreur quadratique moyenne (MSE) et l'erreur quadratique moyenne pénalisent les grandes erreurs de prédiction par rapport à l'erreur absolue moyenne (MAE). Cependant, la RMSE est largement utilisé que la MSE pour évaluer les performances du modèle de régression avec d'autres modèles aléatoires car elle a les mêmes unités que la variable dépendante (axe Y).

La MSE est une fonction différentiable qui facilite l'exécution d'opérations mathématiques par rapport à une fonction non différentiable telle que la MAE. Par conséquent, dans de nombreux modèles, la RMSE est utilisé comme métrique par défaut pour calculer la fonction de perte bien qu'il soit plus difficile à interpréter que la MAE.

Des valeurs basses des MAE, MSE et RMSE impliquent une plus grande précision d'un modèle de régression. Cependant, une valeur plus élevée de R carré est considérée comme souhaitable.

R carré et R carré ajusté sont utilisés pour expliquer dans quelle mesure les variables indépendantes du modèle de régression linéaire expliquent la variabilité de la variable dépendante. La valeur R au carré augmente toujours avec l'ajout des variables indépendantes, ce qui pourrait conduire à l'ajout des variables redondantes dans notre modèle. Cependant, le R-carré ajusté résout ce problème.

Le R carré ajusté prend en compte le nombre de variables prédictives et est utilisé pour déterminer le nombre de variables indépendantes dans notre modèle. La valeur du R carré ajusté diminue si l'augmentation du R carré par la variable supplémentaire n'est pas suffisamment significative.

Pour comparer la précision entre différents modèles de régression linéaire, RMSE est un meilleur choix que R au carré.

A5. Variance et biais :

La variance

La variance est la variabilité de la prédiction du modèle pour un certain ensemble de données ou une valeur qui nous indique la propagation de nos données. Une prédiction à variance élevée signifie que les valeurs de prédiction sont très variées. Un modèle à forte variance accorde beaucoup d'attention aux données d'entraînement et ne généralise pas sur les données qu'il n'a jamais vues auparavant. En conséquence, ces modèles fonctionnent très bien sur les données d'entraînement, mais présentent des taux d'erreur élevés sur les données de test.

Mathématiquement, l'équation de l'erreur de variance dans le modèle prend la forme est :

$$\text{Variance}[f(X)] = E[X^2] - E^2[X]$$

Le biais

Le biais est la différence entre la prédiction moyenne de notre modèle et la valeur correcte que nous essayons de prédire. Un modèle fortement biaisé signifie que ses valeurs de prédiction (moyennes) sont loin des valeurs réelles, un tel modèle accorde très peu d'attention aux données d'entraînement et simplifie excessivement le modèle. Cela conduit toujours à une erreur élevée sur les données d'entraînement et de test.

Mathématiquement, supposons que les variables d'entrée soient X et une variable cible Y. On note la relation entre les deux à l'aide d'une fonction f.

$$Y = f(X) + e$$

Ici, "e" est l'erreur qui est normalement distribuée. Le but de notre modèle $f'(x)$ est de prédire des valeurs aussi proches que possible de $f(x)$. Ici, le biais du modèle est :

$$Bias[f'(X)] = E[(f'(X) - f(X))]$$

A6. Compromis variance-biais :

Si notre modèle est trop simple et comporte très peu de paramètres, il peut avoir un biais élevé et une faible variance. D'autre part, si notre modèle a un grand nombre de paramètres, il aura une variance élevée et un faible biais. Nous devons donc trouver le bon équilibre sans sur-ajuster et sous-ajuster les données.

Ce compromis de complexité est la raison pour laquelle il existe un compromis entre biais et variance. Un algorithme ne peut pas être plus complexe et moins complexe à la fois.

Les résultats de prédiction d'un modèle d'apprentissage automatique se situent quelque part entre :

- faible biais, faible variance.
- faible biais, forte variance.
- fort biais, faible variance.
- fort biais, forte variance.

Un modèle à faible biais et à forte variance est appelé surajustement et un modèle à fort biais et à faible variance est appelé sous-ajustement. Par généralisation, on trouve le meilleur compromis entre sousajustement et surajustement pour qu'un modèle entraîné obtienne les meilleures performances.

Donc, Un modèle sur-ajusté obtient un score de prédiction élevé sur les données vues et un faible score sur les de données jamais vues auparavant. Tandis qu'un modèle sous-ajusté à de faibles performances dans les ensembles de données vues et jamais vues.

En termes de complexité du modèle, nous pouvons utiliser le diagramme dans la figure (II.1) pour décider la complexité optimale de notre modèle.

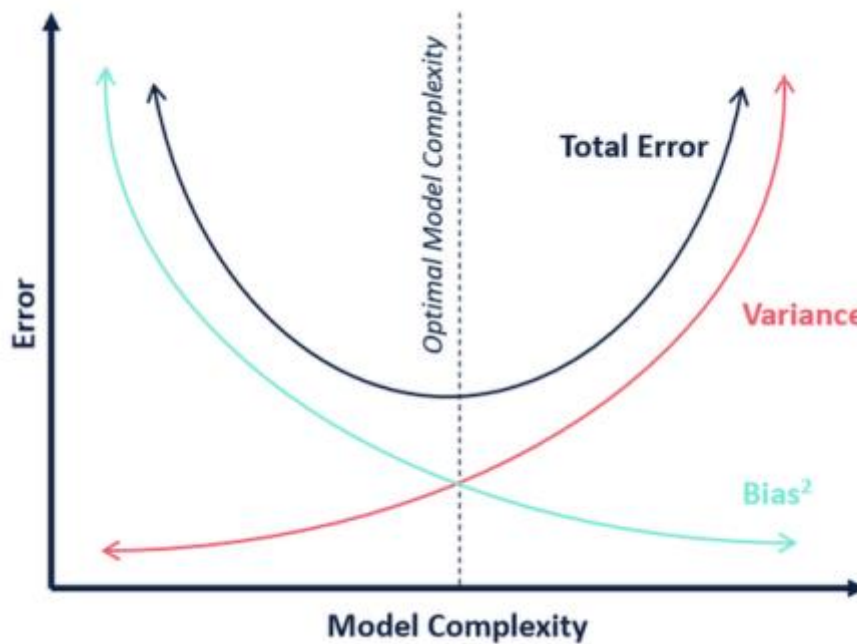


Figure : Erreur/complexité optimale.

A6. Sous-ajustement et surajustement :

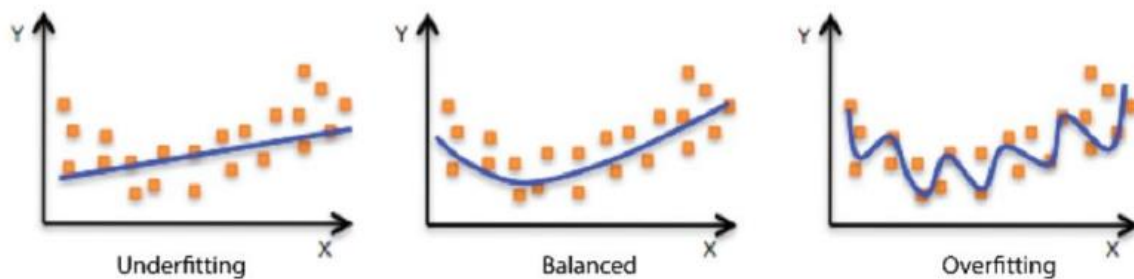


Figure : Les trois cas d'ajustement du modèle.

La figure illustre les trois cas. À gauche, la ligne bleue représente un modèle sous-ajusté. Le modèle arrive à déterminer qu'il existe une certaine tendance dans les données, mais qu'elle n'est pas suffisamment précise pour saisir les informations pertinentes. Il est incapable de faire des prédictions précises pour la formation ou de nouvelles données.

Au milieu, la ligne bleue représente un modèle équilibré. Ce modèle arrive à déterminer qu'il existe une tendance dans les données et la modélise avec précision. Ce modèle pourra se généraliser avec succès.

À droite, la ligne bleue représente un modèle en surajustement. Le modèle arrive à déterminer une tendance dans les données et modélise avec précision les données d'entraînement, mais il est trop spécifique. Il ne parviendra pas à faire des prédictions précises avec de nouvelles données car il a trop bien appris les données de formation.

A7. Valeurs manquantes :

Les données manquantes sont définies comme des valeurs ou données de variables qui ne sont pas représentées. Cela se produit lorsque la collecte des données est mal effectuée ou lorsqu'il y a des erreurs lors de la saisie des données. Dans certaines situations, le gouvernement ou les entités privées ne peuvent pas partager certaines informations. Les données manquantes présentent divers problèmes :

- Des biais dans l'estimation des paramètres.
- Complique l'analyse de l'étude ce qui peut provoquer par la suite des conclusions fausses.
- Plusieurs algorithmes peuvent échouer dans la présence de valeurs manquantes.

Il existe trois types de données manquantes :

- Manquantes complètement aléatoires.
- Manquantes aléatoires.
- Manquantes par omission prévisible.

A8. Valeurs aberrantes:

Une valeur aberrante est une observation qui se situe très loin de la moyenne de la variable aléatoire correspondante. Cette distance est mesurée par rapport à un seuil donné, généralement l'écart type. Ces valeurs aberrantes produisent de grandes erreurs lors de la formation. Les valeurs aberrantes vont être rejetées si leur nombre est très faible. Si ce n'est pas le cas, nous devons adopter des fonctions de coût qui ne sont pas très sensibles aux valeurs aberrantes, par exemple le critère de carré minimum est très sensible en présence de valeurs aberrantes, car les grandes erreurs dominent la fonction de coût en raison de quadrature des termes.

Les valeurs aberrantes peuvent être causées par une erreur de mesure, erreur d'exécution ou variabilité inhérente.

Pour supprimer les valeurs aberrantes du Dataset il faut d'abord les détecter.

A9. Validation croisée :

Afin d'évaluer les performances d'un modèle de Machine Learning, il est nécessaire de le tester sur de nouvelles données. En fonction des performances des modèles sur des données inconnues, on peut déterminer s'il est sous-ajusté, sur-ajusté, ou bien généralisé.

La Cross-Validation permet de tester les performances d'un modèle prédictif de Machine Learning. Cette méthode consiste à découper le Dataset en plusieurs échantillons puis à évaluer le modèle d'apprentissage automatique à l'aide des autres échantillons de données pour connaître la précision du modèle en assurant la diversité des données. Il existe trois méthodes pour calculer la validation croisée :

- Validation
- Leave one out cross validation (LOOCV)
- La validation croisée K-fold

Pour notre étude, on a utilisé la validation croisée K-fold :

C'est la méthode la plus utilisée pour l'évaluation d'un modèle d'apprentissage automatique. Cette méthode consiste à diviser les données en k -parties. Chaque fractionnement des données est appelé fold (un pli). Ici, nous formons tous les k sous-ensembles de données au modèle, la valeur de k varie entre 5 et 10, puis nous laissons de côté un $(k-1)$ sous-ensemble pour effectuer une évaluation sur le modèle formé. On note que la variance de l'estimation résultante est réduite pour les valeurs élevées de k .

Cette méthode permet de s'assurer que toutes les observations de l'ensemble de données original aient une chance d'apparaître dans l'ensemble d'entraînement et dans l'ensemble de test. Cette méthode se traduit par une grande précision et produit des données avec moins de biais. L'avantage de cette méthode est que la manière dont les données sont divisées est moins importante. Chaque point de données se trouve exactement une fois dans un ensemble de test et $k-1$ fois dans un ensemble d'entraînement.

L'inconvénient de cette méthode est que l'algorithme d'apprentissage doit être ré exécuté à partir du début k fois, ce qui signifie qu'il faut k fois plus de calculs pour faire une évaluation.

4-fold validation ($k=4$)



Figure : La validation croisée K-fold avec $K=4$