REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université Mohamed Khider – BISKRA
Faculté des Sciences Exactes, des Sciences de la Nature et de la Vie
Département d'informatique

N° d'ordre : :IA12/M2/2021

Mémoire

Présenté pour obtenir le diplôme de master académique en

Informatique

Parcours : **Intelligence artificielle (IA)**

Détection des sols salés basée sur les images satellitaires et les techniques d'apprentissage automatique

---

Par :

MOHAMMEDI AHLAM

Soutenu le 04/07/2021 devant le jury composé de :

-	-	Président
- Ayad soheyb	MCA Université de Biskra	Rapporteur
-	-	Examinateur

Année universitaire 2020-2021

Dédicaces

Mes parents sont la joie de ma vie,

mes frères,

et toute ma famille,

Et tous mes chers professeurs et amis,

je dédie ce modeste travail.

Remerciements

Avant tout, Nous tenons à remercier le bon DIEU le tout puissant et clément de nous avoir illuminé le chemin du savoir et de nous avoir donné le courage, la puissance et la volonté pour accomplir ce modeste projet.

Nous tenons particulièrement à exprimer notre profonde gratitude à notre encadreur Dr Ayad Soheyb pour ses conseils et encouragements durant toute la période d'encadrement.

Nous désirons témoigner notre reconnaissance et nos remerciements les plus sincères au Chef de département d'informatique et le Doyen de la faculté de la Science de la Nature et la vie.

Nous tenons aussi à remercier vivement les examinateurs pour avoir accepté d'examiner ce travail et leurs participations au jury.

Un énorme merci à nos familles et amies pour leurs éternel soutient et la confiance qu'ils ont en nos capacité.

Résumé

La salinisation est un processus d'accumulation des sels à la surface du sol et dans la zone racinaire des plantes qui occasionne des effets nocifs sur les végétaux et le sol; il s'en suit une diminution des rendements et, à terme, une stérilisation du sol. Ainsi la détection et la délimitation de ces zones restent le seul moyen pour faire face à cette menace. Il existe bon nombres de techniques et de méthodes pour délimiter ces aires, la prospection sur terrain, analyses au laboratoire, ...etc. Cependant, ces opérations représentent des processus complexes. Sur ce, la détection et la délimitation des sols salés représente un véritable challenge qui suscite l'intérêt des chercheurs en ce domaine.

Dans ce projet, nous avons créé une base de données d'informations spectrales sur le sol (précisément la région de Biskra) en se basant sur des images satellitaires, ensuite nous avons étudié et implémenté des modèles d'apprentissage automatique pour prédire la salinité du sol. Après comparaisons entre les modèles développés, les résultats ont montré que la Gradient Boosting Regression est l'algorithme qui surpasse les autres modèles étudiés.

Mots clés : la salinité des sols , télédétection, apprentisage automatique , bandes spéctrales , Conductivité Électrique .

Abstract

Salinization is a process of salt accumulation on the soil surface and in the plant root zone that causes adverse effects on plants and soil, resulting in reduced yields and eventually sterilization of the soil. Thus, the detection and delimitation of these areas remains the only way to deal with this threat. There are many techniques and methods to delineate these areas, field surveys and laboratory analyses, etc. However, these operations represent complex processes. Therefore, the detection and delineation of salty soils is a real challenge that arouses the interest of researchers in this field.

In this project, we created a database of spectral information on the ground (precisely the Biskra region) based on satellite images, then, we studied and implemented machine learning models to predict the salinity of the soil. After comparisons between the models developed, the results showed that the Gradient Boosting Regression is the algorithm that outperforms the other models studied.

Keywords: soil salinity, remote sensing, machine learning, spectral bands , electrical conductivity .

ملخـص

التملح هو عملية تراكم الملح على سطح التربة وفي منطقة جذر النباتات مما يسبب آثارًا ضارة على النباتات والتربة ؛ وهذا يؤدي إلى انخفاض في الغلال ، وعلى المدى الطويل ، تعقيم التربة. وبالتالي ، فإن الكشف عن هذه المناطق وتعيين حدودها يظل السبيل الوحيد لمواجهة هذا التهديد. هناك عدد لا بأس به من الأساليب والأساليب لتحديد هذه المجالات ، والتنقيب الميداني ، والتحليلات المعملية ، ... إلخ. ومع ذلك ، فإن هذه العمليات تمثل عمليات معقدة. في هذا الصدد ، يمثل الكشف عن التربة المالحة وتحديد حدودها تحديًا حقيقيًا يثير اهتمام الباحثين في هذا المجال.

في هذا المشروع ، أنشأنا قاعدة بيانات للمعلومات الطيفية على الأرض (منطقة بسكرة بالتحديد) بناءً على صور الأقمار الصناعية ، ثم درسنا ونفذنا نماذج التعلم الآلي للتنبؤ بملوحة التربة. بعد المقارنات بين النماذج المطورة ، أظهرت النتائج أن الانحدار المعزز للتدرج هو الخوارزمية التي تتفوق في الأداء على النماذج الأخرى المدروسة.

الكلمات المفتاحية : ملوحة التربة . الاستشعار عن بعد . التعلم الالي . النطاقات الطيفية . التوصيل الكهربائي .

Table des matières

Table des figures

Liste des tableaux

Introduction générale

Sommaire

0.1 Introduction générale

La salinité est l'un des plus grands problèmes, dans les environnement sarides et semi-arides du monde . Elle est considérée comme l'une des principales contraintes d'environnement aux quelles l'agriculture moderne est confrontée. Elle est aussi reconnue comme l'une des menaces principales à la durabilité des périmètres irrigués de notre siècle [1].

Les sels issus des sols et des nappes peuvent modifier temporairement ou en permanence les états de surface. Ces changements affectent la végétation et la surface du sols nu. La dynamique spatiale et temporelle des sols salés, particulièrement en zone semi-aride ou aride, nécessite un suivi au sol facilité par la télédétection aérienne et satellitaire [2].

La télédétection spatiale est une discipline scientifique qui intègre un large éventail de compétences et de technologies utilisés pour l'observation, l'analyse et l'interprétation des phénomènes terrestres et atmosphériques , ses principales sources sont les mesures et les images obtenues à l'aide de plates-formes aériennes et spatiales. Les systèmes de télédétection actuels, contrairement à ceux du début du développement de ces technologies ont connu des changements importants, en particulier dans la dernière décennie, avec une technologie essentielle dans le suivi des processus multiples qui affectent la surface et l'atmosphère de la Terre .Un impact important, en particulier sur notre planète, tels que le changement climatique, la déforestation, la désertification, etc.

En Algérie, les facteurs qui contribuent à l'extension du phénomène de salinisation des terres sont liés à l'aridité du climat qui porte sur plus de 95 % du territoire, la qualité médiocre des eaux d'irrigation, le système de drainage souvent inexistant ou non fonctionnel, et la conduite empirique des irrigations[3].

La production agricole, en Algérie est limitée par de faibles ressources hydrauliques, une mauvais erépartition des précipitations et par des teneurs élevées en sels solubles dans les sols et les eaux . La majorité des études existantes concerne, les mécanismes et les processus de salinisation des sols irrigués ont peu abordé les variations spatiales de la

salinité et les dynamiques temporelles associées. Le manque d'une cartographie précise des sols salés et d'un suivi de l'évolution temporelle de la salinité dans les zones irriguées ne permet pas de juger des risques, ni d'ailleurs des efforts entrepris de restauration des sols salés, et encore moins d'anticiper le phénomène avec l'extension « forcée » de l'irrigation [4].

À mesure que la technologie se développe, des méthodes intelligentes de prévision de l'état des sols sont utilisées pour éviter les problèmes et les pertes agricoles, afin de répondre aux divers besoins des utilisateurs.

L'apprentissage automatique est la méthode la plus courante pour prédire l'avenir ou classifier l'information afin d'aider les gens à prendre les décisions nécessaires. Les algorithmes d'apprentissage automatique sont formés sur des exemples ou des exemples à partir desquels ils apprennent des expériences passées et aussi l'analyse de données historiques. Ainsi, comme il répète des exemples, encore et encore, il peut identifier des modèles afin de faire des prédictions sur l'avenir.

Les objectifs de ce travail sont de proposer des modèle basés sur l'apprentissage automatique pour la prédiction les sols salés dans la région de biskra en utilisé la télédetection pour éviter les risques aux cultures agricoles et les effets néfastes sur la croissance des plantes, en utilisant des modèles d'apprentissage automatique.

Dans ce cadre , qui vise à détecter des sols salés par images satellitaires basée sur l'apprentissage automatique , nous avons organisés notre mémoire en 4 chapiter :

Le 1er chapitre , explique les notions de base de la télédétection et du rayonnement électromagnétique , il également le concept de salinités du sols et ses caractéristiques.

Le 2ème chapitre , fournit les conceptes et principes de l'apprentissage automatique aussi que la présentation de quelque travaux connexes.

Le 3ème chapitre , fournit la conception de notre système et il présente la problématique et les objectifs de ce travail, puis, décrit l'architecture de système proposée .

Le 4ème chapitre est dédié à la présentation de la partié exprémentale aussi que la disscussion des résultats obtenus en utilisant l'apprentissage automatique (ML).

Chapitre 1

Généralités sur les sols salés

Sommaire

1.1 Introduction

La 0 des sols est une forme de dégradation des terres qui pose des défis à la productivité agricole et au développement durable. Pour résoudre ce problème, nous utilise sur

les images satellites, qui sont un outil puissant pour cartographier et suivre l'évolution de la salinité grâce à la sensibilité du signal électromagnétique aux paramètres du sols.[1][2]

Dans ce chapitre, nous présentons les notions fondamentales la télédetection de la télédétection et ses types, en plus de cela, nous apprenons le rayonnement magnétique, puis nous expliquons en général la salinité du sols, ses propriétés et ses niveaux. Enfin, nous donnons un aperçu des images satellites et la comparaison entre leurs types.

1.2 Généralités sur la télédétection

1.2.1 Qu'est-ce que la télédétection ?

La télédétection est la technique qui permet d'obtenir de l'information sur la surface de la Terre sans contact direct avec celle-ci. La télédétection englobe tout le processus qui consiste à capter et à enregistrer l'énergie d'un rayonnement électromagnétique émis ou réfléchi, à traiter et à analyser l'information, pour ensuite mettre en application cette information

Dans la plupart des cas, la télédétection implique une interaction entre l'énergie incidente et les cibles. Le processus de la télédétection au moyen de systèmes imageurs comporte les sept étapes que nous élaborons ci-après. Notons cependant que la télédétection peut également impliquer l'énergie émise et utiliser des capteurs non-imageurs .[5]

1.2.2 Processus de la télédétection :

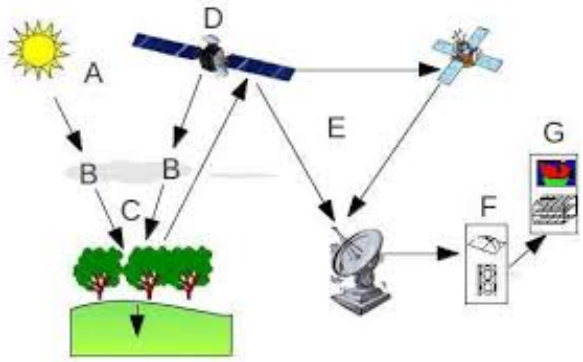

FIGURE 1.1 – Processus de la télédétection [5]

1. **Source d'énergie ou d'illumination (A)** À l'origine de tout processus de télédétection se trouve nécessairement une source d'énergie pour illuminer la cible.

2. **Rayonnement et atmosphère (B)** Durant son parcours entre la source d'énergie et la cible, le rayonnement interagit avec l'atmosphère. Une seconde interaction se produit lors du trajet entre la cible et le capteur.

3. ***Interaction avec la cible (C)*** Une fois parvenue à la cible, l'énergie interagit avec la surface de celle-ci. La nature de cette interaction dépend des caractéristiques du rayonnement et des propriétés de la surface.

4. ***Enregistrement de l'énergie par le capteur (D)*** Une fois l'énergie diffusée ou émise par la cible, elle doit être captée à distance (par un capteur qui n'est pas en contact avec la cible) pour être enfin enregistrée.

5. ***Transmission, réception et traitement (E)*** L'énergie enregistrée par le capteur est transmise, souvent par des moyens électroniques, à une station de réception où l'information est transformée en images (numériques ou photographiques).

6. ***Interprétation et analyse (F)*** Une interprétation visuelle et/ou numérique de l'image traitée est ensuite nécessaire pour extraire l'information que l'on désire obtenir sur la cible.

7. ***Application (G)*** La dernière étape du processus consiste à utiliser l'information extraite de l'image pour mieux comprendre la cible, pour nous en faire découvrir de nouveaux aspects ou pour aider à résoudre un problème particulier.[6]

1.2.3 Le rayonnement électromagnétique

Une source d'énergie sous forme de rayonnement électromagnétique est nécessaire pour illuminer la cible, à moins que la cible ne produise elle-même cette énergie. le rayonnement électromagnétique a comme vecteur le photon, particule dépourvue de masse. Le photon est le boson associé à la force électromagnétique.[7] Du fait de la dualité onde-corpuscule, les rayonnements électromagnétiques peuvent se modéliser de deux manières :

- **Onde électromagnétique :** le rayonnement est une variation des champs électriques et magnétiques, l'analyse spectrale permet de décomposer cette onde en ondes monochromatiques de longueurs d'onde et fréquences différentes ;

- **Photon :** la mécanique quantique associe à une radiation électromagnétique monochromatique un corpuscule de masse nulle nommé photon

  Les échanges d'énergie portée par le rayonnement électromagnétique qui ont lieu entre le soleil et le système terre-océan-atmosphère ne se font pas de manière continue, mais de façon discrète, sous forme de paquets d'énergie, véhiculés par des corpuscules élémentaires immatériels, les photons. Chaque photon transporte ainsi un quantum d'énergie proportionnel à la fréquence de l'onde électromagnétique considérée ; cette énergie est d'autant plus grande que la fréquence est élevée.

  La relation suivante exprime la quantité d'énergie associée à un photon en fonction de la fréquence de l'onde :

$\mathbf{E = h\nu}$

E : l'énergie de l'onde électromagnétique

ν : $la fréquence de l'onde$

h : la constante de Planck (6,625.10-34 J.s)

L'impulsion p du photon est égale à $\mathbf{p = E / c = h / c.}$

1.2.4 Le rayonnement électromagnétique et les différentes réponses spectrales

La télédétection spatiale est une mesure de l'énergie émise par la surface de la Terre sous forme de rayonnement électromagnétique (REM) (Figure 1). La plupart des applications de la télédétection utilisent les domaines du visible (0.4 à 0.7 m) et de l'infrarouge (0.7 à 100 m) (Tableau 1). La source d'énergie peut être passive (soleil) ou active (radar)

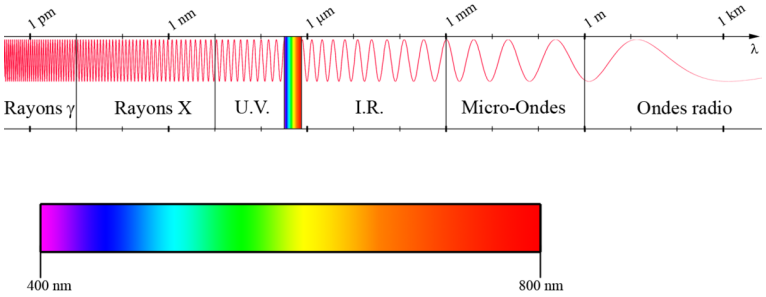

FIGURE 1.2 – Le rayonnement électromagnétique[25]

Region	Longeur d'onde(um)	Designation
isible!isible!Visiblept<	0.40-0.45	Violet
	0.45-0.50	Bleu
	0.50-0.55	Vert
	0.55-0.60	Jaune
	0.60-0.65	Orange
	0.65-0.70	Rouge
rarouge !rarouge !frarouge pt<	0.7-1.0	Proche infrarouge
	1.0-2.5	Moyen infrarouge
	2.5-1000	infrarouge thermique

TABLE 1.1 – Longueurs d'onde du visible et de l'infrarouge [26]

1.2.5 Le spectre électromagnétique

Le spectre électromagnétique s'étend des courtes longueurs d'onde (dont font partie les rayons gamma et les rayons X) aux grandes longueurs d'onde (micro-ondes et ondes radio). La télédétection utilise plusieurs régions du spectre électromagnétique.[8] Le

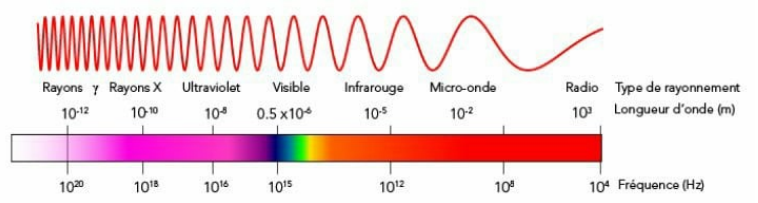

FIGURE 1.3 – Le spectre électromagnétique [8]

spectre électromagnétique représente la répartition des ondes électromagnétiques en fonction de leur longueur d'onde, de leur fréquence ou bien encore de leur énergie .

En partant des ondes les plus énergétiques, on distingue successivement :

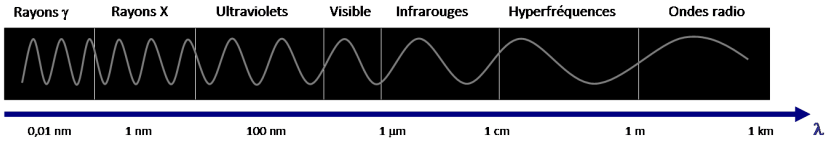

FIGURE 1.4 – Le spectre électromagnétique[8]

- ***Les rayons gamma (γ)*** :

  ils sont dus aux radiations émises par les éléments radioactifs. Très énergétiques, ils traversent facilement la matière et sont très dangereux pour les cellules vivantes.

  Leurs longueurs d'onde s'étendent d'un centième de milliardième (10-14 m) à un milliardième (10-12 m) de millimètre.

- ***Les rayons X :***

  rayonnements très énergétiques traversant plus ou moins facilement les corps matériels et un peu moins nocifs que les rayons gamma, ils sont utilisés notamment en médecine pour les radiographies, dans l'industrie (contrôle des bagages dans le transport aérien), et dans la recherche pour l'étude de la matière (rayonnement synchrotron).

  Les rayons X ont des longueurs d'onde comprises entre un milliardième (10-12 m) et un cent millième (10-8 m) de millimètre.

- ***Les ultraviolets :***

rayonnements qui restent assez énergétiques, ils sont nocifs pour la peau. Heureusement pour nous, une grande part des ultraviolets est stoppée par l'ozone atmosphérique qui sert de bouclier protecteur des cellules.

Leurs longueurs d'onde s'échelonnent d'un cent millième (10-8 m) à quatre dixièmes de millième (4.10-7 m) de millimètre.

- ***Le domaine visible :***

  correspond à la partie très étroite du spectre électromagnétique perceptible par notre œil. C'est dans le domaine visible que le rayonnement solaire atteint son maximum (0,5 m) et c'est également dans cette portion du spectre que l'on peut distinguer l'ensemble des couleurs de l'arc en ciel, du bleu au rouge.

  Il s'étend de quatre dixièmes de millième (4.10-7 m) - lumière bleue - à huit dixièmes de millième (8.10-7 m) de millimètre - lumière rouge.

- ***L'infrarouge :***

  rayonnement émis par tous les corps dont la température est supérieure au zéro absolu (-273°C).

  En télédétection, on utilise certaines bandes spectrales de l'infrarouge pour mesurer la température des surfaces terrestres et océaniques, ainsi que celle des nuages.

  La gamme des infrarouges couvre les longueurs d'onde allant de huit dixièmes de millième de millimètre (8.10-7 m) à un millimètre (10-3 m).

- ***Les ondes radar ou hyperfréquences :***

  Cette région du spectre est utilisée pour mesurer le rayonnement émis par la surface terrestre et s'apparente dans ce cas à la télédétection dans l'infrarouge thermique, mais également par les capteurs actifs comme les systèmes radar.

  Un capteur radar émet son propre rayonnement électromagnétique et en analysant le signal rétrodiffusé, il permet de localiser et d'identifier les objets, et de calculer leur vitesse de déplacement s'ils sont en mouvement. Et ceci, quelque soit la couverture nuageuse, de jour comme de nuit.

  Le domaine des hyperfréquences s'étend des longueurs d'onde de l'ordre du centimètre jusqu'au mètre.

- ***Les ondes radio :***

  Ce domaine de longueurs d'onde est le plus vaste du spectre électromagnétique et concerne les ondes qui ont les plus basses fréquences. Il s'étend des longueurs d'onde de quelques cm à plusieurs km.

Relativement faciles à émettre et à recevoir, les ondes radio sont utilisées pour la transmission de l'information (radio, télévision et téléphone). La bande FM des postes de radio correspond à des longueurs d'onde de l'ordre du mètre. Celles utilisées pour les téléphones cellulaires sont de l'ordre de 10 cm environ.

fenêtres spectrales sont principalement utilisées en télédétection spatiale :

- Le domaine du visible
- Le domaine des infrarouges (proche IR, IR moyen et IR thermique)
- Le domaine des micro ondes ou hyperfréquences (pas abordé ici, même si elles ont une importance considérable en télédétection RADAR notammen

1.2.6 Interactions rayonnement-cible

FIGURE 1.5 – Interaction rayonnement-cible

L'absorption (A) se produit lorsque l'énergie du rayonnement est absorbée par la cible, la transmission (T) lorsque cette énergie passe à travers la cible et la réflexion (R) lorsque la cible redirige l'énergie rayonnante.

L'énergie incidente totale interagira avec la surface selon l'une ou l'autre de ces trois modes d'interaction ou selon leur combinaison. La proportion de chaque interaction dépendra de la longueur d'onde de l'énergie, ainsi que de la nature et des conditions de la surface [9]

En télédétection, est mesuréle rayonnement réfléchipar une cible Il existe deux modes limites de réflexion de l'énergie : Réflexion Diffuse et Réflexion Spéculaire

a distinction entre réflexion diffuse et réflexion spéculaire n'est pas liée aux propriétés chimiques d'une surface réfléchissante, mais à sa morphologie, envisagée à une échelle comparable à la longueur d'onde d'observation.

Un exemple

Un exemple simple permet d'illustrer cette dichotomie :l'eau, H20, sous forme liquide, est un miroir plan se comportant de façon spéculaire vis-à-vis de la lumière visible. Son équivalent cristallisé, la neige, qui possède pourtant les mêmes propriétés optiques dans le domaine visible, est une poudre qui réfléchit de façon dt@ùse la lumière.

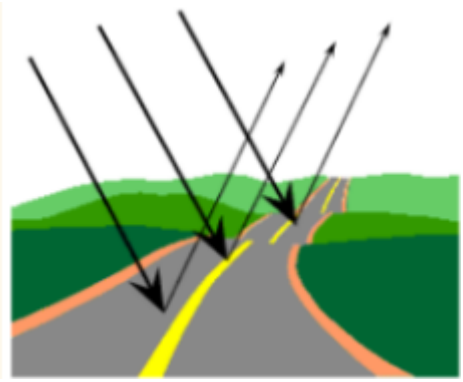

FIGURE 1.6 – Réflexion Spéculaire

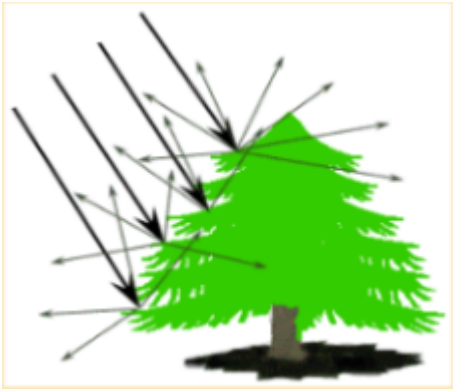

FIGURE 1.7 – Réflexion Diffuse

Or, tout télêdétecteur a constaté, sur une image SPOT, que la mer est très sombre, donc de réflexion spêculaire très faible, alors que la neige est d'un blanc éclatant, donc de réflexion diffuse très élevée.

1.2.6.1 Signatures spectrales des surfaces naturelles :

L'analyse visuelle ou statistique des réflectances nous permet de discriminer des objets dont la réponse spectrale (combinaisons d'intensité d'énergie réfléchie par chaque cible à la surface de la Terre dans des longueurs d'ondes variées) est différente .

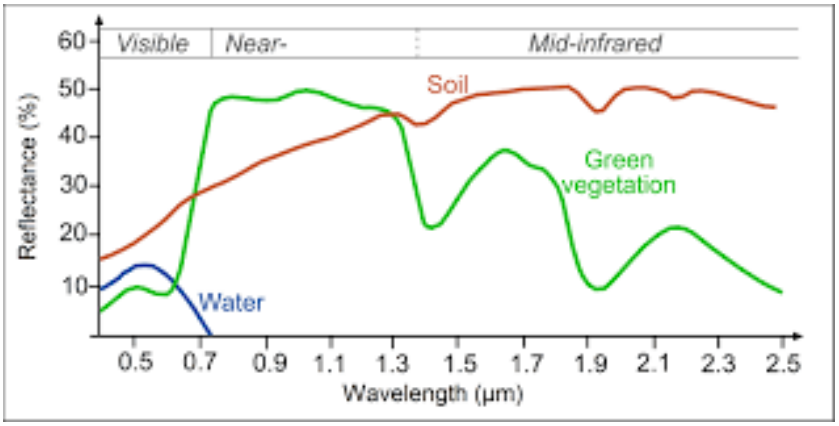

FIGURE 1.8 – Réponses spectrales typiques : sol nu, végétation et eau

1.2.7 Détection passive et active

Une autre possibilité de distinguer les satellites d'observation de la Terre est de comparer les capteurs utilisés. En général, il existe des capteurs passifs qui mesurent la lumière solaire réfléchie ou le rayonnement thermique, et des capteurs actifs qui utilisent leur propre source de rayonnement.[10]

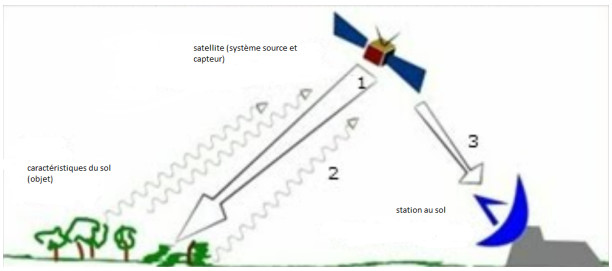

FIGURE 1.9 – Capteur active

1.2.7.1 capteur active

Les capteurs active (par exemple les radars et les scanners laser) émettent un rayonnement artificiel pour surveiller la surface de la Terre ou les caractéristiques atmosphériques. Les radars sont des instruments d'imagerie tandis que les altimètres radar et les diffusiomètres ne sont pas des images. Radar est l'abréviation de Radio Detection and Ranging, une méthode de détection et de télémétrie des caractéristiques de la surface terrestre. Les satellites radar utilisent de courtes impulsions de rayonnement électromagnétique dans la gamme spectrale des micro-ondes, ils ne dépendent donc pas de la lumière du jour et ne sont guère affectés par les nuages, la poussière, le brouillard, le vent et les mauvaises conditions météorologiques. Ils mesurent les impulsions radar réfléchies par le sol, analysent l'intensité du signal afin de récupérer des informations sur la structure de la surface terrestre, et détectent le temps écoulé entre l'émission et le retour des impulsions. Les résultats peuvent être utilisés pour mesurer les distances. Selon la mission du satellite, différentes opérations et procédures sont utilisées pour traiter les signaux en informations viables.[10]

1.2.7.2 capteur passive

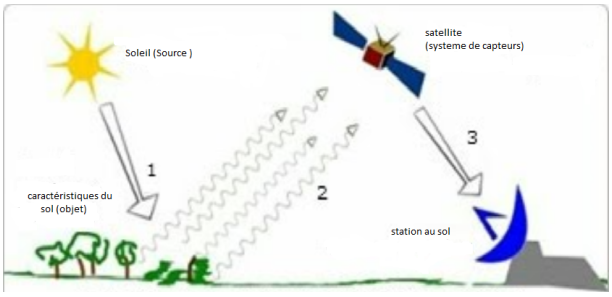

FIGURE 1.10 – Capteur passive

Les capteurs passifs détectent le rayonnement solaire réfléchi par la terre et le rayonnement thermique dans le visible et l'infrarouge du spectre électromagnétique. Ils n'émettent pas leur propre rayonnement, mais reçoivent la lumière naturelle et le rayonnement thermique de la surface de la terre. La plupart des capteurs passifs utilisent un scanner pour l'imagerie, par ex. LANDSAT. Équipés de spectromètres, ils mesurent des signaux sur plusieurs bandes spectrales simultanément

1.2.8 Caractéristiques des images

L'imagerie satellitaire(aussi appelée imagerie spatiale) désigne la prise d'images depuis l'espace, par des capteurs placés sur des satellites. Visuellement, les images satellitaires ressemblent beaucoup à des photos, mais elles contiennent bien plus d'informations.

Les caractéristiques fondamentales des images de télédétection sont :

- la résolution spectrale
- la résolution spatiale

1.2.8.1 Images satellitaires et résolution spatiale

Lorsque l'on prend une photographie classique, l'information est traduite par des formes et des couleurs, qui correspondent à des groupes de pixels plus ou moins homogènes. Sur une même scène photographiée pri !se par deux appareils, plus les pixels seront nombreux dans l'image plus la résolution spatiale sera élevée. On le voit aisément lorsque l'on souhaite faire un agrandissement et que l'on voit apparaître les pixels en zoomant sur une image .

Il en va de même pour une image satellitaire : selon les caractéristiques du capteur, l'altitude du satellite (donc son orbite autour de la Terre),les images seront composées de pixels couvrant une surface au sol plus ou moins grande du sol.[11]

1.2.8.2 Images satellitaires et résolution spectrale

Dans une image satellitaire, l'information sur les couleurs est décomposée en différents canaux ou bandes spectrales. Chaque bande est une image en niveaux de gris, composée de pixels ayant chacun une valeur de réflectance pour un intervalle de longueur d'ondes donné. On parle ainsi de "bande du bleu", du "rouge, du proche infrarouge", etc. Chaque bande va couvrir une portion plus ou moinslarge du spectre électromagnétique. Par exemple, la bande du bleu correspond à des longueurs dans un intervalle autour de 480 nm, celle du rouge autour de 600 nm.

Pour reprendre l'analogie avec une photographie classique,dans une photo, l'information sur les couleurs est contenue dans 3 bandes : la bande des longueurs d'ondes correspondant à la couleur bleue (B pour bleu ou blue), verte (V pour vert ou G pour green) et rouge (R pour rouge ou red). On voit ainsi souvent les acronymes RVB et RGB dans leslogiciels de traitement de photos. Chacune de ces trois bandes est en niveau de gris.[11]

Une image satellitaire, selon les caractéristiques du capteur embarqué sur le satellite, peut contenir en plus des trois bandes du visible (RVB) quelques bandes supplémentaires (par exemple infrarouge, proche infrarouge), et jusqu'à des centaines de bandes. Ces bandes vont couvrir des intervalles plus ou moins large du spectre électromagnétique. On parle ainsi

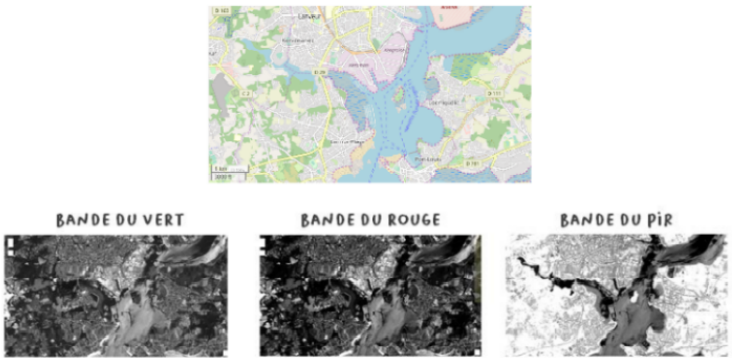

FIGURE 1.11 – Bandes spectrales de l'image satellitaire

d'images multi-spectrales(quelques bandes) ou d'images hyper-spectrales(des dizaines à des centaines de bandes

On peut représenter les valeurs de réflectance des bandes selon un graphique, avec en abscisse les longueurs d'onde du spectre électromagnétique, et en ordonnée les valeurs de réflectance en %. On illustre ainsi la nature de l'information contenue dans une images satellitaire composée de plusieurs bandes ou canaux. Sur le premier graphique, le capteur embarqué sur le satellite va enregistrer peu d'information sur les caractéristiques spectrales des cibles (visible sur les pixels) : l'image seracomposée de 4 bandes, qui couvriront chacune un intervalle de longueur d'onde assez large. Dans le second cas, le capteur va enregistrer une grande quantité d'information spectrale : pour chaque couleur, plusieurs bandes sont enregistrées. Non seulement le nombre de bandes est plus élevé, mais en plus chaque bande traduit des valeurs de réflectance pour de petits intervalles de longueur d'onde.

Généralement, une très haute résolution spectrale est possible au détriment de la résolution spatiale. Les images hyperspectrales sont par exemple très intéressantes pour discriminer des espèces végétales différentes. Tout l'intérêt est de pouvoir combiner les 2 types d'images.

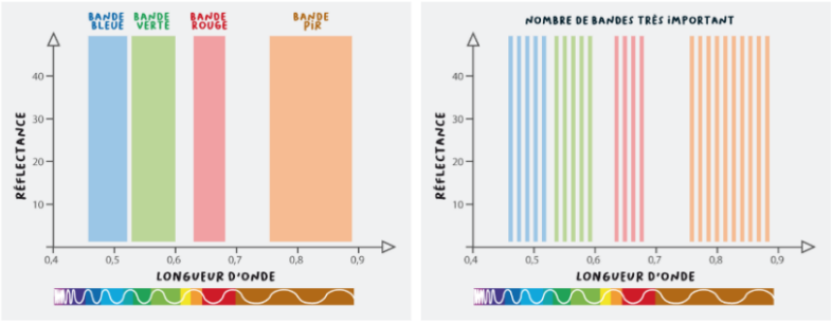

FIGURE 1.12 – Les valeurs de réflectance des bandes d'image satellitaire

1.3 Généralités sur les sols salés

Dans les régions semi-arides et ari-des, la pénurie et la variabilité de la pluie et la forte évaporation affectent l'eau et l'équilibre des sels dansle sols. Les facteurs climatiques sont trèsfavorables à l'ascension des sels, à la concentration de la solution du sols et à la précipitation des sels dans la zone raci-naire et l'horizon superficiel entraînant la salinisation. Cette salinisation peut être naturelle ou anthropogénique. La gestion des sols affectés par les sels requiert unecombinaison de pratiques agronomiques basée sur une bonne définition des caractéristiques hydropédologiques et hydrauli-ques et des conditions locales incluant leclimat, la culture et l'environnement socioéconomique. Plusieurs pratiques sontgénéralement combinées dans un système intégré pour éviter la salinisation excessive et la baisse de la fertilité des sols .[12]

1.3.1 Definition des sols salés

La salinisation des sols est l'accumulation excessive des sels très solubles (chlorures, sulfates, carbonates, de sodium et le magnésium) dans la partie superficielle des sols, ce qui se traduit par une diminution de la fertilité des sols. L'alimentation en eau des plantes est rendue plus difficile ; certains éléments peuvent avoir en outre un effet toxique spécifique (Na, Cl, B, Se) ; le sodium enfin peut se fixer sur les argiles et modifie du même coup leur comportement en présence d'eau. Les propriétés physiques globales du sols (capacité d'infiltration, conductivité hydraulique) sont alors dégradées [13]

1.3.2 les sols salés en Algerie

Les sols salés occupent de vastes superficies (3.2 millions d'hectares de la superficie totale). Ils sont localisés du Nord au sud,l'isohyètes 450mm semble être la limite supérieure des sols fortement sodiques [14].On rencontre plusieurs types de sols salés en Algérie localisés surtout dans les étages bioclimatiques arides et semi-arides.

En Algérie, il n'est recensé aucune étude cartographique fiable et précise permettant de délimiter les zones touchées par la salinité des terres et la quantification de la teneur des sels dans le sol. Néanmoins il existe quelques données fragmentaires qui donnent une idée générale sur le phénomène de salinité et de la dégradation des terres.

D'après SZABLOCS 3,2 million d'hectares subissent à des degrés de sévérité variable, le phénomène de salinisation dont une bonne partie se trouve localisée dans les régions steppiques où le processus de salinisation est plus marqué du fait des températures élevées durant presque toute l'année, du manque d'exutoire et de l'absence de drainage efficient.[15]

Ce phénomène est observé (voir carte ci-jointe) dans les plaines et vallées de l'Ouest du pays dans les hautes plaines de l'Est et dans le grand Sud .[15]

Dans le cadre de projets de création de périmètres d'irrigation (menés généralement

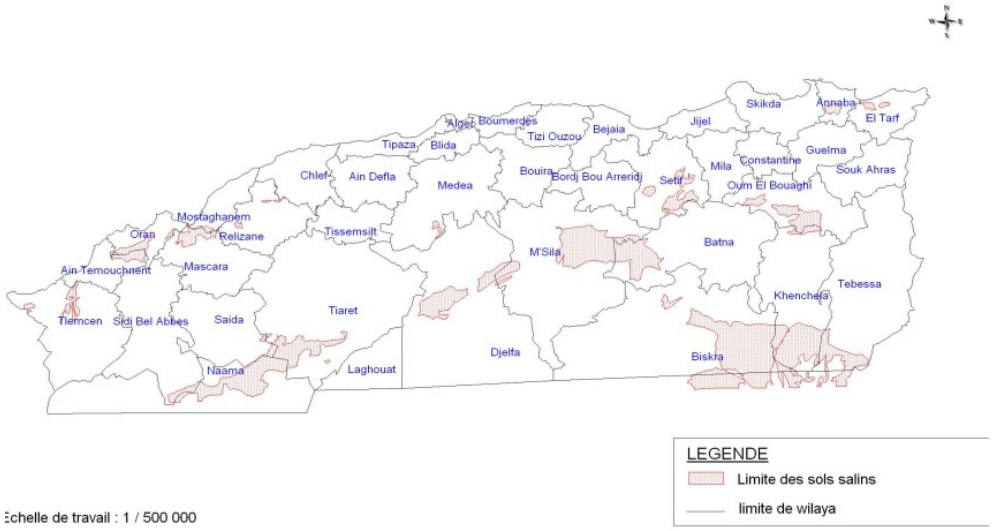

FIGURE 1.13 – Les zones touchées par la salinité des terres de l'Ouest du pays

par le MRE) il est donné quelque fois une idée sur les superficies affectées par la salinité. Par contre les niveaux de salinisation des terres n'est approché que par certaines études spécialisées notamment par l'INSID pour le Bas Chéliff (40 000 hectares ont été couverts par une campagne de mesures du niveau de salinité des terres menée en 1997 1) et dans le cadre de la préparation de thèses et autres .

1.3.3 Les formes de salinisation du sols

La salinisation des milieux naturels, ou salinisation dite «primaire», existe sur tous les continents et sous tous les climats. Les sels, solubles et cristallisés, participent aux cycles telluriques (hydrologiques, biologiques, climatiques...) à des échelles de temps et d'espace variables. L'extension et l'intensification des activités humaines provoquent une salinisation dite «secondaire» qui accentue la salinisation primaire, dégrade les sols non salinisés et plus globalement les écosystèmes et amplifie la désertification

1.3.3.1 Salinisation primaire

La salinisation primaire se produit naturellement là où la roche mère du sols est riche en sels solubles ou bien en présence d'une nappe phréatique proche de la surface. Dans les régions arides et semi-arides, où les précipitations sont insuffisantes pour lixivier les sels solubles du sols et où le drainage est restreint, des sols salins vont se former avec des concentrations élevées de sels («les sols salinisés»). Plusieurs processus géochimiques peuvent également avoir comme conséquence la formation de sols salinisés. La sodisation désigne un excès de sodium à l'origine du processus de salinisation .[16]

1.3.3.2 Salinisation secondaire

Les activités humaines qui induisent une salinisation dite «secondaire» sont nombreuses : irrigation mal conduite, pratiques d'anciennes techniques d'irrigation, irriga-tion avec des eaux riches en sels déforestation intensive , engrais contenant des sels de potassium et d'azote, dépôts atmosphériques près des sites industriels. ces apports entraînent une augmentation de la teneur en sels des sols, ce qui diminue leur productivité. La capacité des cultures à capter l'eau et les micronutriments est réduite. Des ions toxiques sse concentrent dans les végétaux et peuvent dégrader la structure du sols [16]

1.3.3.3 Caractéristiques des sols salés

1.3.3.3.1 Caractéristiques chimiques

1. **La conductivité électrique CE :** La conductivité électrique d'une solution est la conductance de cette solution mesurée entre des électrodes de 1 cm2 de surface. Elle permet de déterminer la salinité globale de l'extrait de pâte saturée.(BAISE, 1988). De plus la connaissance de la conductivité est nécessaire pour l'étude du complexe adsorbant des sols salés.[17]

2. **La réaction du sols, pH :** Le pH se mesure sur une suspension de terre fine[17]. Le pH des sols salés dont la salinité est de type neutre c'est à dire quand elle est due à des sels de bases et d'acides forts (chlorures, sulfates, de sodium, de calcium, de magnésium), reste inférieur à 8,5 et les sols est basique. Si la salinité est en revanche due à des sels de bases fortes et d'acides faibles, ce qui est le cas des bicarbonates ou des carbonates de sodium, le pH est au-dessus de 8,5 et peut atteindre 10, et le sol est alcalin. Le pH peut dépasser 10 après une précipitation du carbonatede calcium, les ségrégations salines sont fortement sodiques et renferment des sols alcalins (NaHCO3, Na2CO3, Na2SO4). Un pH compris entre 8 et 9 est retenu, généralement, comme limite de la dégradation de la structure.[18].

3. **La composition ionique de la solution du sols :** Afin de connaître la concentration en anions solubles (Cl-, SO4–et HCO3-) et en cations solubles (Na+, Ca++, Mg++, K+), une analyse chimique est effectuée sur extrait de pâte saturée ou sur extrait aqueux dilué. Elle sert à classer le type de salinisation selon le diagramme de PIPER ou autre classification. C'est ainsi qu'on peut utiliser le rapport Cl-/SO4–pour classer les solutions du sol [19].Elle sert aussi à calculer le SAR (Sodium Adsorption ratio) qui exprime le pouvoir de sodisation de la solution du sols.

4. **Le SAR « Sodium Adsorption Ration » :** Dans l'étude de mécanisme de sodisation, utilisé paramètre précis pour définir la composition des solutions du sols ou des nappes salées ; il s'agit de SAR « Sodium Adsorption Ration [20]. Le SAR

donne des indications sur le risque d'alcalisation du milieu. Les risques sont faibles si SAR < 10, moyen si SAR est compris entre 10 et 18,élevés si SAR > 18 et très élevés si SAR > 26.

5. **Le taux de sodium échangeable (ESP) :** Il exprime le taux de saturation du complexe absorbant en sodium échangeable par rapport à tous autres cationséchangeables.

1.3.3.3.2 Les caractéristique physique

1. **Capacité de rétention en eau :** Les sols salés peuvent raster humides même en saison sèche grâce à leur richesse en élément minéraux hygroscopique cette richesse hydrique n'est pas toujours disponible pour les plante à cause de dépression osmotique élevée de la solution du sols, la capacité de rétention en eau diminue en fonction de la nature du cation dans l'ordre suivent : Na + >Mg ++ >Ca ++ >K + . [21]

2. **Structure et perméabilité :** La stabilité structurale d'un sol diminué des que le taux de sodium échangeable atteint 12à15%. [22]

   Le gonflement et la dispersion dépendent tous deux de la minéralogie des argiles et des unités totales de Na+ adsorbée sur les sites d'échange .[23]

   Les sols qui ont un taux élève de sodium échangeable ont une structure dense, compacte et sont difficiles à labourer quand ils sont secs, et ont une faible perméabilité pour l'eau quand ils sont mouillés .[24]

1.4 La Comparaison entre les images Landsat 8 et les images Sentinel-2

L'imagerie satellite désigne la prise d'images de la Terre ou d'autres planètes à partir de satellites artificiels. Compte tenu des méthodes utilisées, il s'agit bien d'imagerie et non de photographie

Les images satellites sont capturées par deux satellites LANDSAT-8 et SENTINEL-2 , Les données Sentinel-2 ont des bandes spectrales très similaires à celles de Landsat 8 (à l'exclusion des bandes thermiques du capteur infrarouge thermique de Landsat 8) , Sentinel-2 avec ses 2 satellites est supérieur au système LANDSAT sur tous les aspects, sauf en ce qui concerne l'absence de bandes dans l'infra-rouge thermique qui sont présentes sur LANDSAT .[27] Le placement spécifique des bandes Sentinel-2, par rapport aux bandes Landsat 8, Comme sur l'image ci-dessous .

Les images Landsat 8 Operational Land Imager (OLI) et Thermal Infrared Sensor (TIRS) se composent de neuf bandes spectrales avec une résolution spatiale de 30 mètres

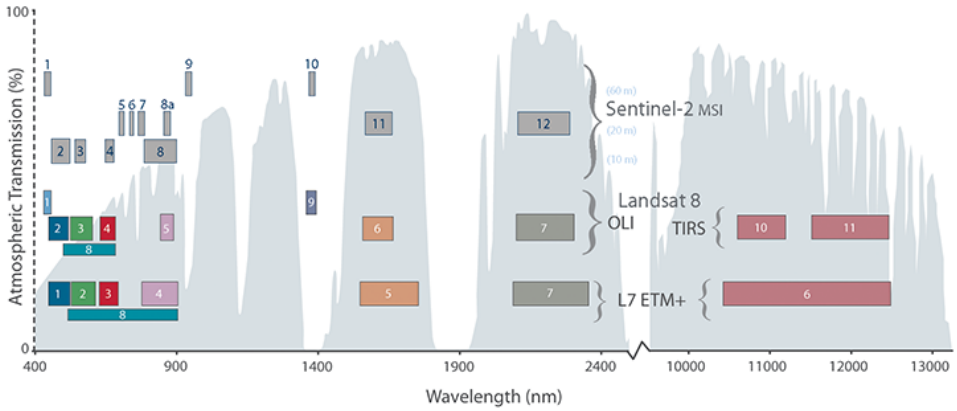

FIGURE 1.14 – Comparaison entre image landsat8 et sentinel-2 [29]

pour les bandes 1 à 7 et 9. La nouvelle bande 1 (ultra-bleu) est utile pour les zones côtières et les aérosols , La nouvelle bande 9 est utile pour la détection des cirrus. La résolution pour la bande 8 (panchromatique) est de 15 mètres. Les bandes thermiques 10 et 11 sont utiles pour fournir des températures de surface plus précises et sont collectées à 100 mètres.

Pendant que Sentinel-2 en est la composante optique, à résolution décamétrique, avec une revisite systématique de 5 jours ,les images Sentinel-2 dispose de 13 bandes spectrales dont 3 dans le moyen infrarouge (MIR). Les images ont un champ de vue de 290 km de large, et une résolution de 10m, 20m ou 60m en fonction des bandes spectrales.[28]

1.5 Conclusion

Dans ce chapitre, nous avons concentrés sur la définition du domaine télédetection, y compris utilise la technique Le rayonnement électromagné , permettant ainsi cartographie la salinité des sols par l'images satellitaires .

Dans le chapitre suivant, nous présenterons les approches d'apprentissage automatique en tant qu'approches alternatives pour la prédiction la salinité des sols .

Chapitre 2

L'apprentissage automatique

Sommaire

2.1 Introduction

Dans ce chapitre, nous discuterons la notion l'apprentissage automatique et leurs différence avec l'apprentissage en profondeur. De plus, nous présenterons les applications les plus utilisées du machine learning. Nous mentionnerons des les models de regression utilisées en apprentissage automatique afin de prédire la salinité des sols . Finalement, nous présentons une synthèse de quelques travaux connexes

2.2 L'apprentissage automatique

2.2.1 La notion d'apprentissage automatique

apprentissage automatique ou Le machine learning en anglais est un concept qui fait de plus en plus parler de lui dans le monde de l'informatique, et qui se rapporte au domaine de l'intelligence artificielle. Encore appelé « apprentissage statistique », ce terme renvoie à un processus de développement, d'analyse et d'implémentation conduisant à la mise en place de procédés systématiques. Pour faire simple, il s'agit d'une sorte de programme permettant à un ordinateur ou à une machine un apprentissage automatisé, de façon à pouvoir réaliser un certain nombre d'opérations très complexes.

L'objectif visé est de rendre la machine ou l'ordinateur capable d'apporter des solutions à des problèmes compliqués, par le traitement d'une quantité astronomique d'informations. Cela offre ainsi une possibilité d'analyser et de mettre en évidence les corrélations qui existent entre deux ou plusieurs situations données, et de prédire leurs différentes implications.[30]

2.2.2 Les différents procédés d'apprentissage automatique

L'apprentissage automatique implique deux principaux systèmes d'apprentissage qui définissent ses différents modes de fonctionnement. Il s'agit de :[31]

2.2.2.1 L'apprentissage supervisé

la machine s'appuie sur des classes prédéterminées et sur un certain nombre de paradigmes connus pour mettre en place un système de classement à partir de modèles déjà

catalogués. Dans ce cas, deux étapes sont nécessaires pour compléter le processus, à commencer par le stade d'apprentissage qui consiste à la modélisation des données cataloguées. Ensuite, il s'agira au second stade de se baser sur les données ainsi définies pour attribuer des classes aux nouveaux modèles introduits dans le système, afin de les cataloguer eux aussi.

2.2.2.2 L'apprentissage non-supervisé

Dans ce mode de fonctionnement du machine learning, il n'est pas question de s'appuyer sur des éléments prédéfinis, et la tâche revient à la machine de procéder toute seule à la catégorisation des données. Pour ce faire, le système va croiser les informations qui lui sont soumises, de manière à pouvoir rassembler dans une même classe les éléments présentant certaines similitudes. Ainsi, en fonction du but recherché, il reviendra à l'opérateur ou au chercheur de les analyser afin d'en déduire les différentes hypothèses.

2.2.3 La Comparaison entre l'apprentissage profond et l'apprentissage automatique

- Le Machine learning s'appuie sur un algorithme qui adapte lui-même le système à partir des retours faits par l'humain. La mise en place de cette technologie implique l'existence de données organisées. Le système est ensuite alimenté par des données structurées et catégorisées lui permettant de comprendre comment classer de nouvelles données similaires. En fonction de ce classement, le système exécute ensuite les actions programmées. Il sait par exemple identifier si une photo montre un chien ou un chat et classer le document dans le dossier correspondant.

- Le Deep learning n'a pas besoin de données structurées. Le système fonctionne à partir de plusieurs couches de réseaux neuronaux, qui combinent différents algorithmes en s'inspirant du cerveau humain. Ainsi, le système est capable de travailler à partir de données non structurées.

- Avec le Deep learning, le système identifie lui-même les caractéristiques discriminantes des données, sans avoir besoin d'une catégorisation préalable. Le système n'a pas besoin d'être entraîné par un développeur. Il évalue lui-même le besoin de modifier le classement ou de créer des catégories inédites en fonction des nouvelles données ,Tandis que le Machine learning fonctionne à partir d'une base de données contrôlable, le Deep learning a besoin d'un volume de données bien plus considérable. Le système doit disposer de plus de 100 millions d'entrées pour donner des résultats fiables.

- La technologie nécessaire pour le Deep learning est plus sophistiquée. Elle exige plus de ressources IT et s'avère nettement plus coûteuse que le Machine learning , elle

n'est donc pas intéressante pour une utilisation de masse par les entreprises.

- Dans l'étape de l'extraction de caractéristiques. Dans les algorithmes de ML traditionnelles l'extractionde caractéristiques est faite manuellement, c'est une étape difficile et coûteuse en temps et requiertun spécialiste en la matière alors qu'en Deep Learning cette étape est exécutée automatiquement parl'algorithme.[32],[33]

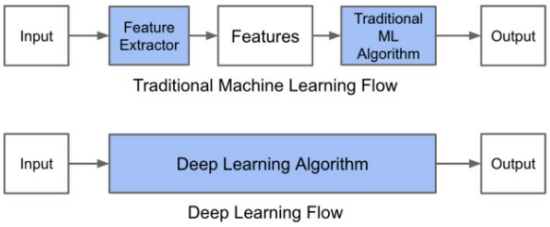

FIGURE 2.1 – Le procède du ML classique comparé à celui du Deep Learning[34]

2.2.4 Les applications d'apprentissage automatique

1. **Marketing en ligne** utilise des outils d'analyse marketing qui s'appuient sur l'apprentissage automatique . Ils évaluent des données définies et peuvent fournir des diagnostics fiables à propos du type de contenu capable d'aboutir à une conversion, des contenus que les clients veulent lire et des canaux marketing les plus efficaces pour conclure une vente.

2. **Support client** les chatbots peuvent s'appuyer sur l'apprentissage automatique. Ils s'orientent en fonction des mots-clés trouvés dans la question de l'utilisateur et, par des questions pour obtenir plus d'informations ou prendre des décisions, dialoguent avec l'utilisateur jusqu'à lui apporter la réponse désirée.

3. **Vente** ce qui fonctionne pour Netflix et Amazon est aussi idéal pour la vente. Grâce au Machine learning, les systèmes peuvent anticiper avec précision les produits et services qui pourraient intéresser les clients sur leur site. Ils peuvent ainsi faire des recommandations détaillées, ce qui facilite la vente avec des gammes de produits très large ou des produits hautement personnalisables.

4. **Informatique décisionnelle** le Machine learning peut aussi servir à visualiser les données importantes de l'entreprise et à rendre différentes prévisions compréhensibles pour les décideurs humains.[35]

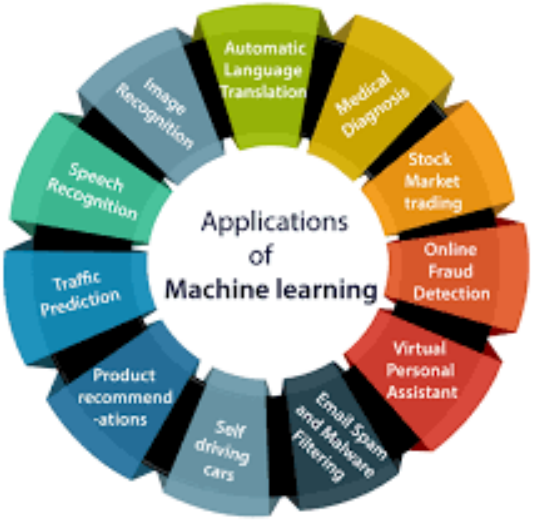

FIGURE 2.2 – Les application d'apprentissage automatique [36]

2.3 Algorithmes du regression d'apprentissage automatique et leurs applications

Les algorithmes de régression font partie de la famille des algorithmes d'apprentissage automatique supervisé, qui est un sous-ensemble des algorithmes d'apprentissage automatique. L'une des principales caractéristiques des algorithmes d'apprentissage supervisé est qu'ils modélisent les dépendances et les relations entre la sortie cible et les caractéristiques d'entrée pour prédire la valeur des nouvelles données. Les algorithmes de régression prédisent les valeurs de sortie en fonction des caractéristiques d'entrée des données introduites dans le système. La méthodologie d'apprentissage supervisé est la suivante : l'algorithme construit un modèle sur les caractéristiques des données d'apprentissage et utilise ce modèle pour prédire la valeur des nouvelles données.

Nous allons rappeler les algorithmes du regression le plus utilisés en Machine Learning.[37]

2.3.1 Modèle de régression linéaire simple

La régression linéaire simple est une méthode statistique qui permet aux utilisateurs de résumer et d'étudier les relations entre deux variables continues (quantitatives). La régression linéaire est un modèle linéaire, c'est-à-dire un modèle qui suppose une relation linéaire entre les variables d'entrée (x) et la variable de sortie unique (y). Ici, la variable y peut être calculée à partir d'une combinaison linéaire des variables d'entrée (x). Lorsqu'il y a une seule variable d'entrée (x), la méthode est appelée régression linéaire simple. Lorsqu'il y a plusieurs variables d'entrée, la procédure est appelée régression linéaire multiple.

Application

Certaines des applications les plus populaires de l'algorithme de régression linéaire sont

les prévisions de portefeuille, les prévisions salariales, les prévisions immobilières et les heures d'arrivée des prévisions de trafic.

2.3.2 Régression Lasso

LASSO est l'abréviation de Least Absolute Selection Shrinkage Operator (opérateur de rétrécissement de la sélection la moins absolue), où le rétrécissement est défini comme une contrainte sur les paramètres. L'objectif de la régression lasso est d'obtenir le sous-ensemble de prédicteurs qui minimise l'erreur de prédiction pour une variable de réponse quantitative. L'algorithme fonctionne en imposant une contrainte sur les paramètres du modèle qui entraîne un rétrécissement des coefficients de régression pour certaines variables vers un zéro.

Application

Les algorithmes de régression Lasso ont été largement utilisés dans les réseaux financiers et en économie. En finance, son application est vue dans la prévision des probabilités de défaut et les modèles de prévision basés sur Lasso sont utilisés dans l'évaluation du cadre de risque à l'échelle de l'entreprise. Les régressions de type Lasso sont également utilisées pour réaliser des plateformes de test de stress afin d'analyser plusieurs scénarios de stress.

2.3.3 Régression logistique

Il s'agit de l'une des techniques de régression les plus couramment utilisées dans le secteur, qui est largement appliquée à la détection des fraudes, au scoring des cartes de crédit et aux essais cliniques, lorsque la réponse est binaire. L'un des principaux avantages de cet algorithme populaire est que l'on peut inclure plus d'une variable dépendante qui peut être continue ou dichotomique. L'autre avantage majeur de cet algorithme d'apprentissage automatique supervisé est qu'il fournit une valeur quantifiée pour mesurer la force de l'association en fonction du reste des variables. Malgré sa popularité, les chercheurs ont souligné ses limites, citant un manque de technique robuste et aussi une grande dépendance du modèle.

Application

Aujourd'hui, les entreprises déploient la régression logistique pour prédire la valeur des maisons dans le secteur de l'immobilier, la valeur de la durée de vie des clients dans le secteur de l'assurance et sont utilisées pour produire un résultat continu

2.3.4 Machines à vecteurs de support(SVM)

La machine à vecteurs de support (SVM) est un autre algorithme très puissant qui repose sur des bases théoriques solides basées sur la théorie de Vapnik-Chervonenkis, Cet algorithme d'apprentissage automatique supervisé dispose d'une forte régularisation et peut

être utilisé à la fois pour la classification et la régression. Ils sont caractérisés par l'utilisation de noyaux, la rareté de la solution et le contrôle de la capacité obtenu en agissant sur la marge, ou sur le nombre de vecteurs de support, etc. La capacité du système est contrôlée par des paramètres qui ne dépendent pas de la dimension de l'espace des caractéristiques. Comme l'algorithme SVM opère nativement sur des attributs numériques, il utilise une normalisation de type z-score sur les attributs numériques. En régression, les algorithmes de Machines à Vecteur de Support utilisent la fonction de perte epsilon-insensibilité (marge de tolérance) pour résoudre les problèmes de régression.

Application

les algorithmes de régression des machines à vecteurs de support ont trouvé plusieurs applications dans l'industrie pétrolière et gazière, la classification d'images et de textes et la catégorisation d'hypertextes. Dans les champs pétrolifères, il est spécifiquement exploité pour l'exploration afin de comprendre la position des couches de roches et de créer des modèles 2D et 3D comme représentation du sous-sol.

2.3.5 Algorithme de régression multivariée

Cette technique est utilisée lorsqu'il y a plus d'une variable prédictive dans un modèle de régression multivariée et le modèle est appelé régression multiple multivariée. Considéré par les chercheurs comme l'un des algorithmes d'apprentissage automatique supervisé les plus simples, cet algorithme de régression est utilisé pour prédire la variable de réponse pour un ensemble de variables explicatives. Cette technique de régression peut être mise en œuvre efficacement à l'aide d'opérations matricielles et en Python, elle peut être mise en œuvre via la bibliothèque "numpy" qui contient des définitions et des opérations pour l'objet matrice.

Application :

L'application industrielle de l'algorithme de régression multivariée est très présente dans le secteur de la vente au détail où les clients font un choix sur un certain nombre de variables telles que la marque, le prix et le produit. L'analyse multivariée aide les décideurs à trouver la meilleure combinaison de facteurs pour augmenter la fréquentation du magasin.

2.3.6 Algorithme de régression multiple

Cet algorithme de régression a plusieurs applications à travers l'industrie pour la tarification des produits, la tarification immobilière, les départements de marketing pour découvrir l'impact des campagnes. Contrairement à la technique de régression linéaire, la régression multiple est une classe plus large de régressions qui englobe les régressions linéaires et non linéaires avec plusieurs variables explicatives.

Application

Certaines des applications commerciales de l'algorithme de régression multiple dans l'industrie sont dans la recherche en sciences sociales, l'analyse comportementale et même dans l'industrie des assurances pour déterminer la valeur des réclamations.

2.3.7 Algorithme des forêts aléatoires

La forêt aléatoire est un autre algorithme utilisé très couramment. Cet algorithme construit plusieurs arbres de classification et de régression (CART, Classification and Regression Tree), chaque arbre étant associé à différents scénarios et différentes variables initiales. L'algorithme est randomisé, ce qui n'est pas le cas des données. Ce type d'algorithme est utilisé pour la modélisation prédictive de classification et de régression.

Par exemple, vous disposez de 1000 observations sur une population, avec 10 variables. Pour construire le modèle CART à utiliser, l'algorithme de forêt aléatoire va prendre un échantillon aléatoire de 100 observations et cinq variables au hasard. L'algorithme répète ce processus à de nombreuses reprises, puis il fait une prédiction finale sur chaque observation. La prédiction finale n'est qu'une fonction correspondant à la somme des différentes prédictions.[38]

2.3.8 Algorithme Ridge Régression

Ridge Regression est un autre algorithme de régression linéaire couramment utilisé dans le Machine Learning. Si une seule variable indépendante est utilisée pour prédire la sortie, elle sera qualifiée d'algorithme de régression linéaire ML , Les experts en ML préfèrent la régression Ridge car elle . les valeurs de sortie sont prédites par un estimateur de crête dans la régression de crête.

La complexité du modèle ML peut également être réduite via la régression de crête.Il convient de noter que tous les coefficients ne sont pas réduits dans la régression de crête, mais cela réduit les coefficients dans une plus grande mesure par rapport aux autres modèles. La régression de crête est représentée par :

$$y = X + \in \tag{2.1}$$

y : est le vecteur N*1 définissant les observations du point de données variable dépendant X est la matrice des régresseurs.

β : est le vecteur N*1 constitué de coefficients de régression

$\in$: est le vecteur (N*1) d'erreurs.

L'algorithme Ridge est également utilisé pour la régression dans l'exploration de données par des experts en informatique en plus du ML.[39]

2.4 Travaux Connexes

2.4.1 Articla 1 :Soil salinity prediction using a machine learning approach through hyperspectral satellite image , Salim KLIBI , September 2020

2.4.1.1 L'objectif

Le but de cet article : améliorer les résultats de la classification d'images basée sur les pixels en utilisant différentes méthodes telles que SVM, KNN et DT et également utiliser l'AE pour améliorer l'efficience.

2.4.1.2 La méthode utilisé

Les modèles ont été appliqués dans deux expériences :

- La première expérience, appliques des modèles sur les valeurs de signature spectrale.
- la seconde expérience, les valeurs de signature spectrale ont été combinées avec features de vecteur. ensuit appliqué EA aux deux expériences.

2.4.1.3 Les résultats

le resultats après la comparaison entre les deux experiences , le résultat est le suivant :

	SS	SS + FV
SVM	76.7%	90.3%
AE +SVM	58.11%	98.7%
KNN	54.2%	69.5%
AE +KNN	41.9%	78.8%
DT	63.6%	84.7%
AE +DT	51.3%	89.4%

Table 2.1 – Comparaison entre les deux experiences

2.4.1.4 Les defis

- L'obtention d'une bonne précision de classification Lors de l'application AE
- Le modèle dans lequel la prédiction avec une bonne précision avant et apres application de AE est SVM .

2.4.1.5 Les limites

- Il y des problèmes de classification en modele SVM
- Il y quelques points mal classifiés pour ce modele KNN .
- Les valeurs de précision pour les trois classes ont été sous-estimées dans modele DT .

2.4.2 Article 2 : Machine learning and multispectral data-based detection of soil salinity in an arid region Central Iran ,Vahid Habibi, October 2020

2.4.2.1 L'objectif

L'objectif est de prédire la salinité du sol en utilisant l'arbre de régression et d'un réseau neuronal artificiel.

2.4.2.2 La methode utiliseé

Ils ont utilisé la méthode du supercube pour localiser les échantillons. Cette méthode est un schéma de classification aléatoire qui aboutit à un échantillonnage efficace au moyen d'une distribution multivariée.

2.4.2.3 Les résultats

les résultats de cette étude ont montré que la précision du modèle d'optimisation du réseau neuro-génétique GFF-GA dans la prédiction de la salinité du sol est plus élevée que les méthodes CART et M5 .

2.4.2.4 Les defis

- Minimiser l'erreur de prédiction et l'efficacite dans estimation et approximation
- L'optimisation des entrées, des tailles de pas et du nombre de nœuds dans chaque couche du réseau neuronal
- La capacité d'augmenter la précision des paramètres du modèle

2.4.2.5 Les limites

- Malgre le modèle de réseau neuronal ait une plus grande capacité , il exige des données de meilleure qualité et une grande taille d'échantillon, ces techniques n'ont pas un débit élevé et présentent de faibles performances.

2.4.3 Article 3 :Soil Salinity Mapping Using Machine Learning Algorithms with the Sentinel-2 MSI in Arid Areas China ,Jiaqiang Wang , December 2020

2.4.3.1 L'objectif

Le but de cette étude est de combiner les données Sentinel-2 Multispectral Imager (MSI) avec des données de salinité du sol mesurées et d'appliquer trois algorithmes d'apprentissage automatique dans la modélisation pour estimer et cartographier la salinité du sol dans la zone de l'étude.

2.4.3.2 La Methode utiliseé

appliquées trois methodes pour évaluer la salinité du sol par rapport à la CE du sol mesurée. à savoir, les machines à vecteurs de soutien (SVM), l'algorithme Random Forest (RF) et l'algorithme ANN.

2.4.3.3 Les résultats

Plus la valeur de RSquared est proche de 1, plus la valeur de précision d'ajustement du modèle est élevée alors Plus la prévisibilité est forte

Les résultats requis n'ont pas été obtenus avec le modèle RF ni le modèle ANN. mais le modele SVM, R2 était la valeur la plus élevée de 0,88, donc le SVM était le plus robuste des trois modèles.

2.4.3.4 Les limites

Il y a des valeurs aberrantes apparentes dans les valeurs estimées des échantillons de sol dans le modèle RF et ANN

2.4.3.5 Les defis

Facilité à traiter des données de haute dimension.

2.5 Conclusion

Dans ce chapitre nous avons présenté une vue globale sur l'apprentissage automatique , Nous avons comencé par une brève definition de l'apprentissage automatique , ensuite on l'a comparaison entre le apprentissage automatique et l'apprentissage profonde , Après la comparaison, nous avons conclu que les performances des algorithmes d'apprentissage automatique sont plus faibles que les performances des algorithmes d'apprentissage profond par rapport l'efficacité ..Nous avons cité Les principaux algorithmes de régression

du machine learning et leurs application et aussi présentrent quelques travaux connexes . Nous allons présenter dans le chapitre suivant notre conception de système proposée pour la prediction des sols salé .

Chapitre 3

Conception

Sommaire

3.1 Introduction

Le chapitre précédent nous a permis de comprendre et de situer clairement les notions de base nécessaire pour la conception et la réalisation de notre projet (la réalisation des predictions la salinite des sols par des modèles l'apprentissage automatique). Dans ce chapiter , on va d'écrire une conception générale et détaillée de notre système pour la prédiction de l'état des sols et délimiter les différents niveaux de salinité.

3.2 Collecte et l'analyse des données

Dans cette section , ont été traitées les images pour obtenir divers facteurs de modélisation, dont 10 bandes spectrales , aprés avoir effectuer aprés avoir effectuer l'analyse en composantes principales (ACP) sur les 10 bandes , et utilisé 13 indices, ces deniers sont calculer appartires des bandes spectrale par des oppérations arithmétique , De plus Un modèle d'élévation numérique ont été qui représente les alttitudes des différents points , tableau suivant (Tableau 1) illustre les indices utilisés.

Indice spectral	Acronyme
Soil Adjust Vegetation Index	SAVI
Normalized Difference Water Index	NDWI
Normalized Difference Vegetation Index	NDVI
Brightnesse Indes	BI
Moister Stresse Index	MSI
Normalized Difference Salinity Index	NDSI
Salinity Index 1	SI1
Salinity Index 2	SI2
Salinity Index 3	SI3
Salinity Index 4	SI4
Salinity Ratio	SR
Vegetation Soil Salinity Index	VSSI
Degital Elevation Model	DEM

TABLE 3.1 – Les formules de préparation des données [40]
Sentinel-2 MSI

3.3 L'architecture générale

Globalement, on peut représente l'architecture de notre système pour la prédiction des sols salés comme suit (Figure 3.1) :

3.3.1 La création de la base de donnée

Dans ce module on ramasse des images satellitaire,Cette image a été capturé à willaya de Biskra qui est ramassé à partir capteurs satellite sentinel-2 pour entraıner les modèles qui va nous aider à predire la salinité des sols .

3.3.2 La prétraitement

Les images capturées sont collectées et traitées afin d'obtenir une plus grande information et avec une grande précision , afin de prédire la salinité du sol .

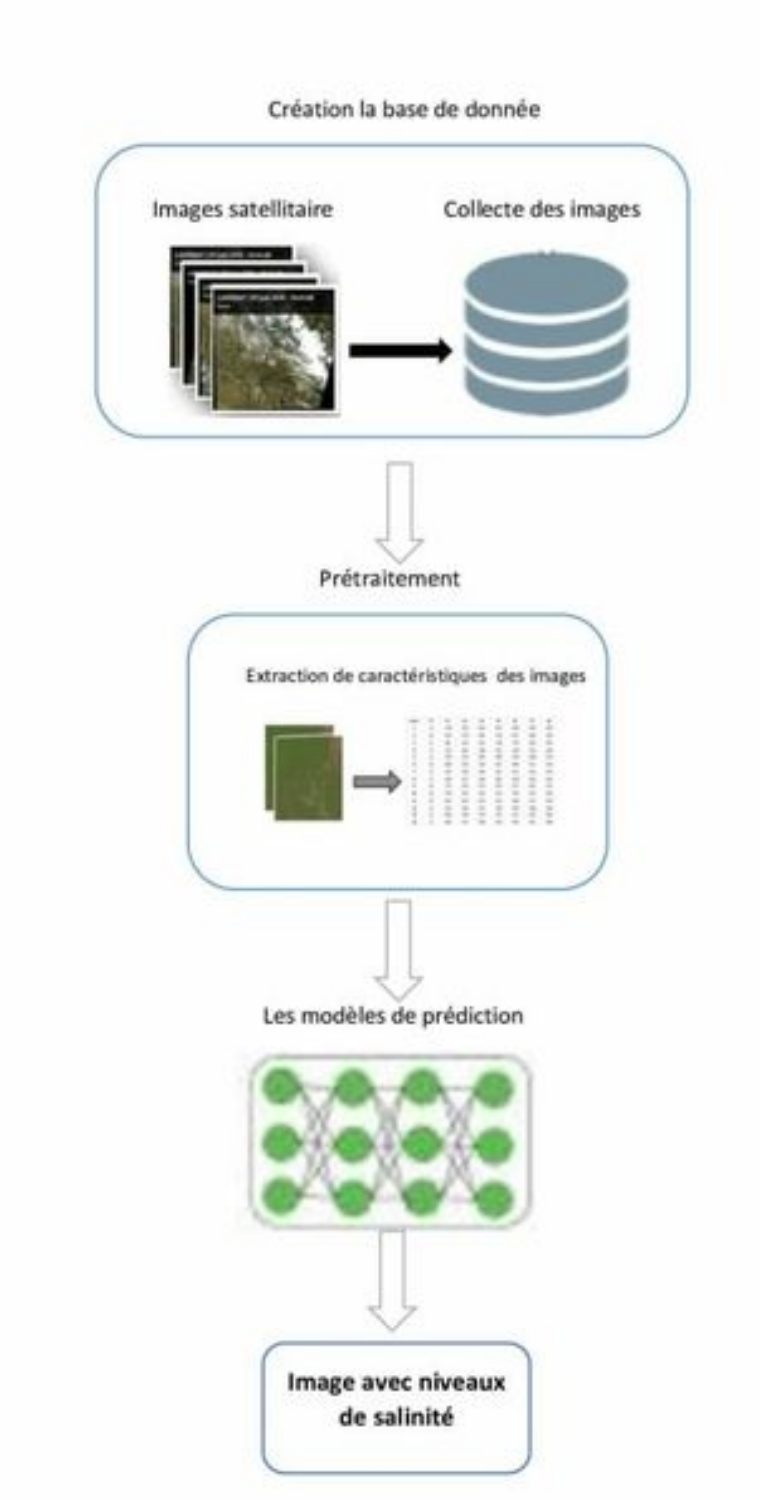

FIGURE 3.1 – L'Architecture globale du System

3.3.3 Les modèles de prédiction

Dans cette section , nous allons appliquant des algorithmes d'apprentissage automatique sur ces données et en comparant pour obtenir le modèle la plus précise après comparaison . Où le sol est connu comme les régions les plus salinité , cela son grace a l'étude de conductivité électrique (CE) , plus la valeurs de conductivité électrique est élevé, plus le sol salié est élevé

3.4 L'architecture Détaillée

Notre système est composé d'un ensemble des processus (étapes) .Il prent en entrée collection des images satellitaires pour prédire les zones salés par des models régression d'apprentissage automatique . La Figure ci-dessous(Figure 3.2) présente l'architecture detaillée de notre système qui est composée essentiellement de trois étapes suivantes :

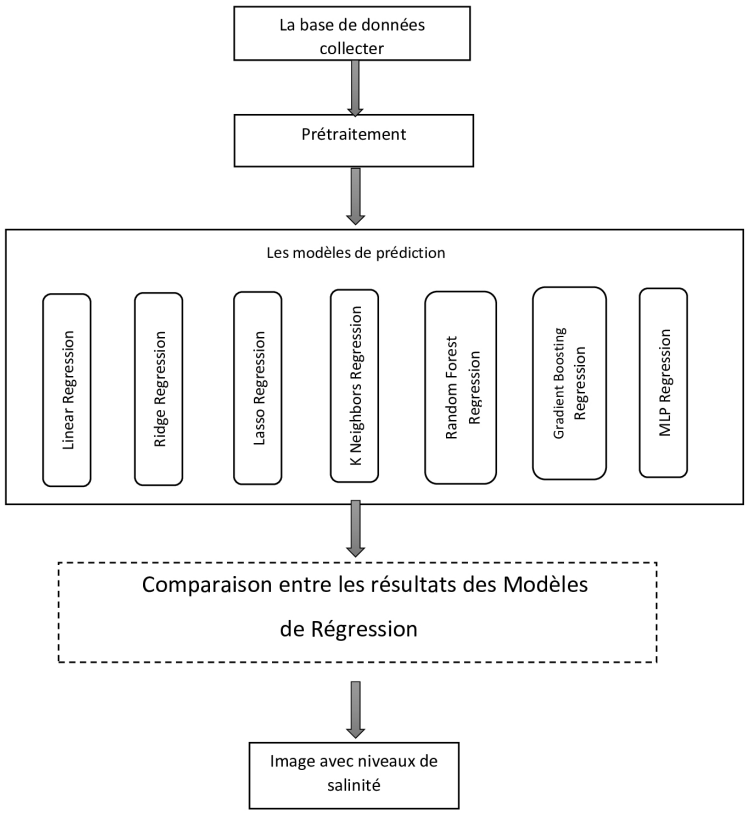

FIGURE 3.2 – L'Architecture Détaillée du System

3.4.1 L'etape 1 : le base de données (Dataset)

Dans cette etude , on a utilisé les images satellite sentinel-2 qui a été expliqué dans le premier chapitre qui peuvent être obtenues à partir de la source suivante de lESA (european space agency) , Cette image a été capturé à willaya de Biskra localisée au Sud-Est algérien, située à 115 Km au Sud-Ouest de Batna , à 222 Km au Nord de Touggourt et à 400 Km environ au Sud-Est d'Alger. et s'étend sur une superficie de prés de 21 509.80 km , sa latitude est de 34° 48' et sa longitude est de plus de 5° 44' . elle est située à une altitude moyenne de 87 m par rapport au niveau de la mer .Le climat de la région est de type

saharien,nous avons étudié cette zone en juillet et septembre et aout. Nous avons choisi cette période en raison de la situation géographique de cette zone . la pluie tombe et il contient des sels et s'accumuler sur le sols et lorsque la température est élevée , ce qui conduit à l'évaporation et à l'assèchement de la terre alors cela crée une couche de salinité des sols , La tableau représente l'image de la zone étudiée (Figure 3.3) [41].

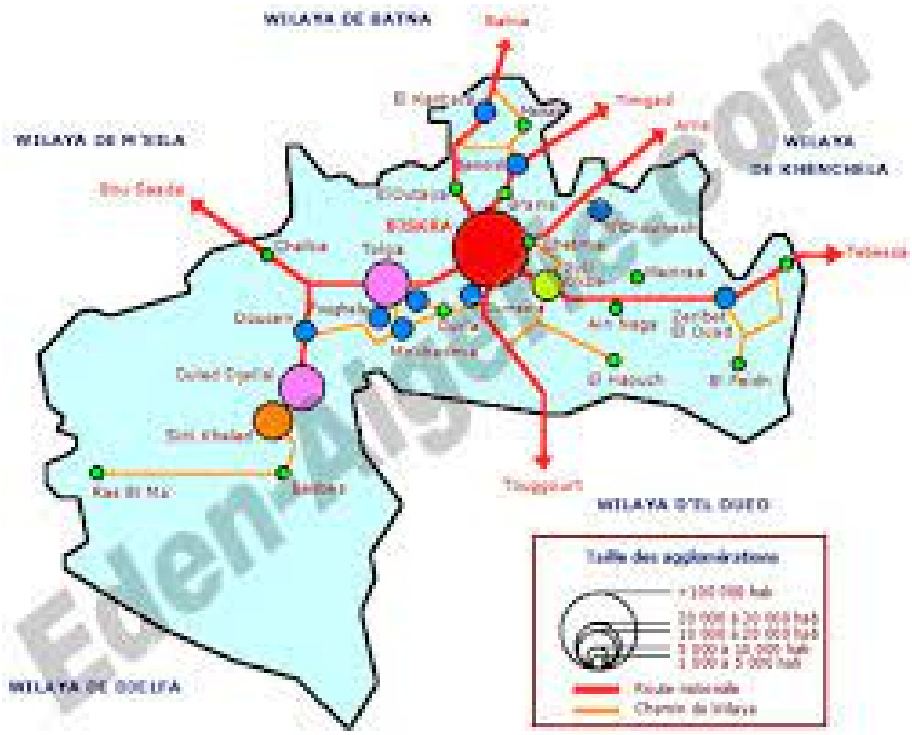

FIGURE 3.3 – Zone D'etude - Biskra [42]

3.4.2 L'etape 2 : Le prétraitement :

Dans cette phase, les données collectées ont été traitées pour obtenir divers facteurs de modélisation , nous avons extrait les informations de chaque points des bandes des images collectées afin d'obtenir des informations précises. et cela en calculent les indices de sols salé de chaque point pour chaque bande , et aussi calculé la conductivité electrique de chaque point . illustre l'image ci-dessous(Figure 3.4).

Comme nous l'avons évoqué précédent, le Sentinel-2 est équipé de l'imageur multispectral (MSI). Ce capteur délivre 13 bandes spectrales . Le tableau ci-dessous représente les caractéristiques de ces bands (longueur d'onde, résolution d'image) (Tableau 3.2).

3.4.2.1 Les Bandes d'images Sentinel 2 et principe de fusion

Pour mieux comprendre les caractéristiques des images, nous pouvons extraire des informations des images par fusion des bandes . illustre l'image ci-dessous(Figure 3.5).[44]

Points	CE	B2	B3	B4	B5	B6	B7	B8	B8A	B11	B12
1	2,3	1738	2670	3632	4107	4180	4318	4196	4421	6040	5470
2	3,4	1970	2750	3482	3746	4017	4066	4160	4290	5146	4400
3	3,5	883	1410	1632	2248	2778	3020	2956	3221	3162	2265
4	3,7	1324	1876	2388	2921	3390	3526	3400	3747	4094	3299
5	2,5	2330	3120	3892	4288	4334	4484	4396	4611	5419	3785
6	1,8	1868	2700	3704	4057	4099	4227	4140	4329	5270	3954
7	2,3	2310	3238	4196	4441	4422	4532	4736	4656	5675	4240
8	3,3	1380	2072	2940	3987	4078	4180	3416	4336	5457	4752
9	3,6	1722	2522	3574	4374	4471	4615	4132	4745	5825	5066
10	3,8	1730	2610	3564	4226	4303	4435	4100	4589	5663	4551
11	3,8	2050	3034	4124	4553	4580	4731	4704	4862	5821	4782
12	3,4	1674	2558	3574	3768	3993	4188	4140	4349	4924	3877

FIGURE 3.4 – Collection des Bands (Dataset)

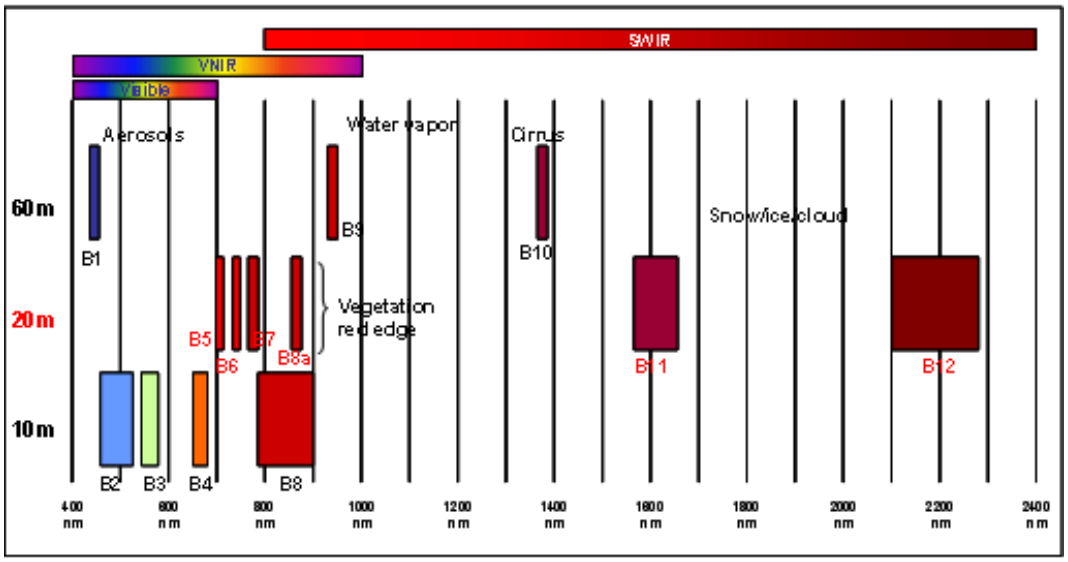

FIGURE 3.5 – Combinaisons des bands [45]

bandes sentinelle-2	région spectrale	l'onde centrale	résolutions(m)
Band1	aérosol côtier	0.443	60
Band2	bleu	0.490	10
Band3	vert	0.560	10
Band4	rouge	0.665	10
Band5	la végétation "red edge"	0.705	20
Band6	la végétation "red edge"	0.740	20
Band7	la végétation "red edge"	0.783	20
Band8	NIR	0.842	10
Band8A	la végétation "étroit"	0.865	20
Band9	vapeur d'eau	0.945	60
Band10	infrarouge à ondes courtes (SWIR)-cirrius	1.375	60
Band11	infrarouge à ondes courtes 1 (SWIR)	1,610	20
Band12	infrarouge à ondes courtes 1 (SWIR)	2.190	20

TABLE 3.2 – Les bandes spéctrales [43]

- **Couleur naturelle (B4, B3, B2)**

  La fusion de bandes des couleurs naturelles nous utilise les bandes rouge (B4), vert (B3) et bleu (B2). Son but est d'afficher des images de la même manière que nos yeux voient le monde.

- **Couleur Infrarouge (B8, B4, B3)**

  La combinaison de bandes infrarouges de couleur est destinée à souligner la végétation saine et malsaine. En utilisant la bande proche infrarouge (B8) .

- **Infrarouge à ondes courtes (B12, B8A, B4)**

  La combinaison de bandes infrarouges à ondes courtes utilise le SWIR (B12), le NIR (B8A) et le rouge (B4). Ce composite montre la végétation dans diverses nuances de vert.

- **Agriculture (B11, B8, B2)**

  Il est principalement utilisé pour surveiller la santé des cultures en raison de la façon dont il utilise les ondes courtes et le proche infrarouge. Ces deux bandes sont particulièrement efficaces pour mettre en évidence une végétation dense qui apparaît en vert foncé.

- **Géologie (B12, B11, B2)**

  La combinaison de bandes de géologie est une application intéressante pour trouver des caractéristiques géologiques. Cela comprend les failles, la lithologie et les formations géologiques.

- **Bathymétrique (B4, B3, B1)**

  la combinaison de bandes bathymétriques est bonne pour les études côtières. La combinaison de bandes bathymétriques est utilisée pour estimer les sédiments en suspension dans l'eau.

3.4.2.2 La Conductivité électrique (CE)

Cette technique est utilisée pour mesurer la salinité des sols. Ainsi, les sols non salés ont des conductivité variant entre 0 et 50mS/m et les sols salés entre 100 et 200mS/m.

La conductivité électrique du sol peut être mesurée pour chaque point d'un terrain en prélevant un échantillon du sol à chaque point et elle est mesurée en laboratoire. Nous avons également utilisé des indicateurs spectraux tels que la salinité NDVI, en liant mesures terrestres et données satellitaires, afin de classer la salinité des sols.[46]

3.4.3 L'etape 3 :les algorithmes d'apprentissage automatique

Dans cette phase , nous appliquerons des algorithmes de regression de l'apprentissage automatique aux données utilisées . Pour que tous les algorithmes fonctionnent sur les mêmes entrées représentées dans dataset image satellitaire pour la zone étudiée, dont nous avons parlé dans la phase de pretraitement et la prédiction est faite sur la même sorties qui est la conductivité électrique (CE).

L'objectif de tout algorithme d'apprentissage supervisé est de définir une fonction de perte et de la minimiser .

3.4.3.1 Algorithme MLP regression

MLP est un algorithme d'apprentissage tel que la rétropropagation de gradient qui est appliqué pour ajuster les poids en réduisant au minimum une fonction d'erreur RMSE [47], illustre l'image ci-dessous(Figure 3.6).

principalement il est composé de :

- Couche d'entrée : représente les bandes spectrales
- La couche cachée : dans laquelle la fonction d'activation et la minimisation des erreurs RMSE sont calculées.
- Couche de sortie : représentée la prédiction de CE.

3.4.3.2 Algorithme Gradient Boosting Régression

Le but de tout algorithme d'apprentissage supervisé est de définir et de réduire la fonction de perte F0(x). Nous voulons que nos prédictions soient précises pour que la fonction de perte (MSE) est minime [49] . illustre l'image ci-dessous(Figure 3.7)

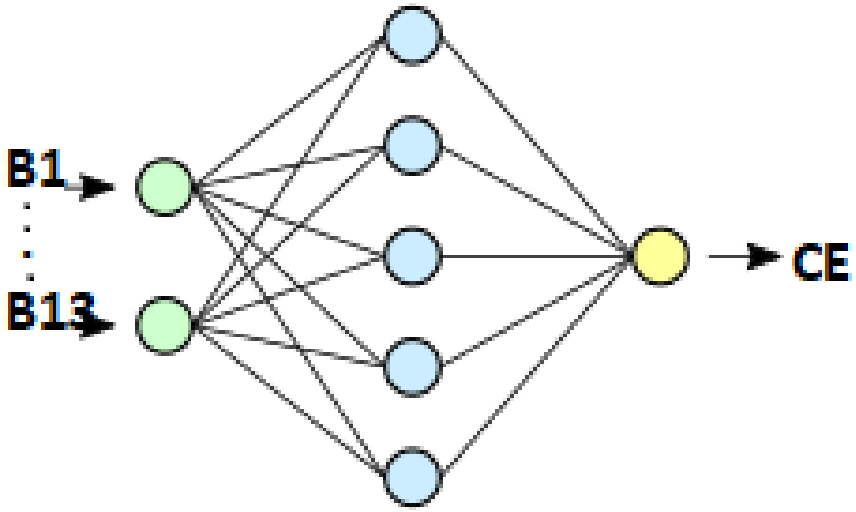

FIGURE 3.6 – Chemin algorithme MLP régression [48]

Algorithm 1: Gradient Boost

1 $F_0(x) = \arg\min_\rho \sum_{i=1}^{N} L(y_i, \rho)$

2 For m = 1 to M do:

3 $\tilde{y}_i = -[\frac{\partial L(y_i F(x_i))}{\partial F(x_i)}]_{F(x)=F_{m-1}(x)}, i = 1, N$

4 $a_m = \arg\min_{a,\beta} \sum_{i=1}^{N} [\tilde{y}_i - \beta h(x_i; a)]^2$

5 $\rho_m = \arg\min_{a,\beta} \sum_{i=1}^{N} L(y_i, F_{m-1}(x_i) + \rho h(x_i; a_m))$

6 $F_m(x) = F_{m-1}(x) + \rho_m h(x; a_m)$

7 end For

8 end Algorithm

FIGURE 3.7 – Algorithme Gradient Boosting Régression [50]

* F(x) : fonction des erreurs(Loss)
* yi : les bands spectrale
* ρ : CE

3.4.3.3 Algorithme Régression linéaire

Dans cette algorithme , utiliser les valeurs de X permet de prévoir les valeurs de Y , le but de minimiser l'erreur de prévision [51].

$$Y = mX + C \tag{3.1}$$

$$MSE = 1/n \cdot \sum_{i=1}^{n} (X - Y)^2 \tag{3.2}$$

MSE : Fonction de perte X :Les entrées qui sont les bandes spéctrales .

Y : C'est la sortie qui est les données prévu (CE)

3.4.3.4 Algorithme Ridge Régression

Dans cet algorithme , les valeurs de sortie sont prédites par un estimateur de crête dans la régression de crête.

Il convient de noter que tous les coefficients ne sont pas réduits dans la régression de crête . La régression de crête est représentée par :[52]

$$y = X\beta + \epsilon \tag{3.3}$$

X :représente les bands spéctrales .

Y : est les données prévu (CE)

3.4.3.5 Algorithme Lasso Régression

le but de lasso est minimiser la complexité du modèle en limitant la somme des valeurs absolues des coefficients du modèle (L1).[53]

La fonction de perte pour la régression au lasso peut être exprimée comme suit :

Fonction de perte est :

$$y_i = W_0 + \sum_{i=1}^{m} X_j W j[15] \tag{3.4}$$

X :représente les bands spéctrales .

yi : C'est la sortie qui est la donnée prévu (CE)

3.4.3.6 Algorithme Random Forest Régression

Random Forest est une technique qui combine les prédictions de plusieurs algorithmes d'apprentissage automatique pour faire des prédictions plus précises que n'importe quel modèle individuel.illustre l'image ci-dessous(Figure 3.8)

data : représente les bandes spectrales.[54]

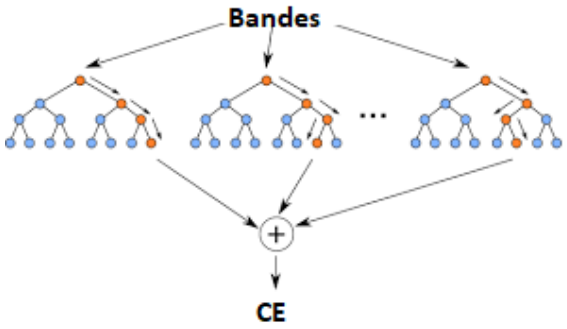

FIGURE 3.8 – Chemin algorithme random forest régression .[55]

3.4.3.7 Algorithme K Nearest Neighbors

Dans cet algorithme, la prédiction par un vote la majorité des voix de ses voisins, le cas étant attribué à la classe la plus courante parmi ses K voisins les plus proches mesurés par une fonction de distance [56]., illustre l'image ci-dessous(Figure 3.9 et 3.10)

où Xi représente les entrée (les bandes spéctrales) et Yi le sortie prévu (CE)

$$\textbf{Euclidean} : \sqrt{\sum_{i=1}^{k}(x_i - y_i)^2} \tag{3.5}$$

$$\textbf{Manhattan} : \sum_{i=1}^{k}|x_i - y_i| \tag{3.6}$$

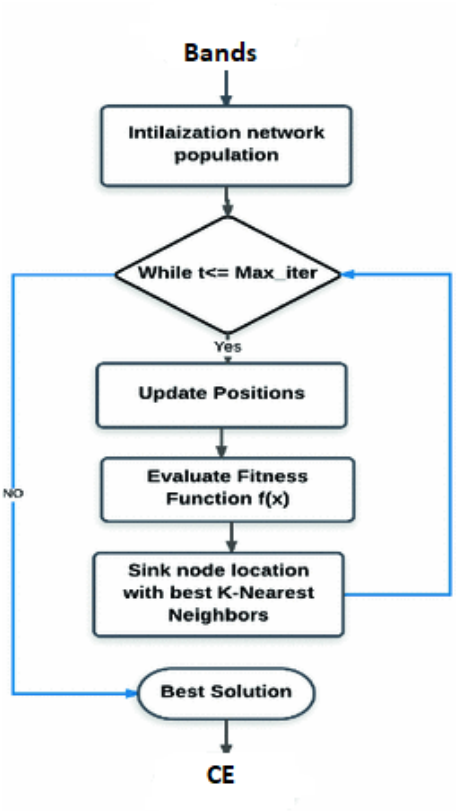

FIGURE 3.9 – Algorithme KNN [58]

3.4.4 L'etape 4 :La Comparaison entre les Algorithmes de regression

Dans cette étape, nous comparerons les algorithmes du regression en termes d'efficacité et de précision dans la partie test et entraînement, nous conclurons par une comparaison des resultats de precision de ces algorithmes dans le chapitre suivant .

3.4.5 L'etape 5 :image avec niveaux de salinité

Dans cette étape , on obtient les images avec niveaux de salinité ça dépend de la valeur de la conductivité électrique (CE)

- si CE inférieur à 4 d/m, le sol peut être salé
- si CE entre 4 dc/m et 8 dc/m, sol salé
- si CE supérieur à 8 d/m, sol très salé (ds/m = critique par mètre, c'est l'unité de mesure utilisée pour la conductivité électrique)

3.5 Conclusion

Dans ce chapitre, nous avons détaillé les étapes à suivre pour la réalisation mettre directant l'objectif de projet . Nous avons aussi présenté les algorithmes utilisé et comment les appliquer pour prédire le niveau de la salinite basés sur image satellite. Le chapitre suivant est consacré à l'implémentation du notre système ainsi qu'aux résultats obtenus .

Chapitre 4

Implémentation

Sommaire

4.1 Introduction

Dans ce chapitre, nous présentons la mise en œuvre de notre application. En commençant tout d'abord par une présentation des outils de développement ; langage de programmation choisi (python) et le matériel (PC). Nous proposons les implémentations des models regression du machine learning sur notre dataset . Ensuite nous présenterons résultats obtenus . Nous terminerons le chapitre par une étude comparative entre les algorithmes proposées.

4.2 Présentation des environnements de développement utilisés

4.2.1 Environnement logiciel

- **Google Colab :** est un service cloud, offert par Google (gratuit), basé sur Jupyter Notebook et destiné à la formation et à la recherche dans l'apprentissage automatique. Cette plateforme permet d'entraîner des modèles de Machine Learning directement dans le cloud.[59]

4.3 Principaux outils utilisés

4.3.1 Python

Python est un langage de programmation orienté objet, interprété et interactif. Il est souvent comparé (favorablement bien sûr) à Lisp, Tcl, Perl, Ruby, C , Visual Basic, Visual Fox Pro, Scheme ou Java ... etc. Python combine un pouvoir remarquable avec une syntaxe très claire. Il comporte des modules, des classes, des exceptions, des types de données dynamiques de très haut niveau et le typage dynamique. Il existe des interfaces vers de nombreux appels systèmes et bibliothèques, ainsi que vers différents systèmes de fenêtrage .[60]

4.3.2 Tensorflow

Tensorflow est un outil open source lancé en 2015 par Google pour aider les développeurs à concevoir, créer et entrainer des modèles de deep learning. Une version stable 1.0 a été annoncée en février 2017. Tensorflow est multiplateforme, il fonctionne sur les GPU, les CPU (central processing unit) et même les TPU (TensorFlow Processing Unit) qui sont des matériels spécialisés pour effectuer des calculs sur des tenseurs et permettant ainsi de faire profiter dune importante accélération .[61]

- Avantages
  - supporté par Google
  - Une très grande communauté
  - Le support du multi-GPU
- Inconvénients
  - Plus lent que les autres framework dans de nombreux benchmarks, bien que Tensorflow se rattrape.

4.3.3 Keras

Keras est une API de rseaux neuronaux de haut niveau, écrite en Python et capable de s'exécuter sur TensorFlow, CNTK ou Theano. Il a ét dveloppé dans le but de permettre une expérimentation rapide. Être en mesure de passer de l'idée au résultat le plus rapidement possible, est la clé pour faire de la recherche :

- Permet un prototypage facile et rapide (grâce à la convivialité, à la modularité et à l'extensibilité).
- Prend en charge les réseaux convolutionnels et les réseaux récurrents ainsi que les combinaisons des deux.
- Fonctionne de manière transparente sur le processeur et le processeur graphique. [62]
- Avantages
  - Python
  - Le backend par excellence pour TensorFlow —Interface haut niveau, intuitive
- Inconvénients
  - Moins flexible que les autres API

4.3.4 Pandas

Pandas est une librairie python qui permet de manipuler facilement des données à analyser [63] :

- manipuler des tableaux de données avec des étiquettes de variables (colonnes) et d'individus (lignes).
- ces tableaux sont appelés DataFrames, similaires aux dataframes sous R.

- on peut facilement lire et écrire ces dataframes à partir ou vers un fichier tabulé.
- on peut faciler tracer des graphes à partir de ces DataFrames grâce à matplotlib.

Pour utiliser pandas : import pandas

4.3.5 NumPy

NumPy est une bibliothèque pour le langage de programmation Python, ajoutant un support pour les matrices et matrices multidimensionnelles de grande taille, ainsi qu'une grande collection de fonctions mathématiques de haut niveau pour fonctionner sur ces matrices. L'ancetre de NumPy, Numeric, a été crée à l'origine par Jim Hugunin avec des contributions de plusieurs autres d´eveloppeurs. En 2005, Travis Oliphant a créé NumPy en incorporant les fonctionnalités de Numarray en Numeric, avec de nombreuses modifications. NumPy est un logiciel open-source et compte de nombreux contributeur .[64]

4.3.6 Matplotlib

Matplotlib est une bibliothèque de traçage disponible pour le langage de programmation Python en tant que composant de NumPy, une ressource de traitement numérique de données volumineuses. Matplotlib utilise une API orientée objet pour intégrer des tracés dans les applications Python.[65]

4.4 L'implémentation

4.4.1 Importe les bibliothèques et le modèles

Pour construire des réseaux de neurones profonds avec Keras et la création les modèles de regression d'apprentissage automatique avec sklearn.linear_model , nous importons d'abord les différentes bibliothèques et modules ,comme illustré tableau 1 et (Figure 4.1) .

Le tableau 1 montre les fonctions et les couches utilisée dans cette modèle , où la première section définit la série à travers laquelle le modèle est créé, aussi il montre les couches ce utilisées qui sont (Dense , Dropout), Ce dernier est la fonction d'activation permet de changer notre manière de voir une donnée dans ce modèle ,

```
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np
from sklearn import preprocessing
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split
from keras.callbacks import EarlyStopping, ModelCheckpoint
from keras.layers import Dense, Dropout
from keras.layers.core import Activation
from keras.models import Sequential
import seaborn as sns
from fractions import Fraction
import keras
import keras.backend as K
from math import sqrt
from keras import Input, Model
from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score
from keras.models import load_model
from sklearn.metrics import mean_squared_error, r2_score
from sklearn.model_selection import KFold, cross_val_score
from sklearn.neural_network import MLPRegressor
from sklearn.linear_model import LinearRegression, Ridge, Lasso
from sklearn.neighbors import KNeighborsRegressor
from sklearn.tree import DecisionTreeRegressor
from sklearn.ensemble import (RandomForestRegressor, GradientBoostingRegressor,
                              AdaBoostRegressor)
from sklearn.inspection import permutation_importance
```

FIGURE 4.1 – Importer le code des bibliothèques

modules	Descriptions
Sequential	Pour créer des modèles d'apprentissage en profondeur où une instance de la classe Sequential est créée et des couches de modèle sont créées et ajoutées à celleci. Un modèle séquentiel est approprié pour une pile simple de couches où chaque couche a exactement un tenseur d'entrée et un tenseur de sortie.
Couches Keras	* **Dense :** Est utilisé pour instancier une couche dense * **Dropout :** Applique dropout à l'entrée. Dropout consiste à définir de manière aléatoire un taux de fraction d'unités d'entrée à 0 à chaque mise à jour pendant le temps d'entraînement, ce qui permet d'éviter le surapprentissage
Activation	Permet d'ajouter une fonction d'activation. à la séquence de couches

TABLE 4.1 – Les formules de préparation des données

4.4.2 La basee de donnée Utilisé :

Elle se compose de 13 bands (B1, B2 , B3 , B4 , B5 , B6 , B7 , B8 , B8A , B9 , B10 , B11 , B12) en niveau de gris different , une colone conductivité électrique (CE) et ensemble d'indices VSSI,SR,SMI,SI1,SI2,SI3 SI4,SAVI NDWI, NDVI,NDSI,MSI,BI,DEM (présentée dans chapiter 3 section 1) .comme illustré (Figure 4.2)

```
#df = pd.read_excel("/content/drive/MyDrive/MAT_24_09_19.xlsx")
df = pd.read_excel("/content/drive/MyDrive/MATRICE.xlsx")
print(df)
```

FIGURE 4.2 – Chargement la base de donnée

Nous avons utilisé la classe sklearn.preprocessing.MinMaxScaler pour transformez les entités en adaptant chaque entité à une plage donnée et la transmettre à notre ensemble de données par la fonction fit_transform() pour créer une version transformée de notre ensemble de données , la figure suivant (Figure 4.3) illustre l'importation de la base de donnée .

Pour tester les modèles utilisés , nous allons utilisé la fonction x_train, x_test, train_test_split() , dont permet de défnir la proportion de l'ensemble de données de test et les données de train . On obtient ainsi l'ensemble d'entraînement X_train avec ses étiquettes y_train et celui de test X_test avec les siennes y_test , comme illustré (Figure 4.4).

```
min_max_scaler = preprocessing.MinMaxScaler(feature_range=(0, 1))
Y=min_max_scaler.fit_transform(Y.reshape(-1,1))
print(Y)
```

FIGURE 4.3 – Chargement la base de donnée

```
x_train, x_test, y_train, y_test = train_test_split(X, Y, test_size=0.25,
                                   random_state = 1,shuffle=True)
print(x_train_scaled.shape)
print(x_train_scaled.shape)
print(val_X.shape)
print(val_Y.shape)
```

FIGURE 4.4 – Les données test et train

4.5 Les processus des Modèles de prédiction

4.5.1 Création du modèle MLP :

Après avoir chargé les ensembles de données et initialisé les hyper-paramètres, nous avons utilisé Tensorflow pour la création de modèle , le figure suivant décrit les couches du modèle MLP (Figure 4.5).

```
model = Sequential()
model.add(Dense(28, input_dim=28 , activation='relu'))
model.add(Dropout(0.25))
model.add(Dense(60, activation='relu'))
model.add(Dense(1, activation='sigmoid'))
```

FIGURE 4.5 – Création de modèle

4.5.1.1 Compilation du modèle

une fois le modèle créé , on peut configurer le modèle à l'aide de la fonction Model.compile() qui prend trois arguments : l'optimiseur, la fonction de perte et les métriques,on a utilisé :

- keras : l'algorithme d'optimisation a été utilisé pour entraîner le réseau. Il s'agit d'une méthode de descente de gradient stochastique, combinée à un taux d'apprentissage de 0,001.
- loss : la fonction de perte du coefficient est notre fonction de perte qui mesure les performances de la segmentation du modèle, la sortie est comprise entre 0 et 1.

- Le coefficient Dice est la métrique adoptée pour évaluer les performances du modèle.

```
model.compile(loss='mean_absolute_error',  optimizer=keras.optimizers.RMSprop(),
              metrics=['mae',r2_keras, root_mean_squared_error,'mean_squared_error'])
history =model.fit(x_train,y_train, epochs=200, batch_size=16, validation_data=(x_test,y
                   shuffle=True, verbose=2, callbacks = [early_stopping, checkpoint])
print(history.history.keys())
```

FIGURE 4.6 – Illustration de la fonction de modèle

```
Epoch 1/200
16/16 - 2s - loss: 0.2768 - mae: 0.2768 - r2_keras: -2.3921e-01 - root_mean_squared_error: 0
INFO:tensorflow:Assets written to: MLP/assets
Epoch 2/200
16/16 - 0s - loss: 0.2342 - mae: 0.2342 - r2_keras: 0.0717 - root_mean_squared_error: 0.2342
INFO:tensorflow:Assets written to: MLP/assets
Epoch 3/200
16/16 - 0s - loss: 0.1998 - mae: 0.1998 - r2_keras: 0.2308 - root_mean_squared_error: 0.1998
INFO:tensorflow:Assets written to: MLP/assets
Epoch 4/200
16/16 - 0s - loss: 0.1811 - mae: 0.1811 - r2_keras: 0.2721 - root_mean_squared_error: 0.1811
INFO:tensorflow:Assets written to: MLP/assets
Epoch 5/200
16/16 - 0s - loss: 0.1733 - mae: 0.1733 - r2_keras: 0.1985 - root_mean_squared_error: 0.1733
INFO:tensorflow:Assets written to: MLP/assets
Epoch 6/200
```

FIGURE 4.7 – Compilation du modèle

4.5.1.2 Visualisation des résultats

dans nos résultats expérimentaux , nous avons le tracé des courbes qui représentent le développement des métriques au cours du processus d'apprentissage (coefficients de loss , MAE et RMSE) à la fois pour les étapes d'entraînement et de test (voir Figure 4.8).

D'après la Figure 8 , l'erreur train et test diminue considérablement dans intervalle de nombre d'époque [0-5] , mais dans l'intervalle [5-34], le test et le train commence à diminuer progressivement, proche de la stabilité , et dans l'intervalle [34-40] le test reste stationnaire dans le niveau (0.14) et train se stabilise au niveau (0.13) .

4.5.1.3 Le prédiction du modèle MLP

—**La Prédiction :** cette phase permet de faire des prédictions sur les données de test, afin d'estimer et d'évaluer le modèle s'il est performant et a bien appris ,la figure suivante illustre les resultats de prédiction (Figure 4.9) .

Les fonctions utilisées dans la phase de prédiction sont présentées comme suit :

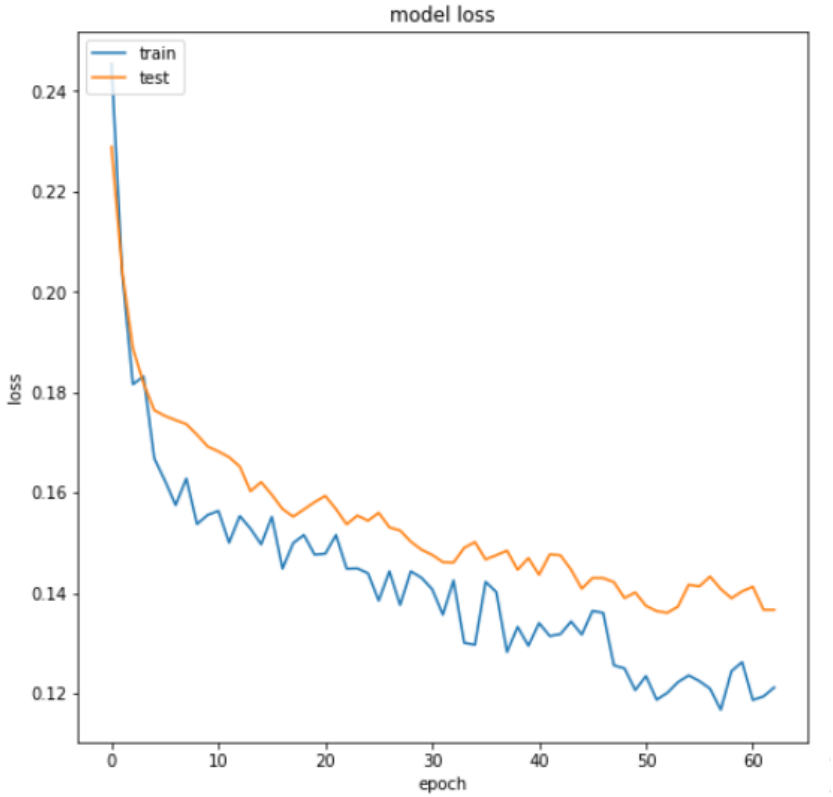

FIGURE 4.8 – Illustration resultat loss de modèle MLP

```
y_val_pred = model.predict(x_test)
y_train_pred = model.predict(x_train)
print(len(y_train_pred))
print(len(y_val_pred))
```

FIGURE 4.9 – Illustration de l'étape de prédiction

- Fonction changement de modèle : nous chargeons le modèle le mieux entraîné avec les poids enregistrés de l'étape précédente .
- Lfonction Predict : elle prend en entrée l'ensemble de test ou les données que l'on voulait prédire., les époques et l'étiquette exacte sur laquelle on souhaite également prédire.

— **resultat de La Prédiction :** pour entraîner le modèle MLP, nous avons formé le modèle sur 250 points , meilleur coefficient d'erreur obtenu était de 0.09 pour l'entraînement VS 0.13 pour les tests ,contrairement à la valeur de $R\hat{2}$, c'est a dire le meilleur coefficient $R\hat{2}$ obtenu était de 0.50 pour le test VS 0.74 pour entrainent , comme illustré (Tableau 2) .

	Train	Test
RMSE	0.09	0.13
$R\hat{2}$	0.74	0.50

TABLE 4.2 – Illustration de RMSE et $R\hat{2}$ de prédiction

4.5.2 Création des modèle de régression

les fonctions du modèle de regression varient selon le type de relation entre les variables , nous allons utiliser les modèles de regression et leur implémenter sur nos données . Après avoir importé les modèles de regression comme le montre (Figure). Nous nous avons crée une fonction append() pour ajouter les scores de validation croisée des algorithmes .

4.5.2.1 L'évaluation train des modèles de régression

Les modèles de régression sont entraînés et testés a partire du coefficient d'erreur RMSE et à partir de mesure r-squared qui représente la qualité de l'ajustement d'un modèle de régression , Plus la valeur de r-squared est proche de 1, meilleur est le modèle ajusté. et aussi le RMSE . Plus la valeur est proche de 0, meilleur est le modèle ajusté , comme illustré figure (Figure 4.10).

	Model	RMSE	R Squared
0	Linear Regression	2.866181	0.668560
1	Ridge Regression	2.847333	0.674256
2	Lasso Regression	4.270636	0.276187
3	K Neighbors Regressor	3.597249	0.478345
4	Decision Tree Regressor	2.820964	0.700297
5	Random Forest Regressor	2.217076	0.819914
6	Gradient Boosting Regressor	1.738495	0.881055
7	Adaboost Regressor	1.505324	0.888594
8	MLP Regressor	3.880014	0.400867

FIGURE 4.10 – Illustration RMSE et r-squared du Train

Le figure 10 montre le coefficient d'erreur RMSE et R-squared pour chaque modèle . et à travers laquelle déterminée la précision de modèle . à travers les résultats obtenus. La précision de prédiction la plus élevée est pour les modèles 'Gradient Boosting Regressor' et 'Adaboost Regressor' . Où le coefficient d'erreur est faible et le R-squared est proche de 1 . Où les résultats de 1.73 pour les premier modèle et deuxième modèle étaient de 1,50 pour l'erreur.

4.5.2.2 Gradient Boosting Régression avec autres paramètres

La coefficient d'erreur obtenu dans cette algorithme est 2.40 pour le test et lorsque la valeur R$\hat{2}$ obtenue était est 0.63 , comme illustré (Figure 11 et 12).

```
fig = plt.figure(figsize=(6, 6))
plt.subplot(1, 1, 1)
plt.title('Deviance')
plt.plot(np.arange(params['n_estimators']) + 1,
        reg.train_score_, 'b-',label='Training Set Deviance')
plt.plot(np.arange(params['n_estimators']) + 1,
        test_score, 'r-',label='Test Set Deviance')
plt.legend(loc='upper right')
plt.xlabel('Boosting Iterations')
plt.ylabel('Deviance')
fig.tight_layout()
plt.show()
```

FIGURE 4.11 – Illustration de la fonction de visualisation du courbe de Déviance

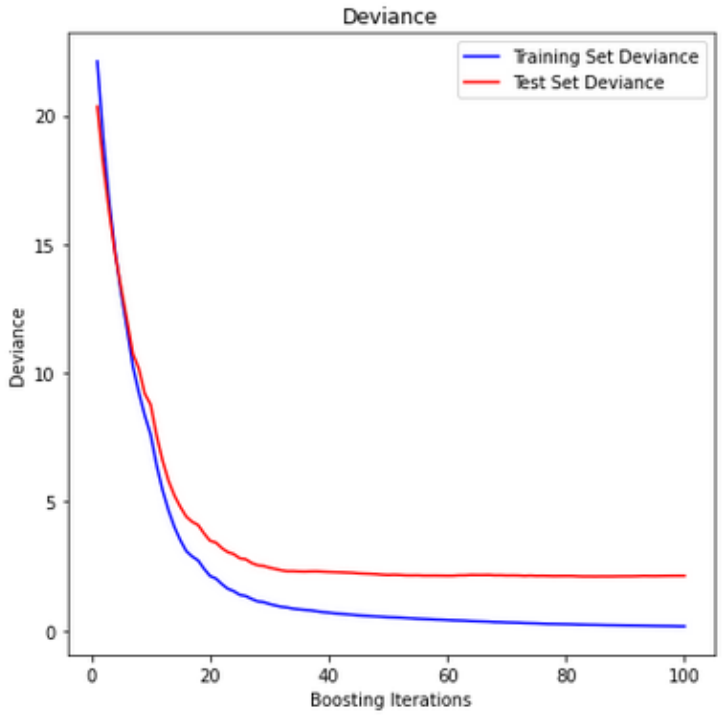

FIGURE 4.12 – Illustration de visualisation du courbe de Déviance

D'après la Figure 11 , déviance de train et de test diminue dans l'intervalle du nombre d'itérations [0-35] et dans l'intervalle [35-100], le test se stabilise dans le niveau (2.5) et train reste stationnaire au niveau (0.6) jusqu'à il prend la valeur zéro dans le point 100 .

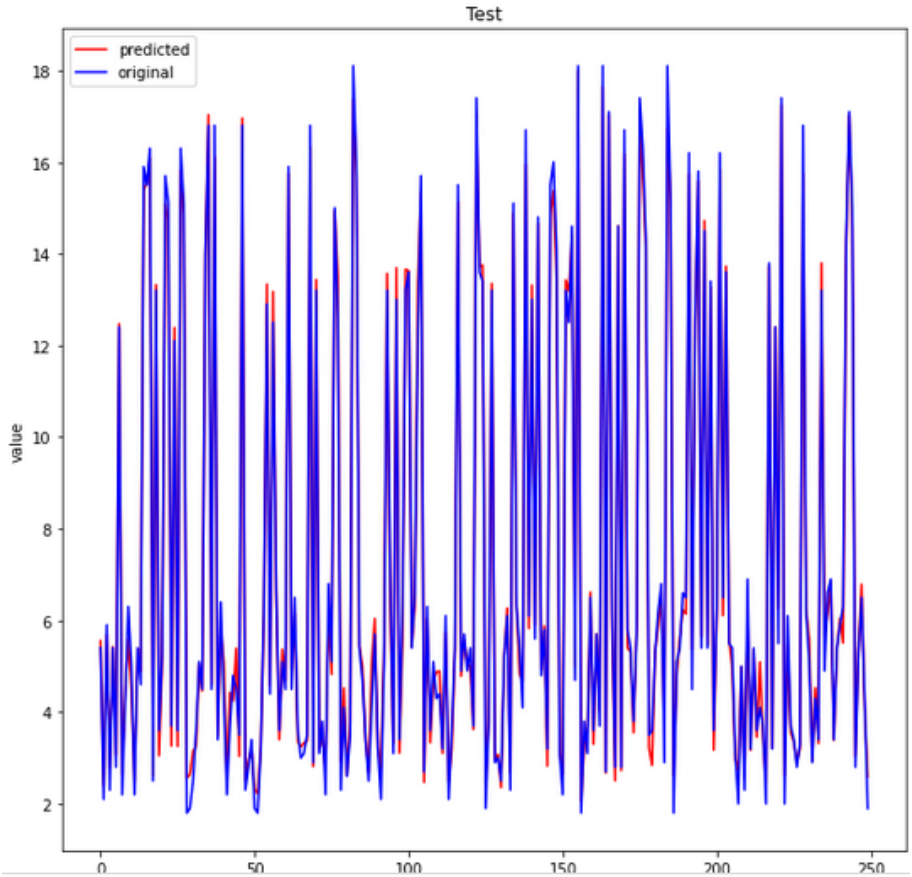

FIGURE 4.13 – Illustration de la prédiction du test

La figure 13 représente la précision de prédiction du test pour les modèles de régression. Où nous remarqué que la précision de la prédiction est très proche de l'original et donc le résultats de ce modèles plus précision .

4.5.3 l'importance des caractéristiques et la permutation des données

L'importance des caractéristiques peuvent donner un aperçu de l'ensemble de données. Les scores relatifs peuvent mettre en évidence quelles caractéristiques peuvent être les plus pertinentes pour le but . On remarque sur la figure que CP2 est le plus pertinent pour le but par rapport au reste des points, sa valeur atteint 0.2 . et aussi pour le reste des points, la valeur est proche de 0 jusqu'à ce qu'elle soit nulle en SI1.

l'importance de permutation est définie comme la diminution du score d'un modèle lorsqu'une seule valeur de caractéristique est mélangée au hasard ,figure suivant (Figure 14) illustre l'importance(MDI) et permutation de données .

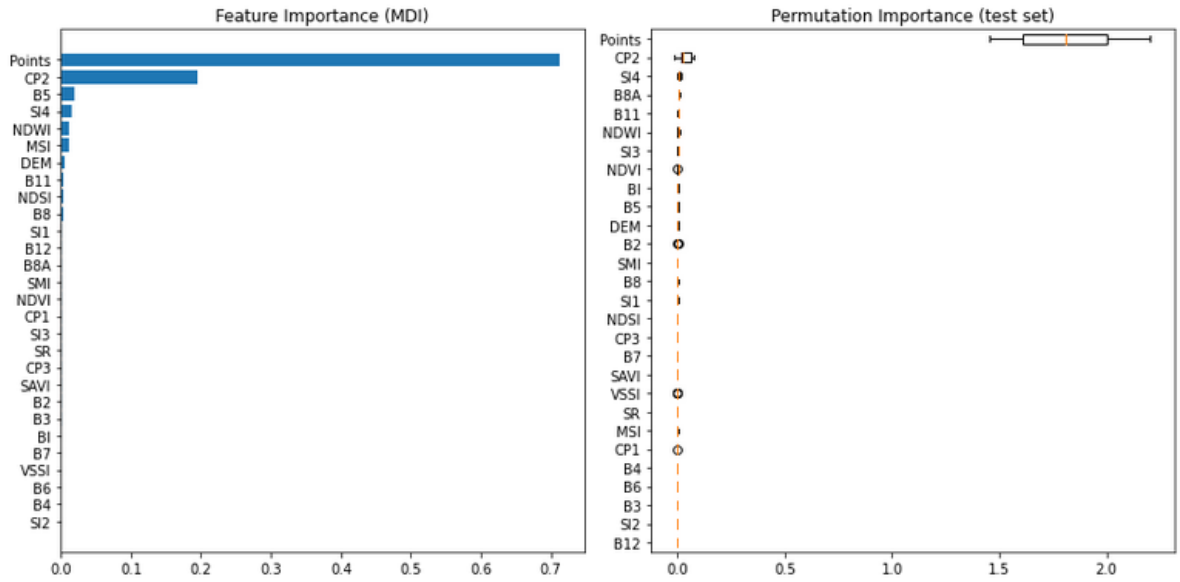

FIGURE 4.14 – illustre l'importance(MDI) et permutation de données

4.5.4 Discussion des résultats et comparison

Afin d'évaluer les performances de notre modèles, nous comparons les résultats obtenus avec nos travaux et les participants au même défi ; En général, nos modèles sont performants , comme présenté dans le tableau suivant (Tableau 2).

modèles	RMSE	R Squared
Linear Regression	5.019798	-1.445458
Ridge Regression	2.786217	0.59
Lasso Regression	3.925501	0.09
K Neighbors Regressor	3.895352	0.12
Decision Tree Regressor	4.226136	-0.22
Random Forest Regressor	3.056686	0.46
Gradient Boosting Regressor	2.475651	0.63
Adaboost Regressor	2.501502	0.54
MLP Regressor	3.944316	0.18

TABLE 4.3 – Comparaison entre les resultats modèles proposés

Des expériences ont montré que plus la valeur d'erreur RMSE de l'algorithme est petit alors plus le taux de précision de la prédiction est élevé. Par conséquent, notre algorithme efficace est le gradient. où le coefficient d'erreur est la valeur la plus basse 2.40 et la valeur de squred est de 0.63 . Par rapport aux autres modèles, où il nous apparaît que chacun des algorithmes " Lasso Regression " et" Decision Tree Regressor ", " Random Forest Regressor " ,"MLP Regressor " et KNN ont des valeurs d'erreur supérieures a 3 , cela signifie que la précision de ces algorithmes est moyenne , on remarque aussi que les algorithmes Linear

Regression et Decision Tree Regressor ont la plus grande valeur d'erreur avec des valeurs de $R\hat{2}$ négatives, cela signifie que cette modèles est pire .

4.6 Conclusion

Dans ce chapitre, nous avons présenté l'implémentation de notre système, où nous avons montré l'environnement et les outils de d´eveloppement qu'on a utilisé. Ensuite nous avons expliqué l'implémentation de modèle MLP , les paramètres utilisé et le test, et aussi nous avons montré les modèles de régression que nos avons utilisé, La meilleur méthode pour notre travail était le modèles Gradient Boosting régression lorsque nous travaillions avec les bandes spéctrales et les indices en entrée , comme nous avons aussi . à la fin nous avons comparaison entre les résultats obtenues.

Conclusion Générale

La salinité des sols est étudiée par la méthode de Télédétection, qui est une technique fréquemment utilisée, qui permet de contrôler de manière significative la salinité sur un plus grand volume de sols en cartographiant le sol et en diagnostiquant ses couvertures.

Pour estimer et prédire la salinité des sols dans la région de Biskra, nous avons proposé un modèle basé sur l'apprentissage automatique, où nous avons utilisé des modèles de régression et des images satellites sentinelle-2, qui se caractérisent par une très grande précision : précision spatiale, précision spectrale et bonne précision temporelle. Nous avons aussi calculé la conductivité électrique d'échantillons de sol dans cette zone pour connaître les niveaux de salinité.

Dans cette étude, généré un ensemble de données pour cette région , collecté en juillet et septembre et août, afin d'entraîner nos modèles à prédire le niveau de salinité, on avez appliqué ce modèles de regression ('Linear Regression', 'Ridge Regression', 'Lasso Regression', 'K Neighbors Regressor', 'Decision Tree Regressor', 'Random Forest Regressor', 'Adaboost Regressor' ,MLP Regressor) , parceque les résultats étaient insuffisants et cela est dû aux caractéristiques de nos base de donnée , où les résultats ont montré que ces modèles ont une précision différente dans la prédiction , le modèle gradient boosting régression avait une précision plus élevée que les autres modèles où le taux d'erreur était très faible par rapport aux autres modèles. Nous avons atteint une précision de 89 %, ce qui est un excellent résultat sur le terrain, et qui nous incite à nous améliorer .

Bien que les résultats de nos modèles soient bons. Cependant, il faut plus de données pour qu'ils soient plus précis. Comme travaux futurs, on envisage de tester nos modèles sur d'autres types d'images satellitaires (informations spectrales) ainsi que des modèles d'apprentissage profond.

Bibliographie

[1] Fan G Qiang. Xiaoyi S. Zhenglong Y , Study on dynamic changes of the soil salinization in the upper stream of the Tarim river based on RS and GIS , Procedia Environmental Sciences, 11 : 1135-1141 , 2011

[2] Farifteh J. Farshad A and George R , Assessing salt-affected soils using remote sensing olute modelling geophysics. Geoderma, 130 : 191-206 , 2006

[3] Daoud et Halitim,Restauration écologique des sols forestiers Cas de la forêt Algérie , 1994

[4] Lahmar etRuellan, Du mulch terreux au mulch organique. Revisiter le dry-farming pourassurer une transition vers l'agriculture durable dans les Hautes Plaines Sétifiennes , 2007

[5] habbane et mohammed ,Teledetection passive et processus decisionnel a reference spatiale , 1997

[6] HERINIRINA N , UNE OBSERVATION D'UN GOITRE SUR DYSGENESIE. THYROIDIENNE PRESTERNALE , 2009

[7] , [8] Caloz R Et Puech C, Hydrologie et imagerie satellitaire. In Précis de télédétection , 1996

[9] Bonn F, Rochon G ,Précis de télédétection Volume 1 : Principes et méthodes. AUPELF-UREF , 1992

[10] Passive vs. Active Sensing , https ://www.nrcan.gc.ca ,Date modified : 7/2020

[11] GeoBretagne ,Comprendre une image satellitaire ,Date modified :2018
[12] ,[13] Jean-Pierre MONTOROI ,La salinisation des écosystèmes De la dégradation insidieuse à la remédiation continuelle par les hommes , 2006

[14] DJILI,Prospection de zones pour la production de semences de pommes de terre Projet SAGRODEV , 2000 .

[15] BOULAINE J , hydropedologie, des écoles nationales de génie rurale, , des eaux et des forets. alger , 1971

[16] Brady N.C ,The Nature and Properties of Soils , 2002

[17] AUBERT G, Les sols de la zone arride ,étude de leure formation ,de leur conservation actes coll .unesco de paris sur le probleme de la zone arride, Paris , 1960

[18] AUBERT G , observation sur les caracteristique la denomination et la classification des sols salés ou sales sodiques . Cashors Tom .ser.Pd Volxxx n°1, 1983

[19] SERVANT, SERVAT, J., t.F , Introduction à l'etude des sols salés litteraux des langaux doc , Montepellier : service d'étude des sols centre de recherches agronomique du MIDI34, 1966

[20] MATHIEU, PIELTAIN, C., F , Analyse chimique des sols. Paris : tec et doc ,lavoisier. 2003

[21] HALITIM, Etude experementation de l'amelioration des sols sodique d'Algerie en vie de leur mise en culture , 1973

[22] DUTHIL, nutrition azotée et la productivité d'une culture de blé dur , 1973

[23] DERMOCH, M . Influence des soulitions salines sur les proprietés physique et l'évolution des mineraux phylliteux de sols systeme. 1976

[24] , [25] BOULAINE J. , hydropedologie, des écoles nationales de génierurale, , des eaux et des forets, 1971

[26] Government of canada , http ://www.ccrs.nrcan.gc.ca/ccrs
, Date modified :2019-08-06

[27] Dallery, Donatien , Suivi des prairies par mesures hyperspectrales et séries temporelles d'images satellitaires à haute résolution spatiale : influence des modes de gestion sur le signal spectral , 2016

[28] Astola, Heikki , Comparison of Sentinel-2 and Landsat 8 imagery for forest variable prediction in boreal region , 2019

[29] Dallery, Donatien , Suivi des prairies par mesures hyperspectrales et séries temporelles d'images satellitaires à haute résolution spatiale : influence des modes de gestion sur le signal spectral , 2019

[30] Rebala, Gopinath and Ravi, Ajay and Churiwala, Sanjay , Machine Learning Definition and Basics , 2019

[31] Mifdal, Rachid , Application des techniques d'apprentissage automatique pour la prédiction de la tendance des titres financiers , 2019

[32] Rachid MIFDAL , Application des techniques d'apprentissage automatique pour la prédiction de la tendance des titres financiers , 2019

[33] Sujatha, R and Chatterjee, Jyotir Moy and Jhanjhi, NZ and Brohi, Sarfraz Nawaz , Performance of deep learning vs machine learning in plant leaf disease detection , 2021

[34] Dias, Sofia B and Hadjileontiadou, Sofia J and Diniz , DeepLMS : a deep learning predictive model for supporting online learning in the Covid-19 era , Scientific reports , 2020

[35], [36] Watson, David S and Krutzinna, Jenny and Bruce, Ian N and Griffiths, Christopher EM and McInnes, Iain B and Barnes, Michael R and Floridi, Luciano , Clinical applications of machine learning algorithms : beyond the black box , Bmj , 2019

[37] Lee, Yena and Ragguett, Renee-Marie and Mansur, Rodrigo B and Boutilier, Justin J and Rosenblat , Applications of machine learning algorithms to predict therapeutic outcomes in depression : a meta-analysis and systematic review , Journal of affective disorders , 2018

[38] Jr, Thomas C and Beard, Karen H and Cutler , Random forests for classification in ecology , Cutler, D Richard and Edwards , Ecology, 2007

[39] Ogutu, Joseph O and Schulz-Streeck, Torben and Piepho, Hans-Peter , Genomic selection using regularized linear regression models : ridge regression, lasso, elastic net and their extensions . BMC proceedings . 2012

[40] Jingzhe Wang ,Machine learning-based detection of soil salinity in an arid desert region,Northwest China , December 2019 .

[41] , [42] Nacer Madjid Tiar Khaled ,Impact de la salinité due au traitement de sel sur l'environnement , biskra , 2012

[43] Contribution of Sentinel-2 data for applications in vegetation monitoring , Addabbo, Pia and Focareta, Mariano and Marcuccio, Salvo and Votto, Claudio and Ullo , 2016

[44] , [45] Wang, Qunming and Shi, Wenzhong and Li, Zhongbin and Atkinson, Peter M , Fusion of Sentinel-2 images , Remote sensing of environment , 2016

[46] J.P. Montoroi , Conductivité électrique de la solutiondu sol et d'extraits aqueux de sol , Centre ORSTOM d'Ile-de-France , 1997)

[47] ,[48] Djamel Belhaouci , Démystifier le Machine Learning : les Réseaux de Neurones artificiels , 2018

[49] Huabin Chen Xiaoqiang Zhang , A novel material removal prediction method based on acoustic sensing and ensemble XGBoost learning algorithm for robotic belt grinding of Inconel 718 , November 2019 The International Journal of Advanced Manufacturing Technology November 2019).

[50] Huabin Chen Xiaoqiang Zhang ,novel material removal prediction method based on acoustic sensing and ensemble XGBoost learning algorithm for robotic belt grinding of Inconel 718 , The International Journal of Advanced Manufacturing Technology , November 2019 .

[51] Ricco RAKOTOMALALA , echnique ensembliste pour l'analyse prédictiveIntroduction explicite d'une fonction de coût ,Université Lumière Lyon 2 , 2018

[52] Nicolas BERNARD , AhmedFETI , Le LASSOLeast Absolute Shrinkage and Selection Operator , 2018

[53] Ranstam, J and Cook, JA , LASSO regression , Journal of British Surgery , 2018

[54] Afroz Chakure , Random Forest Regression , Jun 2019

[55] Chaya Bakshi , Random Forest Regression , Jun 2020

[56] Algorithme des k plus proches voisins pondérés etapplication en diagnostic , Eve Mathieu-Dupas ,arseille, France , Submitted on 24 Jun 2010

[59] Karan Kashyap , Machine Learning : Google Colab- Why, When and How to Use it , May 28, 2020 .

[60] Dave Kuhlman. A python book : Beginning python, advanced python, and python exercises. Dave Kuhlman Lutz, 2009.

[61] Martın Abadi et al. TensorFlow : Large-Scale Machine Learning on Heterogeneous Systems. Software available from tensorflow.org . url : https ://www.tensorflow.org/ , 2015

[63] , [64] The Scipy community. Numpy Manual. Numpy organization. Jan. 2020.

[65] J. D. Hunter. “Matplotlib : A 2D graphics environment”. In : Computing in Science Engineering 9.3 (2007), pp. 90–95. doi :10.1109/MCSE.2007.55.