

الجمهورية الجزائرية الديمقراطية الشعبية
République algérienne démocratique et populaire
وزارة التعليم العالي والبحث العلمي
Ministère de l'enseignement supérieur et de la recherche scientifique
جامعة عين تموشنت بلحاج بوشعيب
Université –Ain Temouchent- Belhadj Bouchaib
Faculté des Sciences et de Technologie
Département Électronique et Télécommunication



Projet de Fin d'Etudes
Pour l'obtention du diplôme de Master en : Réseau et
télécommunication
Domaine : Sciences et Technologie
Filière : Télécommunication
Spécialité : Réseau et télécommunication
Thème :

**Développement d'un système de localisation basé sur les
signaux GSM**

Présenté Par :

- 1) Melle ANBERI Roumaissa
- 2) Melle BAKHTI Fouzia

Devant le jury composé de :

Dr. HAMLILI Heyem	MAB	UAT.B. B (Ain Temouchent) Président
Dr. BENOSMAN Mourad	MCB	UAT.B. B (Ain Temouchent) Examinateur
Dr. YAGOUB Reda	MCB	UAT.B. B (Ain Temouchent) Encadrant

Année Universitaire 2023/2024

Remerciements

Nous tenons à remercier en premier lieu Dieu, le tout puissant et miséricordieux, qui nous a donné la force et la patience d'accomplir ce modeste travail.

*Nous avons l'honneur et le plaisir de présenter notre profonde gratitude et nos sincères remerciements à notre encadreur, **Dr. YAGOUB Reda** qui n'a jamais hésité à nous prodiguer ses conseils et ses orientations précieux.*

*Un grand merci aux membres du jury qui ont accepté d'examiner ce travail : **Dr. HAMLILI Heyem** qui nous a fait l'honneur de présider le jury de soutenance. On exprime nos vifs remerciements au **Dr. BENOSMAN Mourad** qui nous a fait l'honneur d'examiner notre travail.*

Enfin, nous tenons également à exprimer ma sincère gratitude à tous nos enseignants du département d'Électronique et télécommunication, qui nous ont fourni connaissances et savoir tout au long de nos années d'études. Nos proches et amis, qui nous ont toujours soutenu et encouragé au cours de la réalisation de ce mémoire.

Dédicace

Je dédie ce modeste travail à

Mon cher père, décédé trop tôt, qui m'a toujours poussé et motivé dans mes études, je ne saurais exprimer mon grand chagrin en ton absence.

J'aurais aimé que tu sois à mes côtés ce jour.

Ma très chère mère, aucune dédicace ne saurait exprimer mon grand respect et ma reconnaissance pour les sacrifices que tu as consentis pour mon éducation. J'implore Dieu le tout puissant de t'accorder bonne santé et longue vie.

*Mon cher frère **Mohammad**, qui était comme un père, soutiens-moi et aide-moi à chaque étape avec son soutien et ses encouragements constants.*

*À mon binôme et à ma chère amie **Maïssa**, avec laquelle j'ai vécu des souvenirs inoubliables durant les cinq ans d'études à l'université. Avec elle, j'ai partagé le stress et la pression des études, la saveur de la réussite et beaucoup de fou rire. Je te souhaite du succès dans ta vie.*

À toute ma famille, du plus grand au plus petit.

*À toutes les personnes chères à mon cœur : **Khadija, Houaria, Amel k, Amel M, Djihan, Fatima, Aïcha, Safaa, Yamna, Souhila, Youssra.***

À tous les étudiants de la promotion Télécommunication (2023/2024).

À toutes les personnes qui m'ont aidé, soutenu et contribué de près ou de loin à la réalisation de ce modeste travail.

Fouzia

Dédicace

Je dédie cet humble travail à

*Qui m'a soutenu, m'a encouragé et m'a appris ce qu'il faut à mon cher
père.*

*À celle qui m'a appris la fermeté et l'espoir, à la personne la plus
grande et la plus compatissante qui existe, à ma mère bien-aimée.*

*J'exprime mon respect et ma gratitude pour leur soutien, leurs
sacrifices et leurs encouragements tout au long de mes études.*

*À mes frères **Amine** et **Boualem**, et à ma sœur unique **Nadia**, je vous
souhaite un avenir radieux et plein de réussite.*

*À mon binôme et à ma chère amie **Fouzia**, avec laquelle j'ai vécu des
souvenirs inoubliables durant les cinq ans d'études à l'université. Avec
elle, j'ai partagé le stress et la pression des études, la saveur de la
réussite et beaucoup de fou rire. Je te souhaite du succès dans ta vie.*

À toute ma famille, du plus grand au plus petit.

*À toutes les personnes chères à mon cœur : **Khadija**, **Amel**, **Djihan**,
Fatima, **Yamna**, **Youssra**, **Assia**, **Hayat**.*

À tous les étudiants de la promotion Télécommunication (2023/2024).

*À toutes les personnes qui m'ont aidé, soutenu et contribué de près ou
de loin à la réalisation de ce modeste travail.*

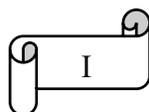
Roumaïssa

Abstract

There are several location methods based on GSM networks, such as Time of Arrival (TOA), Time Difference of Arrival (TDOA), Received Signal Strength Indicator (RSSI), and triangulation.

The objective of this final study project is to develop a localization system based on GSM signals using a machine learning model. This system represents an important and interesting challenge. We first presented how to collect and analyze a large set of data received from GSM signals, then we implemented a machine learning model to improve the prediction accuracy. The model achieved encouraging performance and good accuracy.

Keywords : location, GPS, GSM signals, artificial intelligence, machine learning.



Résumé

Il existe plusieurs méthodes de localisation basées sur les réseaux GSM, telles que : le Time of Arrival (TOA), le Time Difference of Arrival (TDOA), Received Signal Strength Indicator (RSSI), et la triangulation.

L'objectif de ce projet de fin d'étude est de développer un système de localisation basé sur les signaux GSM en utilisant un modèle d'apprentissage automatique. Ce système représente un défi important et intéressant. Nous avons d'abord présenté comment collecter et analyser un large ensemble de données reçues des signaux GSM, nous avons ensuite implémenté un modèle d'apprentissage automatique pour améliorer la précision des prédictions. Le modèle a obtenu des performances encourageantes et une bonne précision.

Mots-clés : localisation, GPS, signaux GSM, intelligence artificielle, apprentissage automatique.

ملخص

هناك العديد من طرق تحديد الموقع المعتمدة على الشبكة جي إس إم مثل: وقت الوصول، وفارق وقت الوصول، ومؤشر قوة الإشارة المستلمة، والتثليث.

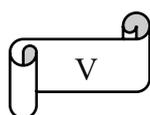
الهدف من مشروع الدراسة النهائية هذا هو تطوير نظام المواقع القائم على الإشارة جي إس إم باستخدام نموذج التعلم الآلي، ويمثل هذا النظام تحديًا مهمًا ومثيرًا للاهتمام. قدمنا أولاً كيفية جمع وتحليل مجموعة كبيرة من البيانات الواردة من الإشارات جي إس إم، ثم قمنا بتنفيذ نموذج التعلم الآلي لتحسين دقة التنبؤات. حقق النموذج أداءً مشجعاً ودقة جيدة.

الكلمات المفتاحية: الموقع، نظام تحديد المواقع العالمي، إشارات جي إس إم، الذكاء الاصطناعي، التعلم الآلي.

Tables des matières

Abstract	I
Résumé	II
ملخص.....	III
Tables des matières.....	IV
Liste des figures	VII
Liste des tableaux	VIII
Liste des abréviations	IX
Introduction Générale.....	1
chapitre 1 : Techniques et technologies de localisation	3
1. Introduction	3
2. Le système de navigation GPS.....	4
2.1 Fonctionnement du système GPS	5
2.2 Les avantages et les inconvénients de GPS	6
3. Système de réseau GSM.....	6
3.1 Historique :.....	8
3.2 Architecture du réseau GSM.....	10
3.2.1 La Mobile Station (MS).....	10
3.2.2 Sous-système de station de base (BSS)	11
3.2.3 Sous-système de commutation de réseau (NSS)	12
3.2.4 Le sous-système opérationnel ou d'exploitation et de maintenance (OSS) ...	13
4. La localisation par réseau GSM	13
4.1 Les techniques de localisation dans le réseau cellulaire GSM.....	14
4.1.1 Identifiant de la cellule (Cell-ID)	14
4.1.2 Temps d'arrivée (TOA)	14
4.1.3 Angle d'Arrivé (AOA)	15
4.1.4 Différence de temps d'arrivée TDOA	16
4.1.5 Indicateur de force du signal reçu (RSSI)	17
4.2 Synthèse des techniques de localisation dans le réseau cellulaire GSM.....	18
5. Les algorithmes de localisation	19
5.1 Les algorithmes pour le scénario LOS.....	19
5.1.1 La triangulation.....	19
5.1.2 La trilatération	20

5.1.3	La multilatération	21
5.2	Les algorithmes de localisation en scénario NLOS	22
5.2.1	Empreinte digitale de force du signal	22
6.	Les avantages et les inconvénients de localisation.....	22
7.	Conclusion.....	22
chapitre 2	: Généralités sur l'intelligence artificielle.....	24
1.	Introduction	25
2.	Intelligence artificielle.....	25
3.	Historique	26
4.	Les types.....	27
4.1	Basé sur les capacités	28
4.1.1	IA étroite (IA faible)	28
4.1.2	IA générale (IA forte)	28
4.1.3	IA super intelligente.....	29
4.2	Basé sur les fonctionnalités.....	29
4.2.1	Machine réactive.....	29
4.2.2	Mémoire limitée	30
4.2.3	Théorie de l'esprit	31
4.2.4	Consciente d'elle-même	31
5.	Les techniques de l'IA.....	31
5.1	Apprentissage automatique (Machine Learning).....	31
5.1.1	Les types	34
5.1.1.1	Apprentissage supervisé.....	34
5.1.1.2	Apprentissage non supervisé.....	36
5.1.1.3	Apprentissage semi-supervisé.....	36
5.1.1.4	Apprentissage par renforcement	37
5.1.2	Les algorithmes d'apprentissage automatique supervisé	37
5.1.3	Le Différence entre apprentissage supervisé et non supervisé	41
6.	Conclusion.....	42
chapitre 3	: Implémentation et Résultats.....	43
1.	Introduction	42
2.	Les outils de développement	42



2.1	Google Colab	42
2.2	Python	42
2.3	Bibliothèques utilisées	43
2.3.1	Pandas	43
2.3.2	Scikit-learn.....	44
3.	La collection de données	44
4.	Traitement des données	48
5.	L'Algorithme Random Forest Regression.....	49
6.	Résultats	51
7.	Conclusion.....	54
	Conclusion générale	55
	Bibliographie.....	56

Liste des figures

Figure 1 : Les segments du système GPS.	5
Figure 2 : Antenne GSM.	8
Figure 3 : Architecture du réseau GSM.	10
Figure 4 : Identifiant de la cellule (Cell-ID).	14
Figure 5 : Localisation par TOA.	15
Figure 6 : Localisation par AOA.	16
Figure 7 : Localisation par TDOA.	17
Figure 8 : Localisation par RSSI.	17
Figure 9 : Le principe de triangulation.	20
Figure 10 : Le principe de trilatération.	21
Figure 11 : Schéma de l'intelligence artificielle.	26
Figure 12 : Le programme IBM Chess, qui a vaincu le champion du monde Garry Kasparov en 1997.	30
Figure 13 : IA à mémoire limitée.	30
Figure 14 : Les types d'apprentissage automatique.	34
Figure 15 : La classification et la régression.	36
Figure 16 : Logo de Google Colab.	42
Figure 17 : Logo de Python.	43
Figure 18 : Capture d'écran de l'application.	45
Figure 19 : Enregistrez et stockez les données tabulaires dans un fichier CSV.	45
Figure 20 : Quelques points des données collectées.	46
Figure 21 : Importation des bibliothèques.	47
Figure 22 : Définir les fonctionnalités et la cible.	48
Figure 23 : Remplacement des valeurs manquantes.	48
Figure 24 : Division des données en ensembles de formation et de test.	48
Figure 25 : Traitement des caractéristiques numériques et catégorielles.	49
Figure 26 : Définir le modèle.	50
Figure 27 : Formation et évaluation du modèle.	51
Figure 28 : Prédiction pour un seul record.	51
Figure 29 : Phase de test des données.	52

Liste des tableaux

Tableau 1 : historique de GSM.	9
Tableau 2 : Synthèse des techniques de localisation dans le réseau cellulaire GSM.	18
Tableau 3 : un résumé des étapes clés de l'évolution de l'IA.	27
Tableau 4 : La différence entre apprentissage supervisé et non supervisé.	41
Tableau 5 : pourcentage d'enregistrements.	52
Tableau 6 : Comparaison de la précision de différentes techniques de positionnement à l'aide des données du réseau GSM.	53

Liste des abréviations

GPS : Global Positioning System.

GNSS : Global Navigation Satellites System.

DoD : Développée par le département américain de la Défense.

GSM : Global System for Mobile Communications.

CEPT : Conférence Européenne des Postes et Télécommunications.

SMS : Short Message Service.

MMS : Multimedia Messaging Service.

ETSI : European Telecommunications Standards Institute.

UMTS : Universal Mobile Telecommunications System.

MS : Mobile Station.

BSS : Base Station Subsystem.

NSS : Network Sub System.

OSS : Operating Support System.

SIM : Subscriber Identity Module.

IMEI : International Mobile Equipment Identity.

TAC : Type Approval Code.

FAC : Final Assembly Code.

MCC : Mobile Country Code.

MNC : Mobile Network Code.

MSIN : Mobile Station Identification Numbers.

BTS : Base Transceiver Station.

BSC : Base Station Controller.

MSC : Mobile Switching Center.

HLR : Home Location Register.

VLR : Visitor location register.

EIR : Equipment Identity Register.

AUC : Authentication Center.

LAC : Location Area Code.

Cell ID : Cellule Identification.

AOA : Angle of Arrival.

TOA : Time of Arrival.

TDOA : Time Difference of Arrival.

RSSI : Received Signal Strength indicator.

LOS : Line of Sight.

NLOS : Non-Line of Sight.

MLAT : Multilateration.

IA : Intelligence Artificielle.

AGI : Intelligence Générale Artificielle.

ASI : Artificial Super intelligence.

IBM : International Business Machines.

ML : Machine Learning.

NB : Naïve Bayes.

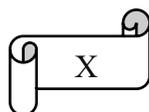
SVM : support vector machine.

KNN : K-Nearest Neighbours.

CSV : Comma-Separated Values.

LTE : Long Term Evolution.

CDMA : Code Division Multiple Access



Introduction Générale

Depuis l'Antiquité, les humains ont eu recours à diverses méthodes pour déterminer leur position géographique, à l'aide d'étoiles, de points de repère, de boussoles et de cartes. Au fil du temps, les techniques de localisation se sont développées grâce à l'émergence de la technologie des communications sans fil, ce qui a conduit à l'émergence de nouvelles techniques, d'horizons plus avancés et efficaces dans ce domaine. Il y a plusieurs techniques pour déterminer la localisation géographique d'un objet ou d'une personne à la surface de la Terre, telles que les systèmes satellitaires en tant que le système de positionnement global (GPS), les systèmes basés sur des réseaux sans fil tels que les réseaux cellulaires en tant que le GSM et l'UMTS, et les réseaux sans fil locaux tels que le Wi-Fi. En bref, la localisation joue un rôle décisif dans de nombreux domaines de la vie quotidienne.

Le système GSM est l'un des systèmes importants utilisés pour déterminer des emplacements, car il contribue à fournir des services fiables et efficaces aux utilisateurs du monde entier. Cela utilise des techniques et des algorithmes avancés pour déterminer l'emplacement des appareils grâce à leur interaction avec le réseau de la tour de téléphonie cellulaire. Parmi ces techniques figurent l'identification de cellules (Cell-ID) et des méthodes basées sur des mesures temporelles telles que l'heure d'arrivée (TOA), la différence de temps d'arrivée (TDOA), l'angle d'arrivée du signal (AOA) et enfin l'indice de force du signal reçu (RSSI).

À l'ère de la technologie moderne, l'intelligence artificielle (IA) représente une révolution dans de nombreux domaines. Telle que les techniques de positionnement, l'intelligence artificielle a offert d'énormes possibilités d'analyse des données volumineuses grâce à l'utilisation de techniques d'apprentissage automatique et de réseaux de neurones artificiels. Malgré les développements des techniques et des algorithmes GSM pour déterminer l'emplacement des appareils, on ne peut pas s'y fier pleinement pour obtenir des résultats précis en raison de facteurs environnementaux ou de réflexions de signaux. Pour résoudre ce problème, nous avons utilisé un modèle d'apprentissage automatique (machine learning) qui s'appuie sur un vaste ensemble de données pour améliorer fortement le positionnement.

Ce mémoire est divisé en trois chapitres. Les deux premiers chapitres sont théoriques, tandis que le dernier chapitre décrit le travail effectué pour créer le modèle d'apprentissage automatique proposé.

Le premier chapitre décrit les principaux systèmes de positionnement tels que la localisation GPS et GSM ainsi que les techniques et méthodes utilisées pour déterminer la localisation d'un téléphone mobile.

Le deuxième chapitre définit les principes et les concepts de base de l'intelligence artificielle, mentionnant ses techniques et ses différents types.

Dans le troisième chapitre, nous présenterons les étapes de base pour créer et évaluer le modèle et discuterons de ses résultats.

Enfin, nous terminerons par une conclusion générale sur le travail effectué.

chapitre 1 : Techniques et technologies de localisation

1. Introduction

La localisation est un processus de localiser un appareil ou la position d'un être humain à la surface de la terre précisément ou approximativement en temps réel, à l'aide de ses coordonnées géographiques (latitude, longitude) ou de divers outils tels que des cartes ou des systèmes de géolocalisation, par exemple GPS.

De nos jours, la création d'outils tels que les cartes numériques, les applications et les globes virtuels (Google Earth et Google Maps) a contribué à faciliter la circulation des individus et des groupes. Ce suivi peut être bénéfique dans plusieurs domaines tels que la sécurité, dans le domaine médical, la navigation... etc. [1]

De nombreux systèmes de localisation ont également été développés ces dernières années. Dans cette partie, nous examinerons en détail certains des systèmes les plus performants.

Les systèmes de positionnement par satellite de navigation sont également appelés GNSS. Ce système fournit des informations précises, continues et globales sur la position et la vitesse aux utilisateurs équipés de l'équipement de réception approprié. Voici quelques systèmes de localisation par satellites en service ou en développement :

- GPS : développé par les États-Unis, instauré dans les années 1990, mais il permet une localisation précise à une marge d'erreur de 4 à 5 mètres.
- GLONASS : dispositif militaire russe mis en service en 1982. Sa précision est aujourd'hui proche de celle du GPS.
- GALILEO : dispositif civil européen mis en service en 2011 qui offre la plus grande précision.
- Beidou : Le système chinois Beidou est opérationnel depuis 2003 et a été amélioré avec Beidou-3 en 2020. Il assure une couverture régionale et mondiale. La précision varie en fonction du nombre de signaux captés, mais tous reposent sur la trilatération des signaux émis par les satellites. [2]

Dans ce chapitre, nous allons présenter les différents systèmes de localisation, les différentes techniques de localisation (AOA, TOA, TDOA et RSSI) et les algorithmes les plus connus qui sont la triangulation et la trilatération.

2. Le système de navigation GPS

Le système de positionnement global (GPS) est un système de navigation par satellite qui appartient aux États-Unis et qui assure des services de positionnement, de navigation et de référence temporelle, dits « services PNT » (position, navigation, and timing). Développé par le département américain de la Défense (DoD) au début des années 1970 à des fins militaires, il a ensuite été mis à la disposition des civils et constitue désormais un système à double usage accessible à la fois aux utilisateurs militaires et civils. Il permet de déterminer les coordonnées géographiques de n'importe quel point situé à la surface du globe. Sa précision peut atteindre 1 mètre [3]. Le système GPS se compose de trois segments principaux, comme le montre la figure 1 :

- **Segment spatial** : Le segment spatial du système GPS comprend une constellation de 24 satellites. Chaque satellite émet un signal composé de plusieurs éléments, notamment deux ondes sinusoïdales (ou fréquences porteuses), deux codes numériques et un message de navigation. Les porteuses et les codes sont principalement utilisés pour calculer la distance entre le récepteur utilisateur et les satellites GPS. Le message de navigation contient diverses informations, dont les coordonnées spatiales des satellites en fonction du temps. Ces signaux émis sont régulés par des horloges atomiques de haute précision à bord des satellites. [4]
- **Segment de contrôle** : Le segment de contrôle du système GPS est constitué d'un réseau mondial de stations de suivi. Sa principale responsabilité est de suivre les satellites GPS pour déterminer et prédire leur position, assurer l'intégrité du système, collecter des données atmosphériques, maintenir l'almanach des satellites et gérer d'autres aspects. Ces informations sont ensuite compilées et téléchargées vers les satellites GPS. [4]
- **Segment utilisateur** : il comprend les récepteurs GPS, comme ceux que l'on trouve dans les smartphones ou les systèmes de navigation des véhicules, qui reçoivent les signaux des satellites et calculent la position tridimensionnelle et l'heure. [4]

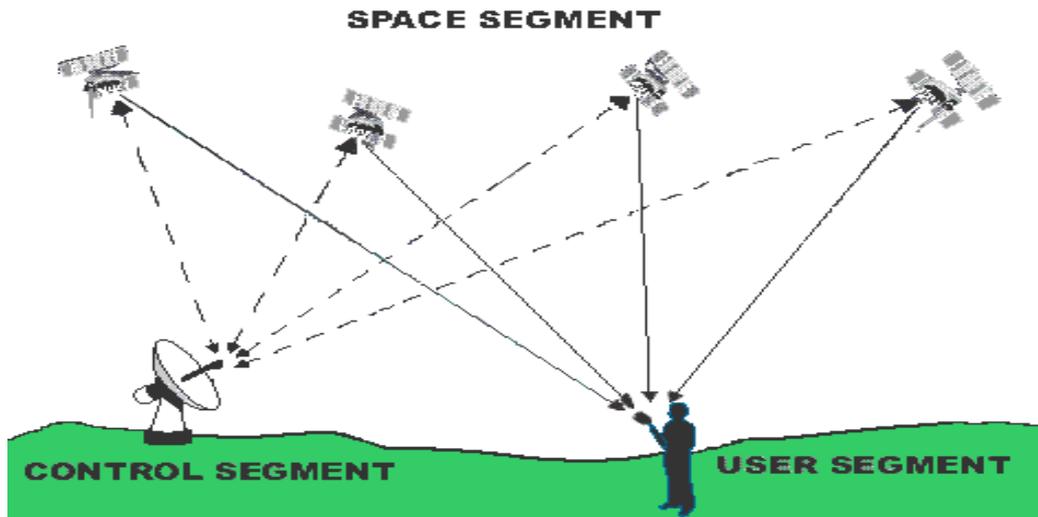


Figure 1 : Les segments du système GPS. [5]

2.1 Fonctionnement du système GPS

Le fonctionnement du GPS est basé sur un réseau de satellites qui orbitent autour de la Terre et transmettent des signaux que les récepteurs GPS utilisent pour déterminer leur position géographique. Les satellites dédiés au système GPS orbitent autour de la Terre deux fois par jour suivant des trajectoires bien définies. Chaque satellite émet un signal distinct ainsi que des informations orbitales qui permettent aux dispositifs GPS de déchiffrer et de calculer avec précision la localisation du satellite.

Le principe de fonctionnement du GPS repose sur la trilatération, une méthode qui calcule la position d'un objet en se basant sur la distance plutôt que sur les angles, bien que le terme triangulation soit souvent utilisé de manière interchangeable. Le récepteur GPS évalue sa distance par rapport à chaque satellite en mesurant le temps que leurs signaux mettent pour l'atteindre. Lorsque votre récepteur GPS a les informations sur les positions et les distances d'au moins trois satellites, il dispose des données nécessaires pour calculer votre position. Il doit également connaître la position exacte de chaque satellite dans le ciel, ce qu'il fait en utilisant des données orbitales précises stockées dans sa mémoire. Ce processus de calcul peut être complexe, mais c'est une tâche prise en charge par votre récepteur GPS. Pour que cela fonctionne, votre récepteur doit avoir une synchronisation précise pour savoir quand chaque signal satellite a été émis et quand il a été reçu. Ainsi, l'exactitude de l'heure est cruciale pour le bon fonctionnement du système.

De nos jours, le GPS est intégré dans une multitude d'appareils variés, allant des montres intelligentes aux dispositifs de communication par satellite, en passant par les automobiles et les bateaux. [6]

Le système de positionnement global (GPS) fournit d'autres informations utiles après avoir déterminé l'emplacement de l'utilisateur, telles que :

- La distance parcourue.
- La vitesse de déplacement.
- La distance restante jusqu'à la destination.

2.2 Les avantages et les inconvénients de GPS

Le GPS (système de positionnement global) offre plusieurs avantages, mais il présente également quelques inconvénients :

Les Avantages

- **Localisation en temps réel** : elle permet d'identifier la position actuelle des véhicules, des produits et des utilisateurs avec précision et sans délai.
- **Amélioration de la navigation et de la planification des itinéraires** : Le système GPS contribue à simplifier la navigation et l'élaboration des parcours pour les utilisateurs individuels et les sociétés, ce qui permet d'économiser du temps et du carburant.
- **Assistance et suivi en cas d'urgence** : Il sert à surveiller la sûreté des travailleurs dans des zones isolées ou périlleuses.
- **Réduction de la probabilité de se perdre** : l'utilisation du GPS facilite la navigation et l'organisation des déplacements, ce qui réduit les risques de se perdre, notamment dans des lieux inconnus.
- **Réduction des erreurs dans la collecte et l'analyse des données** : Le GPS permet une collecte de données avec une grande précision, ce qui peut améliorer la qualité de l'analyse des données et de la prise de décision.
- **Capacité accrue à effectuer des tâches à distance** : Le GPS autorise l'exécution de tâches à distance par les individus et les entreprises, comme la surveillance et le contrôle des équipements. Cela économise du temps, de l'argent tout en améliorant la productivité.

- **Réduction du besoin de travail manuel :** La technologie GPS facilite l'automatisation de diverses tâches comme le suivi des biens et la planification des trajets, ce qui peut diminuer la dépendance au travail manuel. [7]

Les inconvénients

- **Fonctionnalité limitée dans certains domaines :** dans des lieux spécifiques tels que les canyons urbains, les forêts denses ou les endroits isolés sans ligne de visée vers le ciel, les signaux peuvent être obstrués ou affaiblis, entravant ainsi les performances des dispositifs GPS.
- **Risque d'interférence ou de perturbation du signal :** les signaux émis par les satellites peuvent également être perturbés par divers facteurs, tels que les éruptions solaires, les variations atmosphériques, ce qui peut entraîner des erreurs dans les données de localisation.
- **Surveillance et suivi :** L'usage du GPS pour le suivi de personnes ou de biens peut engendrer des inquiétudes quant à la protection de la vie privée et aux droits individuels.
- **Risque de compromission des données personnelles :** le GPS permet la collecte et stocke des données personnelles qui risquent d'être compromises par des pirates informatiques.
- **Dépenses permanentes pour l'entretien et les mises à jour :** l'utilisation continue de la technologie GPS implique des frais constants pour la maintenance, les mises à jour logicielles, le remplacement des pièces défectueuses et la réparation des dommages.
- **Risque de défaillance de l'équipement :** Le GPS repose sur un réseau élaboré d'appareils susceptibles de dysfonctionnements, ce qui peut provoquer des inexactitudes dans les informations de géolocalisation.
- **Absence de plan de secours en cas de panne du système :** En cas de défaillance totale du système GPS, il n'y aurait pas de plan de secours sur lequel s'appuyer. [7]

3. Système de réseau GSM

Le GSM (Global System for Mobile Communications) est la première norme de téléphonie cellulaire numérique, reconnue et globale pour la communication, lancée par la CEPT (Conférence Européenne des Postes et Télécommunications). Il s'agit d'une technologie mobile avancée utilisée pour offrir des administrations basées sur la voix et améliorer la portabilité et réduire la taille des appareils mobiles. GSM est le nom d'un groupe

d'institutionnalisation créé en 1982 pour créer une norme européenne typique de téléphonie mobile, offrant des avantages en termes de confort d'utilisation, de coût et de compatibilité avec les téléphones fixes.

Le GSM fonctionne sur les groupes de correspondance polyvalents 900 MHz et 1800 MHz dans de nombreuses régions du monde. Le GSM a été un succès majeur, marquant le passage à la deuxième génération de réseaux cellulaires avec un débit théorique de 9,6 kbit/s, permettant la transmission de données numériques telles que les SMS et les MMS. La figure 2 illustre une antenne GSM. [8]



Figure 2 : Antenne GSM. [9]

3.1 Historique :

Le GSM, système mondial de communication mobile, a joué un rôle essentiel dans la révolution des communications sans fil. Son développement a été marqué par des jalons importants, de sa création dans les années 1980 à son adoption mondiale. Voici un résumé

chronologique de son évolution. Le tableau 1 ci-dessous représente les étapes de développement de GSM :

Tableau 1 : historique de GSM. [10]

Année	Événements
1982	– Le Groupe spécial mobile (GSM) est formé par la Confédération européenne des postes et télécommunications (CEPT) pour concevoir une technologie mobile paneuropéenne.
1986	– Les dirigeants des États membres de l'Union Européenne ont donné leur approbation au projet GSM. L'initiative de la Commission Européenne propose de réserver une bande de fréquences de 900 MHz au GSM.
1987	– 13 pays ont signé un protocole d'accord visant à créer un système de téléphonie cellulaire commun à l'échelle paneuropéenne dans cette même bande de fréquences.
1989	– La responsabilité des spécifications GSM est transférée à un comité technique de l'Institut européen des normes de télécommunications (ETSI).
1990	– Les travaux d'adaptation GSM commencent pour la bande DCS1800.
1991	– Une première communication expérimentale est faite par le groupe GSM.
1992	– Lancement effectif du service commercial et élargissement des pays signataires du GSM. Premier SMS envoyé.
1993	– Plusieurs pays non européens d'Amérique du Sud, d'Asie et d'Australie ont adopté le GSM.
1995	– Les services de fax, de données et de SMS sont démarrés, la vidéo sur GSM est démontrée.
1997	– 15 réseaux GSM sont en ondes aux États-Unis utilisant la bande 1900 MHz.
Début des années 2000	– Transition vers la 3G avec l'avènement de l'UMTS, offrant des vitesses de données encore plus rapides et de nouvelles fonctionnalités.
2010	– Le GSM reste largement utilisé, mais est progressivement remplacé par des technologies de communication mobile plus avancées, telles que l'UMTS (3G) et le LTE (4G).
Aujourd'hui	– GSM est l'acronyme de Global System for Mobile Communications, ce qui souligne son objectif d'être une norme mondiale.

3.2 Architecture du réseau GSM

Le réseau GSM est composé de quatre sous-ensembles :

- La station mobile (MS).
- Sous-système de station de base (BSS).
- Le sous-système d'acheminement (NSS).
- Le sous-système d'exploitation et de maintenance (OSS). [8]

La figure suivante nous présente l'architecture d'un réseau GSM :

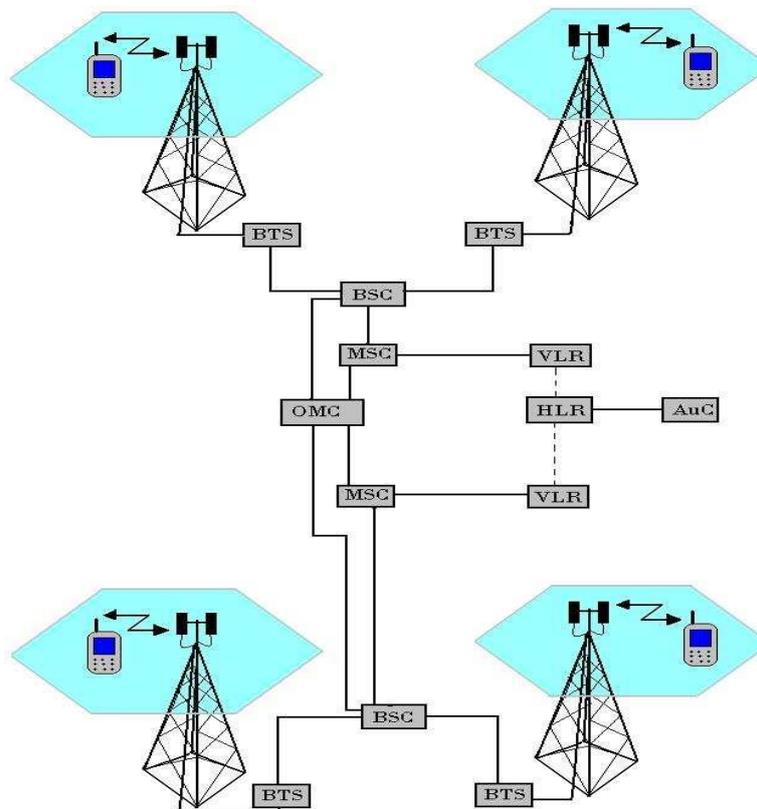


Figure 3 : Architecture du réseau GSM. [11]

3.2.1 La Mobile Station (MS)

Mobile Station composée du Mobile Equipment (le terminal GSM) et du Module d'identité de l'abonné (SIM), une petite carte dotée de mémoire et de microprocesseur, qui peut être utilisée pour passer ou recevoir des appels, permet d'identifier l'abonné indépendamment du terminal utilisé. La carte SIM peut stocker des messages courts et des informations de facturation, et elle dispose d'une fonction d'annuaire téléphonique et d'une courte liste de numéros d'appel stockant les noms et les numéros de téléphone. Ainsi, il est possible de recevoir et d'émettre des appels, ainsi que d'utiliser tous les services simplement en insérant la carte SIM dans n'importe quel terminal.

Le Mobile Equipment est identifié exclusivement à l'intérieur de n'importe quel réseau GSM par l'International Mobile Equipment Identity (IMEI), c'est une sorte de numéro de série. Il permet d'identifier les équipements obsolètes, volés ou non fonctionnels et peut refuser le service si nécessaire. [12]

L'IMEI est un numéro à 15 chiffres qui suit la structure suivante : $IMEI = TAC / FAC / SNR / SP$.

Où :

- **TAC** : (Type Approval Code), déterminé par le corps central du GSM est composé de 6 chiffres.
- **FAC** : (Final Assembly Code), identifie le constructeur et est constitué de 2 chiffres.
- **SNR** : (Serial Number) Le numéro de série du mobile, composé de 6 chiffres.
- **SP** : (chiffre supplémentaire de réserve) Un chiffre supplémentaire réservé à des fins spécifiques.

Donc, l'IMEI est une combinaison unique de ces éléments qui permet d'identifier de manière précise chaque appareil mobile dans le réseau GSM.

La carte SIM contient l'**Identité internationale d'abonné mobile (IMSI)**, qui sert à identifier l'abonné dans n'importe lequel des systèmes GSM. Il est composé de trois parties :

- **MCC** : (Mobile Country Code) Un code à 2 ou 3 chiffres qui représente le pays de l'opérateur.
- **MNC** : (Mobile Network Code) Un code à 2 chiffres qui identifie l'opérateur mobile au sein du pays.
- **MSIN** : (Mobile Station Identification Numbers) pouvant comporter jusqu'à 10 chiffres.

Donc la carte SIM est bien plus qu'une simple puce ; elle joue un rôle crucial dans l'identification de l'abonné et la sécurisation des communications mobiles. [13]

3.2.2 Sous-système de station de base (BSS)

Ce sous-système de station de base gère la communication avec les mobiles du réseau à un niveau fondamental. Il se compose de deux éléments, la station de base BTS (Base Transceiver Station) et le contrôleur de station de base (BSC).

La station de base est un élément essentiel des réseaux de téléphonie mobile, contient des émetteurs et des récepteurs radio qui desservent une zone spécifique appelée cellule, Elle communique directement avec les téléphones mobiles en utilisant des signaux radio et transmet les signaux des appels téléphoniques et des données vers les appareils mobiles.

La BTS est connectée au BSC via une interface spécifique et transmet les données des téléphones mobiles vers le MSC et reçoit des instructions de gestion et de contrôle du BSC.

Les BSC sont des équipements centralisés qui contrôlent et supervisent les activités des BTS. Comprennent le contrôle de la puissance des signaux radio émis par les BTS, et la gestion des livraisons (transferts d'appels d'une cellule à une autre), et la planification des fréquences radio pour éviter les interférences, la collecte de données de performance et la coordination des ressources radio entre les BTS. [14]

3.2.3 Sous-système de commutation de réseau (NSS)

Le sous-système de commutation de réseau (NSS) est une composante clé dans le réseau GSM. Il assure la gestion des appels, la commutation des données et la gestion des abonnés dans un réseau mobile. Le NSS est responsable de tâches telles que l'enregistrement des abonnés, l'acheminement des appels, la gestion de la mobilité des abonnés, etc.

Les composants importants à l'intérieur du NSS : [14]

- **Le Centre de Commutation Mobile (MSC)**

Il occupe une position centrale dans le réseau de télécommunications mobiles. Il contrôle tous les appels au sein des BSS connectés et gère toutes les commutations d'appels dans l'ensemble de la zone, y compris entre deux appels sous la même cellule dans la même station de base. Il est connecté aux éléments de réseau de support tels que le VLR, HLR, EIR et AUC. [14]

- **Registre de localisation du domicile (HLR)**

Le registre de localisation domestique est une grande base de données dans un réseau de communication sans fil. Qui contient des informations sur le profil de service et l'emplacement actuel (adresse VLR) des abonnés (SIM) appartenant au réseau enregistré sur le HLR spécifique. [14]

- **Registre de localisation des visiteurs (VLR)**

Le registre de localisation domestique est une base de données utilisée dans les réseaux de télécommunications mobiles. Il contient des informations sur tous les mobiles (cartes SIM)

actuellement connectés au réseau desservi par le MSC. Dans le VLR, nous trouvons une copie de toutes les informations HLR concernant le profil de service de l'utilisateur, ainsi que des informations sur la localisation actuelle (zone de localisation LAC) du mobile. [14]

- **Registre d'authentification (AUC)**

Le registre d'authentification (AUC) contient la clé d'identification individuelle de chaque abonné, également présente sur la carte SIM de l'abonné. Lorsqu'un appareil mobile tente de se connecter au réseau, le HLR ou le VLR peut interroger l'AUC pour obtenir des informations d'authentification. Ces informations servent à vérifier l'identité de l'abonné et à sécuriser les communications en les cryptant. [14]

- **Registre d'identité des équipements (EIR)**

Le registre d'identité de l'équipement est une base de données employée dans les réseaux de télécommunications mobiles pour enregistrer les données relatives aux appareils mobiles individuels. En cas de vol d'un téléphone, son identifiant unique matériel IMEI peut être ajouté à une liste noire dans l'EIR, bloquant ainsi les appels entrants et sortants de ce téléphone. [14]

3.2.4 Le sous-système opérationnel ou d'exploitation et de maintenance (OSS)

Le sous-système opérationnel est un composant de l'ingénierie générale de l'organisation GSM. Il est utilisé pour la gestion du réseau et le contrôle du système GSM, et il est également utilisé pour contrôler le tas de mouvements du BSS. Il intervient à plusieurs niveaux, notamment pour la détection des pannes, la mise en service des sites, la modification des paramètres et la réalisation de statistiques. [8]

4. La localisation par réseau GSM

La géolocalisation GSM utilise le réseau de téléphonie mobile pour établir la position d'un téléphone portable. Le principe de fonctionnement de la localisation Gsm est assez simple, en fait.

Le réseau cellulaire est divisé en zones appelées "cellules", chacune desservie par une station de base. Ces cellules se superposent partiellement pour assurer une transition fluide de la couverture réseau lorsqu'un utilisateur se déplace d'une zone à l'autre. Un téléphone portable, équipé d'une carte SIM contenant un numéro unique, communique périodiquement avec les antennes du réseau pour maintenir sa connectivité. Lorsqu'un utilisateur se déplace, le réseau détermine quelle station de base gèrera la communication. Pour localiser un téléphone, le

réseau GSM utilise plusieurs techniques de géolocalisation. La précision varie en fonction du milieu, qu'il soit urbain ou rural. [15]

4.1 Les techniques de localisation dans le réseau cellulaire GSM

Les méthodes de positionnement mesurent les distances et les angles en utilisant diverses techniques de mesure, notamment l'identifiant de cellule (Cell-ID), le temps d'arrivée (TOA), la différence de temps d'arrivée (TDOA), l'angle d'arrivée (AOA) et la puissance du signal reçu (RSS – Received Signal Strength).

4.1.1 Identifiant de la cellule (Cell-ID)

La méthode de localisation « à la cellule près » est simple et économique à mettre en place, car elle fonctionne avec tous les terminaux existants. Elle repose sur l'émission d'un signal aller-retour de signalisation entre l'utilisateur et l'antenne. Le téléphone mobile est localisé en identifiant la cellule à laquelle appartient l'antenne utilisée pour la communication.

Le calcul de la position est rapide, se limitant à la recherche dans la base de données à partir de l'identifiant de la cellule. Cependant, cette technique manque de précision : en zone urbaine dense, elle situe une personne à environ 250 mètres près, tandis qu'en milieu rural, la marge d'erreur peut atteindre une dizaine de kilomètres. Cependant, cette méthode reste moins précise dans les zones où une seule station de base couvre un vaste territoire. Dans les villes où les opérateurs ont installé plusieurs stations de base pour mieux desservir les utilisateurs, les cellules couvrent une zone plus restreinte, améliorant ainsi la précision du système. Cette technique est illustrée par la figure 4 . [16]

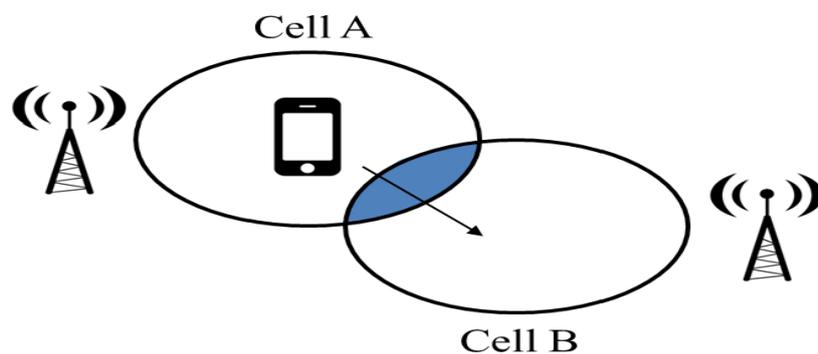


Figure 4 : Identifiant de la cellule (Cell-ID). [17]

4.1.2 Temps d'arrivée (TOA)

La technique TOA est une méthode de localisation utilisée dans les réseaux GSM. Il s'agit de mesurer le temps nécessaire à un signal pour parcourir une station de base et un téléphone

mobile ou inversement. Pour déterminer la position d'un appareil, TOA mesure le temps d'aller-retour d'un signal transmis de l'émetteur au récepteur, puis renvoyé en écho à l'émetteur. En connaissant la vitesse de propagation du signal (la vitesse de la lumière) et le temps de propagation, on peut calculer la distance entre l'appareil et chaque station de base. Ainsi, l'utilisation de plus de trois émetteurs est préférable pour localiser des objets en mouvement. Illustré par la figure 5. [18]

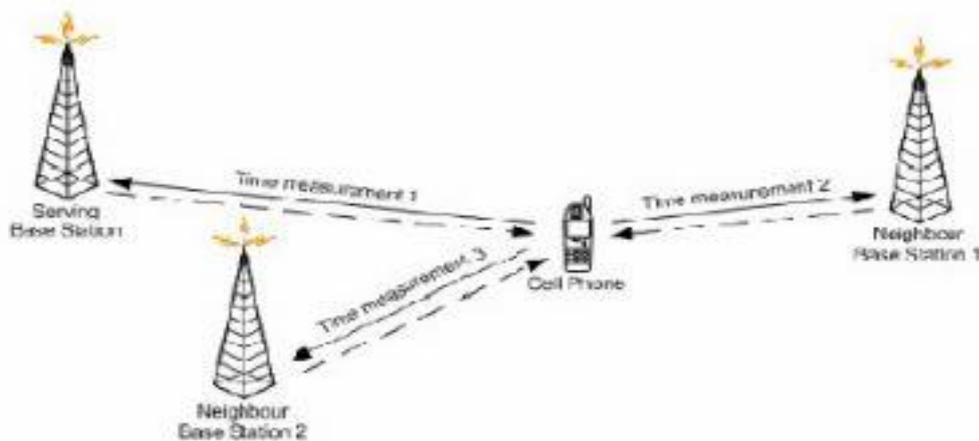


Figure 5 : Localisation par TOA. [19]

4.1.3 Angle d'Arrivé (AOA)

L'angle d'arrivée est mesuré dans deux sens : soit de la station de base vers le téléphone portable, soit en mesurant l'angle d'arrivée d'un signal provenant du téléphone portable vers la station de base. Cette technique permet de calculer les angles auxquels un signal arrive à deux stations de base à partir d'un combiné, utilisant la triangulation pour déterminer l'emplacement. La mesure du signal génère une ligne droite entre la station de base et le téléphone portable. De manière similaire, une autre mesure de l'angle d'arrivée fournit une deuxième ligne droite. L'intersection de ces deux lignes permet d'obtenir une position fixe pour ce système angle-angle. Comme présenté dans la figure 6 . [18]



Figure 6 : Localisation par AOA. [19]

4.1.4 Différence de temps d'arrivée TDOA

La différence de temps d'arrivée est utilisée pour déterminer la position d'un équipement mobile sans nécessiter une synchronisation précise entre l'équipement mobile et les stations de base. Cette méthode de localisation est applicable uniquement aux mobiles situés à proximité des stations de base. L'équipement mobile émet un signal qui arrive à des instants différents au niveau de chaque BTS. Un serveur central collecte ces signaux provenant des BTS et calcule la différence de temps d'arrivée entre eux. Il est essentiel de noter que dans les systèmes TDOA (Time Difference of Arrival), la synchronisation des stations de base est cruciale selon le mode de fonctionnement du système :

- Pour le mode d'autopositionnement, les stations de base sont traitées comme des émetteurs, et le signal transmis doit quitter chaque station de base simultanément. Sinon, des erreurs de biais peuvent se produire dans les mesures TDOA, affectant la précision du positionnement hyperbolique.
- Pour le mode de positionnement à distance, les stations de base sont considérées comme des récepteurs. Étant donné que la transmission du téléphone mobile est détectée par plusieurs stations de base, il est nécessaire d'avoir une relation temporelle connue entre les horloges des récepteurs de ces stations. Sinon, des erreurs de biais peuvent survenir, affectant la précision du positionnement. Comme le montre la figure suivante « figure 7 ». [18]

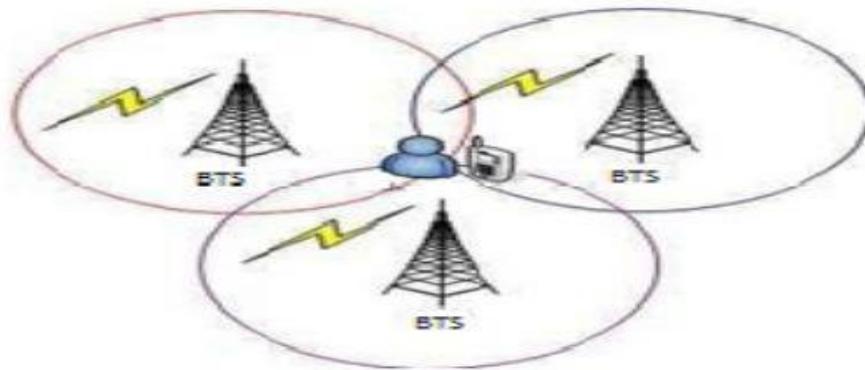


Figure 7 : Localisation par TDOA. [19]

4.1.5 Indicateur de force du signal reçu (RSSI)

Le RSSI (Received Signal Strength Indicator) est une mesure de la force du signal reçu par un appareil Wi-Fi. Celle-ci dépend de la puissance d'émission et de la distance qui sépare l'émetteur du récepteur. L'unité de mesure RSSI est le dBm (décibel milliwatt). En général, un RSSI plus élevé est plus adapté à une meilleure qualité de communication. Cependant, il est important de noter que le RSSI ne reflète pas directement la qualité du signal. D'autres facteurs, tels que les obstacles physiques (murs, etc.) et les interférences peuvent également influencer la qualité réelle du signal. La méthode de RSSI est basée sur le calcul de la distance entre l'émetteur et le récepteur à partir de la puissance du signal reçu associée à un modèle de propagation dans l'environnement. Cette technique est illustrée par la figure 8 . [15]

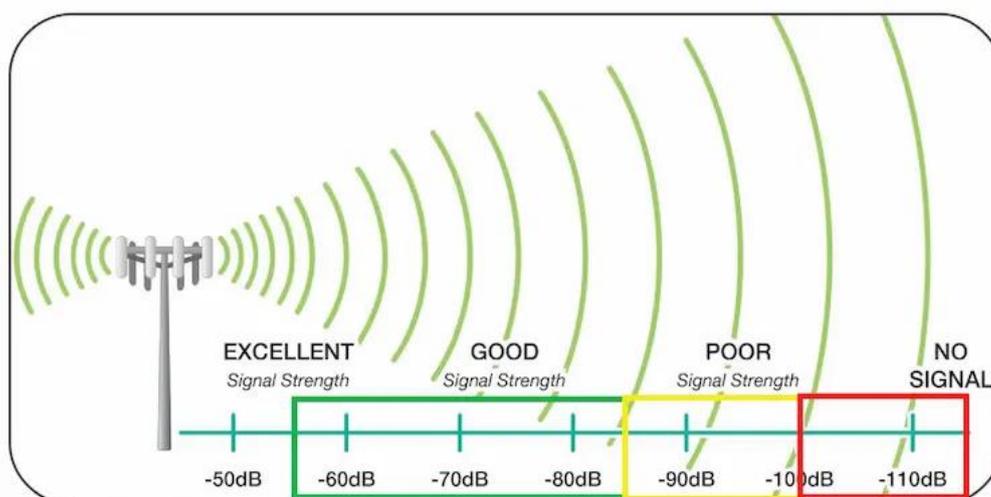


Figure 8 : Localisation par RSSI. [19]

4.2 Synthèse des techniques de localisation dans le réseau cellulaire GSM

Le tableau 2 montre les avantages et les inconvénients de chaque technique de localisation en utilisant le réseau GSM.

Tableau 2 : synthèse des techniques de localisation dans le réseau cellulaire GSM.

Les techniques	Avantage	Inconvénient
Identifiant de la cellule (Cell-ID)	<ul style="list-style-type: none"> – En évaluant la force du signal, la cellule permet aux appareils de sélectionner la meilleure cellule à laquelle se connecter. – L'identification des cellules consomme moins d'énergie. 	<ul style="list-style-type: none"> – La qualité du service est affectée dans les zones urbaines denses.
Temps d'arrivée (TOA)	<ul style="list-style-type: none"> – L'algorithme de positionnement est facile à mettre en œuvre. – La méthode offre une précision accrue dans les environnements confinés. [20] 	<ul style="list-style-type: none"> – Bien que TOA puisse offrir une grande précision, sa mise en œuvre peut être coûteuse en raison de la nécessité d'équipements sophistiqués. [20]
Angle d'arrivée (AOA)	<ul style="list-style-type: none"> – Un nombre réduit de stations de base fixes est requis. – L'algorithme de détermination de position est facile à appliquer. [20] 	<ul style="list-style-type: none"> – Il est essentiel de bénéficier d'une ligne de transmission directe. – Les frais d'installation élevés. – La précision est limitée. [20]
Différence de temps d'arrivée (TDOA)	<ul style="list-style-type: none"> – L'algorithme de positionnement est simple à mettre en œuvre. – Précision plus élevée en milieu confiné. – Aucune synchronisation d'horloge n'est requise entre les appareils mobiles et les stations de base. [20] 	<ul style="list-style-type: none"> – Une résolution temporelle élevée au niveau du récepteur est nécessaire. – Son efficacité est limitée dans des environnements spécifiques tels que les zones urbaines où les immeubles peuvent entraîner des interférences et ainsi rendre la localisation moins précise. [20]

L'indicateur de force du signal reçu (RSSI)	<ul style="list-style-type: none"> – Algorithme simple – Le coût d'implantation est faible. 	<ul style="list-style-type: none"> – Précision peu élevée.
---	---	---

5. Les algorithmes de localisation

Dans les systèmes de localisation, en particulier ceux utilisant GSM, les algorithmes varient en fonction des conditions de visibilité directe (LOS : Line-of-Sight) et des conditions sans visibilité directe (NLOS : Non-Line-of-Sight).

5.1 Les algorithmes pour le scénario LOS

Dans ce scénario, il existe une visibilité directe entre l'appareil à localiser et les antennes de téléphonie mobile.

5.1.1 La triangulation

La triangulation est une méthode de localisation qui utilise les signaux de trois satellites pour déterminer la position exacte d'un récepteur GPS. Chaque satellite, équipé d'une horloge atomique très précise, transmet des signaux dont le temps de parcours jusqu'au récepteur est mesuré. La distance « d » entre un satellite et le récepteur est calculée en fonction de ce temps de parcours. Et la surface sur laquelle le récepteur peut se trouver est représentée par une sphère centrée sur le satellite et de rayon « d ». Lorsqu'un deuxième satellite est impliqué, l'intersection des deux sphères crée un cercle qui représente toutes les positions potentielles du récepteur. Cependant, deux satellites ne suffisent pas pour une localisation précise. L'introduction d'un troisième satellite permet de réduire les possibilités à deux points. Si l'on considère que le récepteur se trouve sur la surface de la Terre, un seul de ces points est viable. La position exacte du récepteur est alors déterminée en éliminant le point non pertinent, ce qui permet de localiser précisément l'objet ou la personne sur la carte.

La localisation mobile utilise des principes similaires, mais avec des différences en termes de couverture et de précision dues aux caractéristiques des réseaux de téléphonie mobile. La précision de la localisation mobile est généralement inférieure à celle du GPS en raison de la variabilité des signaux GSM et de la densité des cellules de réseau dans différentes régions. La localisation GSM, similaire à la triangulation par satellite, détermine la position d'un téléphone mobile en utilisant les signaux des antennes relais. La localisation GSM s'appuie sur le principe de triangulation, utilisant les signaux émis par trois antennes relais pour

localiser un téléphone mobile. Ces antennes, qui émettent et reçoivent des informations, varient en fonction des mouvements du. La triangulation GSM est considérée comme l'une des méthodes de localisation les plus précises, avec une marge d'erreur relativement faible : environ 120 à 130 mètres en milieu urbain où le réseau est dense, et jusqu'à 5 kilomètres en zone rurale. Le processus de localisation prend généralement entre 4 et 6 secondes. Les téléphones peuvent être localisés en utilisant les antennes GSM, qui sont normalement utilisées pour transmettre des données. La précision de cette méthode de localisation varie de moins de 200 mètres à plusieurs kilomètres, en fonction de la densité des antennes relais. En milieu urbain, où les antennes sont plus nombreuses, la localisation est généralement plus précise que dans les zones rurales. Sans antennes à proximité, la géolocalisation devient impossible. La figure ci-dessous montre le principe de triangulation. [21]

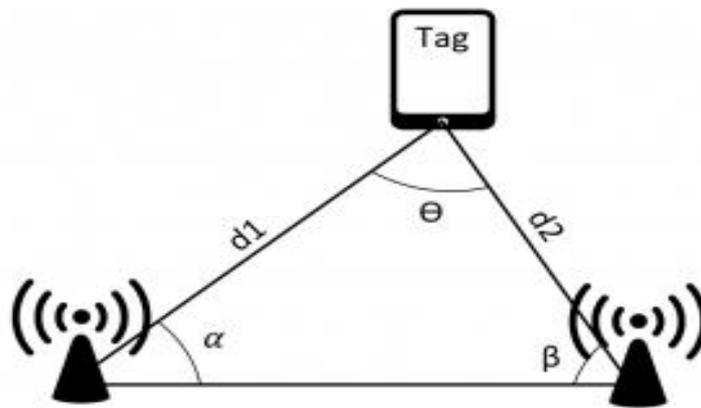


Figure 9 : Le principe de triangulation. [22]

5.1.2 La trilatération

La trilatération est une méthode mathématique utilisée pour déterminer avec précision la position d'un point en se basant sur la géométrie des triangles et la mesure des distances. Contrairement à la triangulation, qui implique la mesure des angles. Ce procédé est largement utilisé dans divers domaines tels que l'arpentage, la navigation et les systèmes de positionnement global.

La trilatération repose souvent sur l'utilisation de signaux radio. Elle exploite la vitesse de propagation des ondes pour calculer les distances. Les points environnants émettent des ondes qui sont ensuite reçues et renvoyées par l'objet dont on cherche la position. En mesurant le temps mis par l'onde pour parcourir cette distance, on peut calculer la distance entre l'objet et l'émetteur. Le calcul devient alors simple (en théorie) :

$$V=d/t \leftrightarrow d=v*t$$

Où :

- V : la vitesse
- d : la distance
- t : le temps

Une fois les distances mesurées, on trace des cercles centrés sur les points connus, dont les rayons correspondent à la distance mesurée entre chaque point et l'objet inconnu. L'intersection de ces cercles indique la position recherchée du point recherché. La figure 10 illustre le principe de trilatération. [23]

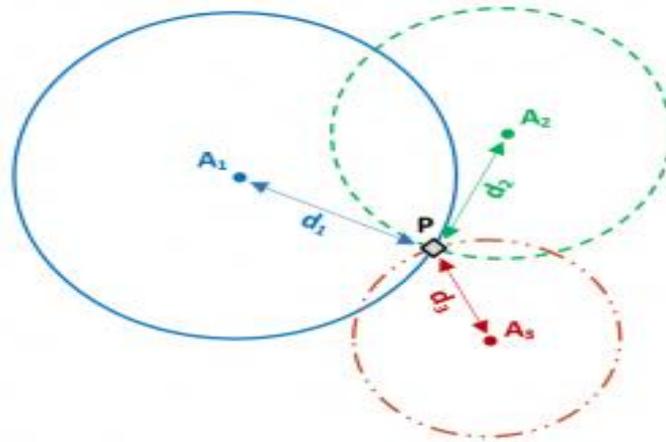


Figure 10 : Le principe de trilatération. [22]

5.1.3 La multilatération

Les systèmes de multilatération (MLAT) sont une technique de localisation utilisée pour déterminer les coordonnées d'objets en déplacement, comme des avions ou des installations d'aérodrome, qui émettent des signaux radio. Son principe de base repose sur le concept multilatéral, également employé dans les systèmes de radionavigation à longue distance basés sur le décalage horaire. En utilisant les distances mesurées à partir de plusieurs antennes, des sphères sont créées autour de chaque antenne, avec le centre de la sphère à l'emplacement de l'antenne et un rayon égal à la distance mesurée. La position de l'objet est ensuite calculée en déterminant le point d'intersection de toutes ces sphères. [24]

5.2 Les algorithmes de localisation en scénario NLOS

Dans ce scénario, il y a des obstacles entre l'appareil et les antennes, ce qui rend la visibilité directe limitée ou absente.

5.2.1 Empreinte digitale de force du signal

L'empreinte digitale est le moyen le plus courant d'identifier l'emplacement des appareils mobiles en créant une base de données d'empreintes digitales de force du signal. La prise d'empreintes digitales est prise en charge non seulement dans les environnements intérieurs, mais également dans les environnements extérieurs ou dans les réseaux d'accès radio cellulaire.

Le processus d'empreinte digitale se compose de deux étapes : l'étape d'entraînement et l'étape de positionnement. À l'étape d'entraînement, il collecte un ensemble de modèles de signaux RF, et à l'étape de positionnement, l'algorithme estime la position de l'utilisateur en trouvant un emplacement de référence. Il peut atteindre une meilleure précision que la simple triangulation ou les méthodes basées sur l'identification des cellules, mais nécessite une collecte et une maintenance de données importantes. [25]

6. Les avantages et les inconvénients de localisation**Les avantages**

- Détermine avec précision l'emplacement des objets, des personnes ou des véhicules, ce qui est important pour de nombreuses applications.
- Permet aux personnes à risque de les localiser rapidement.
- Améliore les opérations logistiques et prévient le vol en suivant et en gérant les véhicules et les équipements.

Les inconvénients

- La collecte continue de données sur le site peut être considérée comme une atteinte à la vie privée des individus.
- Les pannes de ces systèmes peuvent entraîner des perturbations majeures.
- La maintenance de ces systèmes nécessite souvent un investissement financier énorme de la part des entreprises.

7. Conclusion

Dans ce chapitre, nous avons présenté certains systèmes de positionnement majeurs tels que le système de positionnement global (GPS). Ainsi, nous avons introduit la localisation par le

réseau GSM et les différentes techniques associées (Cell ID, AOA, TOA et TDOA), ainsi que les algorithmes de positionnement utilisés dans les scénarios LOS et NLOS.

L'intelligence artificielle est un domaine qui chaque jour évolue un peu plus en s'appropriant des capacités cognitives humaines, en les développant et parfois même en surpassant ce que les meilleurs êtres humains dans leur domaine sont capables de réaliser. Elle a donné naissance à plusieurs techniques qui sont utilisées globalement dans plusieurs domaines des sciences et des technologies. Dans le prochain chapitre, nous présenterons les concepts de l'intelligence artificielle et ses techniques et les types les plus importants.

chapitre 2 : Généralités sur l'intelligence artificielle

1. Introduction

L'intelligence artificielle (IA) représente une avancée révolutionnaire dans le domaine technologique, redéfinissant fondamentalement notre manière de vivre, de travailler et d'interagir avec la technologie. Cette discipline fascinante se concentre sur la capacité des machines à reproduire et même à améliorer l'intelligence humaine, en englobant des capacités telles que le raisonnement, l'apprentissage par l'expérience et la résolution de problèmes.

En étudiant comment le cerveau humain traite l'information, apprend et prend des décisions, les chercheurs ont développé des méthodologies pour émuler ces processus au sein de logiciels et de systèmes intelligents. Cette émulation forme le fondement d'une vaste gamme d'applications, allant des programmes informatiques traditionnels aux produits et services de pointe. L'avènement de l'apprentissage profond (Deep Learning) et de l'apprentissage automatique (Machine Learning) a propulsé l'IA vers de nouveaux horizons de capacité et de fonctionnalité. Ces technologies permettent aux ordinateurs d'absorber d'énormes volumes de données, d'en discerner des motifs significatifs et d'en tirer des insights pour accomplir des tâches spécifiques. Des chatbots facilitant les interactions client aux voitures autonomes naviguant dans des environnements routiers complexes, l'IA démontre son potentiel à travers l'intégration fluide de ces techniques avancées.

Alors que l'IA continue d'évoluer, son impact sur la société promet d'être profond, ouvrant la voie à une ère d'avancées technologiques sans précédent et de changements transformateurs. En embrassant les possibilités offertes par l'IA, tant les individus que les organisations se positionnent pour débloquer de nouvelles opportunités et ouvrir la voie vers un avenir défini par l'intelligence, l'innovation et un potentiel infini. [26]

2. Intelligence artificielle

L'intelligence artificielle (IA) se compose de multiples algorithmes qui dotent les machines de la faculté d'analyser et de prendre des décisions. Cette technologie permet aux machines de s'ajuster de manière réfléchie aux divers contextes en prédisant des issues basées sur des données préalablement collectées. L'IA offre de nombreux bénéfices, tant sur le plan individuel que professionnel, en apportant des réponses à des enjeux complexes et en nous libérant de certaines obligations. Toutefois, il est essentiel d'instaurer des mécanismes de contrôle rigoureux pour garantir la sécurité et l'efficacité de l'IA. En tant que domaine en constante évolution, l'IA a le potentiel de transformer profondément notre société. Un seul algorithme, capable de digérer des informations et de formuler des choix avisés, peut débloquer un vaste spectre d'applications concrètes aussi bien pour les entreprises que pour

les usages personnels. Par ailleurs, malgré les avancées technologiques récentes, le développement de systèmes d'IA plus performants représente un défi technique pour les chercheurs.

L'IA englobe une gamme de technologies. Parmi elles figurent l'apprentissage automatique et l'apprentissage profond, comme le montre la figure suivante « figure 11 », qui consistent à développer des algorithmes imitant les processus de prise de décision du cerveau humain. Ces technologies sont capables d'intégrer des informations et d'améliorer les compétences de classification ou de prédiction au fur et à mesure de leur apprentissage. [27]

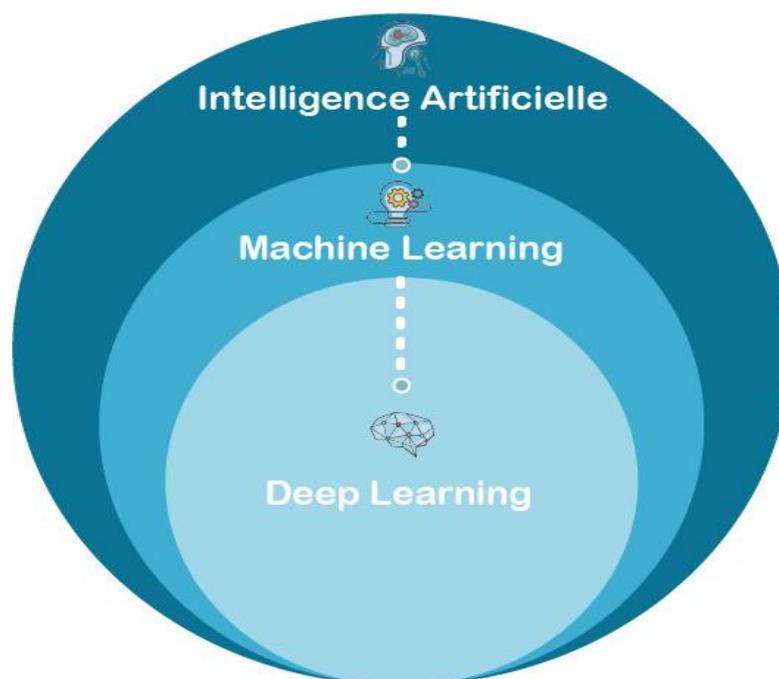


Figure 11 : Schéma de l'intelligence artificielle. [26]

3. Historique

Cela remonte à des centaines d'années. Mais ce n'est que dans les années 1900 que l'idée est devenue plus raisonnable pour les scientifiques. Dès 1921, les dramaturges écrivaient sur les robots et les scientifiques envisageaient l'idée que les ordinateurs soient capables de penser par eux-mêmes. Mais les travaux sur ce qui est devenu l'IA moderne n'ont commencé qu'en 1950. Le tableau 3 résume les événements historiques du développement de l'IA.

Tableau 3 : un résumé des étapes clés de l'évolution de l'IA. [28]

Année	Événements
[1950-1979]	Période de questionnement et de fondation de l'IA, marquée par la création du premier langage de programmation dédié et le développement de programmes capables de réaliser des tâches simples comme répondre à des questions ou classer des informations. C'est aussi à cette époque que le premier prototype de véhicule autonome a vu le jour.
[1980-1987]	Premier âge d'or de l'IA, caractérisé par des avancées significatives et une augmentation des investissements, tant publics que privés. Cette période a également été témoin des premières conférences dédiées à l'IA et de l'introduction des premiers produits d'IA sur le marché.
[1987-1993]	Après une période d'enthousiasme, l'IA a connu un ralentissement, souvent appelé le premier hiver de l'IA, dû à une réduction des investissements face aux attentes non comblées. Malgré cela, les chercheurs ont continué à travailler avec des moyens plus limités.
[1993-2011]	Un renouveau de l'intérêt pour l'IA a mené à de nouvelles avancées et à un regain de financement. Durant cette période, un programme d'IA a battu un champion du monde d'échecs, la NASA a déployé des véhicules sur Mars, et Apple a introduit Siri, le premier assistant numérique.
À aujourd'hui	Nous vivons l'ère de l'intelligence artificielle généralisée. L'IA est omniprésente et continue d'attirer des investissements massifs. Les systèmes d'IA actuels sont capables de jouer à des jeux complexes, de créer des langages originaux et de reconnaître des images avec une précision étonnante.

Les systèmes d'intelligence artificielle fonctionnent en émulant trois capacités : l'apprentissage, le raisonnement et l'autocorrection. L'apprentissage consiste à collecter une énorme quantité de données et à les appliquer à des modèles mathématiques ou à des algorithmes. Ces informations sont utilisées pour faire des prédictions. Le raisonnement

consiste à déterminer l'algorithme approprié pour obtenir le résultat souhaité. L'autocorrection est utilisée pour entraîner les algorithmes qui apprennent et s'adaptent constamment à de nouvelles données pour obtenir des résultats plus précis. Cela permet aux systèmes d'intelligence artificielle de résoudre des problèmes et d'adapter leurs actions en fonction des situations. Par exemple : un chatbot alimenté en données textuelles, peut parler à un humain.

4. Les types

L'intelligence artificielle peut être catégorisée en plusieurs types selon ses capacités, ses fonctionnalités et les technologies. Voici un résumé des principales classifications des différents types d'IA :

4.1 Basé sur les capacités

4.1.1 IA étroite (IA faible)

L'intelligence artificielle étroite, également connue sous le nom d'IA faible, est le seul type d'IA que nous avons réussi à développer jusqu'à présent. Ces systèmes sont spécialisés dans des tâches spécifiques. Puisqu'il n'est équipé que pour une mission particulière, l'AI étroit ne peut pas travailler en dehors de sa zone ou de ses limites. Elle se contente de simuler le comportement humain en se basant sur un ensemble restreint de paramètres et de contextes.

Voici quelques exemples d'applications de l'IA étroite :

- Logiciels de reconnaissance d'images et de visages.
- Outils de cartographie et de prédiction des maladies
- Robots de fabrication et drones.
- Voitures autonomes. [29]

4.1.2 IA générale (IA forte)

L'intelligence générale artificielle (AGI), parfois désignée sous le terme d'IA forte ou d'IA profonde, vise à construire une machine capable d'apprendre et de penser comme une personne seule. Cette forme d'IA est capable d'apprendre et d'appliquer son intelligence pour résoudre n'importe quel problème, agissant de manière indiscernable d'un être humain dans une situation donnée.

Actuellement, les chercheurs du monde entier se concentrent désormais sur la conception de machines avec AI général. Étant donné que les systèmes généraux d'AI sont encore à l'étude, la conception de tels systèmes prendrait beaucoup de temps et d'efforts.

L'AGI s'appuierait sur un cadre théorique appelé théorie de l'esprit, qui consiste à comprendre les besoins, les émotions, les croyances et les processus de pensée d'autres êtres intelligents. [29]

4.1.3 IA super intelligente

La super intelligence artificielle (ASI) est un niveau d'intelligence système auquel les ordinateurs peuvent déjouer les humains et effectuer n'importe quelle tâche mieux que les humains dotés de propriétés cognitives. Il représente une étape où les machines deviennent conscientes d'elles-mêmes et ont des compétences telles que la capacité de réfléchir et de résoudre des problèmes et de prendre des décisions.

En plus de reproduire divers aspects de l'intelligence humaine, l'ASI serait théoriquement supérieure dans tous les domaines, des mathématiques aux relations émotionnelles. Elle disposerait d'une mémoire étendue et d'une capacité de traitement des données bien plus rapide que celle des humains. La conception de tels systèmes dans le monde réel reste un défi qui change le monde. [29]

4.2 Basé sur les fonctionnalités

4.2.1 Machine réactive

L'IA machine réactive fait référence aux machines qui s'appuient uniquement sur les données actuelles pour prendre des décisions, sans la possibilité de déduire ou de prédire des actions futures sur la base de ces données. Ces systèmes d'IA fonctionnent dans un périmètre limité de tâches prédéfinies et n'intègrent pas de mécanismes d'apprentissage ou de planification. Un exemple emblématique de l'IA réactive est le programme IBM Chess, qui a vaincu le champion du monde Garry Kasparov, comme le montre ci-dessous « figure 12 ». [29]



Figure 12 : Le programme IBM Chess, qui a vaincu le champion du monde Garry Kasparov en 1997. [30]

4.2.2 Mémoire limitée

Les systèmes d'IA à mémoire limitée sont capables de prendre des décisions plus éclairées en exploitant les données passées qu'ils ont collectées et stockées pendant une courte période de temps. Une différence cruciale entre ce type d'IA et des versions plus avancées réside dans le fait qu'une fois programmées et entraînées, elles ne peuvent pas s'améliorer par elles-mêmes. Voici quelques exemples d'applications d'IA à mémoire limitée : Les voitures autonomes collectent des données sur les conditions de conduite, la circulation et les piétons environnants pour prendre des décisions et éviter les accidents. Comme dans l'exemple décrit dans la « figure 13 ».

Les robots autonomes utilisent des données limitées sur leur environnement pour prendre des décisions dans leurs actions. [29]



Figure 13 : IA à mémoire limitée. [30]

4.2.3 Théorie de l'esprit

Il s'agit d'un type d'intelligence artificielle plus avancé. Il vise à comprendre les croyances, les pensées et les émotions humaines. En se concentrant sur l'intelligence émotionnelle, ces systèmes pourraient interagir de manière plus naturelle et plus empathique avec les humains. Bien que ce type de machine AL n'ait pas encore été développé, les chercheurs travaillent dur pour améliorer et construire de telles machines IA. [29]

4.2.4 Consciente d'elle-même

Cela représente l'avenir envisagé de l'intelligence artificielle, où les machines posséderaient conscience, sensibilité et conscience d'elles-mêmes. Ce type théoricien d'IA serait capable de comprendre les émotions, de former des croyances et même de développer des désirs. Cependant, cela reste purement spéculatif à ce stade, et des progrès significatifs en matière de technologie et de compréhension seraient nécessaires pour atteindre un tel niveau de sophistication de l'IA. Ces machines seront plus intelligentes que les esprits humains. [29]

5. Les techniques de l'IA

L'intelligence artificielle peut être utilisée pour résoudre des problèmes du monde réel en mettant en œuvre la technique suivante :

5.1 Apprentissage automatique (Machine Learning)

L'apprentissage automatique, ou ML, est une branche de l'intelligence artificielle (IA) axée sur la création d'algorithmes capables de s'auto-améliorer par l'expérience et l'analyse de données. L'objectif de l'apprentissage automatique n'est pas seulement une collecte efficace de données, mais également d'utiliser les quantités toujours croissantes de données collectées en les manipulant et en les analysant sans intervention humaine importante. L'apprentissage automatique se développe d'algorithmes qui aident à prendre ces décisions ou à faire ces prédictions. Ces algorithmes sont programmés pour affiner leur précision et leur efficacité en traitant un volume croissant de données. Il est à la base de plusieurs innovations technologiques actuelles, telles que les assistants personnels intelligents, les systèmes de recommandations personnalisés, les véhicules autonomes et les analyses prédictives avancées. [31]

Les étapes de l'apprentissage automatique (Machine Learning) sont généralement structurées en plusieurs phases clés :

– **Collecte de données** : l'acquisition de données de qualité et en quantité suffisante est essentielle, car elle influence directement l'efficacité du modèle d'apprentissage automatique.

La première étape dans le développement de notre modèle consiste à rassembler des données adéquates pour l'entraînement. Cette phase est le pilier de l'apprentissage automatique. Des erreurs telles que le choix de caractéristiques incorrectes ou la concentration sur des types limités d'entrées pour l'ensemble de données peuvent rendre le modèle totalement inefficace. [31]

– **Préparation des données** : une fois que les données de formation ont été collectées, nous passons à la préparation des données. Ici les données sont chargées dans un emplacement approprié et préparées pour l'entraînement du modèle. Pendant cette phase, nous traitons les données collectées à l'étape précédente, en les nettoyant pour supprimer les doublons, corriger les erreurs, gérer les valeurs manquantes et effectuer des conversions de types de données si nécessaire. De plus, la visualisation des données est souvent utilisée pour identifier les relations entre les différentes variables et détecter les déséquilibres potentiels. Une autre étape importante est la division des données en deux ensembles distincts : environ 80 % des données sont utilisées pour former le modèle, tandis que les 20 % restants servent à évaluer ses performances. Cette séparation est essentielle pour garantir une évaluation impartiale des performances du modèle dans des conditions réelles, car l'utilisation des mêmes données pour la formation et l'évaluation peut conduire à des résultats biaisés. [31]

– **Choisir un modèle** : Une fois les étapes centrées sur les données terminées, notre prochaine étape est de choisir le type de modèle à utiliser. Les data scientists ont développé différents modèles adaptés à diverses fins, chacun étant efficace pour modéliser les caractéristiques spécifiques des ensembles de données. Chaque modèle est élaboré en considérant des buts particuliers en tête. Par exemple, certains modèles sont plus adaptés au traitement des textes tandis qu'un autre modèle peut être mieux équipé pour gérer les images. [31]

– **Formation** : l'entraînement du modèle est au cœur du processus d'apprentissage automatique. C'est durant cette phase que se déroule la majorité de l'apprentissage. Cet entraînement exige de la persévérance et de la pratique et une compréhension approfondie du domaine d'application. L'entraînement peut être particulièrement gratifiant lorsque le modèle commence à bien fonctionner dans sa fonction.

Le processus de formation consiste à initialiser certaines valeurs aléatoires pour, par exemple, X et Y de notre modèle, à prédire le résultat avec ces valeurs, puis à le comparer avec la prédiction du modèle, puis à ajuster les valeurs afin qu'elles correspondent aux prédictions

faites précédemment. Ce processus se répète ensuite et chaque cycle de mise à jour est appelé une étape de formation. [31]

– **Évaluation** : après avoir été formé, le modèle doit être évalué pour déterminer s'il fonctionnerait efficacement dans des situations réelles. C'est pourquoi la partie de l'ensemble de données créée pour l'évaluation est utilisée pour vérifier la compétence du modèle. Permettant ainsi de tester le modèle dans des conditions qu'il n'a pas rencontrées pendant son entraînement. L'évaluation revêt une importance cruciale, car elle permet aux data scientistes de vérifier si les objectifs fixés ont été atteints. En cas de résultats insatisfaisants, les étapes précédentes doivent être réexaminées pour identifier et corriger les lacunes du modèle. [31]

– **Réglage des hyper paramètres** : après l'évaluation initiale, des améliorations supplémentaires peuvent être apportées à notre formation en ajustant les paramètres. Certains paramètres étaient implicites lors de la phase de formation initiale, comme le taux d'apprentissage, qui détermine l'ampleur du décalage à chaque étape en se basant sur les informations de l'étape de formation précédente. Ces valeurs influent sur la précision du modèle et la durée de la formation. [31]

– **Prédiction** : la phase finale du processus d'apprentissage automatique est la prédiction, marquant le moment où le modèle est considéré prêt pour une application pratique. C'est là que la véritable valeur de l'apprentissage automatique se manifeste. Ici, nous pouvons enfin utiliser notre modèle pour prédire le résultat de ce que nous voulons. [31]

5.1.1 Les types

La figure suivante montre un schéma synoptique des types d'apprentissage automatique.

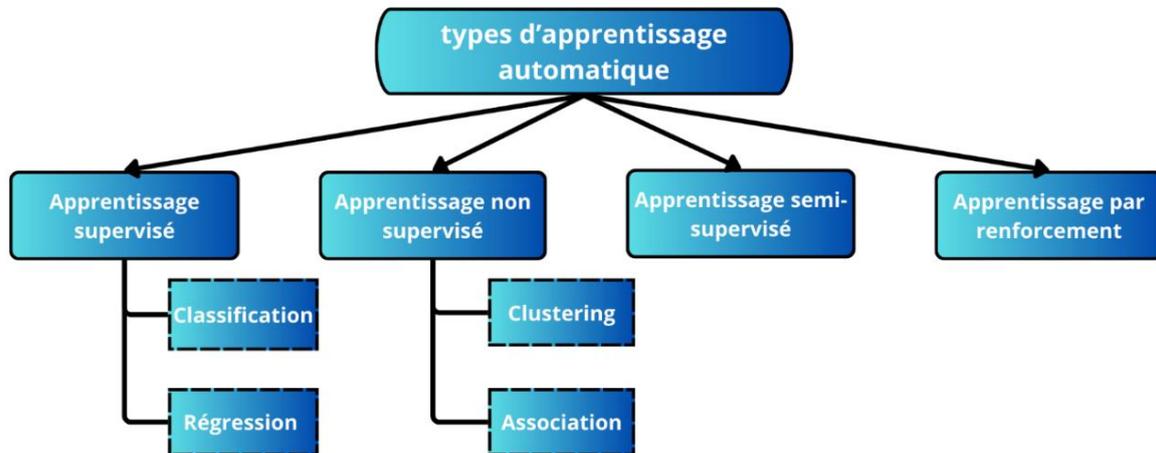


Figure 14 : les types d'apprentissage automatique.

5.1.1.1 Apprentissage supervisé

Les méthodes d'apprentissage supervisé sont parmi les plus répandues en apprentissage automatique. Leur objectif est de tirer parti d'un ensemble de données existant pour construire un modèle capable de prédire des données non étiquetées à venir. En fournissant à un algorithme d'apprentissage supervisé un ensemble de données d'entraînement, celui-ci est en mesure de développer un modèle à partir de ces données. Parmi les algorithmes couramment utilisés en apprentissage supervisé, on retrouve les réseaux neuronaux, les arbres de décision, la régression linéaire et les machines à vecteurs de support. Cette appellation, « apprentissage supervisé », indique que le processus d'apprentissage est guidé, où nous fournissons à l'algorithme des informations pour l'aider à apprendre. Les données étiquetées sont utilisées pour fournir les résultats attendus, tandis que les autres informations sont utilisées comme caractéristiques d'entrée. [32]

L'apprentissage supervisé peut être séparé en deux types de problèmes lors de l'exploration de données : la classification et la régression :

- **Classification**

La classification est le processus permettant de regrouper la sortie en différentes classes en fonction d'une ou plusieurs variables d'entrée. Pour ce faire, un modèle est entraîné sur des exemples étiquetés, afin d'apprendre les relations entre les caractéristiques d'entrée et les classes de sortie. Dans le contexte de la classification, la variable cible est une valeur catégorielle. Par exemple, on peut classer les e-mails comme étant du spam ou non.

L'objectif du modèle est de généraliser cet apprentissage pour effectuer des prédictions précises sur de nouvelles données invisibles.

Classification peut être séparée en deux types :

- **Classification binaire** : elle fait référence aux tâches de classification ayant deux étiquettes de classe telles que « vrai et faux » ou « oui et non ».
- **Classification multi classe** : traditionnellement, cela fait référence aux tâches de classification ayant plus de deux étiquettes de classe. [31]

- **Régression**

La régression est un type d'apprentissage supervisé dont le but est de prédire des valeurs numériques continues en fonction des caractéristiques d'entrée. Cela implique d'établir une relation entre la variable d'entrée et la variable de sortie. Il est utilisé lorsque la valeur de la variable de sortie est continue ou réelle. Par exemple, dans le contexte de la prévision des prix de l'immobilier, nous pouvons utiliser des caractéristiques telles que la taille de la maison, le nombre de chambres et l'emplacement.

L'objectif principal de la régression est de minimiser la différence entre les valeurs prédites et les valeurs réelles. Divers algorithmes peuvent être utilisés pour les tâches de régression, notamment la régression linéaire, les arbres de décision et les réseaux de neurones. Ces modèles visent à capturer les modèles sous-jacents dans les données, nous permettant ainsi de faire des prédictions précises. Le graphique suivant, « Figure 15 », nous montre la différence entre la classification et la régression. [31]

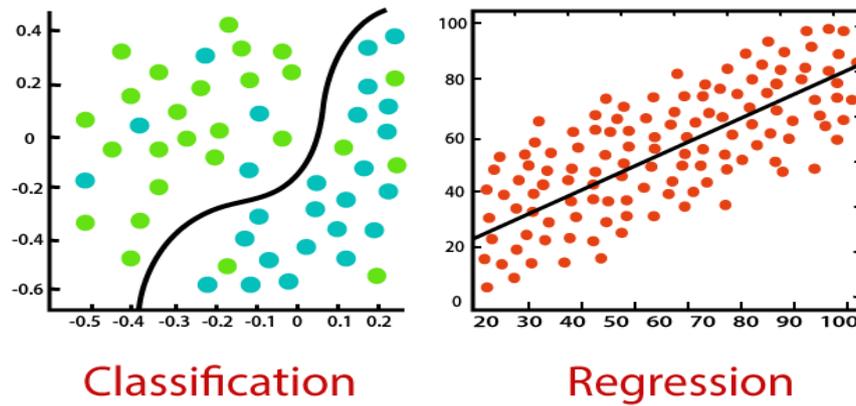


Figure 15 : la classification et la régression. [33]

5.1.1.2 Apprentissage non supervisé

L'apprentissage non supervisé implique l'utilisation d'algorithmes d'apprentissage automatique pour analyser et regrouper des ensembles de données non étiquetés. Ces algorithmes identifient des modèles inconnus dans les données sans intervention humaine, ce qui signifie qu'il y a des entrées sans sorties prédéterminées. Contrairement à l'apprentissage supervisé, dans l'apprentissage non supervisé, il n'est pas nécessaire de guider ou d'enseigner au modèle. Les principales tâches de l'apprentissage non supervisé comprennent le clustering, l'association et la réduction de dimensionnalité. [31]

Le clustering, une méthode d'apprentissage automatique non supervisée. Les clusters sont identifiés sur la base de mesures de similarité ou de différences, en fonction du modèle de machine utilisé. Les algorithmes de clustering sont conçus pour regrouper des données similaires en clusters distincts, permettant de découvrir des structures intrinsèques dans les données sans que les étiquettes de classe ou de catégorie antérieures soient fournies. [34]

L'association est une technique d'apprentissage automatique non supervisée qui emploie des règles spécifiques pour déceler des liens entre variables au sein d'un vaste corpus de données. Cette méthode d'association améliore l'efficacité des stratégies de commercialisation. Par exemple, les magasins commerciaux utilisent des algorithmes basés sur cette technique pour découvrir la relation entre la vente d'un produit et les ventes d'un autre en fonction du comportement du client. [31]

5.1.1.3 Apprentissage semi-supervisé

L'apprentissage semi-supervisé est un type d'apprentissage automatique qui se situe entre l'apprentissage supervisé et non supervisé. Il utilise à la fois des données étiquetées en petite quantité et des données non étiquetées en grande quantité pour former un modèle. Son objectif

est de développer une fonction capable de prédire précisément la variable de sortie en se basant sur les variables d'entrée. Contrairement à l'apprentissage supervisé, l'algorithme est formé sur un ensemble de données contenant à la fois des données étiquetées et non étiquetées. Cette approche s'avère particulièrement bénéfique lorsque de nombreuses données non étiquetées sont disponibles, mais que l'étiquetage complet de toutes ces données est trop coûteux ou difficile. [35]

5.1.1.4 Apprentissage par renforcement

L'apprentissage par renforcement est un type d'algorithme d'apprentissage automatique qui permet aux agents logiciels et aux machines d'évaluer automatiquement le comportement optimal dans un contexte ou un environnement particulier pour améliorer son efficacité. Ce type d'apprentissage repose sur le système de récompense ou de pénalité, où l'objectif ultime est d'optimiser les actions en fonction des rétroactions reçues. Il s'agit d'un outil puissant pour former des modèles d'IA qui peuvent aider à accroître l'automatisation ou à optimiser l'efficacité opérationnelle de systèmes sophistiqués tels que la robotique, les tâches de conduite autonome, la logistique de fabrication et de chaîne d'approvisionnement. Cependant, il n'est pas toujours approprié pour résoudre des problèmes simples. [36]

5.1.2 Les algorithmes d'apprentissage automatique supervisé

L'apprentissage supervisé, une branche de l'intelligence artificielle, se décline en diverses catégories, chacune possédant des attributs et des domaines d'application spécifiques. Voici une sélection des variantes les plus répandues d'algorithmes d'apprentissage supervisé :

- **Réseaux neuronaux**

Les réseaux de neurones artificiels sont au cœur de ce que l'on appelle désormais l'intelligence artificielle et jouent un rôle crucial dans l'avancement de la technologie d'apprentissage automatique et simulent la complexité du cerveau humain en utilisant des couches interconnectées de nœuds. Chaque nœud est composé d'entrées, de poids, d'un biais (seuil) et d'une sortie. Lorsque la sortie franchit un certain seuil, le nœud s'active, envoyant l'information à la prochaine couche. Ces réseaux s'entraînent à cartographier les données grâce à l'apprentissage supervisé, en affinant leurs paramètres par la minimisation de la fonction de coût via la descente de gradient. Un coût minimal ou nul indique une haute précision du modèle. [34]

- **Naïve Bayes (NB)**

L'algorithme Naïve Bayes, grâce à sa simplicité et son efficacité, s'est avéré être un outil puissant dans le domaine de l'apprentissage automatique pour la classification des données. L'algorithme de Bayes naïf repose sur le théorème de Bayes et suppose que les caractéristiques sont indépendantes les unes des autres. Il est efficace et adapté aux classifications binaires et multi-classes dans divers contextes pratiques. On distingue trois variantes de classificateurs bayésiens naïfs : multinomial, Bernoulli et gaussien. Cette technique est couramment employée pour la classification de textes, la détection de pourriels et les systèmes de recommandation. Un atout majeur de cette méthode est sa capacité à déduire les paramètres nécessaires avec un volume réduit de données d'entraînement, comparativement aux méthodes plus complexes. Cependant, grâce à sa facilité de mise en œuvre et à son efficacité informatique, Naïve Bayes reste un choix populaire pour les projets nécessitant une solution rapide et relativement précise pour la classification des données. [34]

- **Régression linéaire**

La régression linéaire est une technique statistique de base utilisée pour analyser et préparer la relation entre des variables quantitatives (d'entrée et de sortie). C'est une méthode simple mais largement utilisée dans le domaine de l'apprentissage supervisé, où la prédiction de la sortie se fait en combinant de manière pondérée les caractéristiques d'entrée. C'est la simplicité de ce modèle qui en fait son succès auprès des data scientists. En effet, il nécessite peu de paramètres et peut être facilement représenté graphiquement, ce qui le rend également compréhensible pour les décideurs. [34]

- **Régression logistique**

La régression logistique est un modèle statistique répandu fondé sur la probabilité, principalement utilisé pour relever les défis de classification dans le domaine de l'apprentissage automatique. Cette approche exploite typiquement une fonction logistique pour calculer des estimations de probabilité. Bien qu'elle puisse conduire à un surajustement pour les jeux de données à dimensions élevées, elle est efficace lorsque les données peuvent être distinguées de manière linéaire. Bien que la régression logistique soit suffisamment polyvalente pour s'attaquer à la fois aux tâches de classification et de régression, sa principale application se situe dans le domaine de la classification. [37]

- **L'arbre de décision**

La méthode de l'arbre de décision est une technique d'analyse prédictive et d'apprentissage supervisé visualisée sous la forme d'un diagramme arborescent. Pour mettre en œuvre cet

algorithme, il est essentiel de disposer de jeux de données dédiés à l'entraînement et au test. Cet algorithme facilite la prise de décisions en établissant des règles pour déterminer la classe ou l'attribut cible. Les branches de l'arbre symbolisent les décisions à prendre ou les tests à réaliser, et les feuilles représentent les prédictions ou les résultats finaux.

Un arbre de décision est un type d'algorithme utilisé à la fois pour les tâches de classification et de régression.

– **Régression des arbres de décision** : les arbres de décision peuvent être utilisés pour des tâches de régression en prédisant la valeur liée à un nœud feuille.

– **Classification des arbres de décision** : Forêt aléatoire est un algorithme d'apprentissage automatique qui utilise plusieurs arbres de décision pour améliorer la classification et éviter le sur apprentissage. [34]

- **Forêts aléatoires (Random Forest)**

L'algorithme forêt aléatoire est une technique robuste et efficace en apprentissage automatique. Il s'agit de l'une des méthodes d'apprentissage les plus populaires couramment utilisées pour l'exploration de données. Il se distingue par sa capacité à créer plusieurs arbres de décision lors de la phase d'apprentissage, chacun utilisant un sous-ensemble aléatoire des données et des caractéristiques. C'est l'un des algorithmes d'apprentissage les plus précis disponibles. Lors de la phase de prédiction, les résultats de chaque arbre sont combinés, soit par un vote majoritaire pour les tâches de classification, soit par une moyenne pour les tâches de régression. Cette agrégation des prédictions provenant de multiples arbres permet d'obtenir des résultats robustes et précis. La forêt aléatoire dispose d'une méthode efficace pour estimer les données manquantes et maintient l'exactitude lorsqu'une grande proportion de données est manquante.

Les forêts aléatoires sont largement utilisées pour leurs performances supérieures en classification et en régression. Elles sont appréciées pour leur capacité à gérer des ensembles de données de grande taille, leur adaptabilité à différents types de problèmes, ce qui en fait un choix populaire parmi les praticiens de l'apprentissage automatique. [38]

- **Machine à vecteurs de support (SVM)**

La machine à vecteurs de support (SVM) est un modèle d'apprentissage supervisé largement utilisé, développé par Vladimir Vapnik. Il est couramment utilisé pour la classification et la

régression des données. L'algorithme vise à maximiser les distances entre la ligne de séparation et les différents échantillons de données, connus sous le nom de vecteurs supports.

En essence, l'objectif est de trouver l'hyperplan optimal, celui qui sépare parfaitement les différentes classes de données pour minimiser les erreurs de classification. Le SVM offre de nombreuses applications dans une large gamme de domaines et d'options de flexibilité. Pour l'efficacité, la notation en classification des images et la détection des spams. Cependant, lorsque les données deviennent plus complexes, la tâche de classification peut devenir plus difficile. Dans de tels cas, l'algorithme SVM peut ne pas être le plus approprié, car il peut avoir du mal à distinguer efficacement les différentes catégories. [34]

- **K- Voisins les plus proches (K-Nearest Neighbours)**

L'algorithme de classification K-NN est un algorithme très important en apprentissage automatique pour la classification des supervisés. KNN fonctionne en identifiant les k exemples de formation les plus proches d'une entrée donnée et prédit ensuite la classe ou la valeur en fonction de la classe majoritaire ou de la valeur moyenne de ces voisins. L'efficacité de KNN dépend de la sélection de la métrique k et de la distance appropriée. L'algorithme K-NN peut être utilisé pour des problèmes de classification binaire ou multiclasse. Un (KNN) est un type d'algorithme utilisé à la fois pour les tâches de classification et de régression.

- **Régression des K-voisins les plus proches** : elle prédit les valeurs continues en faisant la moyenne des sorties des k voisins les plus proches.

- **Classification des K-voisins les plus proches** : les points de données sont classés en fonction de la classe majoritaire de leurs k voisins les plus proches. [34]

- **Augmentation du dégradé (Gradient Boosting)**

L'augmentation du dégradé est une méthode d'apprentissage automatique largement utilisée pour les modèles de classification et d'association. Cette méthode améliore progressivement la précision en ajoutant séquentiellement des modèles qui corrigent les erreurs des itérations antérieures. Il s'agit d'un type d'apprentissage d'ensemble qui consiste à combiner plusieurs modèles faibles pour créer un modèle fort. En se concentrant sur les erreurs résiduelles des prédictions précédentes, chaque modèle additionnel affine la prédiction globale, résultant en un algorithme performant apte à interpréter des structures de données complexes. Une augmentation du dégradé est un type d'algorithme utilisé à la fois pour les tâches de classification et de régression.

- **Régression boostée par gradient** : elle crée un ensemble d'apprenants faibles pour améliorer la précision des prédictions grâce à une formation itérative.
- **Classification d'amplification de gradient** : crée un groupe de classificateurs pour améliorer continuellement la précision des prédictions par le biais d'itérations. [39]

5.1.3 Le Différence entre apprentissage supervisé et non supervisé

L'apprentissage supervisé et non supervisé sont les deux techniques d'apprentissage automatique. Mais les deux techniques sont utilisées dans différents scénarios et avec différents ensembles de données. Le tableau 4 résume la différence entre apprentissage supervisé et non supervisé.

Tableau 4 : La différence entre apprentissage supervisé et non supervisé. [40]

Apprentissage supervisé	Apprentissage non supervisé
- Les algorithmes d'apprentissage supervisé sont entraînés à l'aide de données étiquetées.	- Les algorithmes d'apprentissage non supervisé sont entraînés à l'aide de données non étiquetées.
- Le modèle utilise des données étiquetées pour apprendre à prédire des résultats corrects. Ces résultats corrects servent de référence pour évaluer la performance du modèle.	- Le modèle n'utilise pas de retour direct pour évaluer ses prédictions ou ajuster son comportement.
- Le modèle utilise les données d'entraînement pour apprendre à prédire correctement les résultats.	- Le modèle cherche à découvrir les structures, ou les regroupements cachés dans les données sans avoir besoin d'étiquettes préalables.
- Les données d'entrée sont présentées au modèle avec les données de sortie correspondantes.	- Seules les données d'entrée sont fournies au modèle.
- L'objectif de l'apprentissage supervisé est de permettre au modèle d'apprendre à prédire les résultats corrects en utilisant les données	- Le but de l'apprentissage non supervisé est de trouver les modèles cachés et les informations utiles à partir d'un ensemble de

d'entraînement.	données inconnu.
– L'apprentissage supervisé nécessite une supervision afin de guider l'entraînement du modèle.	– L'apprentissage non supervisé, il n'y a pas besoin de supervision pour entraîner le modèle.
– L'apprentissage supervisé peut être divisé en problèmes de classification et de régression.	– L'apprentissage non supervisé peut être divisé en problèmes de clustering et d'associations.
– Le modèle d'apprentissage supervisé génère des prédictions précises.	– Le modèle d'apprentissage non supervisé peut parfois fournir des résultats moins précis que ceux obtenus avec l'apprentissage supervisé.
– L'apprentissage supervisé ne représente pas pleinement la véritable intelligence artificielle, car il implique d'entraîner le modèle avec des données étiquetées avant qu'il puisse faire des prédictions précises.	– L'apprentissage non supervisé est souvent considéré comme plus proche de la véritable intelligence artificielle, car il apprend de la même manière qu'un enfant apprend les choses de la routine quotidienne grâce à ses expériences.

6. Conclusion

Dans ce chapitre, nous avons parlé de manière générale sur les principes et concepts de base de l'intelligence artificielle. À travers une analyse détaillée de sa définition, de ses différents types et de ses techniques, nous avons pu saisir la complexité et l'ampleur de cette discipline en évolution.

chapitre 3 : Implémentation et Résultats

1. Introduction

La technologie de positionnement à l'aide du réseau GSM est considérée comme l'un des points clés les plus importants. Dans ce chapitre, nous nous concentrerons sur la présentation des étapes de base pour créer et évaluer un modèle, depuis la collecte et le nettoyage des données, jusqu'à l'apprentissage et le test avec les données. Nous utiliserons l'algorithme Random Forest Regressor, qui établit un équilibre parfait entre précision et efficacité, ce qui le rend idéal pour analyser le Big Data et fournir des estimations de localisation précises.

2. Les outils de développement

2.1 Google Colab

Google Colab, également connu sous le nom de Colaboratory, est un service cloud gratuit proposé par Google, et un environnement de bloc-notes Jupyter basé sur le cloud. Il s'exécute dans votre navigateur Web et permet à toute personne ayant accès à Internet d'expérimenter l'apprentissage automatique et le codage pour l'intelligence artificielle. Vous pouvez écrire et exécuter du code Python, partager votre code et le modifier simultanément avec d'autres membres de l'équipe, et tout documenter en le combinant dans un seul bloc-notes avec du texte, des graphiques, des images. [41]



Figure 16 : Logo de Google Colab.

2.2 Python

Python est le langage de programmation informatique le plus largement adopté, particulièrement dans des domaines comme la Data Science et l'apprentissage automatique. L'une des grandes forces de Python est qu'il est livré avec une bibliothèque standard très complète qui nous permet de faire des choses comme télécharger un fichier depuis Internet. Sa compatibilité multiplateforme, fonctionnant sur différents systèmes d'exploitation tels que

Windows, en fait un choix polyvalent. Contrairement à HTML, CSS ou JavaScript, son usage n'est donc pas limité au développement Web. Python est utilisé dans divers contextes, que ce soit pour le développement back-end d'applications web ou mobiles, la création de logiciels pour PC ou même l'écriture de scripts système pour automatiser des tâches informatiques.

De plus, Python domine le traitement mégadonnées, les calculs mathématiques et l'apprentissage automatique, en raison de sa syntaxe claire et intuitive qui facilite l'apprentissage pour les débutants et améliore la productivité des développeurs expérimentés. Son écosystème de bibliothèques spécialisées contribue à son efficacité dans des domaines tels que l'analyse de données et le développement Web. Python est largement adopté par une variété d'acteurs, des développeurs individuels aux grandes entreprises technologiques, en raison de sa facilité d'utilisation et de sa capacité à répondre à une grande diversité de besoins en programmation et en développement logiciel. [42]

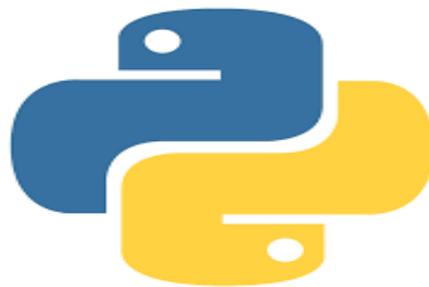


Figure 17 : logo de python.

2.3 Bibliothèques utilisées

2.3.1 Pandas

Pandas est une bibliothèque Python Open Source pour l'analyse de données. Il donne à Python la possibilité de travailler avec des données. Alors que les outils d'apprentissage automatique et de visualisation captivent souvent l'attention, mais le package Pandas constitue l'épine dorsale de la plupart des projets de données. Pour donner à Python ces fonctionnalités améliorées, Pandas introduit deux nouveaux types de données dans Python : Series et DataFrame. Le DataFrame représentera l'intégralité de votre feuille de calcul ou vos données rectangulaires, tandis que la série est une seule colonne du DataFrame. Un Pandas DataFrame peut également être considéré comme un dictionnaire ou une collection de séries.

Pandas est construit sur le package NumPy, donc une grande partie de la structure de NumPy est utilisée dans Pandas. Les données Pandas sont souvent utilisées pour l'analyse statistique dans SciPy, les visualisations avec Matplotlib et les algorithmes d'apprentissage automatique avec Scikit-learn.

Les Notebooks Jupyter fournissent un environnement idéal pour explorer et modéliser les données avec Pandas. Ils permettent d'exécuter du code dans des cellules spécifiques, ce qui est très efficace pour travailler avec de gros ensembles de données et des transformations complexes. Pandas est une bibliothèque Python Open Source pour l'analyse de données. Il donne à Python la possibilité de travailler avec des données. [43]

2.3.2 Scikit-learn

Scikit-learn, également connu sous le nom de Sklearn, représente la bibliothèque de référence en matière de machine learning en Python. Elle offre une gamme complète d'outils pour l'apprentissage automatique et la modélisation statistique, couvrant des tâches telles que la classification, la régression et le regroupement, le tout via une interface cohérente en Python. Cette bibliothèque, largement rédigée en Python, repose sur des piliers essentiels comme NumPy, SciPy et Matplotlib. Scikit-learn est un choix de prédilection dans de nombreux domaines, servant d'outil principal pour développer des modèles prédictifs. Il est également très simple à utiliser et relativement exempt de bugs de programmation.[44]

3. La collection de données

Afin d'implémenter et d'évaluer un modèle d'apprentissage automatique pour localiser des appareils, nous utiliserons la collection de données proposée par l'application Tower Collector. Tower Collector est une application conçue pour recueillir les coordonnées GPS des tours de téléphonie cellulaire GSM/UMTS/LTE/CDMA. L'objectif principal de cette application est de contribuer à la cartographie de la couverture du réseau de téléphonie mobile en collectant des informations sur l'emplacement des tours de téléphonie cellulaire.

Nous avons d'abord installé l'application Tower Collector sur notre téléphone portable. Une fois l'installation terminée, nous l'avons activé et avons commencé à nous déplacer dans l'université (Belhadj Bouchaib). Cette application nous a permis de collecter des données de base sur les tours de téléphonie cellulaire environnantes telles que (Cell ID, LAC, MCC, MNC, force du signal, latitude, longitude...). Comme le montre la figure suivante « figure 18 ». Tower Collector facilite le processus de collecte en enregistrant efficacement les données au fur et à mesure que nous nous déplaçons.

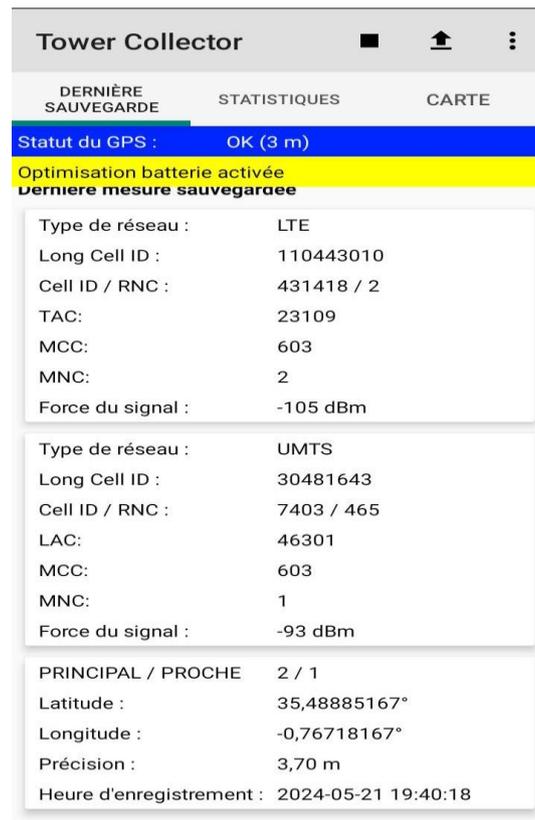
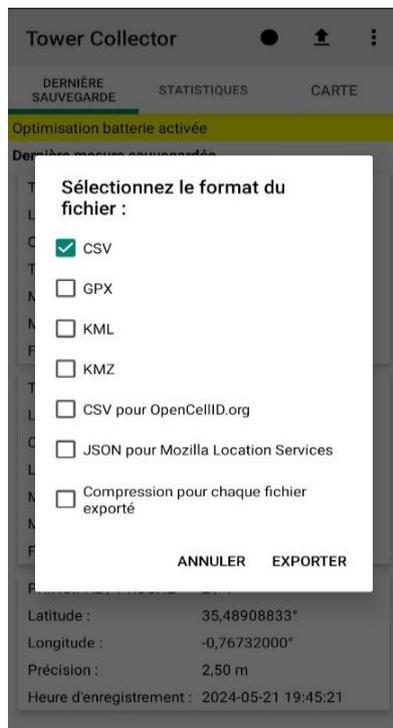


Figure 18: Capture d'écran de l'application.

Nous avons ensuite enregistré et stocké les données tabulaires dans un fichier CSV. Comme le montre la figure suivante, figure 4.



altitude

	A	B	C	D	E	F
1	latitude	longitude	mcc	mnc	lac	cell_id
2	35.29126833	-1.122991667	603	1	46301	30491702
3	35.29126833	-1.122991667	603	2	24603	114212608
4	35.29283333	-1.121561667	603	1	46301	30491702
5	35.29283333	-1.121561667	603	2	24603	114212608
6	35.293165	-1.12105	603	1	46301	30491702
7	35.293165	-1.12105	603	2	24603	114212608
8	35.29327833	-1.12115	603	1	46301	30491702
9	35.29327833	-1.12115	603	2	5602	187913
10	35.29336667	-1.1211	603	1	46301	30481702

Figure 19 : Enregistrez et stockez les données tabulaires dans un fichier CSV.

En utilisant l'application Tower Collector, nous avons recueilli des données. Nous voulions enregistrer des données relatives aux tours de communication à l'intérieur des limites de l'université. Cependant, il nous a fourni certaines données provenant d'emplacements extérieurs à l'université. Nous avons donc soigneusement examiné les données pour identifier les points situés en dehors du périmètre cible. Ensuite, nous avons supprimé ces données inutiles. Cette étape était nécessaire pour garantir l'exactitude et la validité de notre étude. La figure 20 illustre les emplacements au sein de l'Université.



Figure 20 : Quelques points des données collectés.

Après avoir obtenu la base de données, nous créerons un modèle d'apprentissage automatique pour le positionnement à l'aide du réseau GSM.

La figure 21 montre les bibliothèques utilisées. Nous avons utilisé plusieurs bibliothèques spécialisées pour le traitement des données et la modélisation prédictive. Nous avons employé **Pandas** pour la manipulation et l'analyse des données en format tabulaire, offrant des fonctionnalités robustes pour le nettoyage et la préparation des données. Pour diviser nos données en ensembles d'entraînement et de test, nous avons utilisé **train_test_split** de la bibliothèque **Scikit-learn**.

La transformation des colonnes a été gérée par **Column Transformer**, permettant l'application sélective de différentes transformations sur les colonnes pertinentes. Dans ce contexte, **Standard Scaler** a été utilisé pour la normalisation des caractéristiques numériques, tandis que **One Hot Encoder** a été appliqué aux caractéristiques catégorielles pour les convertir en un format utilisable par les algorithmes de machine Learning.

Pour structurer ces étapes de prétraitement et de modélisation de manière systématique et reproductible, nous avons mis en place des **Pipelines**. Ces pipelines facilitent l'enchaînement des transformations et l'entraînement des modèles. Nous avons utilisé un **Random Forest Regressor** pour ses capacités à gérer des relations non linéaires et à réduire le risque de surapprentissage.

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.compose import ColumnTransformer
from sklearn.preprocessing import StandardScaler, OneHotEncoder
from sklearn.pipeline import Pipeline
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_absolute_error
from sklearn.neural_network import MLPRegressor
```

Figure 21 : Importation des bibliothèques.

La figure 22 montre la sélection des entrées (Features) et la définition de la cible (Target) dans le code fourni. Elle démontre l'importance d'une préparation minutieuse des données dans les projets d'apprentissage automatique lorsque nous définissons l'ensemble des variables qui seront utilisées comme entrée pour entraîner le modèle. Elle est représentée par ['mcc', 'mnc', 'lac', 'cell_id' ...]. D'autre part, cible fait référence à la variable de sortie que le modèle tente de prédire en fonction des variables d'entrées disponibles et qui consiste en des coordonnées de latitude et de longitude.

- Features : variables d'entrée utilisées pour la prédiction, incluses dans la variable X.
- Target : La variable de sortie à prédire, représentée par la variable y.

```
# Define features and target
features = ['mcc', 'mnc', 'lac', 'cell_id', 'rnc', 'psc', 'asu', 'dbm', 'accuracy', 'speed', 'altitude',
X = data[features]
y = data[['latitude', 'longitude']]
```

Figure 22: Définir les fonctionnalités et la cible.

Cet ensemble contient beaucoup de données manquantes, nous devons donc le nettoyer pour garantir la qualité de l'ensemble de données avant d'entraîner le modèle d'apprentissage automatique. Nous remplissons toutes les valeurs manquantes avec zéro ('0'). Comme le montre la figure suivante « figure 23 ».

```
# Handle missing values if necessary
# For simplicity, filling numerical missing values with 0, which might not be the best approach.
X.fillna(0, inplace=True)
```

Figure 23 : Remplacement des valeurs manquantes.

Nous divisons les données en deux parties à l'aide de la fonction **train_test_split** de la bibliothèque Scikit-learn. Cette fonction divise aléatoirement l'ensemble des données en deux sous-ensembles : un pour entraîner le modèle et un pour le tester, avec 80 % des données utilisées pour l'entraînement (X_train et Y_train) et 20 % pour les tests (X_test et Y_test). Le paramètre **random_state** est défini sur 42 pour garantir la reproductibilité. Comme le montre la figure ci-dessous « figure 24 ».

```
# Split the data into training and test sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

Figure 24 : Division des données en ensembles de formation et de test.

4. Traitement des données

Après avoir divisé les données, nous définirons les opérations de traitement des données pour les types numériques et catégoriels, comme le montre la figure ci-dessous « figure 25 ». Nous

créons une liste appelée « numerical_features », qui contient les noms des caractéristiques numériques de l'ensemble de données. Ensuite, un objet appelé « numerical_transformer » est créé à l'aide de la classe « Standard Scaler » de la bibliothèque " Scikit-learn". Standard Scaler est une technique courante pour traiter les données et optimiser les fonctionnalités numériques avant de les utiliser dans un modèle d'apprentissage automatique. Il convertit également les données de sorte que la valeur résultante pour chaque fonctionnalité ait une valeur moyenne de "0" et un écart type de "1". Pour définir des fonctionnalités catégorielles qui contiennent des valeurs catégorielles telles que la classe 'net_type', nous utilisons "One Hot Encoder" qui convertit les variables catégorielles en variables binaires. One Hot Encoder crée deux colonnes, l'une représentant « LTE » avec une valeur de "0" et l'autre avec « UMTS » avec une valeur de "1".

Nous utilisons "handle_unknown=ignore" pour ignorer toutes les valeurs inconnues pouvant apparaître pendant le processus de conversion."Categories=auto" détermine automatiquement les catégories à l'aide des données fournies, ce qui facilite le processus de conversion.

```
# Preprocessing for numerical data
numerical_features = ['mcc', 'mnc', 'lac', 'cell_id', 'rnc', 'psc', 'asu', 'dbm', 'accuracy',
numerical_transformer = StandardScaler()

# Preprocessing for categorical data
#LTE=0, UMTS = 1
categorical_features = ['net_type']
#categorical_transformer = OneHotEncoder(handle_unknown='ignore')
categorical_transformer = OneHotEncoder(handle_unknown='ignore', categories='auto')
```

Figure 25 : Traitement des caractéristiques numériques et catégorielles.

5. L'Algorithme Random Forest Regression

La régression de forêt aléatoire est une technique d'apprentissage automatique polyvalente pour prédire des valeurs numériques capables d'effectuer à la fois des tâches de régression et de classification à l'aide de nombreux arbres de décision multiples. Il s'agit d'un type particulier de technique d'apprentissage en groupe qui combine les prédictions de divers modèles pour réduire les dépassements et améliorer la précision. Des données et des caractéristiques sélectionnées au hasard sont utilisées avec substitution pour entraîner chaque arbre de décision dans une forêt aléatoire et la moyenne de toutes les prédictions d'arbres est

utilisée pour obtenir la prédiction finale du déclin aléatoire de la forêt. Les bibliothèques d'apprentissage automatique de Python facilitent et optimisent la mise en œuvre de cette approche. [55]

Déterminer le nombre d'arbres dans la forêt et ajuster l'état d'aléatoire est important pour la performance de la modèle, donc nous définissons les paramètres « n_estimators » : ce paramètre détermine pour nous le nombre d'arbres qui seront dans la forêt, qui est défini sur 100 arbres, et le deuxième paramètre "random_state" qui détermine pour nous la graine utilisée pour itérer les résultats. Lorsqu'une valeur spécifique est définie pour la graine, l'utilisateur obtiendra les mêmes résultats à chaque exécution.

Après cela, nous avons utilisé le pipeline de la bibliothèque "Scikit-learn", qui comprend deux étapes : la première étape est appelée « préprocesseur », qui prépare les données, tandis que la deuxième étape est appelée « modèle », qui contient le modèle utilisé pour s'entraîner. Pipeline effectue une série d'étapes sur les données, intégrant la préparation des données et la formation des modèles en une seule étape pour accélérer le développement et le test des modèles prédictifs. Dans la figure suivante, « figure 26 », nous montrons l'algorithme de régression de forêt aléatoire.

```
# Define the model
model = RandomForestRegressor(n_estimators=100, random_state=42)
#model = MLPRegressor(hidden_layer_sizes=(100, 50), activation='r

# Bundle preprocessing and modeling code in a pipeline
clf = Pipeline(steps=[('preprocessor', preprocessor),
                      ('model', model)])
```

Figure 26 : Définir le Modèle.

La figure 27 représente le processus de formation du modèle et d'évaluation de ses performances dans le développement de modèles prédictifs. Dans l'étape de formation du modèle, nous utilisons les données de formation X_train et Y_train.

« X_train » représente les variables d'entrée, tandis que « y_train » représente la variable que nous voulons prédire. Après entraînement, le modèle prédit les valeurs sur les données expérimentales X_test. Les valeurs prédites sont stockées dans Y_pred. Les performances du

modèle sont évaluées à l'aide de la métrique Mean Absolute Error (MAE), qui mesure la similitude des valeurs prédites avec les valeurs réelles.

```
# Train the model
clf.fit(X_train, y_train)

# Predict on test data
y_pred = clf.predict(X_test)

# Evaluate the model
print(f"Mean Absolute Error: {mean_absolute_error(y_test, y_pred)}")
```

Figure 27 : Formation et évaluation du modèle.

6. Résultats

Afin d'évaluer notre modèle, nous avons créé une nouvelle base de données de test qui contient 571 enregistrements. Nous avons utilisé le modèle entraîné pour prédire la latitude et la longitude attendues à l'aide d'une fonction de prédiction (**predict**), les prédictions sont ensuite enregistrées dans un fichier CSV. La figure 28 montre la prédiction pour un seul record.

```
# 35.292273452505, -1.124959615991
new_input_data = {
    'mcc': [603],
    'mnc': [1],
    'lac': [46301],
    'cell_id': [30480991],
    'rnc': [465],
    'psc': [66],
    'asu': [27],
    'dbm': [-59],
    'accuracy': [3],
    'speed': [0],
    'altitude': [308],
    'net_type': ["UMTS"],
    'rsrp': [0],
    'rsrq': [0],
    'rssi': [0],
    'rssnr': [0],
    'ec_no': [0]
}

# Convert to DataFrame
new_input_df = pd.DataFrame(new_input_data)

### Step 5: Make Predictions with the Trained Model

# Make predictions
new_pred = clf.predict(new_input_df)

# Output the prediction
print(f"Predicted Latitude: {new_pred[0][0]}, Predicted Longitude: {new_pred[0][1]}")
```

Figure 28 : Prédiction pour un seul record.

Nous avons téléchargé le fichier CSV « filtered_testdata4.csv » contenant les données de test, puis l'avons converti en « data frame » à l'aide de la bibliothèque Pandas. Ensuite, nous prétraiterons les données d'entrée en utilisant les mêmes étapes que celles appliquées lors de l'entraînement. Le modèle entraîné (clf) est utilisé pour faire des prédictions sur les données traitées. Les prédictions sont organisées dans un nouveau bloc de données contenant des coordonnées de latitude (predicted_latitude) et de longitude (predicted_longitude) et exportées vers un fichier CSV. Comme le montre la figure ci-dessous, « figure 29 ».

```
# Load the CSV file containing the data points you want to test
test_data_path = '../content/sample_data/filtered_testdata4.csv'
test_data = pd.read_csv(test_data_path)
test_data_df = pd.DataFrame(test_data)

# Preprocess the input data using the same preprocessing steps applied during training

preprocessed_test_data = preprocessor.transform(test_data_df)
preprocessed_test_data = preprocessed_test_data[:, 1:]

# Use the trained model to make predictions on the preprocessed input data
predictions = clf.predict(test_data_df)

# Output the predictions

predictions_df = pd.DataFrame({
    'cell_id': test_data_df['cell_id'],
    'rnc': test_data_df['rnc'],
    'predicted_latitude': predictions[:, 0],
    'predicted_longitude': predictions[:, 1]
})

# Export predictions to a CSV file
predictions_df.to_csv('predictions2.csv', index=False)

# Output the predictions along with cell_id and rnc
for i, (prediction, cell_id, rnc) in enumerate(zip(predictions, test_data_df['cell_id'], test_data_df['rnc'])):
    print(f'Data Point {i+1}: {prediction[0]},{prediction[1]}, Cell ID: {cell_id}, RNC: {rnc}')
```

Figure 29 : Phase de test des données.

Après avoir obtenu les prédictions de 571 enregistrements, nous les avons comparées aux positions GPS déterminées par l'application Tower Collector. Le pourcentage d'enregistrements était le suivant :

Tableau 5 : pourcentage d'enregistrements.

Erreur de distance en mètre	Nombre d'enregistrements
Moins de 10 mètres	396
Moins de 20 mètres	10
Moins de 30 mètres	13
Moins de 40 mètres	9
Moins de 100 mètres	34
Moins de 200 mètres	79
Plus de 200 mètres	30

Nous concluons que notre modèle d'apprentissage automatique fonctionne bien. Sur les 571 enregistrements, nous avons reçu 396 enregistrements dans lesquels les emplacements estimés correspondaient aux emplacements réels avec une précision inférieure à 10 mètres. Ces résultats positifs indiquent une grande efficacité dans l'identification des emplacements à proximité. D'autre part, le modèle nous a montré qu'il existe certains résultats dans lesquels les différences entre les emplacements estimés et réels dépassent plus de 100 mètres. Cette différence pourrait être le résultat de divers facteurs.

Nous comparerons nos résultats avec d'autres techniques de détermination des emplacements utilisant le réseau GSM que nous avons mentionnées dans le premier chapitre. Comme le montre le « tableau 6 ».

Les techniques de positionnement traditionnelles via le système GSM reposent souvent sur la triangulation, où le taux de précision dans les zones urbaines se situe à quelques centaines de mètres en raison de la densité de population, tandis que dans les zones rurales, il atteint plusieurs kilomètres, mais il peut souffrir d'imprécision en raison des réflexions du signal ou d'autres obstacles. Une autre méthode courante consiste à déterminer les emplacements en fonction de l'identification de la cellule, où un emplacement est déterminé en identifiant la cellule à laquelle appartient l'antenne utilisée pour la communication. Cette technologie peut fournir des estimations approximatives de l'emplacement, mais peut avoir des difficultés de précision, en particulier en milieu rural. Dans les zones urbaines densément peuplées, l'emplacement est déterminé de plusieurs centaines de mètres à plusieurs kilomètres. Quant à la technique d'empreinte digitale de force du signal (Signal Strength Fingerprinting), elle compare les caractéristiques du signal reçu avec celles de la base de données pour estimer l'emplacement de l'appareil, mais elle nécessite de collecter et de conserver des données volumineuses pour obtenir une meilleure précision que la simple triangulation ou les méthodes basées sur la cellule.

Comparé aux techniques de positionnement GSM traditionnelles, notre modèle basé sur l'apprentissage automatique semble plus précis, surtout à moins de 100 mètres.

Tableau 6 : Comparaison de la précision de différentes techniques de positionnement à l'aide des données du réseau GSM

Technique	Précision
Apprentissage automatique	-Moins de 10 mètres : 396 enregistrements -Moins de 20 mètres : 10 enregistrements -Moins de 30 mètres : 13 enregistrements -Moins de 40 mètres : 9 enregistrements -Moins de 100 mètres : 34 enregistrements -Plus de 100 mètres : 109 enregistrements
Triangulation	-Généralement à quelques centaines de mètres, selon la densité de la population
Positionnement basé sur l'ID de cellule	-Souvent à plusieurs centaines de mètres ou plusieurs kilomètres selon la densité des tours de téléphonie cellulaire.
Empreinte digitale de force du signal	- La précision peut être atteinte à quelques dizaines de mètres.

7. Conclusion

À la fin de ce chapitre, nous avons présenté les différentes étapes de création d'un modèle. Déterminer des emplacements à l'aide du réseau GSM. Nous avons commencé par mettre en œuvre des étapes de prétraitement sur la collecte des données. Suivi de la création de groupes d'entraînement et de test.

En analysant les résultats et en les comparant avec d'autres technologies, nous avons constaté que le modèle que nous avons créé offre des performances de positionnement encourageantes.

Conclusion générale

Les techniques de positionnement utilisant le réseau GSM jouent un rôle crucial dans divers domaines, allant de la sécurité publique à la navigation personnelle et aux applications commerciales. Les principales méthodes incluent le Time of Arrival (TOA), le Time Difference of Arrival (TDOA), la mesure de la puissance du signal reçu (RSSI) et la triangulation.

Ces techniques présentent plusieurs avantages. Tout d'abord, elles exploitent l'infrastructure existante des réseaux cellulaires, ce qui les rend relativement peu coûteuses à déployer et à utiliser. Elles sont également efficaces dans les environnements urbains denses où les signaux GPS peuvent être atténués ou bloqués.

Cependant, ces méthodes ne sont pas exemptes de limitations. La précision du positionnement GSM peut être affectée par des facteurs tels que la densité des tours cellulaires, les interférences du signal et la propagation multi-trajet dans les environnements urbains. Par rapport au GPS, le positionnement GSM est généralement moins précis, bien qu'il puisse être suffisant pour de nombreuses applications où une précision de quelques dizaines de mètres est acceptable.

Dans ce projet de fin d'étude, nous avons étudié et analysé diverses techniques de positionnement utilisant le réseau GSM. Nous sommes arrivés que la précision fournie par ces techniques n'était pas suffisante, par conséquent nous avons développé un modèle d'apprentissage automatique basé sur l'intelligence artificielle, qui est l'un des facteurs les plus importants contribuant au développement de solutions plus efficaces et plus précises dans ce domaine.

Les résultats du modèle obtenu ont montré des performances encourageantes et une bonne précision par rapport à d'autres techniques. Par conséquent, nous concluons que l'intégration des techniques de positionnement utilisant le réseau GSM avec des modèles d'apprentissage automatique peut ouvrir de nouveaux horizons pour leurs applications dans divers domaines.

Pour améliorer cette recherche et obtenir des résultats plus précis, nous suggérons de collecter une grande quantité de données sur lesquelles entraîner le modèle et d'expérimenter avec d'autres algorithmes.

Bibliographie

- [1] K. P. Prashant Krishnamurthy, Principles of Wireless Access and Localization, Wiley, 2013, p. 598.
- [2] E. D. K. Christopher J. Hegarty, Understanding GPS/GNSS: Principles and Applications, Third Edition, Artech House Publishers, 2017, pp. [1-8].
- [3] Dr. SATYAPRIYA MAHATO, «INTRODUCTION TO GPS,» [En ligne]. [Accès le 20 mars 2024].
- [4] Ahmed El-Rabbany, Introduction to GPS, Artech House, 2002, pp. [2-3].
- [5] R. Frederic G. Snider, «GPS: Theory, Practice and Applications,» [En ligne]. Available: <https://www.pdhone.com/courses/1116/1116content.htm>. [Accès le 02 mars 2024].
- [6] Stephen W. Hinch, Outdoor Navigation with GPS, W. Press, Éd., 2010, p. 5.
- [7] Redaction Team, 23 Mars 2023. [En ligne]. Available: <https://barrazacarlos.com/fr/avantages-et-inconvenients-du-gps/>. [Accès le 27 Mars 2024].
- [8] Z. U. R. ,. A. M. F. Ihtesham ul Haq, «GSM Technology: Architecture, Security and Future Challenges,» *International Journal of Science Engineering and Advance Technology*, Janvier 2017.
- [9] Google, «Les antennes relais GSM en copropriété,» [En ligne]. Available: <https://econhomes.fr/antennes-relais-gsm-en-copropriete/>. [Accès le 03 MARS 2024].
- [10] Sajal Kumar Das, Mobile Handset Design, Wiley, 2013.
- [11] Google, «Etude comparative des technologies GSM et CDMA - Architecture du réseau GSM,» [En ligne]. Available: <https://www.mongosukulu.com/index.php/contenu/informatique-et-reseaux/telecommunications/770-etude-comparative-des-technologies-gsm-et-cdm?start=1>. [Accès le 11 Mars 2024].

- [12] H.-J. V. C. B. C. H. Jrg Eberspcher, GSM - Architecture, Protocols and Services, Wiley, 2008, p. 45.
- [13] Google, «Architecture d'un réseau radio mobile GSM,» [En ligne]. Available: https://www.technologuepro.com/gsm/chapitre_2_GSM.htm. [Accès le 25 Mars 2024].
- [14] Morten Tolstrup, Indoor Radio Planning:A Practical Guide for GSM, DCS, UMTS, HSPA and LTE, Wiley, 2011, p. 56.
- [15] H. BENOUDNINE, *Applications Radars et Réseaux sans fil mobiles.*, 2013.
- [16] Agassi Melikov, Cellular Networks:Positionnement, analyse des performances, fiabilité, InTech, 2011.
- [17] Sofia Nyberg, «Physical Cell ID Allocation in Cellular Networks,» 2016. [En ligne]. [Accès le 17 mars 2024].
- [18] J. S. D. M. N. C. E. F. Jean-Pierre Dubois, «GSM Position Tracking using a Kalman Filter,» 2012.
- [19] M. D. AMMOUCHE Amina, *Localisation d'une station mobile par le filtre Kalman et l'intersection de la covariance*, 2019.
- [20] T. L. CHABA Lydia, *Localisation indoor*, 2017.
- [21] Google, «Géolocalisation par triangulation : c'est quoi exactement ?,» [En ligne]. Available: <https://info-high-tech.com/geolocalisation-par-triangulation-cest-quoi-exactement/>. [Accès le 06 05 2024].
- [22] Rayan Mac, «La géolocalisation par ondes radiofréquences (2/2),» 28 juin 2018. [En ligne]. Available: <https://www.linuxembedded.fr/2018/06/la-geolocalisation-par-ondes-radiofrquences-22>. [Accès le 11 05 2024].
- [23] B. D. Djennane Nabila, *Localisation des mobiles par WIFI*, 2013.
- [24] Oleg Nicolaevich Skrypnik, Radio Navigation Systems for Airports and Airways, Springer Nature Singapore.
- [25] R. C. Robert E.Guinness, Geospatial Computing in Mobile Devices, Artech House, 2014.

- [26] Kumar Abhishek, «Introduction to artificial intelligence,» 10 May 2022. [En ligne]. Available: <https://www.red-gate.com/simple-talk/development/data-science-development/introduction-to-artificial-intelligence/>?. [Accès le 05 05 2024].
- [27] ayajr2001, «Conclusion Et Intro Final,» 2023. [En ligne]. [Accès le 21 04 2024].
- [28] Ronald T.Kneusel, How AI Works, No Starch Press, 2023.
- [29] Prof. MUTTAPPA M MANTUR, ARTIFICIAL INTELLIGENCE WITH PYTHON, Deccan International Academic Publisher, 2023.
- [30] Google, «Understanding Different Types of Artificial Intelligence with Examples,» [En ligne]. Available: <https://www.edureka.co/blog/types-of-artificial-intelligence/#Types%20Of%20Artificial%20Intelligence>. [Accès le 22 April 2024].
- [31] R. K. J. D. R. D. Kamal Kant Hiran, Machine Learning, Bpb Publications, September 16,2021.
- [32] M. C. Fred Nwanganga, Practical Machine Learning in R, Wiley, May 27, 2020.
- [33] Google, «Regression vs. Classification in Machine Learning,» [En ligne]. Available: <https://www.javatpoint.com/regression-vs-classification-in-machine-learning>. [Accès le 10 05 2024].
- [34] Patrice rey, L'intelligence artificielle expliquée simplement, BoD - Books on Demand, 2024, pp. [39-49].
- [35] A. S. U. S. Shanthamallu, Machine and Deep Learning Algorithms and Applications, Springer International Publishing, 2022, p. 33.
- [36] A. G. B. Richard S. Sutton, Reinforcement Learning, MIT Press, 2018, p. 1.
- [37] Scott W.Menard, Logistic Regression:From Introductory to Advanced Concepts and Applications, SAGE Publications, 2010, p. 1.
- [38] I. R. M. Association, Éd., Research Anthology on Machine Learning Techniques, Methods, and Applications, IGI Global, 2022, p. 98.
- [39] Dothang Truong, Data Science and Machine Learning for Non-Programmers, CRC Press,

2024.

[40] Julianna Delua, «Supervised vs. Unsupervised Learning: What's the Difference?,» 12 Mars 2021. [En ligne]. Available: <https://www.ibm.com/think/topics/supervised-vs-unsupervised-learning>. [Accès le 24 04 2024].

[41] Google, «What is Google Colab?,» [En ligne]. Available: <https://www.androidpolice.com/google-colab-explainer/>?. [Accès le 10 may 2024].

[42] M. Summerfield, Programming in Python 3, Addison-Wesley, 2010, p. 1.

[43] Daniel Chen, Pandas for Everyone, Pearson Education, 2022, p. 1.

[44] D. Paper, Hands-on Scikit-Learn for Machine Learning Applications, Apress, 2019, p. 1.

[45] H. M. V. Yuriy Shvets, Éd., Proceedings of the 2nd International Conference on Mathematical Statistics and Economic Analysis, EAI Publishing.