

REPUBLIQUE DU CAMEROUN

Paix-Travail-Patrie

UNIVERSITE DE YAOUNDE I

CENTRE DE RECHERCHE ET DE FORMATION
DOCTORALE EN SCIENCES, TECHNOLOGIES ET
GEOSCIENCES

UNITE DE RECHERCHE ET DE FORMATION DOCTORALE EN PHYSIQUE ET APPLICATIONS

B.P 812 Yaoundé

Email : crfd_stg@uy1.uninet.cm



REPUBLIC OF CAMEROON

Peace-Work-Fatherland

THE UNIVERSITY OF YAOUNDE I

POSTGRADUATE SCHOOL OF SCIENCES,
TECHNOLOGY AND GEOSCIENCES

RESEARCH AND POSTGRADUATE
TRAINING

UNIT FOR PHYSICS AND APPLICATIONS

P.O 812 Yaoundé

Email : crfd_stg@uy1.uninet.cm

LABORATOIRE D'ENERGIE, DES SYSTEMES ELECTRIQUES ET ELECTRONIQUES

ENERGY LABORATORY, ELECTRICALS AND ELECTRONICS SYSTEMS

Option : Energie et Environnement

**PREDICTION DE L'IRRADIATION SOLAIRE GLOBALE À
L'AIDE DES ALGORITHMES MACHINE LEARNING : UN CAS
DE QUELQUES VILLES DU CAMEROUN**

Mémoire rédigé et présenté en vue de l'obtention du Master of Science en Physique

Par :

NDACKSSI NATHAN WILSON

Matricule : 1712579

Licencié en physique



Sous la supervision de :

KAMENI NEMATCHOUA MODESTE

Chargé de cours, Université de Yaoundé I

Année 2024

DÉDICACE :

Je dédie ce modeste travail de mémoire :

À mon père

TAGUIATCHUE PIERRE

Et À ma mère

NGUEDIA ANNIE VIVIANE

REMERCIEMENTS

Ce travail a été effectué au sein du laboratoire d'Energie, des Systèmes Electriques et Electroniques du département de Physique de l'Université de Yaoundé I.

Je tiens à exprimer mes plus profonds remerciements :

- ❖ A Dieu tout puissant pour la santé et l'accompagnement tout au long de mon parcours.
- ❖ Pr NDJAKA JEAN MARIE chef de département de physique pour sa bienveillance et conseil
- ❖ Dr KAMENI NEMATCHOUA Modeste, pour la confiance qu'il m'a accordé en acceptant de diriger ce travail.
- ❖ Aux différents membres du jury, pour m'avoir honoré à travers leur présence et aussi pour avoir accepté d'évaluer ce modeste travail.
- ❖ Au corps enseignant du Département de Physique de la Faculté de Sciences pour leur enseignement reçu tout au long de mon parcours académique.
- ❖ À tous mes camarades de promotion, plus particulièrement MBANI ANDRE, VOUFO LAURA, ABADA DUCLOS, VOPE KENFACK, LONODJI CHRISTABELLE, KENE KENECTO MARDEL pour la sincère collaboration, la solidarité, l'esprit d'équipe et les débats passionnants qui nous ont beaucoup appris.
- ❖ À tous mes amis avec qui je suis depuis le niveau 1, le groupe la G et plus particulièrement à TCHUENKAM SADRACK, TEUMA STEVE DUVAL, TCHATCHOUANG LOIC, KEUBOU MAHIVA, DJOUSSE GAVINI, MBOUYOM FLORA, NNANG OSSE, DJOMO WILLIAM pour la sincère collaboration au cours de notre parcours.
- ❖ Mon papa TAGUIATCHUE PIERRE et ma maman NGUEDIA ANNIE VIVIANE vos encouragements et vos prières m'ont été d'un grand soutien et réconfort.
- ❖ À tous mes amis particulièrement KENGNE ROMAIN, CHEDJOU DANICK, MUNGUNSI COLLINS, MOUNCHAROU JOEL BORIS pour leurs conseils et encouragements.
- ❖ À mes frères et sœurs, ainsi qu'à toute ma famille pour le soutien apporté.
- ❖ À tous ceux qui ont contribué de près ou de loin, d'une manière ou d'une autre à l'aboutissement de ce travail.

TABLE DE MATIERE

DÉDICACE :.....	I
REMERCIEMENTS	II
TABLE DE MATIERE	III
LISTE DES FIGURES	VI
LISTE DES TABLEAUX	VIII
LISTES DES ABREVIATIONS ET ACCRONYMES	IX
RESUME.....	X
ABSTRACT	XI
INTRODUCTION GENERALE.....	XII
CHAPITRE 1 : REVUE DE LA LITTERATURE	2
Introduction	2
1.1 Rayonnement solaire	2
1.1.1 Constante solaire	2
1.1.2 Éclairement solaire extraterrestre	3
1.2 Coordonnées géographiques terrestres	4
1.2.1 Latitude Φ	4
1.2.2 Longitude L	4
1.2.3 Altitude	4
1.3 Coordonnées solaires	4
1.3.1 Coordonnées équatoriales	5
1.3.2 Coordonnées horizontales	5
1.4 Rayonnement solaire au sol sur une surface horizontale	6
1.5 Appareils de mesure du rayonnement solaire	7
1.5.1 Héliographe	7
1.5.2 Pyranomètre	8

1.5.3 Pyrhéliomètre	9
1.6 Intelligence Artificielle	9
1.6.1 Apprentissage automatique	10
1.6.2 Méthode d'apprentissage automatique	10
1.6.3 Historique	11
1.6.4 Quelques applications de l'IA.....	13
1.7 Méthode de prévision du rayonnement solaire	14
Conclusion.....	17
CHAPITRE 2 : MATERIELS ET METHODES	19
Introduction	19
2.1 Sites d'études.....	19
2.2 Collecte des données	20
2.3 Prétraitement des données	21
2.3.1 Profilage des données.....	21
2.3.2 Nettoyage des données	21
2.3.3 Normalisation des données ou mise à l'échelle des caractéristiques	21
2.3.4 Transformation des données.....	21
2.3.5 Séparation des données	22
2.4 Structure de base de l'étude	22
2.5 Algorithmes de machine Learning	22
2.5.1 Régression linéaire	23
2.5.2. L'arbre à décision.....	24
2.5.3. Les forêts aléatoires.....	25
2.5.4. Les machines à vecteur de support.....	26
2.6 Paramètres d'évaluations.....	28
Conclusion.....	29
CHAPITRE 3 : RESULTATS ET DISCUSSIONS	31
Introduction	31

3.1. Performance des algorithmes Machine Learning.....	31
3.2 Discussion des performances	32
3.3 Visualisation des données	35
3.3.1. Irradiation solaire global	35
3.3.2 Performance en termes de RMSE et R^2	37
3.4 Comparaison des modèles de machine Learning	43
3.5 Résultats de la prédiction	47
Conclusion.....	50
CONCLUSION GENERALE ET PERSPECTIVES	51
REFERENCES.....	52

LISTE DES FIGURES

Figure 1: distribution spectrale du rayonnement solaire	3
Figure 2: Coordonnées géographiques	4
Figure 3: Coordonnées équatoriales	5
Figure 4: Coordonnées géographiques	5
Figure 5: Composantes du rayonnement solaire	7
Figure 6: Héliographe.....	8
Figure 7: Pyranomètre	9
Figure 8: Pyrhéliomètre.....	9
Figure 9: Types d'apprentissage automatique.....	11
Figure 10: Apprentissage automatique supervisé.....	11
Figure 11: Domaine d'application de L'IA dans la médecine	13
Figure 12: Irradiation solaire moyenne global horizontale au Cameroun 1994-2018.....	20
Figure 13: Structure de base de l'étude	22
Figure 14: Modèle d'apprentissage automatique utilisé pour estimer le rayonnement solaire	23
Figure 15: Régression linéaire simple	24
Figure 16: Arbre à décision	24
Figure 17: Diagramme du processus des forêts aléatoires	26
Figure 18: Hyperplan de marge maximale et marge minimale pour une SVM entraînée avec des échantillons de 2 classes	27
Figure 19: Irradiation solaire global journalière dans la ville de Bertoua	35
Figure 20: Irradiation solaire global journalière dans la ville de Douala	35
Figure 21: Irradiation solaire global journalière dans la ville de Garoua.....	36
Figure 22: Irradiation solaire global dans la ville de Maroua.....	36
Figure 23: Irradiation solaire global dans la ville de Yaoundé.....	37
Figure 24: Valeurs du RMSE dans le cas des forêts aléatoires pour les villes de Bertoua, Douala, Garoua, Maroua, Yaoundé	37
Figure 25: Valeurs du RMSE pour l'arbre à décision pour les villes de Bertoua, Douala, Garoua, Maroua, Yaoundé	38
Figure 26: Valeurs du RMSE pour les machines à vecteur de support pour les villes de Bertoua, Douala, Garoua, Maroua, Yaoundé	38
Figure 27: valeurs du RMSE du modèle de régression linéaire pour les villes de Bertoua, Douala, Garoua, Maroua, Yaoundé.	39
Figure 28: Tableau récapitulatif des scores en termes de RMSE pour les différents algorithmes de machine Learning pour les différentes villes.....	39

Figure 29: Valeurs du coefficient de détermination R^2 pour la ville de Bertoua.....	40
Figure 30: Valeurs du coefficient de détermination R^2 pour la ville de Douala.....	41
Figure 31: Valeurs du coefficient de détermination R^2 pour la ville de Garoua	41
Figure 32: Valeurs du coefficient de détermination R^2 pour la ville de Maroua.....	42
Figure 33: Valeurs du coefficient de détermination R^2 pour la ville de Yaoundé.....	42
Figure 34: Récapitulatif des valeurs du coefficient de détermination R^2 pour les différentes villes	43
Figure 35: Comparaison entre les scores de R^2 et RMSE pour la ville de Bertoua.	44
Figure 36: Comparaison entre les scores de R^2 et RMSE pour la ville de Douala.	44
Figure 37: Comparaison entre les scores de R^2 et RMSE pour la ville de Garoua.....	45
Figure 38: Comparaison entre les scores de R^2 et RMSE pour la ville de Maroua	46
Figure 39: Comparaison entre les scores de R^2 et RMSE pour la ville de Yaoundé	46
Figure 40: Nuage de points des valeurs de l'irradiation journalière estimées et mesurées pour la ville de Bertoua	47
Figure 41: Nuage de points des valeurs de l'irradiation journalière estimées et mesurées pour la ville de Douala	48
Figure 42: Nuage de points des valeurs de l'irradiation journalière estimées et mesurées pour la ville de Garoua	48
Figure 43: Nuage de points des valeurs de l'irradiation journalière estimées et mesurées pour la ville de Maroua	49
Figure 44: Nuage de points des valeurs de l'irradiation journalière estimées et mesurées pour la ville de Yaoundé	49

LISTE DES TABLEAUX

<i>Tableau I: Résumé des performances des modèles d'apprentissage automatique pour l'estimation de l'irradiation solaire journalière</i>	<i>14</i>
<i>Tableau II: Géolocalisation de certaines villes du Cameroun.....</i>	<i>19</i>
<i>Tableau III: Procédure générale des forêts aléatoires pour la régression</i>	<i>25</i>
<i>Tableau IV: Mesures statistiques</i>	<i>28</i>
<i>Tableau V: Comparaison des performances des algorithmes pour chaque ville.....</i>	<i>32</i>

LISTES DES ABREVIATIONS ET ACCRONYMES

NASA	National Aeronautics and Space Administration
ML	Machine Learning
MAE	Mean Absolute Error
MSE	Mean Square Error
RMSE	Erreur Quadratique Moyenne (Root Mean Square error)
DNI	Rayonnement solaire normal direct
DHI	Rayonnement solaire normal direct
GHI	Irradiation Solaire Globale
HG_H	Irradiation solaire global sur la surface horizontale
IA	Intelligence Artificielle
LM	Linear Model (Modèle Linéaire)
DT	Decision Tree
RF	Random Forest
SVM	Support vector Machine
ARE	Relative Error
R^2	Coefficient de détermination
GBT	Gradient Boost Tree
DL	Deep Learning
MBE	Mean Bias Error
RNA	Réseau de Neurones Artificiel
MAPE	Mean Absolute Percentage Error
PMC	Perception Multicouche
SVR	Support Vector Regression

RESUME

L'objectif principal de cette étude est de prévoir la valeur de l'irradiation solaire globale sur une surface horizontale dans cinq villes du Cameroun, où la distribution du rayonnement solaire est relativement constante. Pour ce faire, quatre algorithmes d'apprentissage automatique ont été utilisés : la régression linéaire, les arbres à décision, les forêts aléatoires et les machines à vecteurs de support. Les paramètres tels que la température moyenne de l'air, l'humidité relative, la vitesse du vent et l'irradiation solaire globale ont été utilisés pour entraîner ces algorithmes. Les données utilisées proviennent du site de la NASA et couvrent la période allant de 2004 à 2006. Pour valider les performances de ces algorithmes, quatre mesures statistiques ont été utilisées : l'erreur relative, l'erreur absolue moyenne, l'erreur quadratique moyenne et le coefficient de détermination R^2 . Les résultats montrent que les valeurs de MAE, MSE, RMSE et R^2 varient respectivement de 0,52 à 0,9, 0,69 à 1,12, 0,83 à 1,06, et de 0,03 à 0,58 kWh/m². Globalement, les quatre algorithmes d'apprentissage automatique utilisés dans cette recherche peuvent être appliqués pour prédire les données quotidiennes de l'irradiation solaire globale avec une bonne précision. Parmi ces algorithmes, le modèle des forêts aléatoires s'est révélé le plus performant pour la prédiction, suivi respectivement par le modèle de régression linéaire, des machines à vecteurs de support et des arbres à décision.

Mots clés : prédiction, irradiation solaire globale, Machine Learning

ABSTRACT

The aim of this study is to predict the value of global solar irradiation on the horizontal surface of 05 cities in Cameroon with relatively uniform solar radiation distribution. Four machine learning algorithms were employed for this task: linear regression model, decision tree model, random forests, and support vector machines. Parameters such as average air temperature, relative humidity, wind speed, and global solar irradiation are used for training these algorithms. The data used is provided by NASA's website and covers the period from 2004 to 2006. To validate the performance of these algorithms, four different statistical measures are used: relative error, mean absolute error, mean squared error, and coefficient of determination R^2 . The results show that the values of MAE, MSE, RMSE, and R^2 vary from 0.52 to 0.9, from 0.69 to 1.12, from 0.83 to 1.06, and from 0.03 to 0.58 kWh/m², respectively. Globally, all the four machine learning algorithms used in this research can be applied to predict the daily global solar radiation data with good accuracy. Among these four algorithms, random forests algorithm is found to be the most suitable for prediction, followed by linear model, support vector machines, and decision tree, respectively.

Keywords: prediction, global solar irradiation, machine learning.

INTRODUCTION GENERALE

Dans le contexte socioéconomique, les besoins énergétiques de l'humanité n'ont cessé de croître avec son évolution. Initialement, ces besoins étaient centrés sur l'utilisation des ressources naturelles telles que le charbon et le bois. Avec les avancées scientifiques, le pétrole, le gaz, l'énergie hydraulique et l'énergie nucléaire sont devenus les principales sources d'énergie. Actuellement, les énergies fossiles (pétrole et gaz) sont consommées beaucoup plus rapidement qu'elles ne se forment dans la nature. Selon les estimations, les réserves mondiales de ces ressources seront épuisées vers 2030 si la consommation ne change pas radicalement. Dans le meilleur des cas, avec des efforts substantiels pour améliorer la production et la consommation, elles pourraient durer jusqu'à 2100. Cette limitation temporelle des énergies fossiles impose la nécessité de trouver des solutions alternatives pour prendre le relais. Il est essentiel d'explorer d'autres sources d'énergie en tenant compte des contraintes économiques et environnementales. L'énergie solaire qui est l'objectif de notre étude se caractérise par une empreinte environnementale nulle et une abondance exceptionnelle en Afrique et en occurrence en Afrique centrale. Au Cameroun le mix énergétique est principalement concentré autour des centrales hydroélectriques (62%) suivis des centrales thermiques (24%) et des centrales à gaz (14%) les centrales solaires ne présentent que 0.1% de la production d'énergie [1]. Le Cameroun, certes producteur de pétrole brut, dépend en grande partie des importations de produits pétroliers liquéfiés et cokéfiés. Les statistiques du commerce extérieur révèlent une augmentation nette des importations passées de 1.026 millions de tonnes en 2018, à 1.62 millions de tonnes en 2020, soit une hausse d'environ 59% [1]. La nécessité de réduire la dépendance du Cameroun aux produits pétroliers transformés importés s'impose. Le pays dispose d'énormes potentialités en termes d'énergie solaire. L'installation d'une usine de fabrication des plaques solaires pour un usage direct par les ménages serait donc une solution compatible avec notre environnement. Cependant le dimensionnement et l'utilisation des équipements liés à cette ressource, et plus particulièrement les dispositifs photovoltaïques ou thermiques, nécessitent la connaissance aussi bien quantitative que qualitative du gisement solaire [2]. Dans notre étude, on s'est intéressé, en particulier, aux modèles mathématiques d'apprentissage automatique (Machine - Learning) basés uniquement sur les paramètres météorologiques couramment mesurés (température de l'air, humidité relative de l'air, vitesse du vent).

L'objectif de ce travail est de prédire les données quotidiennes de l'irradiation solaire à l'aide des algorithmes Machine Learning.

Ce travail s'effectue suivant trois chapitres :

- ✓ Le premier chapitre s'intitule revue de Littérature qui introduit des connaissances de base sur les irradiances solaires et le Machine Learning.
- ✓ Le deuxième chapitre s'intitule matériels et méthodes qui nous présente les différents modèles de Machine Learning, les paramètres d'évaluation et les logiciels utilisés.
- ✓ Le dernier chapitre est dédié à la présentation et une discussion détaillée des résultats.

CHAPITRE 1 : **REVUE DE LA LITTERATURE**

CHAPITRE 1 : REVUE DE LA LITTÉRATURE

Introduction

L'énergie solaire est l'une des formes d'énergie les plus abondantes et durables disponibles sur Terre. Elle provient du rayonnement solaire, qui est émis par le soleil sous forme d'ondes électromagnétiques. Ce rayonnement est composé de différentes longueurs d'ondes, qui sont classées en fonction de leur énergie et de leur longueur d'onde. Elle est fonction de plusieurs paramètres géographiques, météorologiques et astronomiques. L'énergie solaire disponible au sol joue un rôle crucial dans divers domaines : la prévision de la production des systèmes de conversion solaires photovoltaïques et thermiques, les prévisions météorologiques et climatiques, l'efficacité énergétique des bâtiments, ainsi que l'agriculture et l'environnement. Une connaissance précise du rayonnement solaire incident au sol est nécessaire pour concevoir et réaliser des études technico-économiques afin d'évaluer la rentabilité de tout projet d'installation solaire [2].

Ainsi, nous avons consacré ce chapitre pour donner des rappels sur l'énergie solaire, les différents paramètres qui influent sur l'intensité du flux solaire reçu au niveau du sol avant d'introduire la notion de Machine Learning.

1.1 Rayonnement solaire

Le rayonnement solaire désigne l'énergie émise par le Soleil sous forme de rayonnements électromagnétiques. Le rayonnement solaire est l'ensemble des radiations électromagnétiques émises par le Soleil. La distribution spectrale du rayonnement solaire contient une partie visible, située au domaine de longueurs d'onde λ comprises entre 400 nm et 770 nm ($\sim 48\%$), une partie dans le domaine ultraviolet ($\lambda < 400$ nm) ($\sim 8\%$) et une partie infrarouge ($\lambda > 770$ nm) ($\sim 42\%$). La distribution spectrale du rayonnement solaire extraterrestre ressemble approximativement à la distribution d'un corps noir à une température égale à 5777 K (Figure 1). Ce rayonnement varie en fonction de plusieurs paramètres tels que de la situation géographique, le climat, la saison et le moment de la journée.

1.1.1 Constante solaire

La constante solaire est définie comme le flux d'énergie solaire par unité de surface perpendiculaire aux rayons du Soleil, mesuré à la distance moyenne de la Terre du Soleil. Elle est généralement exprimée en watts par mètre carré (W/m^2). D'après les résultats de mesures faites entre 1969 et 1980, la valeur de la constante solaire est estimée à 1367 W/m^2 [3]. Elle

peut être mesurée par des satellites situés à l'extérieur de l'atmosphère terrestre. Ces satellites sont équipés de radiomètres qui peuvent mesurer avec précision le flux de rayonnement solaire. Elle peut varier légèrement en raison des fluctuations de l'activité solaire et de la distance changeante entre la Terre et le Soleil au cours de l'année [4].

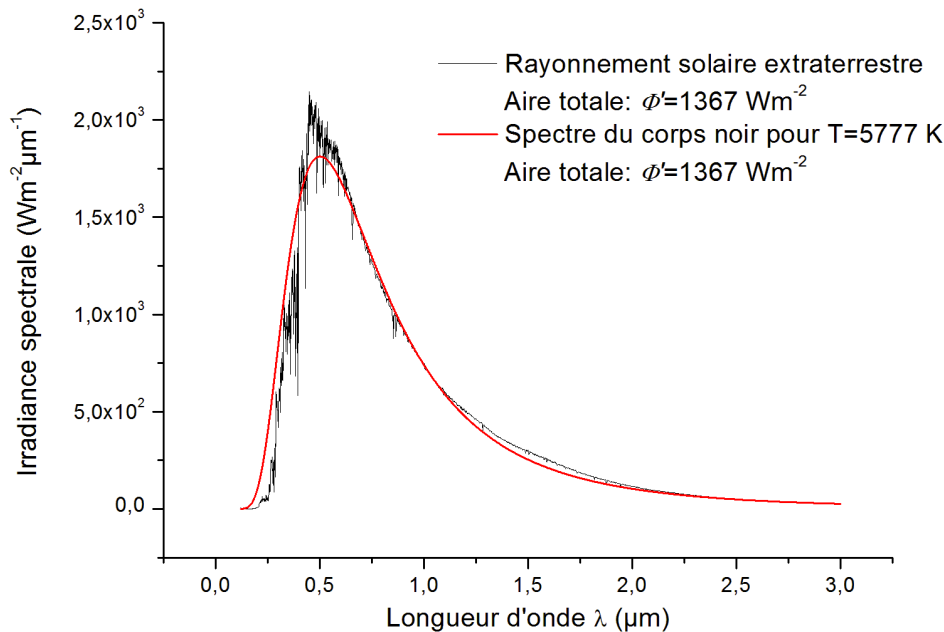


Figure 1: Distribution spectrale du rayonnement solaire [5]

La constante solaire est une mesure fondamentale de l'énergie solaire reçue par la Terre. C'est une donnée clé pour comprendre l'équilibre énergétique de la Terre jouant un rôle crucial dans notre compréhension du climat, de l'énergie solaire et des processus atmosphériques.

1.1.2 Éclairement solaire extraterrestre

L'éclairement extraterrestre est une mesure de l'intensité de la lumière solaire reçue juste à l'extérieur de l'atmosphère terrestre. Il représente la quantité d'énergie solaire par unité de surface atteignant un plan perpendiculaire aux rayons du soleil en dehors de l'atmosphère terrestre. Elle est donnée par la relation suivante :

$$I_{0n} = I_0 \left[1 + 0.033 \cos \left(\frac{360J}{365} \right) \right] \quad [6] \quad (1-1)$$

Où I_0 représente la constante solaire égale à 1367 W/m^2

Pour une surface d'inclinaison quelconque, l'éclairement solaire reçu dépend de l'angle d'incidence θ des rayons solaires :

$$I_{0i} = I_{0n} \cos(\theta) \quad [6] \quad (1-2)$$

En résumé, l'éclairement extraterrestre est une mesure fondamentale pour comprendre l'apport d'énergie solaire avant qu'elle n'interagisse avec l'atmosphère terrestre, influençant ainsi les conditions climatiques et météorologiques de notre planète.

1.2 Coordonnées géographiques terrestres

La position géographique d'un lieu est déterminée par sa latitude, sa longitude, et son altitude.

1.2.1 Latitude Φ

La latitude ϕ est une coordonnée géographique qui représente l'angle d'inclinaison d'un point sur la surface de la terre par rapport au plan de l'équateur. Elle varie de -90 à 90 degrés et est comptée positivement vers le pôle nord et négativement vers le pôle sud.

1.2.2 Longitude L

En géométrie solaire, la longitude L d'un lieu à la surface de la terre représente l'angle que fait le méridien passant par le lieu considéré avec le méridien de Greenwich qui est le méridien de référence de longitude 0° (Figure 2). La longitude varie entre -180° et 180° . Elle est comptée positivement vers l'est et négativement vers l'ouest.

1.2.3 Altitude

L'altitude est l'élévation verticale d'un lieu par rapport au niveau moyen de la mer. Plus l'altitude d'un lieu est importante plus la pression atmosphérique diminue.

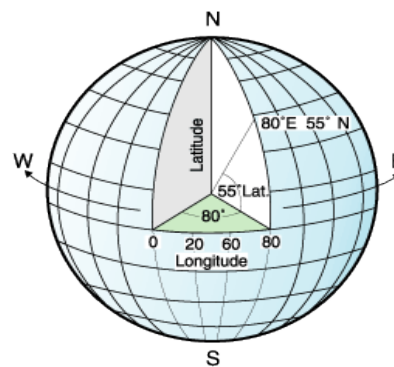


Figure 2: Coordonnées géographiques [7]

1.3 Coordonnées solaires

Les coordonnées solaires sont utilisées pour décrire la position du Soleil dans le ciel par rapport à un observateur sur Terre. Il existe plusieurs systèmes de coordonnées pour définir cette position, chacun étant adapté à différents types d'observations et de calculs. Les deux

systèmes de coordonnées solaires les plus couramment utilisés sont les coordonnées équatoriales et les coordonnées horizontales.

1.3.1 Coordonnées équatoriales

Les coordonnées équatoriales sont un système de coordonnées céleste basé sur l'équateur céleste, analogue au système de latitude et de longitude utilisé pour localiser des points sur la surface de la Terre. Ce système est couramment utilisé en astronomie pour spécifier la position des objets célestes. Les deux principales coordonnées équatoriales sont l'ascension droite (α) et la déclinaison (δ).

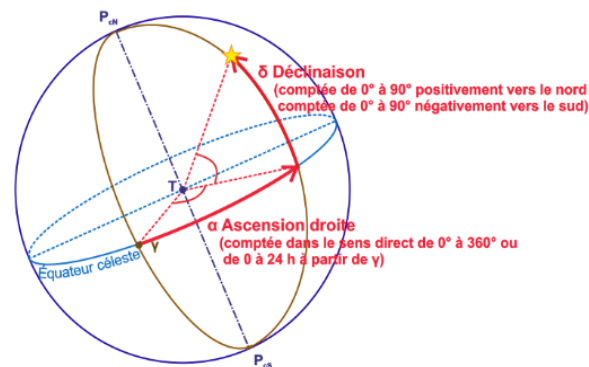


Figure 3: Coordonnées équatoriales [8]

1.3.2 Coordonnées horizontales

Les coordonnées horizontales sont un système de coordonnées célestes utilisé pour spécifier la position des objets dans le ciel par rapport à un observateur situé à un point donné sur la surface de la Terre. Ce système est basé sur l'horizon local de l'observateur et change continuellement avec le temps et la position de l'observateur. Les deux principales coordonnées horizontales sont l'azimut (A_z) et la hauteur (h), également appelée altitude.

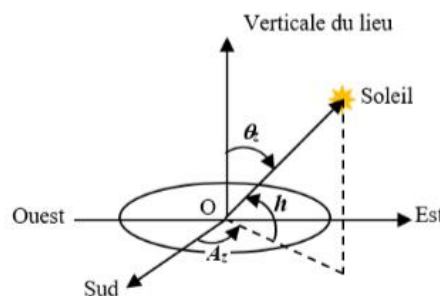


Figure 4: Coordonnées géographiques [9]

- La hauteur h du soleil : aussi appelé élévation ou altitude solaire, est définie par l'angle que forme le vecteur observateur local-soleil avec le plan horizontal du lieu considéré. Elle peut être calculé par la relation :

$$\sin(h) = \sin(\delta)\sin(\phi) + \cos(\delta)\cos(\phi)\cos(\omega) \quad [6] \quad (1-3)$$

L'angle zénithal solaire : C'est l'angle d'incidence des rayons solaires sur une surface horizontale. Sa valeur varie entre 0 et 90°. Un angle de 0° indique que le soleil est au zénith, alors qu'un angle de 90° indique que l'angle zénithal θ_z est l'angle que fait la verticale en un point donné avec la direction du Soleil.

Elle se définit par : $\cos(\theta_z) = \sin(h) \quad [6] \quad (1-4)$

- L'azimut solaire A_z est l'angle compris entre la projection de la direction du soleil sur le plan horizontal et la direction sud du lieu géographique considéré. A_z est compté négativement vers l'Est et positivement vers l'Ouest. Sa valeur varie entre -180 et 180°.

L'azimut solaire A_z peut être calculé à l'aide de la relation :

$$A_z = \frac{\cos \delta \sin \omega}{\cos(h)} \quad [6] \quad (1-5)$$

1.4 Rayonnement solaire au sol sur une surface horizontale

Lorsque le rayonnement solaire n'est ni réfléchi, ni absorbé, il peut être soit transmis au sol (rayonnement solaire direct) soit diffusé par les molécules, les poussières et les aérosols en suspension dans l'atmosphère. Ce qui rend l'énergie solaire si attrayante, c'est qu'elle est disponible chaque jour et partout dans le monde, et qu'elle est non polluante.

En générale, les principales composantes du rayonnement solaire horizontale sont :

- Rayonnement solaire normal direct (DNI) : C'est le rayonnement solaire qui atteint la surface terrestre sans être diffusé par l'atmosphère, comprenant la lumière directe du soleil.
- Rayonnement solaire diffus horizontal (DHI) : Il s'agit du rayonnement solaire qui est dispersé dans l'atmosphère par des particules atmosphériques telles que les molécules d'air, les aérosols ou les nuages. Ce rayonnement peut provenir du soleil ou être réfléchi par la surface terrestre.
- Rayonnement réfléchi : C'est le rayonnement solaire qui est renvoyé dans l'atmosphère après avoir été réfléchi par la surface terrestre. Ce rayonnement peut être dirigé directement vers l'espace ou diffusé dans différentes directions. L'Albédo : est la partie réfléchie par le sol et il dépend de l'environnement du site.

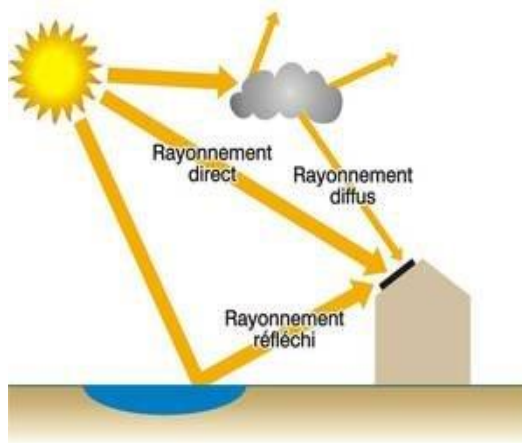


Figure 5: Composantes du rayonnement solaire [10]

- Rayonnement global : est le rayonnement solaire parvenant au niveau du sol sur une surface horizontale, soit directement, soit après diffusion. La composante GHI est la somme de la composante DHI et de la projection horizontale du DNI.

Ces composantes contribuent ensemble à l'énergie totale reçue sur la surface terrestre et influent sur divers processus naturels, tels que le réchauffement de la planète, la formation des nuages et les schémas de circulation atmosphérique.

L'étude de l'irradiation solaire comporte plusieurs applications importantes, notamment la prédiction de la production d'énergie des centrales solaires, les charges de chauffage et de refroidissement des bâtiments, ainsi que la modélisation climatique et la prévision météorologique [11].

1.5 Appareils de mesure du rayonnement solaire

L'utilisation du rayonnement solaire comme source d'énergie pose donc un problème bien particulier. En effet, le rayonnement solaire n'est pas toujours disponible, en outre on ne peut l'emmagasiner ni le transporter. Le concepteur d'un système qui emploie le rayonnement solaire comme source d'énergie doit donc déterminer la quantité d'énergie solaire disponible à l'endroit visé et le moment où cette énergie est disponible. Plusieurs appareils de mesure existent depuis des années et servent à la mesure de différentes composantes du rayonnement solaire.

1.5.1 Héliographe

L'héliographe est un instrument qui permet de mesurer la durée de l'ensoleillement sur un point de la surface de la planète. Plus précisément, il enregistre la durée pendant laquelle le rayonnement solaire est d'une intensité suffisante pour produire des ombres distinctes. Le

premier héliographe dit « héliographe Campbell-Stokes » a été fabriqué par Campbell en 1853. Il était constitué d'une boule de cristal de 10 cm de diamètre, qui, exposée au Soleil, faisant effet de lentille concentrant le rayonnement infrarouge du Soleil, tout en l'atténuant plus qu'une loupe, de manière à brûler une feuille de papier rigide spécial placée à une distance judicieuse sous la boule sur un support incurvé. La position et l'étendue des brûlures sur la carte permettent de déterminer les périodes d'ensoleillement. L'héliographe est encore utilisé aujourd'hui dans des stations météorologiques pour collecter des données sur l'ensoleillement essentielle pour les études climatologiques.



Figure 6: Héliographe [12]

1.5.2 Pyranomètre

Le pyranomètre est un instrument utilisé pour mesurer l'irradiance solaire (directe et diffuse) reçu sur une surface horizontale. Contrairement à l'héliographe, qui mesure la durée d'ensoleillement, le pyranomètre mesure l'intensité du rayonnement solaire. Il fonctionne grâce à un capteur thermopile recouvert d'un dôme en verre ou en quartz comme représenté dans la figure. Ce dôme protège le capteur tout en laissant passer le rayonnement solaire. Lorsque le rayonnement atteint le capteur, il provoque une différence de température entre les jonctions chaudes et froides de la thermopile, générant ainsi une tension électrique proportionnelle à l'intensité du rayonnement. La différence de potentiel proportionnel au flux incident ($0 - 15\mu\text{V} / \text{W.m}^{-2}$). Tous les ans, il doit être étalonné. Il est aussi utilisé pour la surveillance et l'optimisation des performances des panneaux solaires.



Figure 7: Pyranomètre [13]

1.5.3 Pyrhéliomètre

Le pyrhéliomètre est un instrument utilisé pour mesurer le rayonnement solaire direct. Contrairement au pyranomètre, qui mesure le rayonnement solaire global, le pyrhéliomètre se concentre uniquement sur la composante directe du rayonnement solaire. Il fonctionne en dirigeant le rayonnement solaire direct à travers une fenêtre étroite ou un collimateur vers un capteur. L'élément sensible (thermopile) du pyrhéliomètre est constitué de thermocouples montés en série. Le rayonnement solaire direct est ainsi isolé de la lumière diffuse et indirecte. La différence de température induite sur le capteur par le rayonnement solaire direct est convertie en une tension électrique proportionnelle à l'intensité du rayonnement.



Figure 8: Pyrhéliomètre [14]

1.6 Intelligence Artificielle

L'intelligence artificielle (IA) est un domaine de l'informatique qui vise à créer des systèmes informatiques capables d'imiter, de comprendre et d'effectuer des tâches qui nécessitent normalement l'intelligence humaine. En d'autres mots c'est toute technologie algorithmique qui permet de résoudre des problèmes complexes qu'on aurait cru réserver à l'intelligence humaine. L'IA cherche à développer des programmes et des machines qui peuvent raisonner, apprendre, percevoir, planifier, comprendre le langage naturel et résoudre des problèmes complexes.

1.6.1 Apprentissage automatique

L'apprentissage automatique ou Machine Learning est un champ d'étude de l'intelligence artificielle qui vise à créer des algorithmes capables d'apprendre à partir de données, sans qu'ils soient explicitement programmés. En bref, c'est une méthode qui permet aux ordinateurs d'analyser des données, d'identifier des schémas et de prendre des décisions en se basant sur ces données. Dans plusieurs domaines, l'apprentissage automatique est devenu particulièrement important pour résoudre les problèmes de classification, de prédiction, d'optimisation, de reconnaissance des formes, etc. Deux types de techniques d'apprentissage sont utilisés dans les modèles d'apprentissage automatique : l'apprentissage supervisé qui développe un modèle prédictif basé sur des données d'entrée et de sortie connues pour prédire les sorties futures, et l'apprentissage non supervisé, qui regroupe et interprète les données en se basant uniquement sur les données d'entrée.

1.6.2 Méthode d'apprentissage automatique

L'apprentissage automatique comporte généralement deux phases. La première consiste à estimer un modèle à partir de données, appelées observations, qui sont disponibles et en nombre fini, lors de la phase de conception du système. L'estimation du modèle consiste à résoudre une tâche pratique telle que traduire un discours ou participer à la conduite d'un véhicule automobile. Cette phase dite « d'apprentissage » ou « d'entraînement » est généralement réalisée préalablement à l'utilisation pratique du modèle. La seconde phase correspond à la mise en production : le modèle étant déterminé, de nouvelles données peuvent alors être soumises afin d'obtenir le résultat correspondant à la tâche souhaitée.

L'utilisation et la manipulation des données se fait selon le type d'apprentissage automatique. Elles peuvent être regroupées en trois à savoir :

- Apprentissage supervisé : Les algorithmes sont entraînés sur un ensemble de données étiquetées, où chaque exemple est associé à une sortie désirée. Ces algorithmes sont ensuite capables de généraliser et de faire des prédictions sur de nouvelles données. Plus formellement, étant donné un ensemble de données D , décrit par un ensemble de caractéristiques X , un algorithme d'apprentissage supervisé va trouver une fonction f en entrée x et la variable à prédire y . on distingue deux problèmes d'apprentissage supervisé la Classification qui consiste à prédire la valeur ou la classe d'une variable d'entrée discrète et la Régression qui consiste à prédire la valeur ou la classe d'une variable d'entrée continue.

- Apprentissage non supervisé : Les algorithmes sont entraînés sur des données non étiquetées et sont chargés de trouver des modèles ou des structures intrinsèques dans ces données. Il est basé sur plusieurs méthodes tels que le clustering où regroupement et les réseaux de neurones artificiels.
- Apprentissage par renforcement : Les algorithmes apprennent à travers l'interaction avec un environnement afin de construire son propre jeu de données.

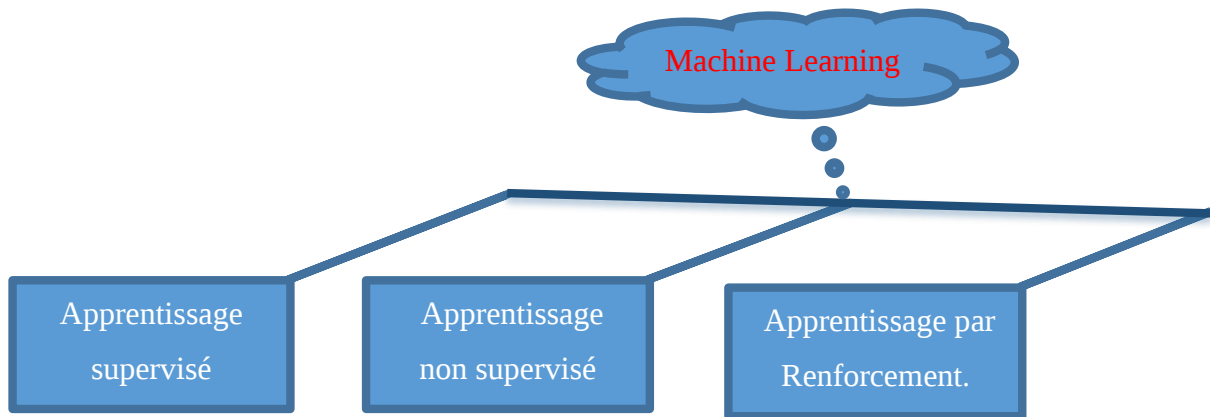


Figure 9: Types d'apprentissage automatique

La seconde phase est la phase d'évaluation (ou de test) : Une fois que le modèle a été entraîné, il doit être évalué pour estimer sa performance et sa capacité à généraliser à de nouvelles données. Pendant cette phase, le modèle est testé sur un ensemble de données de test indépendant, sur lequel il n'a pas été formé. Cela permet de déterminer si le modèle a appris des motifs utiles dans les données ou s'il souffre de sur apprentissage (overfitting), où il mémorise les données d'entraînement au lieu d'apprendre des schémas généraux.

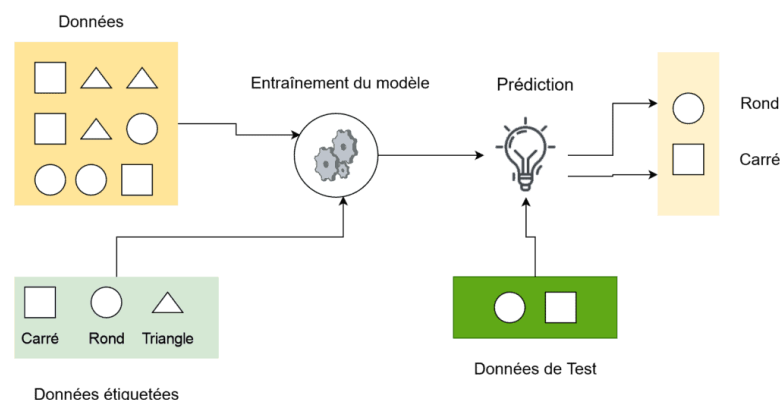


Figure 10: Apprentissage automatique supervisé [15]

1.6.3 Historique

Le domaine du machine Learning a une histoire qui remonte à plusieurs décennies.

- Début du concept (1950 - 1960) : Le concept de machine Learning a émergé dans les années 1950 et 1960, avec des chercheurs comme Arthur Samuel qui ont commencé à explorer les algorithmes capables d'apprendre à partir de données [16]. À cette époque, les premiers travaux se concentraient sur des techniques telles que les arbres de décision et les réseaux neuronaux.
- Premiers succès (1970-1980) : Pendant les années 1970 et 1980, le domaine a connu des développements significatifs dans des domaines tels que les réseaux neuronaux, les arbres de décision et les méthodes de classification statistique [17, 18]. Des progrès ont été réalisés dans des domaines comme la reconnaissance de la parole et la vision par ordinateur.
- Renaissance de l'apprentissage profond (années 2000) : Bien que l'intérêt pour le Machine Learning n'ait jamais complètement disparu, les années 2000 ont été marquées par une renaissance de l'apprentissage profond, une sous-discipline du machine Learning basée sur des réseaux neuronaux avec de multiples couches. Des progrès dans le matériel informatique, l'accès à de grandes quantités de données et de nouveaux algorithmes ont contribué à cette renaissance.
- Boom de l'apprentissage automatique (années 2010) : La décennie suivante a été témoin d'un boom massif de l'apprentissage automatique, alimenté par des avancées dans les domaines de l'apprentissage profond, du Big Data et de l'informatique en nuage. Des entreprises comme Google, Facebook et Amazon ont massivement investi dans cette technologie pour des applications telles que la recherche en ligne, la recommandation de contenu, la reconnaissance d'images et la traduction automatique.
- Vers l'IA généralisée (années 2020) : Au cours des années 2020, le ML a continué à évoluer vers des systèmes plus sophistiqués et généralisés, visant à créer des modèles d'IA capables de performances humaines voire supérieures dans de nombreux domaines, comme la médecine, la finance, l'automobile autonome et plus encore [19, 20].
- Défis et éthique : Parallèlement aux progrès techniques, le domaine du machine Learning a également dû faire face à des défis importants en matière d'éthique, de transparence et de responsabilité. Des questions telles que la protection de la vie privée, la prévention de la discrimination algorithmique et la création de systèmes équitables ont été au centre des préoccupations croissantes [21].

Aujourd'hui, le machine Learning continue de progresser à un rythme rapide, avec des applications dans pratiquement tous les aspects de notre vie quotidienne et une influence croissante sur la société dans son ensemble.

1.6.4 Quelques applications de l'IA

L'intelligence artificielle peut s'appliquer à tous les domaines et s'investir dans notre vie quotidienne à travers :

- La Reconnaissance vocale et traitement du langage naturel : Siri de Apple, Google Assistant et les systèmes de traduction automatique de Google utilisent l'IA pour comprendre et générer du langage humain.
- La Vision par ordinateur : ici l'on fait référence aux systèmes de surveillance vidéo, les voitures autonomes et les applications de réalité augmentée qui utilisent l'IA pour analyser et interpréter les images.
- La Santé : L'IA est utilisée pour aider au diagnostic médical, la découverte de médicaments, la personnalisation des traitements, prévention des maladies, gestion des dossiers médicaux et interventions chirurgicales assistées par ordinateur.



Figure 11: Domaine d'application de L'IA dans la médecine [22]

- La Finance : Les algorithmes d'IA sont utilisés pour l'analyse prédictive des marchés financiers, la détection de fraudes et la gestion de portefeuille.
- Dans les sciences physiques : l'IA offre de nombreuses possibilités pour accélérer la recherche, améliorer la compréhension des phénomènes complexes et optimiser les processus à travers des outils de Modélisation et simulation, analyse des données expérimentales, instrumentation intelligente, prévision et contrôle et découverte des matériaux.

En conclusion, les applications de l'IA sont vastes et touchent de nombreux aspects de notre vie quotidienne, de la médecine aux sciences physiques en passant par les réseaux sociaux. L'IA offre des opportunités exceptionnelles pour accélérer les progrès dans divers domaines,

améliorer l'efficacité des processus, et offrir des expériences plus personnalisées et adaptées aux besoins individuels.

1.7 Méthode de prévision du rayonnement solaire

Ces dernières années, les techniques intelligentes, telles que les réseaux de neurones, la logique floue, les réseaux neuro-flous, et les méthodes d'apprentissage automatique, ont été largement adoptées pour résoudre des problèmes réels lorsque les méthodes classiques se révèlent insuffisantes. La prédiction de l'irradiation solaire est un domaine où ces techniques ont été particulièrement réussies, notamment les réseaux de neurones artificiels et les méthodes d'apprentissage automatique [23] [24] [25] [26].

Tableau I : Résumé des performances des modèles d'apprentissage automatique pour l'estimation de l'irradiation solaire journalière

Références	Composante solaire	Localisation	Modèle	Performances
(Sözen and Arcaklioğlu, 2005)	Irradiation solaire globale journalière	12 villes en Turquie	RNA (PMC)	MAPE $\leq 3.832\%$ $R^2 = 0.998$ RMSE $\leq 1.89\%$
(Benghanem et al., 2009)	Irradiation solaire globale journalière	Al-Madinah (Arabie Saoudite)	RNA (PMC)	MAPE = 2.2077% RMSE = 0.049515 MBE = 1.2514 R = 94.65%
(Behrang et al., 2010)	Irradiation solaire globale journalière	Dezful (Iran)	RNA (PMC)	MAPE = 5.21% $R^2 = 99.57\%$
(Rahimikhoob, 2010)	Irradiation solaire globale journalière	Ahwaz (Iran)	RNA (PMC)	R = 0.97 $R^2 = 0.89$ RMSE = 2.53 MJ/m ² /d RE = 13.83%
(Bilgili and Ozgoren, 2011)	Irradiation solaire globale journalière	Adana (Turquie)	RNA	MAE = 1.00 MJ/m ² MAPE = 9.23% R = 0.9751 $R^2 = 0.9508$
(Linares-Rodríguez et al., 2011)	Irradiation solaire globale journalière	Espagne	RNA	MBE = 0.49 MJ/m ² RMSE = 14.20% R = 0.93
			SVR	MBE = -0.459 MJ/m ² RMSE = 1.056 MJ/m ² $R^2 = 0.975$
		Le Caire (Égypte)	RF	MBE = -0.272 MJ/m ² RMSE = 1.214 MJ/m ² $R^2 = 0.966$

(Linares-Rodríguez et al., 2011)	Irradiation solaire globale journalière	Le Caire (Égypte)	SVR	MBE = -0.235 MJ/m ² RMSE = 0.809 MJ/m ² R ² = 0.98
		Ma'am (Jordanie)	RF	MBE = 0.006 MJ/m ² RMSE = 0.978 MJ/m ² R ² = 0.980
			SVR	MBE = 0.055 MJ/m ² RMSE=1.483 MJ/m ² R ² = 0.962
(Muhammed A Hassan et al., 2017)	Irradiation solaire globale journalière	Tataouine (Tunisie)	Random Forests	MBE = 0.311MJ/m ² RMSE = 1.838MJ/m ² R ² = 0.942
			RNA	MBE = 0.288MJ/m ² RMSE = 0.996 MJ/m ² R ² = 0.973
		Ghardaïa (Algérie)	Boosting	MBE = 0.304 MJ/m ² RMSE = 1.220 MJ/m ² R ² = 0.959
			Bagging	MBE = 0.258 MJ/m ² RMSE = 0.986 MJ/m ² R ² = 0.973
		Tan-Tan (Maroc)	Random Forests	MBE = -0.247 MJ/m ² RMSE = 1.569 MJ/m ² R ² = 0.869
			Boosting	MBE = -0.261 MJ/m ² RMSE = 1.337 MJ/m ² R ² = 0.905
(Leva S et al. 2017)	Irradiation solaire globale journalière	Milano, Italie	ANN	MAE < 15%
(Pierro M et al. 2017)	Irradiation solaire globale journalière	Italie	ANN	RMSE : 7%–7.5%
(Liu W et al. 2018)	Irradiation solaire globale journalière	Beijing, chine	SVM et ANN	MRE : 11.61%
(Fan et al., 2019b)	Irradiation solaire globale journalière	Chine	SVM	-0.437 < MBE < 0.637MJ /m ² /d 0.131 < nRMSE < 0.197 MJ /m ² /d 0.879 < R ² < 0.933
			RF	-0.364 < MBE < 0.744 MJ/m ² /d 0.144 < nRMSE < 0.213MJ /m ² /d 0.839 < R ² < 0.919
			LM	RMSE (kWh/m2): 0.166 AAE: 0.126

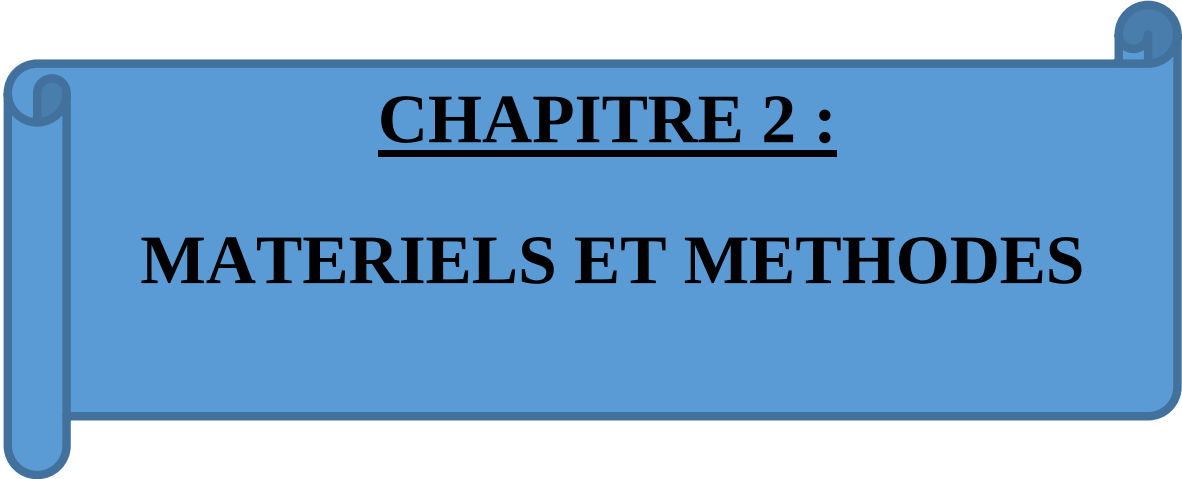
(Modeste Kameni et al, 2021)	Irradiation solaire globale journalière	27 pays en Europe Paris (France)		ARE (%): 54.74 R2: 0.681
			DL	RMSE (kWh/m2): 0.155 AAE: 0.120 ARE (%): 45.80 R2: 0.723
			DT	RMSE (kWh/m2): 0.162 AAE: 0.126 ARE (%):52.84 R2: 0.694
			RF	RMSE (kWh/m2): 0.160 AAE: 0.127 ARE (%):57.53 R2: 0.703
			GBT	RMSE (kWh/m2): 0.163 AAE: 0.130 ARE (%):58.37 R2: 0.693
			SVM	RMSE (kWh/m2): 0.145 AAE: 0.116 ARE (%):44.00 R2: 0.775

(Muhammed A Hassan et al., 2017) ont proposé une étude pour explorer le potentiel des méthodes d'ensemble basées sur les arbres de décision dans la modélisation du rayonnement solaire. Pour cela, trois modèles (Gradient Boosting, Bagging et Random Forest (RF)) ont été développés pour estimer les valeurs journalières et horaires des trois composantes solaires (GHI, DHI et DNI). En général, les résultats ont montré que les techniques proposées sont très fiables et précises, malgré leur relative simplicité. Cependant, les deux algorithmes Bagging et RF ont donné les meilleures estimations par rapport au modèle du Gradient Boosting.

(Modeste Kameni et al., 2021) à effectuer la prédiction de l'irradiation solaire Globale journalière sur 27 villes dans 27 pays en Europe. Six modèles ont été utilisés, les forêts aléatoires, l'arbre à décision, la régression linéaire, les arbres boostés par gradient, le Deep Learning, les machines à vecteur de support. Les résultats ont montré que le modèle SVM est le meilleur modèle parmi les six modèles utilisés suivi respectivement par le DL, LM, GB, RF et DT.

Conclusion

En définitive, nous avons présenté les types de rayonnement solaire ainsi que les différents systèmes de coordonnées permettant de les lire. Nous avons donné un bref aperçu des différentes composantes du rayonnement solaire ainsi que les différents appareils de mesure du dit rayonnement tels que le pyrhéliomètre et le pyranomètre. Nous avons ressorti l'historique du Machine Learning et présenter quelques travaux menés par les chercheurs dans le but de prédire l'irradiation solaire globale à l'aide des algorithmes d'apprentissage automatique. Il en ressort que l'énergie solaire est une source d'énergie gratuite et quantifiable et qui peut se repérer à l'aide de plusieurs systèmes de coordonnées géographiques et ou solaire. Les travaux de recherches menées par plusieurs chercheurs à travers le monde pour une connaissance de ce rayonnement à montrer que l'on peut le prédire à l'aide des modèles d'apprentissage automatique tels que le modèle de régression linéaire, les forêts aléatoires, les réseaux de neurones artificiels, les machines à vecteurs de support et bien d'autre avec une certaine marge d'erreur qui est régulée en tenant compte des performances des algorithmes.



CHAPITRE 2 : MATERIELS ET METHODES

CHAPITRE 2 : MATERIELS ET METHODES

Introduction

L'irradiation solaire dépend de différents facteurs géographiques et météorologiques. A cet égard, plusieurs modèles empiriques et intelligents ont été développés pour calculer et prédire l'irradiation solaire globale à l'aide de divers paramètres météorologiques. Dans ce chapitre, nous présentons en détail la démarche méthodologique et les différents outils utilisés pour prédire l'irradiation solaire globale. Nous allons tout d'abord débiter par la délimitation des sites d'études et la collecte des données ensuite nous parlerons du prétraitement de ces données. Nous allons expliciter les différents modèles d'algorithme de Machine Learning que nous avons utilisés pour l'estimation de l'irradiation solaire globale en fonction de quelques paramètres météorologiques et nous présenterons aussi dans ce qui suit, les principaux critères de performances permettant d'évaluer les différents modèles.

2.1 Sites d'études

Le Cameroun est un pays d'Afrique centrale situé sur la côte ouest du continent Africain. Situé près de l'équateur il bénéficie d'un climat tropical avec des températures élevées et des précipitations abondantes dans certaines régions. Les régions côtières sont plus chaudes et humides, tandis que les zones montagneuses sont plus fraîches. En raison de sa situation géographique près de l'équateur, la durée d'ensoleillement annuelle est estimée à environ 2500 heures par ans pour un rayonnement solaire moyen estimé à 1825 kWh/m^2 [27]. Dans cette étude, 5 villes du pays ont été analysées. Les principaux détails sont donnés respectivement dans le Tableau II et également à la figure 12. Dans le tableau I nous présentons les villes sélectionnées ainsi que leurs coordonnées géographiques et dans La figure 12 nous présentons la variation du rayonnement solaire global du Cameroun.

Tableau II: Géolocalisation de certaines villes du Cameroun [28]

Villes	Latitude	Longitude	Altitude
Bertoua	4.577° N	13.685° E	680m
Douala	4.051° N	9.768° E	103m
Garoua	9.308° N	13.398° E	287m
Maroua	10.5955° N	14.325° E	418 m
Yaoundé	3.848° N	11.502° E	589m

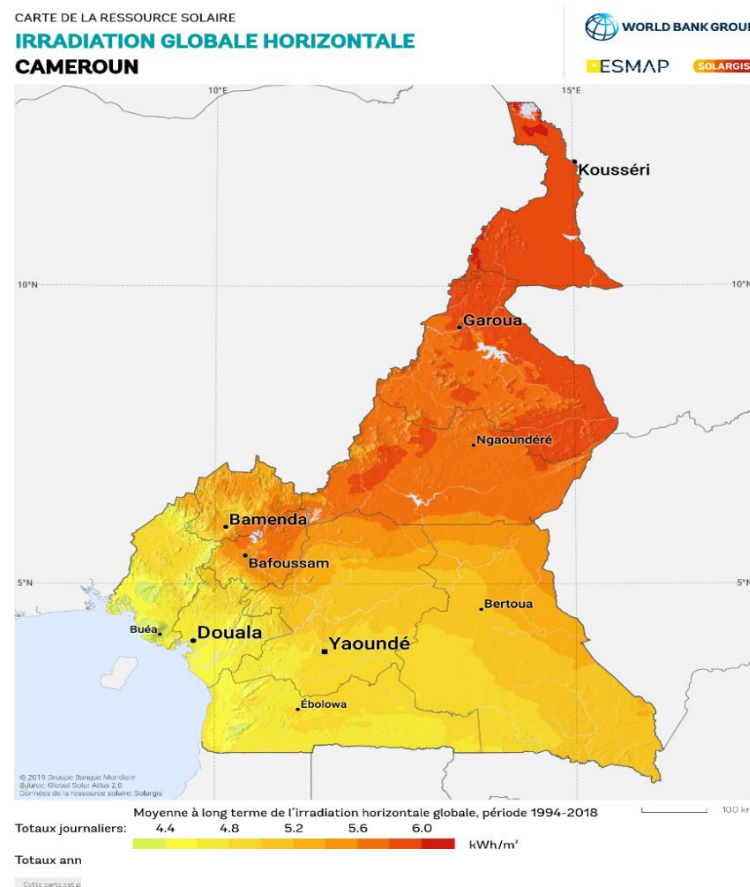


Figure 12: Irradiation solaire moyenne global horizontale au Cameroun 1994-2018 [29]

On remarque que les villes du septentrion (Maroua et Garoua) ont des valeurs plus élevées que celle des autres villes. Nous précisons que les études menées dans ces 5 villes peuvent être appliquées et adaptées aussi bien au type de climat africain qu'asiatique et européen dans le monde.

2.2 Collecte des données

Dans le cadre de notre étude, la collecte de données revêt une importance cruciale pour garantir la qualité et la pertinence des modèles de machine Learning que nous allons développer. Nous avons utilisé des données journalières moyennes mensuelles représentatives et diversifiées tels que la température ambiante et l'humidité relative pris à 2m de la surface du sol, la vitesse du vent quant à elle pris à 10m de la surface du sol. Ces données ont été obtenues sur le site de la NASA et couvrent une période allant de 2004 à 2006 relativement à la situation géographique de nos différents sites d'études.

2.3 Prétraitement des données

Une fois les données collectées, elles sont sous une forme brute elles doivent être traitées pour pouvoir être utilisées. Le prétraitement des données est une étape cruciale dans la formation des algorithmes de machines Learning. Il nous permet de transformer les données brutes dans un format compréhensible et utilisable pour l'analyse et doit être effectué avant la création des modèles d'apprentissage. Il nous permet d'éliminer les erreurs, d'imputer les valeurs manquantes et de rectifier les incohérences. Il implique plusieurs étapes clés tels que :

2.3.1 Profilage des données

Comprendre vos données est la première étape du prétraitement. Le profilage des données implique l'examen des données à l'aide de statistiques récapitulatives et de distributions pour comprendre leur structure, leur contenu et leur qualité. Cette étape peut révéler des modèles, des anomalies et des corrélations cruciales pour un prétraitement éclairé.

2.3.2 Nettoyage des données

Les données peuvent contenir des valeurs manquantes, des valeurs aberrantes ou des erreurs. Avant de les utiliser pour former les algorithmes, il est important de nettoyer les données en supprimant ou en corrigeant ces valeurs en les supprimant ou en faisant un remplissage par interpolation. Pour cela, il faudra tout d'abord charger les données dans un Data Frame. Pour notre travail, après examination, nos données sont complètes et ne possèdent pas de valeurs aberrantes.

2.3.3 Normalisation des données ou mise à l'échelle des caractéristiques

Les différentes caractéristiques des données peuvent avoir des plages de valeurs différentes. Pour garantir une convergence rapide et une meilleure performance des algorithmes, les données sont souvent normalisées pour mettre toutes les caractéristiques à la même échelle. Cela peut être fait en utilisant des techniques telles que la mise à l'échelle min-max ou la standardisation. Pour une meilleure conservation de la distribution des données nous avons opté pour la méthode de la mise à l'échelle min-max.

2.3.4 Transformation des données

Elle aide à modifier les données pour des besoins spécifiques. Le tri, la création de nouveaux paramètres résultant des opérations de conversion d'addition et autre chacune jouant un rôle essentiel dans la compréhension des données. Nous avons effectué des opérations de conversion sur la température initialement en degrés kelvin en degré Celsius.

2.3.5 Séparation des données

Les données sont généralement séparées en ensemble d'entraînement et ensemble de test. L'ensemble d'entraînement est utilisé pour former les algorithmes, tandis que l'ensemble de test est utilisé pour évaluer leurs performances.

En résumé, le prétraitement des données vise à préparer les données brutes pour qu'elles soient adaptées à l'entraînement des algorithmes d'apprentissage automatique, en garantissant leur qualité et affinant la précision, les performances et la fiabilité du modèle.

2.4 Structure de base de l'étude

Le machine Learning permet aux systèmes informatiques de reconnaître des modèles et de prendre des décisions en se basant sur l'expérience accumulée à partir des données. L'apprentissage supervisé est l'une des branches fondamentales du machine Learning, où un modèle est entraîné à partir de données étiquetées afin d'effectuer la prédiction. Dans ce contexte, les algorithmes d'apprentissage supervisé jouent un rôle central en permettant aux systèmes informatiques d'apprendre à partir d'exemples pour effectuer la prédiction de l'irradiation solaire globale. Dans cette étude quatre algorithmes de Machine Learning ont été utilisés comme présenté à la figure 13. Il s'agit du modèle Linéaire (LM), l'arbre à décision (DT), la forêt aléatoire (RF), et les machines à vecteur de support (SVM)

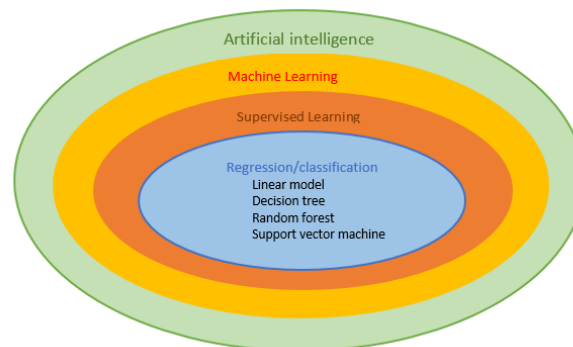


Figure 13: Structure de base de l'étude [11]

2.5 Algorithmes de machine Learning

Une fois la phase de prétraitement des données, le modèle d'algorithme machine Learning peut être construit avant d'être évalué par la suite. La performance d'un algorithme d'apprentissage automatique est liée à la sélection de ses attributs ainsi qu'à la réussite du processus d'apprentissage [30]. Notre apprentissage comme présenter à la figure 13 s'effectue

en trois étapes : le prétraitement des données, la construction du modèle et la prédiction du modèle.

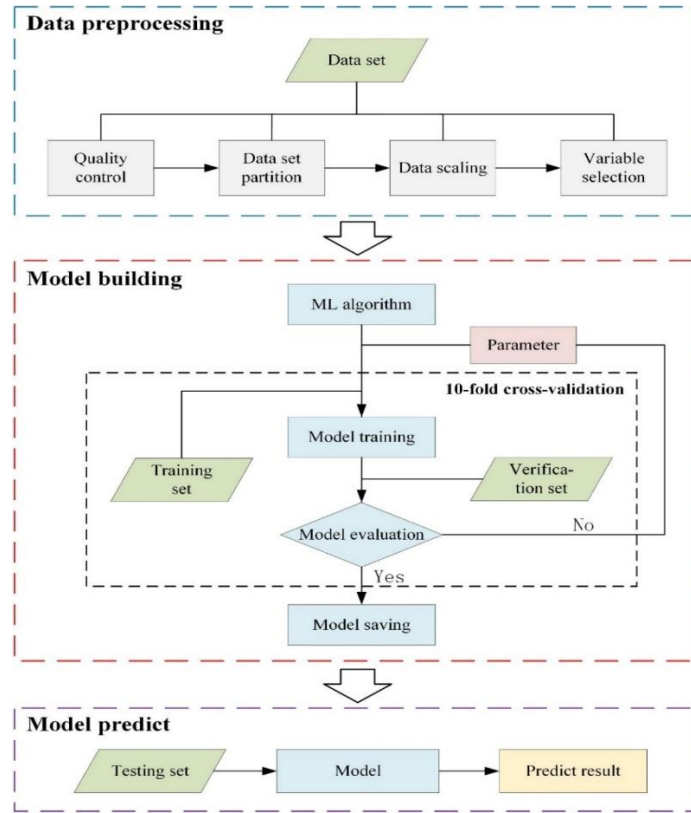


Figure 14: Modèle d'apprentissage automatique utilisé pour estimer le rayonnement solaire [31]

2.5.1 Régression linéaire

La régression linéaire (Linear Model) vise à modéliser la relation entre une variable dépendante ou cible et une ou plusieurs variables indépendantes ou caractéristiques en ajustant une ligne droite à travers les données. La régression linéaire est considérée comme l'un des algorithmes les plus connus et les mieux compris en statistique et en apprentissage automatique [32]. Les relations sont dans la majorité des cas de modèles avec des fonctions de prédiction linéaires dont des paramètres inconnus du modèle sont estimés à partir des données [33, 36]. Parmi ces différents modèles Nous avons opté pour le modèle de régression linéaire multiple en raison du nombre de nos paramètres d'entrée à savoir la température ambiante, la vitesse du vent, et l'humidité relative. Le modèle mathématique utilisé est de la forme :

$$Y = \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + \dots + \beta_n * X_n + \varepsilon \quad (2-1)$$

Les coefficients β et ε sont estimés en ajustant le modèle aux données d'entraînement. Le paramètre de sortie est la donnée que l'on veut prédire dans notre cas il s'agit de l'irradiation solaire global journalière.

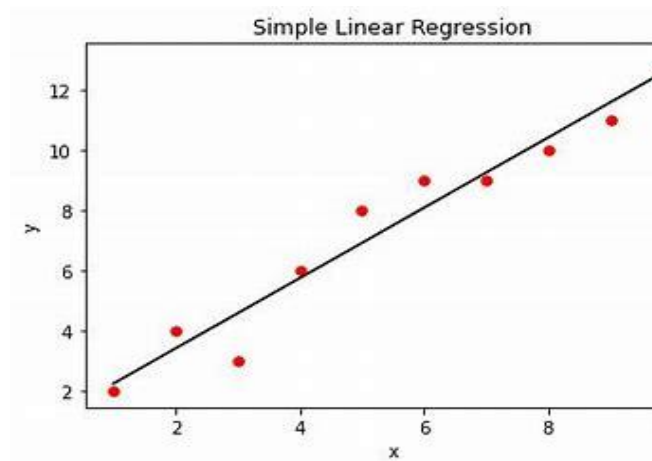


Figure 15: Régression linéaire simple

2.5.2. L'arbre à décision

L'arbre de décision (DT) est un modèle d'apprentissage automatique largement utilisé dans le domaine de l'intelligence artificielle et de l'analyse des données. Il est utilisé pour prendre des décisions ou pour prédire des résultats en fonction d'un ensemble de conditions. Il est aussi utilisé en recherche opérationnelle, en particulier dans l'analyse décisionnelle, pour aider à identifier une stratégie la plus susceptible d'atteindre un objectif [37, 38]. La construction de l'arbre de décision pour les données quantitatives implique la sélection des meilleures caractéristiques et des seuils optimaux pour diviser les données à chaque nœud.

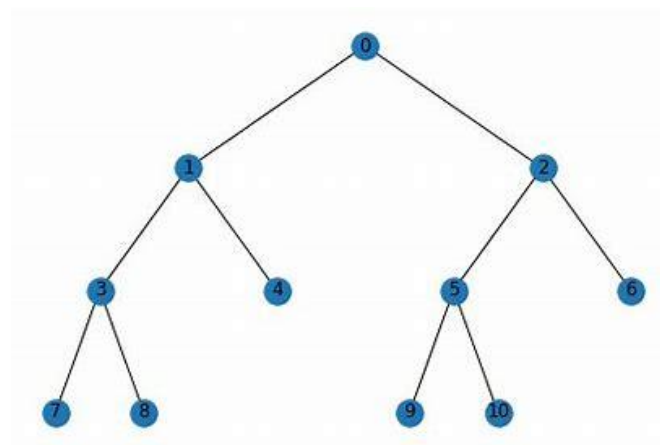


Figure 16: Arbre à décision [39]

Ce type de modèle est efficace et flexible et parfois complexe en raison de la profondeur des nœuds [40]. Une fois que l'arbre de décision est construit, il peut être utilisé pour prendre des décisions ou prédire des résultats en appliquant les tests à chaque nœud jusqu'à atteindre une feuille. Les nœuds de l'arbre représentent les tests sur les caractéristiques quantitatives, les

branches représentent les résultats des comparaisons et les nœuds feuilles contiennent les décisions ou les résultats prédits. Certaines informations concernant ce modèle peuvent être trouvées dans la littérature, et plus précisément dans les références [41, 42].

2.5.3. Les forêts aléatoires

La forêt aléatoire (RF) est une technique d'apprentissage d'ensemble proposée par Breiman qui combine le bagging et la sélection aléatoire de caractéristiques en utilisant un grand nombre d'arbres de décision non élagués [11]. Une forêt aléatoire est constituée en une collection d'arbres de décision [43]. À partir de l'ensemble de données d'entraînement, des échantillons sont sélectionnés de manière aléatoire avec remplacement pour former les ensembles d'entraînement de chaque arbre. Cette technique est appelée Bootstrap. Pour chaque échantillon, un arbre de décision est construit en utilisant un sous-ensemble aléatoire des caractéristiques. La procédure générale des forêts aléatoires pour résoudre un problème de régression est présentée dans le tableau II et la figure 17 nous montre le diagramme du processus des forêts aléatoires.

Tableau III: Procédure générale des forêts aléatoires pour la régression [2]

Soit x_0 à prévoir et

$Z = \{(x_1, y_1), \dots, (x_n, y_n)\}$, un échantillon d'entraînement de taille n

Pour k allant de 1 à K :

- Tirer un échantillon Bootstrap Z_k^* .
- Développer un arbre de régression pour cet échantillon avec randomisation des variables : la recherche de chaque division optimale est précédée d'un tirage aléatoire d'un sous ensemble de m prédicteurs.
- Estimer $\hat{f}_{zk}(x_0)$ pour l'arbre développée.

Fin pour

Calculer l'estimation moyenne $\hat{f}_{zk}(x_0) = 1/K \sum_{k=1}^K \hat{f}_{zk}(x_0)$

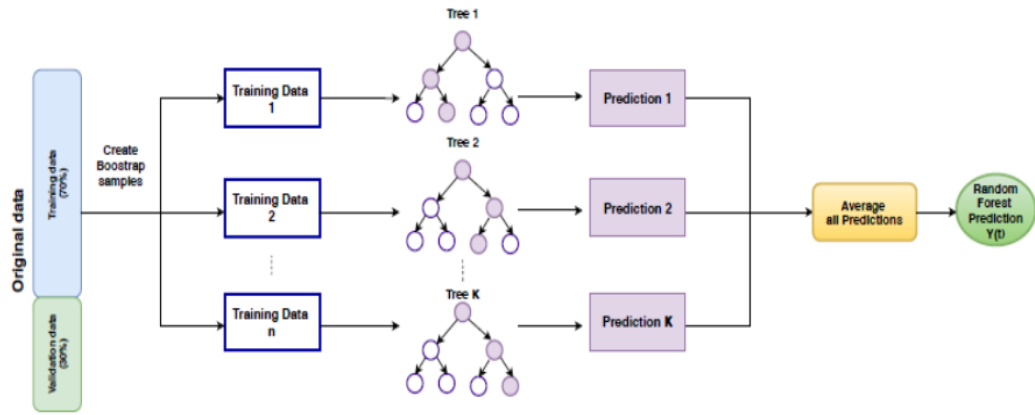


Figure 17: Diagramme du processus des forêts aléatoires [2]

La sélection aléatoire des fonctionnalités pour chaque arbre permet de décorréler les prédictions des différents arbres [44].

En résumé, les forêts aléatoires sont un modèle d'apprentissage automatique qui combine plusieurs arbres de décision en utilisant des échantillons aléatoires et des caractéristiques aléatoires. Elles offrent une meilleure généralisation, une résistance au sur ajustement et sont adaptées à des problèmes complexes de classification et de régression.

2.5.4. Les machines à vecteur de support

L'algorithme d'apprentissage automatique nommé machine à vecteurs de support est couramment utilisé pour gérer les problèmes de classification et de régression [45]. Elles sont particulièrement adaptées aux problèmes avec des données linéairement séparables ou avec une séparation complexe dans un espace de dimension supérieure. Les machines à vecteurs de support sont des outils puissants, mais leurs besoins en calcul et en stockage augmentent rapidement avec le nombre de vecteurs d'entraînement. Le cœur d'une SVM est un problème de programmation quadratique (QP), séparant les vecteurs de support du reste de l'apprentissage [46]. Les SVM cherchent à trouver un hyperplan optimal qui sépare les données en différentes classes. L'objectif est de maximiser la marge, c'est-à-dire la distance entre l'hyperplan et les exemples d'entraînement les plus proches de chaque classe. Elle utilise une entrée x qui est cartographiée dans un espace de caractéristiques à m dimensions. Ensuite, un modèle linéaire est réalisé dans cet espace caractéristique. La notation mathématique pour le modèle linéaire $f(x)$ est donnée par :

$$f(x) = \sum_{j=1}^m w_j g_j(x) + b \quad [47]$$

Où w_j est le vecteur de poids, $g_j(x)$, $j = 1, \dots, m$, désigne les caractéristiques de l'espace de haute dimension qui mappent le vecteur d'espace d'entrée x , et b est le terme du biais

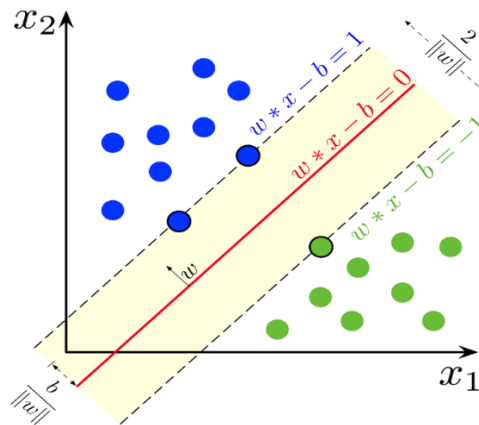


Figure 18: Hyperplan de marge maximale et marge minimale pour une SVM entraînée avec des échantillons de 2 classes [48]

Ils peuvent traiter des données linéairement séparables en utilisant un hyperplan linéaire. Cependant, ils peuvent également utiliser des noyaux (kernels) pour projeter les données dans un espace de dimension supérieure, où elles peuvent devenir linéairement séparables. Les noyaux les plus couramment utilisés sont le noyau linéaire, le noyau polynomial et le noyau gaussien (ou RBF). Les SVM peuvent également utiliser des noyaux pour traiter des données non linéairement séparables et sont résistants au bruit et aux valeurs aberrantes.

La sélection de ces quatre algorithmes ML n'a pas été effectuée au hasard. En effet, ces algorithmes sont fréquemment utilisés dans de nombreuses recherches pour prédire le rayonnement solaire global quotidien dans une région [49-50]. Ces algorithmes peuvent être utilisés pour des fonctions telles que la prédiction et l'ajustement des courbes. L'avantage de la plupart de ces algorithmes, c'est qu'ils n'ont pas, ils ont toujours besoin de la connaissance de calculs mathématiques entre les paramètres, mais ils impliquent Moins d'effort de calcul et fournir une solution compacte pour les problèmes multi variables [11]. Pendant la phase de construction du modèle notre Base de données a été repartie tels que suit 70% pour la phase l'entraînement et 30% pour la phase de test. Pour évaluer les performances de nos algorithmes on utilisera les mesures statistiques. Les différents types d'erreurs détaillés dans le tableau III donnent une vue d'ensemble de la précision entre les données mesurées et simulées.

2.6 Paramètres d'évaluations

Il est important de choisir les mesures d'évaluation appropriées en fonction du contexte et des objectifs du problème de Machine Learning. Dans cette étude, certaines mesures sont appliquées telles que l'AAE (erreur absolue moyenne), le RMSE (erreur quadratique moyenne), ARE (erreur relative moyenne) et R2 (coefficient de détermination), pour comparer le succès des performances de ces modèles de prédiction. Ces différentes mesures statistiques, leurs équations correspondantes et de brèves descriptions sont données dans le Tableau IV.

Tableau IV: Mesures statistiques

Métriques	Equations	Description et référence
L'erreur absolue moyenne (MAE)	$\frac{1}{n} \sum_{i=1}^n y_i - x_i $	MAE est un modèle de prédiction à long terme. Une valeur proche de zéro indique une meilleure performance.[51].
Erreur quadratique Moyenne (RMSE)	$\frac{1}{\sqrt{n}} \sum_{i=1}^n (y_i - x_i)^2$	RMSE donne des nouvelles concernant la précision à court terme. Il doit être positif et proche de 0 pour une bonne précision [51].
Coefficient de détermination R²	$1 - \frac{\sum_{j=1}^n (y_i - x_i)^2}{\sum_{i=1}^n (x_i - \bar{x}_i)^2}$	Valeur de 0 à 1. Une bonne performance signifie que R ² doit être proche de 1 [52].
L'erreur relative % (ARE)	$\sigma = \left \frac{v_a - v_e}{v_e} \right * 100$	Elle varie de 0 à 100 [52].

La méthode R², (également appelé coefficient de détermination), qui est la proportion de variance (%) de la variable dépendante qui peut être expliquée par les variables indépendante [53]. Le R² permet d'évaluer la qualité de la prédiction d'un modèle statistique.

- ◇ Si la valeur du R carré < 0,3, cette valeur est généralement considérée comme une taille d'effet Nulle ou Très faible,
- ◇ Si la valeur du R carré est de 0,3 < r < 0,5, cette valeur est généralement considérée comme une taille d'effet faible ou faible,

- ◇ Si la valeur du R carré est de $0,5 < r < 0,7$, cette valeur est généralement considérée comme une taille d'effet modérée,
- ◇ Si la valeur du R carré $r > 0,7$, cette valeur est généralement considérée comme une taille d'effet forte. [11]

Dans cette étude le logiciel utilisé pour la programmation est le logiciel Jupyter d'Anaconda et parfois VS code. Les différentes figures sont obtenues grâce au logiciel Excel de Microsoft office et également sur Jupyter notebook.

Conclusion

En conclusion Dans ce chapitre il était question pour nous de présenter les différents matériels et méthodes permettant de prédire le rayonnement solaire globale. Pour cela nous avons commencé par délimiter les sites d'études, faire une collecte des données sur l'irradiation solaire, suivis d'une étape de prétraitement de données afin de permettre leur utilisation. Une sélection de choix a été effectué sur les algorithmes machine Learning utilisés pour faire notre prédiction ainsi que sur les performances de nos algorithmes. Le choix de nos logiciels pour la programmation c'est porté sur jupyter d'anaconda par conséquent une préférence pour le langage de programmation python.

CHAPITRE 3 : **RESULTATS ET DISCUSSIONS**

CHAPITRE 3 : RESULTATS ET DISCUSSIONS

Introduction

Les méthodologies présentées dans les deux chapitres précédents nous ont permis d'entreprendre une série de manipulations en vue de la prédiction de l'irradiation solaire globale journalière. Le chapitre 2, a présenté les modèles d'apprentissage automatique qui seront utilisés dans ce mémoire, les méthodes d'élaboration de ceux-ci ainsi les différents critères de performance qui seront par la suite employés pour tester la fiabilité des modèles proposés. La prochaine section de ce chapitre est dédiée à l'étude et l'estimation de l'irradiation solaire globale journalière dont plusieurs approches intelligentes seront développées et testées par le biais d'outils de comparaison et d'évaluation d'erreurs.

Cette partie de notre travail est consacrée à la présentation de nos différents résultats sur les modèles de machines Learning utilisés pour la prédiction, une comparaison de ceux-ci ainsi qu'à la discussion détaillée de ces résultats.

3.1 Performance des algorithmes Machine Learning

Dans ce travail évaluons la prédiction de l'irradiation solaire journalière dans 5 villes du Cameroun à l'aide de 4 algorithmes Machine Learning et nous discutons sur les performances de ceux-ci à l'aide de quatre métriques régulièrement utilisés dans la littérature.

Le tableau V nous montre les différentes valeurs numériques des métriques pour les différentes villes étudiées pour chaque modèle développé. Le temps relatif à l'entraînement du modèle est en moyenne de 2min 20s. On remarque que le coefficient de corrélation R^2 varie entre 0.03 et 0.58.

Tableau V : Comparaison des performances des algorithmes pour chaque ville

Villes	Métriques	Modèle Linéaire	Arbre à décision	Forêts aléatoires	Machines à vecteur de support
BERTOUA	R2	0.27	0.03	0.38	0.27
	MSE	0.98	1.12	0.91	0.99
	MAE	0.8	0.9	0.72	0.79
	RMSE	0.99	1.06	0.95	1
DOUALA	R2	0.54	0.34	0.58	0.31
	MSE	0.72	0.87	0.69	0.89
	MAE	0.6	0.64	0.52	0.71
	RMSE	0.85	0.93	0.83	0.91
GAROUA	R2	0.46	0.25	0.48	0.45
	MSE	0.8	0.94	0.79	0.81
	MAE	0.61	0.7	0.61	0.61
	RMSE	0.9	0.98	0.89	0.9
MAROUA	R2	0.38	0.36	0.47	0.4
	MSE	0.8	0.81	0.74	0.78
	MAE	0.62	0.6	0.55	0.6
	RMSE	0.89	0.9	0.86	0.89
YAOUNDE	R2	0.28	0.14	0.35	0.31
	MSE	0.89	0.97	0.85	0.86
	MAE	0.72	0.76	0.69	0.7
	RMSE	0.94	0.99	0.92	0.93

3.2 Discussion des performances

A Bertoua, le modèle linéaire nous donne un score en termes de R^2 de 0.27. Ce résultat montre une corrélation de 27% entre les variables indépendantes ceci est une taille d'effet faible. En termes de MAE nous avons une valeur de 0.8. Ceci signifie que si l'on prédit la valeur l'irradiation solaire globale journalière ($HG_H = 4.86$), la valeur réelle est comprise entre 4.06 et 5.66.

L'algorithme des forêts aléatoires nous donne un score de 0.38 en termes de R^2 ce qui correspond à une corrélation de 38% entre les variables indépendantes ceci est considéré

comme une valeur faible. En termes de MAE nous avons un score de 0.72. Cela signifie que si on prédit la valeur de l'irradiation solaire journalière ($HG_H = 5.61$), la valeur réelle se situe entre 4.89 et 6.33. -

Le modèle d'arbre à décision nous donne un score de 0.03 en termes de R^2 , Une valeur qui tend vers zéro ce qui correspond à un effet de corrélation nulle entre les variables indépendantes. Cependant pour ce même algorithme on trouve un score de 0.9 en termes de MAE. Cela signifie que si l'on prédit la valeur du rayonnement solaire globale journalier ($HG_H = 5.75$), la valeur réelle se situe entre 4.85 et 6.65.

Dans la ville de Douala, le modèle linéaire nous donne un score en termes de R^2 de 0.54. Ce résultat montre une corrélation de 54 % entre les variables indépendantes ceci est une valeur -significative avec un effet modéré. En termes de MAE nous avons une valeur de 0.6. Ceci signifie que si l'on prédit la valeur l'irradiation solaire globale journalière ($HG_H = 4.86$), la valeur réelle est comprise entre 4.26 et 5.45.

L'algorithme des forêts aléatoires nous donne un score de 0.58 en termes de R^2 ce qui correspond à une corrélation de 58% entre les variables indépendantes ceci est considéré comme ayant un effet modéré. En termes de MAE nous avons un score de 0.52. Cela signifie que si on prédit la valeur de l'irradiation solaire journalière ($HG_H = 5.61$), la valeur réelle se situe entre 5.09 et 6.13.

Le modèle d'arbre à décision nous donne un score de 0.34 en termes de R , ce qui correspond effet de corrélation faible entre les variables indépendantes. Cependant pour ce même algorithme on trouve un score de 0.64 en termes de MAE. Cela signifie que si l'on prédit la valeur du rayonnement solaire globale journalier ($HG_H = 5.75$), la valeur réelle se situe entre 5.11 et 6.39.

A Garoua, le modèle linéaire nous donne un score en termes de R^2 de 0.46. Ce résultat montre une corrélation de 46 % entre les variables indépendantes ceci est une taille d'effet faible. En termes de MAE nous avons une valeur de 0.61. Ceci signifie que si l'on prédit la valeur de l'irradiation solaire global journalier ($HG_H = 4.86$), la valeur réelle est comprise entre 4.25 et 5.47.

L'algorithme des forêts aléatoires nous donne un score de 0.48 en termes de R^2 ce qui correspond à une corrélation de 48% entre les variables indépendantes ceci est considéré comme une valeur faible. En termes de MAE nous avons un score de 0.61. Cela signifie que si

on prédit la valeur de l'irradiation solaire journalière ($HG_H = 5.61$), la valeur réelle se situe entre 5.00 et 6.22.

Le modèle d'arbre à décision nous donne un score de 0.27 en termes de R^2 , cette valeur correspond à un effet de corrélation très faible entre les variables indépendantes. Cependant pour ce même algorithme on trouve un score de 0.7 en termes de MAE. Cela signifie que si l'on prédit la valeur du rayonnement solaire global journalier ($HG_H = 5.75$), la valeur réelle se situe entre 5.05 et 6.45.

Dans la ville de Maroua, le modèle linéaire nous donne un score en termes de R^2 de 0.38. Ce résultat montre une corrélation de 38 % entre les variables indépendantes ceci est une valeur significative avec un effet modéré. En termes de MAE nous avons une valeur de 0.62. Ceci signifie que si l'on prédit la valeur l'irradiation solaire globale journalière ($HG_H = 4.86$), la valeur réelle est comprise entre 4.24 et 5.48. L'algorithme des forêts aléatoires nous donne un score de 0.47 en termes de R^2 ce qui correspond à une corrélation de 47% entre les variables indépendantes ceci est considéré comme ayant un effet faible. En termes de MAE nous avons un score de 0.55. Cela signifie que si on prédit la valeur de l'irradiation solaire journalière ($HG_H = 5.61$), la valeur réelle se situe entre 5.06 et 6.16.

Le modèle d'arbre à décision nous donne un score de 0.36 en termes de R , ce qui correspond à un effet de corrélation faible entre les variables indépendantes. Cependant pour ce même algorithme on trouve un score de 0.6 en termes de MAE. Cela signifie que si l'on prédit la valeur du rayonnement solaire globale journalier ($HG_H = 5.75$), la valeur réelle se situe entre 5.15 et 6.35.

La ville de Yaoundé quant à elle, nous donne pour le modèle linéaire un score en termes de R^2 de 0.28. Ce résultat montre une corrélation de 28 % entre les variables indépendantes. Ceci montre un effet relativement faible. En termes de MAE nous avons une valeur de 0.72. Ceci signifie que si l'on prédit la valeur l'irradiation solaire globale journalière ($HG_H = 4.86$), la valeur réelle est comprise entre 4.14 et 5.45.

L'algorithme des forêts aléatoires nous donne un score de 0.35 en termes de R^2 ce qui correspond à une corrélation de 35% entre les variables indépendantes ceci est considéré comme ayant un effet faible. En termes de MAE nous avons un score de 0.69. Cela signifie que si on prédit la valeur de l'irradiation solaire journalière ($HG_H = 5.61$), la valeur réelle se situe entre 4.92 et 6.3.

Le modèle d'arbre à décision nous donne un score de 0.14 en termes de R, ce qui correspond à un effet de corrélation très faible entre les variables indépendantes. Cependant pour ce même algorithme on trouve un score de 0.76 en termes de MAE. Cela signifie que si l'on prédit la valeur du rayonnement solaire global journalier ($HG_H = 5.75$), la valeur réelle se situe entre 4.99 et 6.51.

3.3 Visualisation des données

Dans cette partie on présente les différentes figures obtenues depuis le traitement des données jusqu'aux métriques permettant de mieux interpréter nos résultats.

3.3.1 Irradiation solaire global

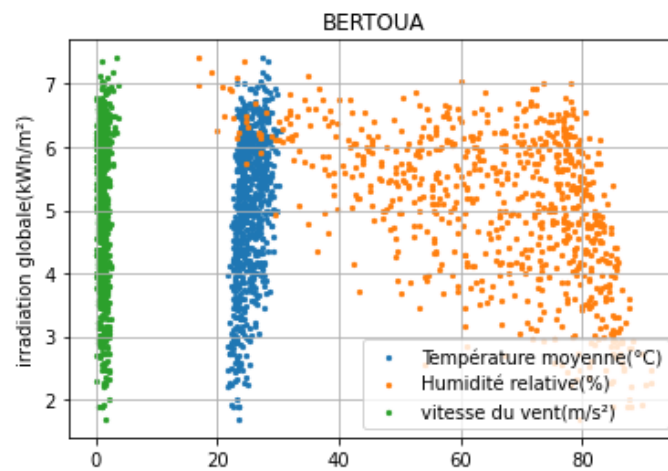


Figure 19: Irradiation solaire global journalière dans la ville de Bertoua

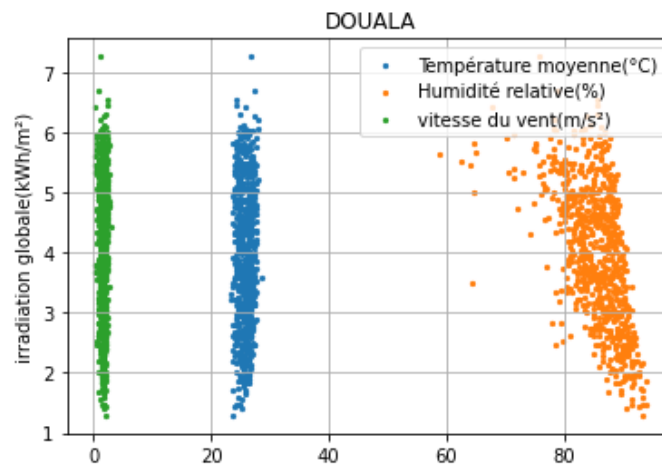


Figure 20: Irradiation solaire global journalière dans la ville de Douala

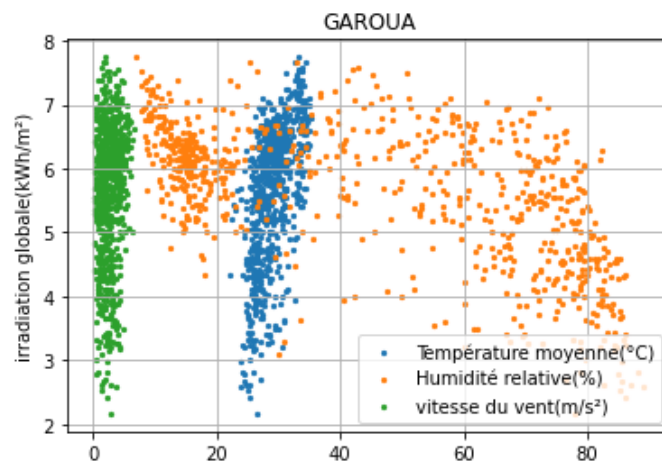


Figure 21: Irradiation solaire global journalière dans la ville de Garoua

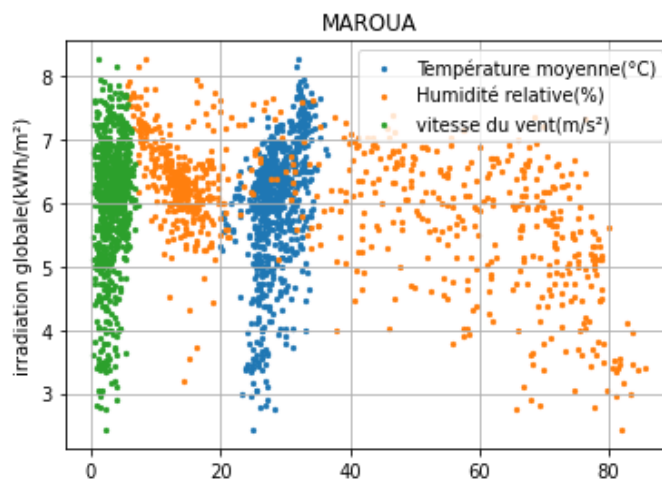


Figure 22: Irradiation solaire global dans la ville de Maroua

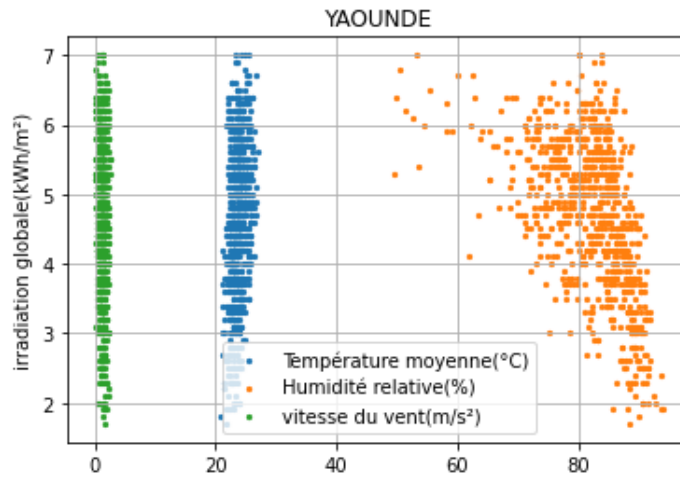


Figure 23: Irradiation solaire global dans la ville de Yaoundé

Dans ces différentes figures, l'irradiation solaire globale sur la surface horizontale est représenté en fonction de la température moyenne en degré Celsius en Bleu, l'humidité relative en Orange, la vitesse du vent en vert pour les différentes villes étudiée Bertoua, Douala, Garoua, Maroua, Yaoundé. On constate des distributions différentes pour chacune des villes malgré que la température et l'irradiation varient très peu contrairement à l'humidité relative qui est beaucoup plus étalé.

3.3.2 Performance en termes de RMSE et R^2

Pour définir le meilleur modèle de Machine Learning adapté pour la prédiction. Nous avons opté sur l'analyse approfondi de deux métriques à savoir le RMSE et la méthode R^2 . Les figure ci-dessous nous permet de visualiser les scores des différents algorithmes de machines Learning pour les différentes villes.

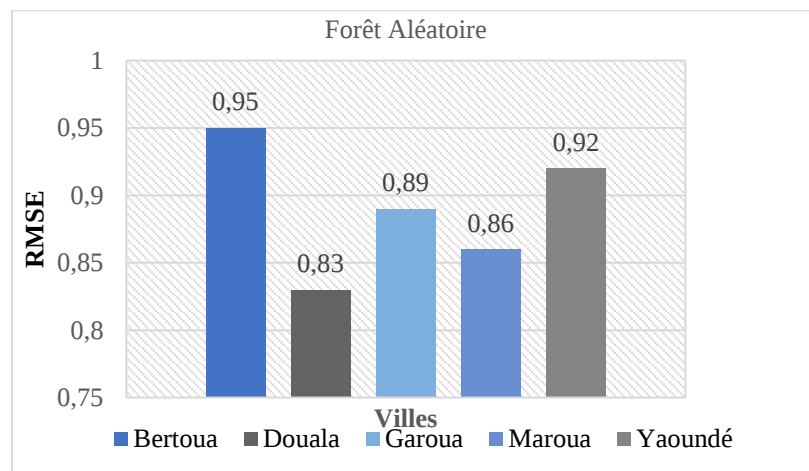


Figure 24: Valeurs du RMSE dans le cas des forêts aléatoires pour les villes de Bertoua, Douala, Garoua, Maroua, Yaoundé

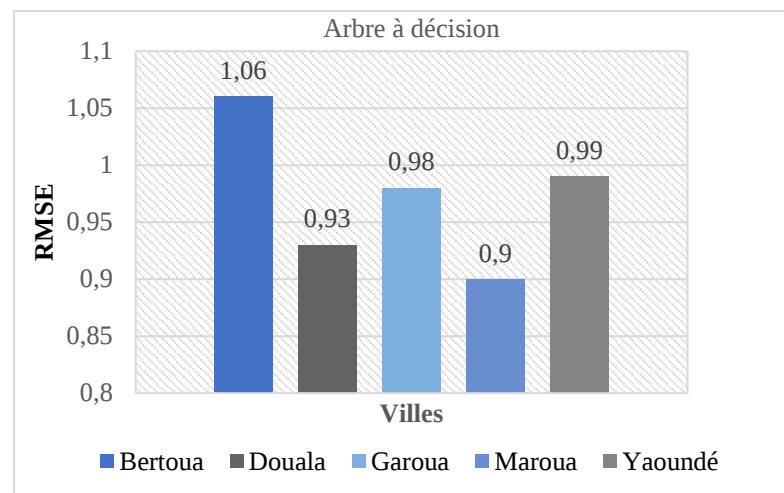


Figure 25: Valeurs du RMSE pour l'arbre à décision pour les villes de Bertoua, Douala, Garoua, Maroua, Yaoundé

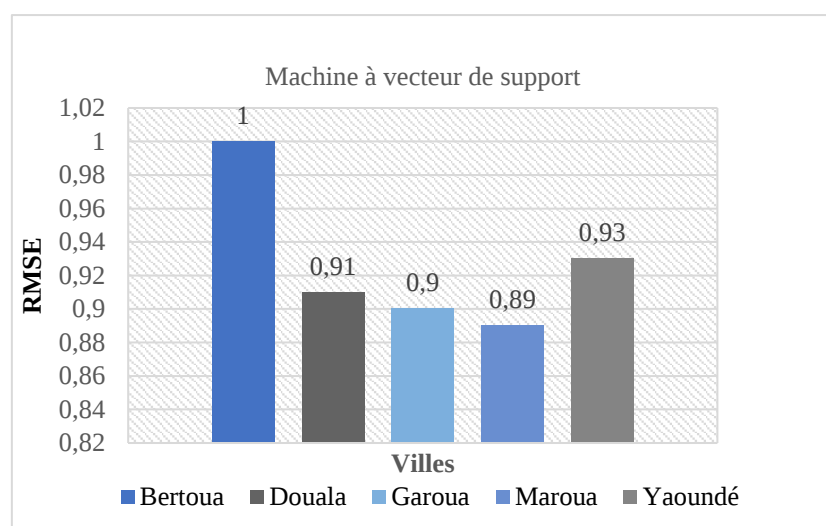


Figure 26: Valeurs du RMSE pour les machines à vecteur de support pour les villes de Bertoua, Douala, Garoua, Maroua, Yaoundé

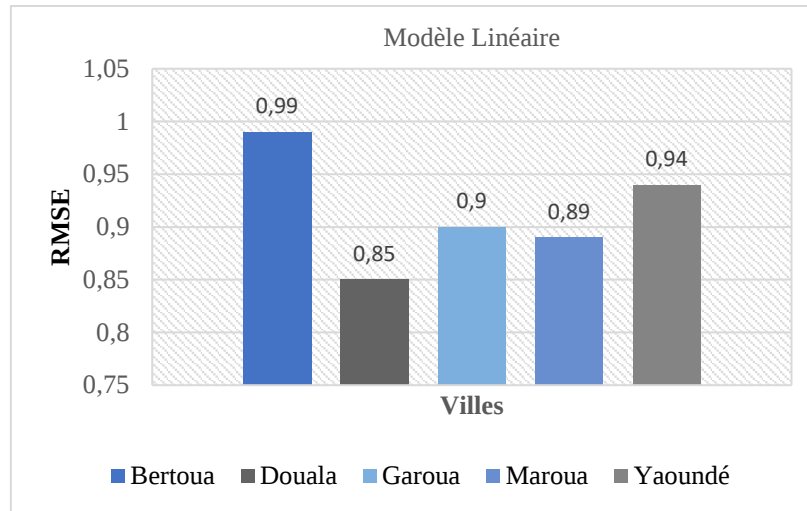


Figure 27: valeurs du RMSE du modèle de régression linéaire pour les villes de Bertoua, Douala, Garoua, Maroua, Yaoundé.

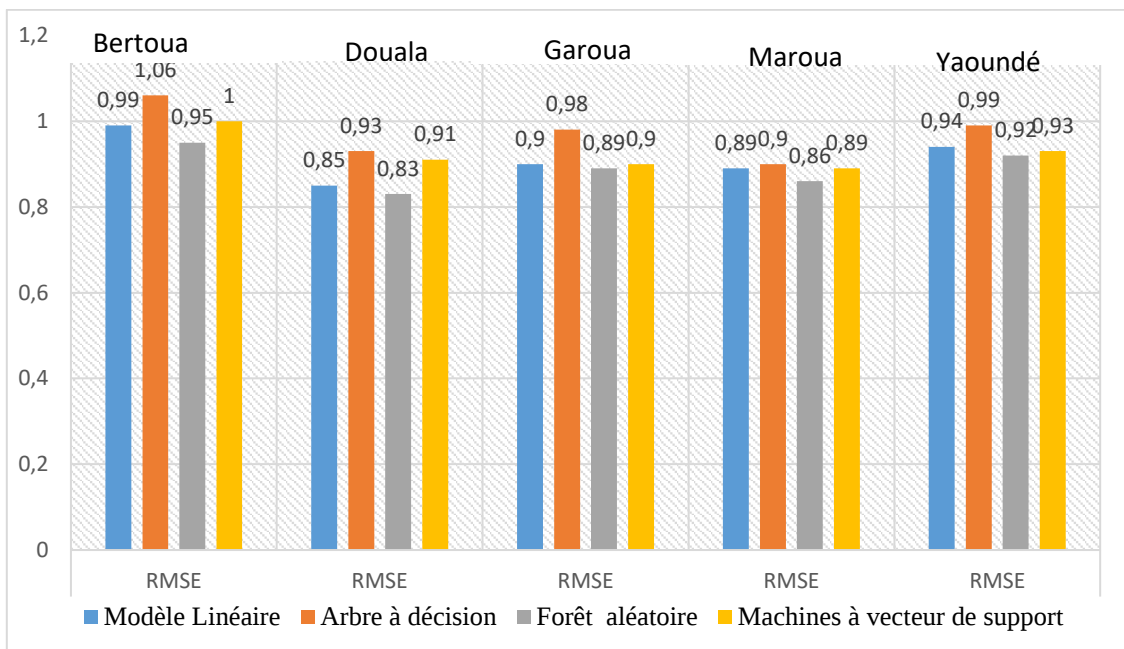


Figure 28: Tableau récapitulatif des scores en termes de RMSE pour les différents algorithmes de machine Learning pour les différentes villes

En termes de RMSE plus la valeur est proche de zéro plus la prédiction est bonne.

La valeur la plus basse du RMSE pour le modèle des forêts aléatoires a été observé pour la ville de Douala suivi respectivement de Maroua, Garoua, Yaoundé et Bertoua. Cela montre

que la meilleure performance de l'algorithme des forêts aléatoires est enregistrée dans la ville de Douala.

Pour le modèle d'arbre à décision, les valeurs du RMSE de manière croissante ont été respectivement observées pour les villes de Douala, Maroua, Garoua, Yaoundé et Bertoua. Cela montre que la meilleure performance de l'algorithme d'arbre à décision est enregistrée dans la ville de Douala

Pour le modèle de machines à vecteur de support, les valeurs sont observées de manière ascendante pour les villes de Maroua, Garoua, Douala, Yaoundé, Bertoua. Cela montre que la meilleure performance de l'algorithme des machines à vecteur de support est enregistrée dans la ville de Maroua.

Pour le modèle de régression linéaire, les valeurs sont observées de manière ascendante pour les villes de Douala, Garoua, Maroua, Yaoundé, Bertoua. Cela montre que la meilleure performance de l'algorithme de régression linéaire est enregistrée dans la ville de Douala.

Les figures ci-dessous nous montre les valeurs des scores en termes de R^2 de chaque algorithme de machines Learning pour différentes villes.

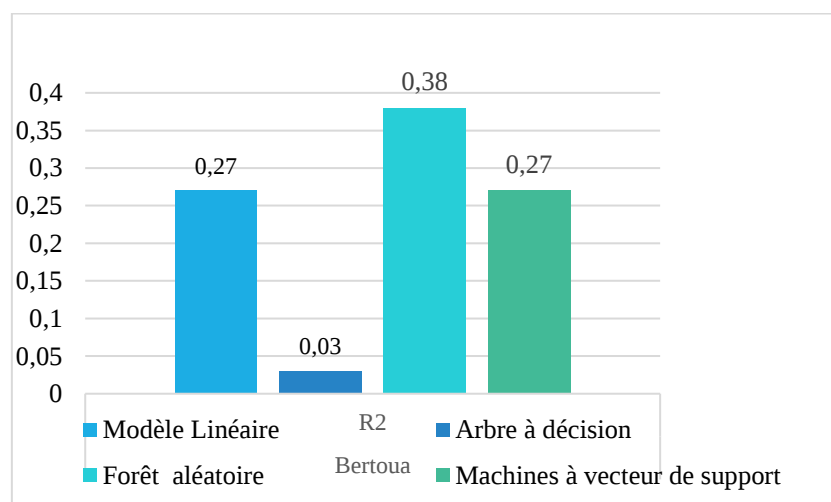


Figure 29: Valeurs du coefficient de détermination R^2 pour la ville de Bertoua

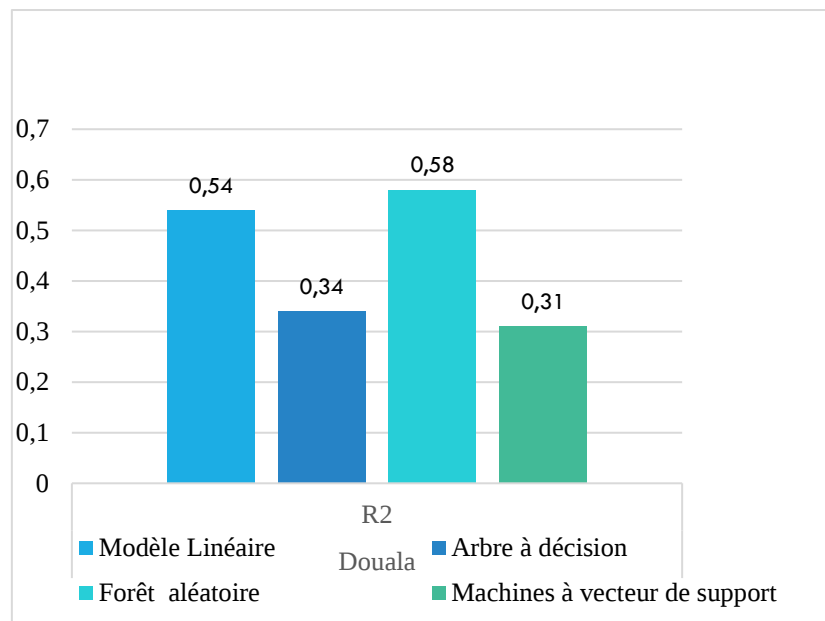


Figure 30: Valeurs du coefficient de détermination R^2 pour la ville de Douala

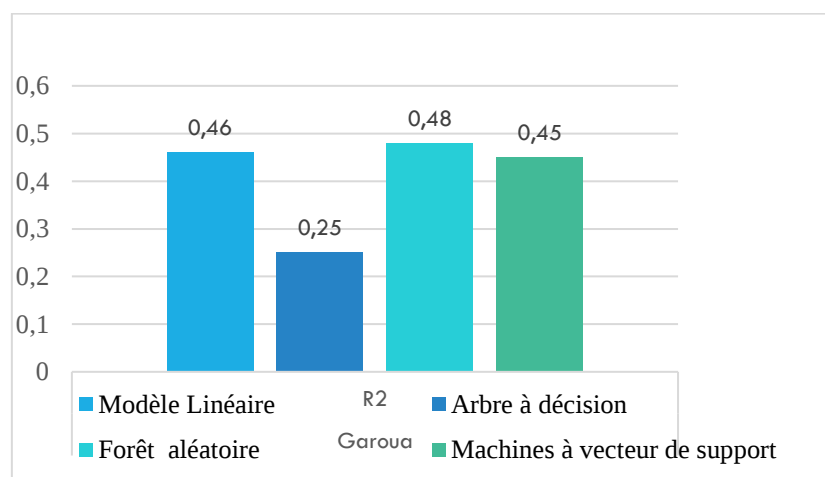


Figure 31: Valeurs du coefficient de détermination R^2 pour la ville de Garoua

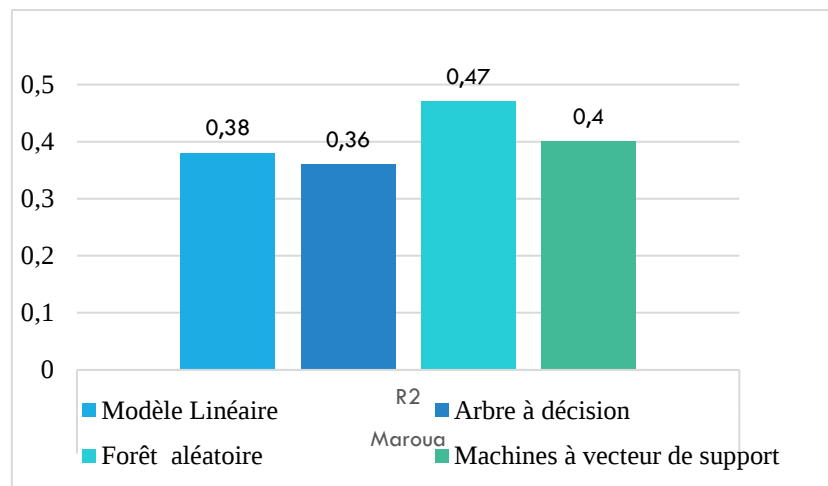


Figure 32: Valeurs du coefficient de détermination R^2 pour la ville de Maroua

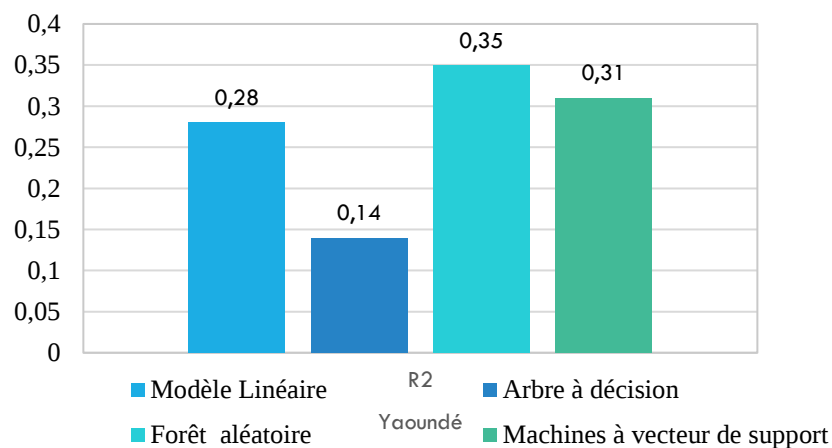


Figure 33: Valeurs du coefficient de détermination R^2 pour la ville de Yaoundé

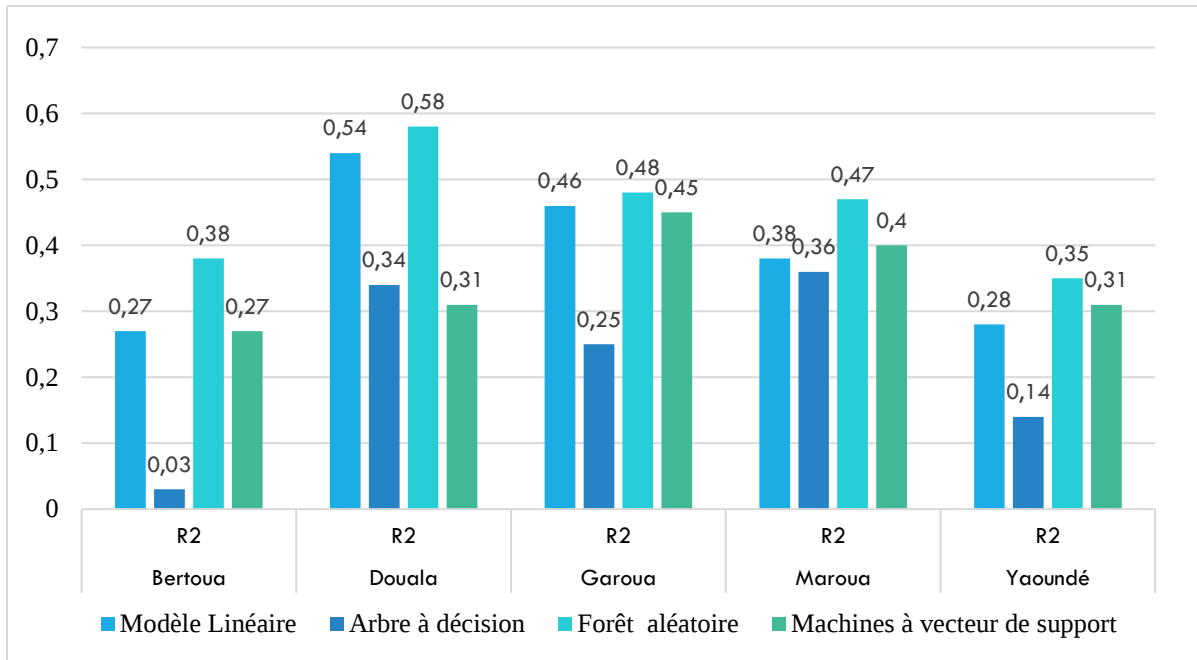


Figure 34: Récapitulatif des valeurs du coefficient de détermination R^2 pour les différentes villes

Ces figures nous montrent qu'en considérant la méthode R^2 pour l'évaluation des performances de nos algorithmes, le modèle des forêts aléatoires obtienne de meilleurs score suivis respectivement du modèle linéaire, des machines à vecteur de support et de l'arbre à décision.

3.4 Comparaison des modèles de machine Learning

Dans cette partie nous effectuons des comparaisons entre les différents algorithmes de Machine Learning afin de déterminer le meilleur modèle adapté pour la prédiction de l'irradiation solaire. Pour définir le meilleur algorithme adapté pour les différentes villes Bertoua, Douala, Garoua, Maroua et Yaoundé. Nous faisons une observation comparative sur les métriques R^2 et MSE.

Les figures ci-contre récapitule les variations de RMSE et R^2 pour les différents modèles de Machine Learning.

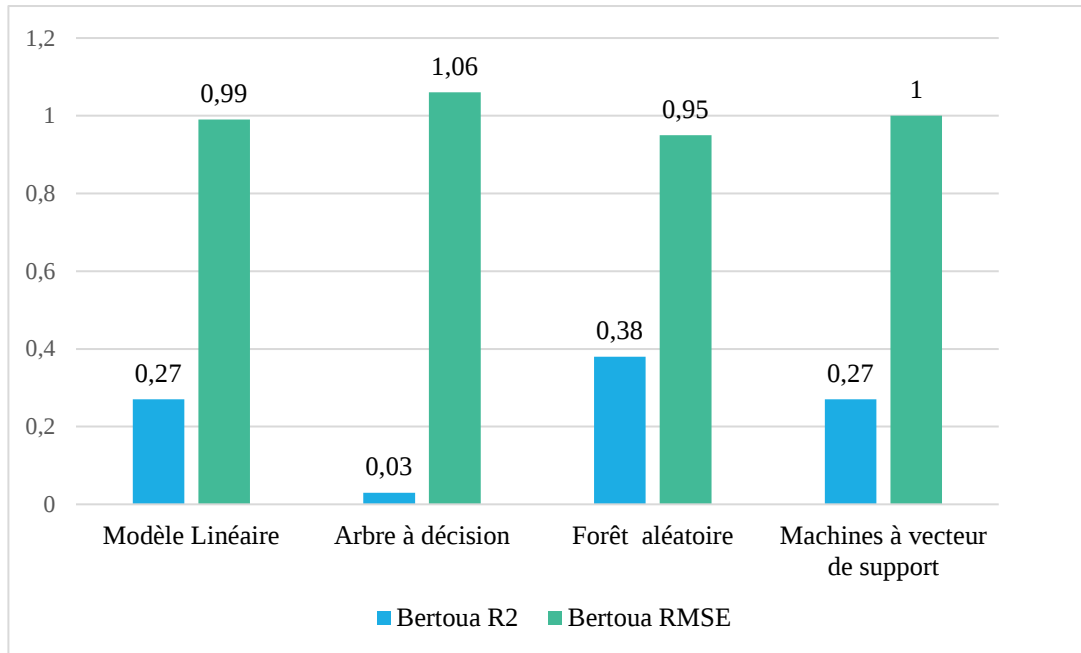


Figure 35: Comparaison entre les scores de R^2 et RMSE pour la ville de Bertoua.

De ces résultats l'on observe que la valeur la plus basse de RMSE est obtenu pour l'algorithme des forêts aléatoires avec un score de 0.95 et la plus haute de R^2 est obtenu pour les forêt aléatoires avec un score de 0.38. Cela montre que pour la ville de Bertoua le meilleur modèle de machine Learning adapté pour la prédiction est le modèle des forêts aléatoires.

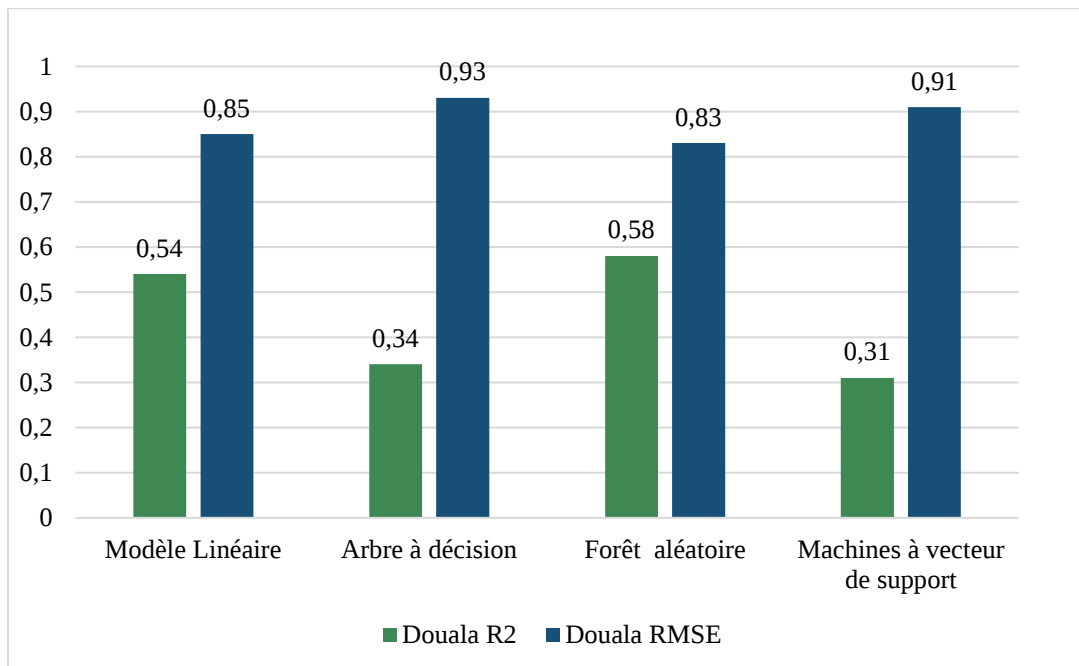


Figure 36: Comparaison entre les scores de R^2 et RMSE pour la ville de Douala.

De ces résultats de la figure 36 l'on observe que la valeur la plus basse de RMSE est obtenu pour l'algorithme des forêts aléatoires avec un score de 0.83 et la plus haute de R^2 est obtenu pour les forêt aléatoires avec un score de 0.58. Cela montre que pour la ville de Douala le meilleur modèle de machine Learning adapté pour la prédiction est également le modèle des forêts aléatoires.

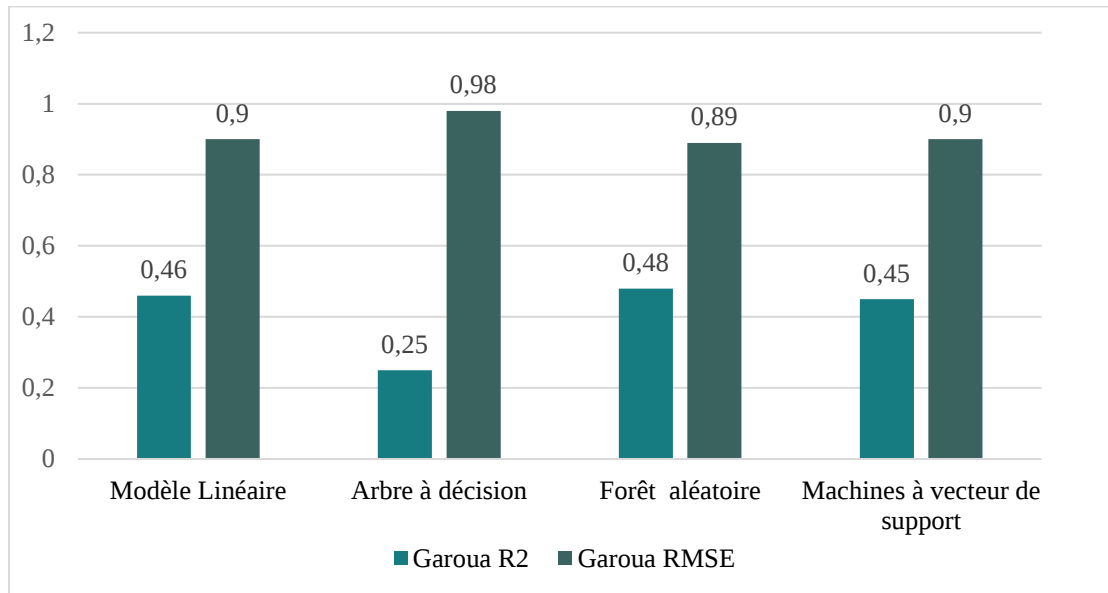


Figure 37: Comparaison entre les scores de R^2 et RMSE pour la ville de Garoua

Des résultats de la figure 37 l'on observe que la valeur la plus basse de RMSE est obtenu pour l'algorithme des forêts aléatoires avec un score de 0.89 et la plus haute de R^2 est obtenu pour les forêt aléatoires avec un score de 0.48. Cela montre que pour la ville de Garoua le meilleur modèle de machine Learning adapté pour la prédiction est le modèle des forêts aléatoires.

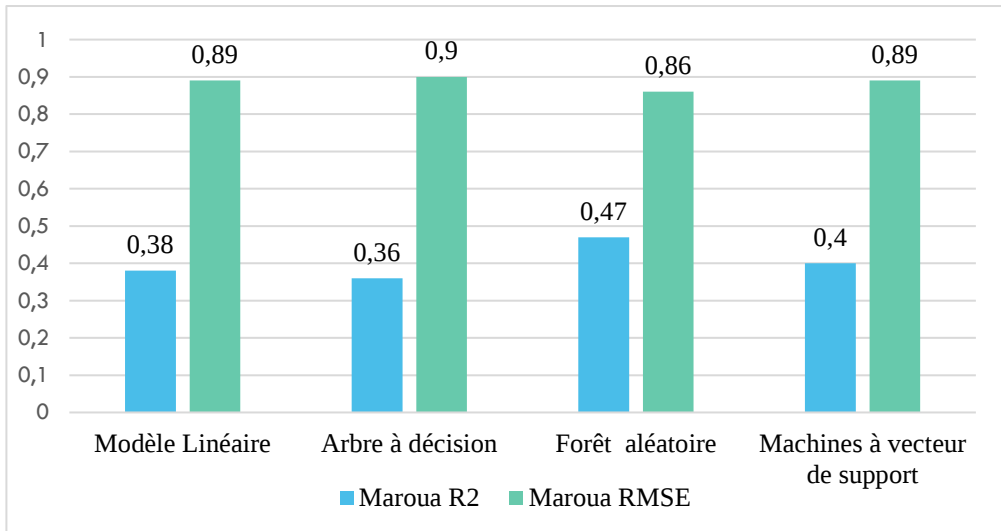


Figure 38: Comparaison entre les scores de R^2 et RMSE pour la ville de Maroua

Des résultats de la figure 38 l'on observe que la valeur la plus basse de RMSE est obtenu pour l'algorithme des forêts aléatoires avec un score de 0.86 et la plus haute de R^2 est obtenu pour les forêt aléatoires avec un score de 0.47. Cela montre que pour la ville de Maroua le meilleur modèle de machine Learning adapté pour la prédiction est le modèle des forêts aléatoires.

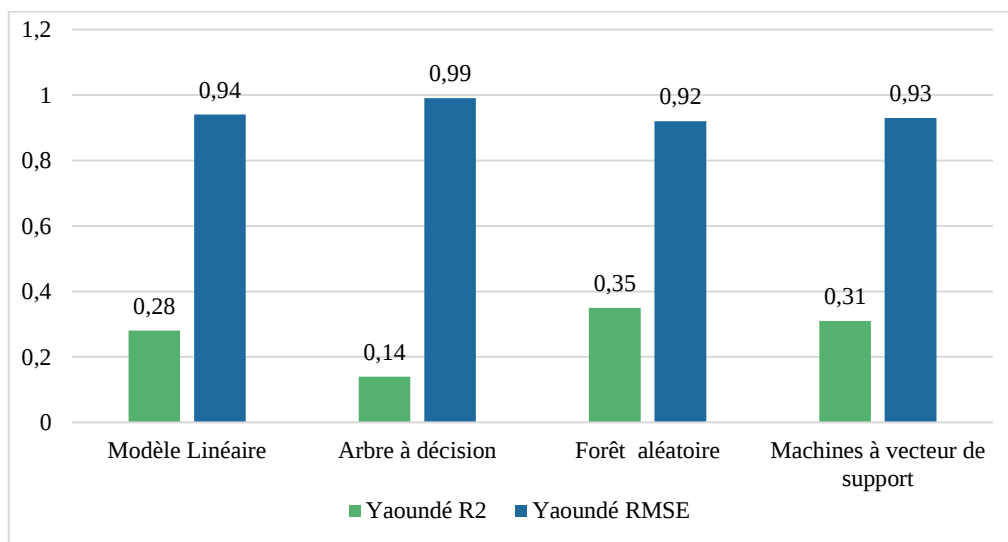


Figure 39: Comparaison entre les scores de R^2 et RMSE pour la ville de Yaoundé

Des résultats de la figure 39 l'on observe que la valeur la plus basse de RMSE est obtenu pour l'algorithme des forêts aléatoires avec un score de 0.92 et la plus haute de R^2 est obtenu pour les forêt aléatoires avec un score de 0.35. Cela montre que pour la ville de Yaoundé le

meilleur modèle de machine Learning adapté pour la prédiction est le modèle des forêts aléatoires.

Les différents résultats obtenus sur les différentes villes nous montrent donc que le meilleur modèle pour prédire l'irradiation solaire globale est le modèle des forêts aléatoires dans toutes les villes. Cependant les résultats obtenus avec le modèle linéaire se rapproche de très près à ceux obtenu avec l'algorithme des forêts aléatoires suivis des machines à vecteur de support. Le pire score est obtenu avec le modèle d'arbre à décision.

3.5 Résultats de la prédiction

Dans cette partie nous présentons les résultats des valeurs réelles et prédite de l'irradiation solaire en kWh/m² obtenu avec le modèle des forêts aléatoires pour nos différents sites d'études.

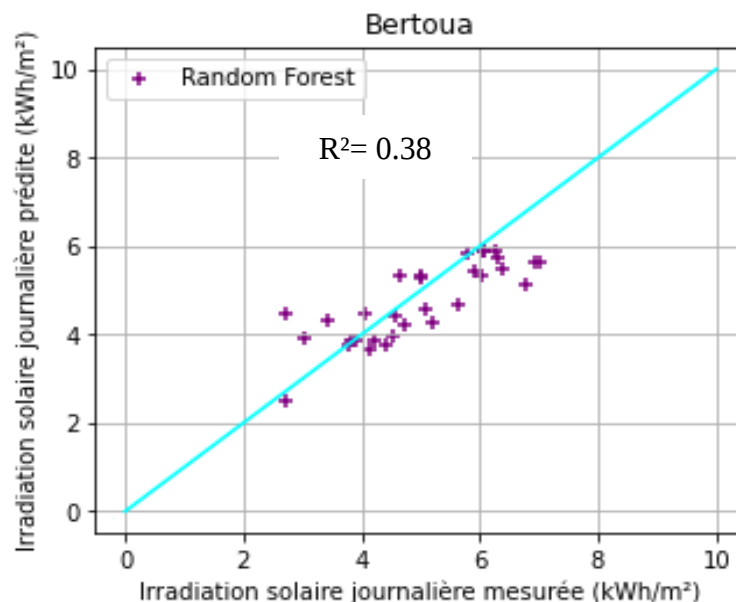


Figure 40: Nuage de points des valeurs de l'irradiation journalière mesurées et estimées pour la ville de Bertoua

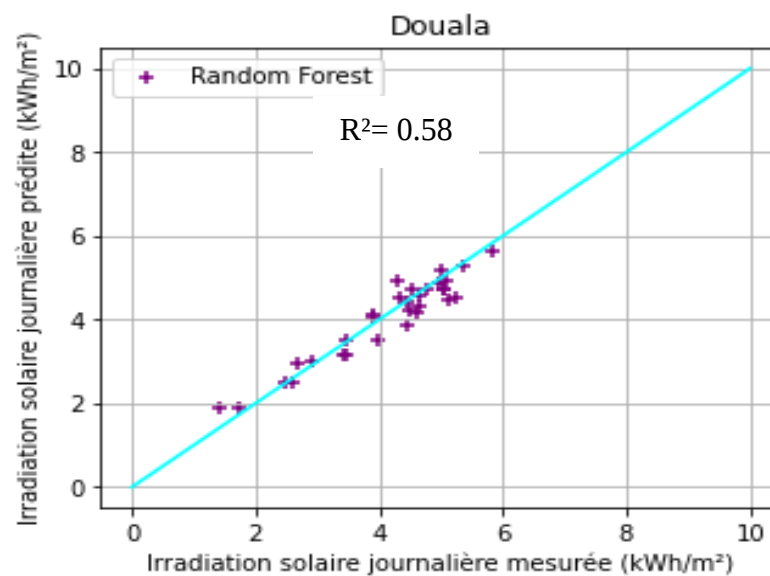


Figure 41: Nuage de points des valeurs de l'irradiation journalière mesurées et estimées pour la ville de Douala

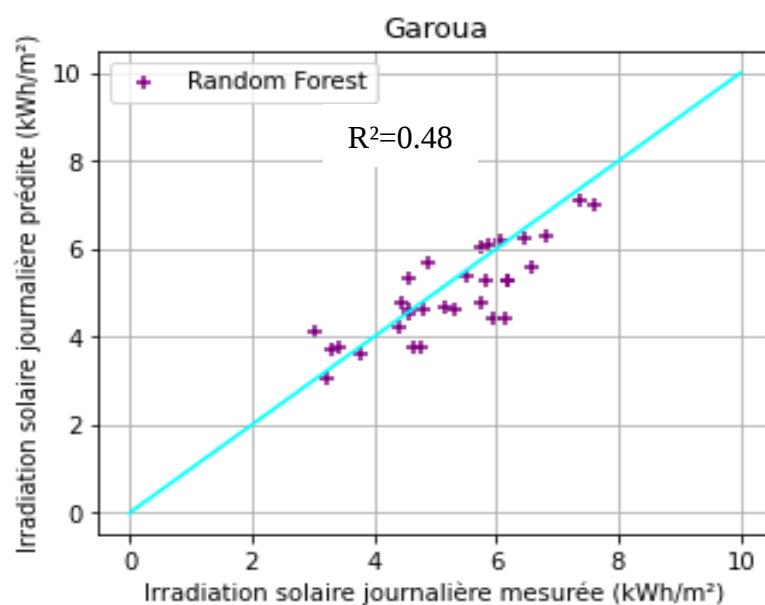


Figure 42: Nuage de points des valeurs de l'irradiation journalière mesurées et estimées pour la ville de Garoua

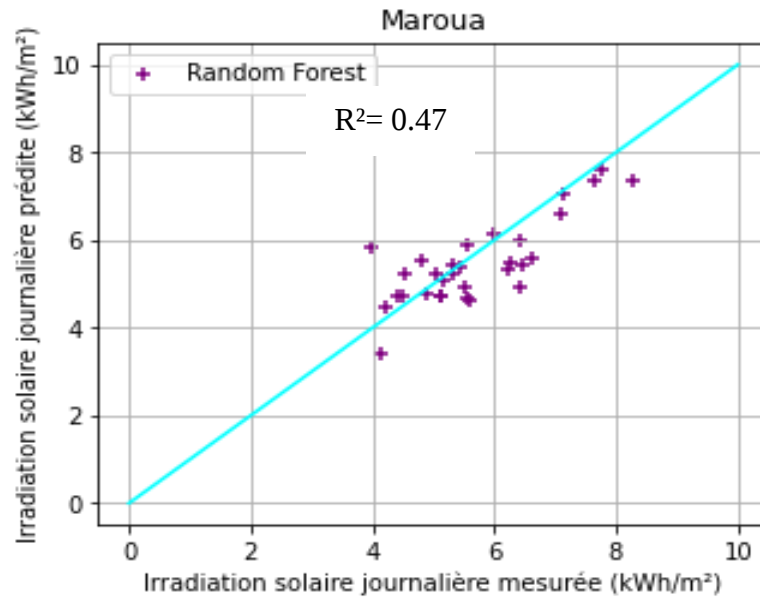


Figure 43: Nuage de points des valeurs de l'irradiation journalière mesurées et estimées pour la ville de Maroua

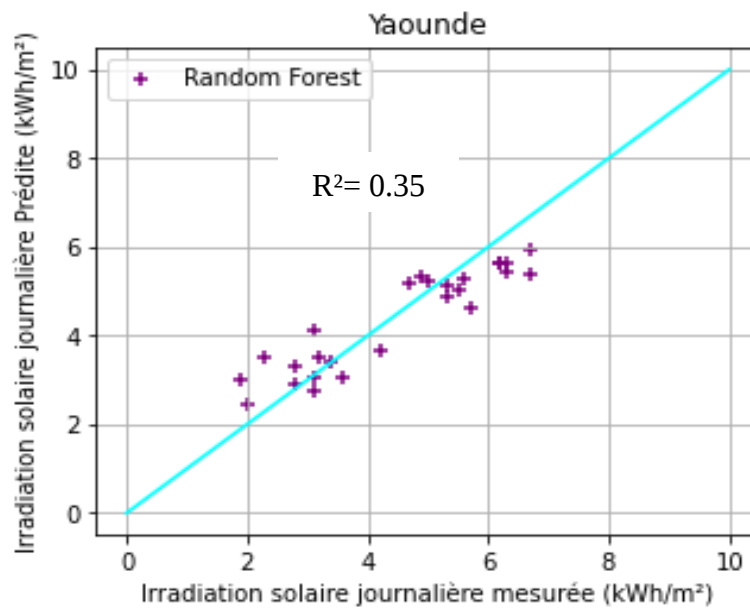


Figure 44: Nuage de points des valeurs de l'irradiation journalière mesurées et estimées pour la ville de Yaoundé

Ici, la comparaison s'est faite avec une plage de 25 à 30 valeurs aléatoires de l'irradiation solaire pour chacune des différentes villes. Ainsi on a obtenu les valeurs de l'irradiation solaire globale journalière prédite pour le modèle des forêts aléatoires qui a enregistré le meilleur score. Dans ces différentes figures, plus les points se rapproche de la droite de régression linéaire plus

le résultat de la prédiction est proche de la valeur mesurée. L'irradiation solaire journalière prédite. On remarque que pour le cas de la ville de Douala, la distribution des valeurs suit la droite de régression l'écart type est relativement faible ce qui est en accord avec les résultats obtenus au tableau IV Plus haut.

Conclusion

En conclusion dans ce chapitre, nous avons évalué les performances de nos algorithmes à l'aide des quatre métriques à savoir le MSE, MAE, RMSE et R^2 . En effectuant des comparaisons sur les différents algorithmes nous avons trouvé que l'algorithme des forêts aléatoires obtient le meilleur score dans toutes les villes ciblées de notre étude. Cela a montré que l'algorithme des forêts aléatoires est le mieux adapté pour la prédiction du rayonnement solaire dans les différentes villes en accord avec notre distribution de données. Par la suite nous avons fait une comparaison entre les résultats obtenus lors de la prédiction et celle mesurée ce qui a été en accord avec les résultats obtenus plus haut.

CONCLUSION GENERALE ET PERSPECTIVES

Les travaux réalisés au cours de ce mémoire ont porté sur la prédiction de la prédiction de l'irradiation solaire global à l'aide des algorithmes machines Learning. Pour ce faire nous avons subdivisé notre travail en trois chapitres.

Dans le premier chapitre, nous avons introduit des connaissances de base sur les irradiances solaires et les modèles de prévisions utilisés dans la littérature. Nous avons présenté les connaissances de base sur l'intelligence artificielle. Nous avons ressorti l'historique du Machine Learning et présenter quelques travaux menés par les chercheurs dans le but de prédire l'irradiation solaire globale à l'aide des algorithmes d'apprentissage automatique.

Le deuxième chapitre a été consacré sur les matériels et méthodes utilisés. Au cours de ce chapitre nous situons nos zones d'études, nous avons présenté les différents modèles de machine Learning utilisés tels que le LM, DT, SVM, RF, les étapes de la conception d'un bon modèle, les paramètres d'évaluation, les logiciels utilisés ainsi que le langage de programmation.

Dans le troisième chapitre, nous avons évalué les performances de nos algorithmes à l'aide des quatre métriques à savoir le MSE, MAE, RMSE et R^2 . En effectuant des comparaisons sur les différents algorithmes nous avons trouvé que l'algorithme des forêts aléatoires obtient le meilleur score dans toutes les villes ciblées de notre étude. Cela a montré que l'algorithme des forêts aléatoires est le mieux adapté pour la prédiction du rayonnement solaire dans les différentes villes en accord avec notre distribution de données. Par la suite nous avons obtenu les résultats de la prédiction pour ce modèle les remarques faites ont été en accord avec les résultats obtenus plus haut.

Au vu de l'ensemble des difficultés auxquelles nous nous sommes heurtées, nous envisageons pour des investigations futures de :

- Optimiser les différents modèles grâce aux techniques avancées d'optimisation afin d'obtenir de meilleurs scores.
- D'utiliser les méthodes d'apprentissage supervisées pour améliorer les résultats de la prédiction.

REFERENCES

- [1] <https://groupe-kedibuild.com/2022/05/05/situation-energetique-au-cameroun/> Consulté le 04 Avril 2024.
- [2] Gairaa, K., Khellaf, A., Messlem, Y., & Chellali, F. (2016). Estimation of the daily global solar radiation based on Box-Jenkins and ANN models: A combined approach. In *Renewable and Sustainable Energy Reviews* (Vol. 57, pp. 238–249). Elsevier Ltd. <https://doi.org/10.1016/j.rser.2015.12.111>
- [3] Bessemoulin, P., & Olivieri, J. (2000). Le rayonnement solaire et sa composante ultraviolette. *La Météorologie*, 8(31). <https://doi.org/10.4267/2042/36135>
- [4] Trouillet, F., Vallée, J.-M., 2008. Mesure de la constante solaire "http://acces.ens-lyon.fr/acces/thematiques/paleo/systemclim/effet-de-serre/data/constante_solaire_flo.pdf" (consulté le 18/06/2024). Vapnik, V., 2013. The nature of statistical learning theory. Springer science & business media.
- [5] Belorizky Elie, Pique Jean-Paul (20 Septembre 2018), Le rayonnement thermique du corps noir, Encyclopédie de l'Environnement. Consulté le 19 Juin 2024 [en ligne ISSN 2555-0950] url : <https://www.encyclopedia-environnement.org/physique/le-rayonnement-thermique-du-corps-noir/>.
- [6] Bounoua, Z., Chahidi, L.O., Mechaqrane, A., 2021. Estimation of daily global solar radiation, using empirical and machine-learning methods: a case study of five Moroccan
- [7] https://www.ibm.com/docs/it/SSEPGG_8.2.0/com.ibm.db2.udb.doc/opt/0sbp5004.gif consulté le 04 Avril 2024.
- [8] [Équateur céleste | Association Méridienne \(meridienne.org\)](https://meridienne.org/) consulté le 14 Juin 2024.
- [9] Modélisation et optimisation économique de micro-réseaux alimentés par des sources photovoltaïques et éoliennes - Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/figure/Parametrisation-de-la-position-du-soleil-en-coordonnees-horizontales-Mat10a_fig20_280494287 [accessed 19 June, 2024].
- [10] <https://pro.arcgis.com/fr/pro-app/latest/tool-reference/spatial-analyst/GUID-2E38BCF6-2101-4BAF-A607-6F1CCC8DE5AF-web.gif>. Consulté le 15 Avril 2024.

- [11] Modeste Kameni Nematchoua et al. Prediction of daily global solar radiation and air temperature using six machine learning algorithms; a case of 27 European countries.
- [12] [f5805071a4_101696_stokes-campbell-heliographe.jpg \(500×375\) \(futura-sciences.com\)](#) consulté le 15 Juin 2024.
- [13] [LP02 : Pyranomètre \(campbellsci.fr\)](#) consulté le 15 Juin 2024
- [14] [What is a pyrhelimeter? | Hukseflux](#) consulté le 15 Juin 2024
- [15] [www.data-transitionnumerique.com](#) consulté le 17 Juin 2024
- [16] Samuel, Arthur. "Some Studies in Machine Learning Using the Game of Checkers." IBM Journal of Research and Development, vol. 3, no. 3, 1959, pp. 210-229.
- [17] Rosenblatt, Frank. "The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain." Psychological Review, vol. 65, no. 6, 1958, pp. 386-408.
- [18] Minsky, Marvin, and Seymour Papert. Perceptrons: An Introduction to Computational Geometry. MIT Press, 1969.
- [19] LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep Learning." Nature, vol. 521, no. 7553, 2015, pp. 436-444.
- [20] Goodfellow, Ian, et al. "Generative Adversarial Nets." Advances in Neural Information Processing Systems, 2014, pp. 2672-2680.
- [21] Bishop, Christopher M. Pattern Recognition and Machine Learning. Springer, 2006.
- [22] [Calaméo - 28.B - Intelligence Artificielle \(calameo.com\)](#) consulté le 19 juin 2024
- [23] Wang L, Kisi O, Zounemat-Kermani M, Salazar G.A, Zhu Z, Gong W. Solar radiation prediction using different techniques: model evaluation and comparison Renew. Sustain. Energy Rev., Vol ,61 page 384-397, 2016.
- [24] Guermoui M, Melgani F, Gairaa K, Mekhalfi M.L. A comprehensive review of hybrid models for solar radiation forecasting J. Clean. Prod., vol 258 ,2020.
- [25] Ağbulut U, Gürel A. E, Biçen Y. Prediction of daily global solar radiation using different machine learning algorithms: evaluation and comparison. Renew. Sustain. Energy Rev., Vol 135 ,2021.

- [26] Obiora C.N, Hasan A.N, Ali A, Alajarmeh N. Forecasting hourly solar radiation using artificial intelligence techniques, IEEE Can. J. Electr. Comput. Eng.Vol 44, page 497–508,2021.
- [27] Samuel DJAOWE Electrification rurale décentralisée par les systèmes photovoltaïques individuels. Cas de la localité de Ntui –Essong. 2014, pp 14.
- [28] [Altitude of a point - SoDa \(soda-pro.com\)](https://soda-pro.com/)
- [29] [https:// worldbank-atlas.s3.us-east-1.amazonaws.com/download](https://worldbank-atlas.s3.us-east-1.amazonaws.com/download) Consulté le 04 Avril 2024.
- [30] Bakirci K. Correlations for estimation of daily global solar radiation with hours of bright sunshine in Turkey. Energy 2009;34(4):485–501.
- [31] Huang Liexing, Kang Junfeng, Wan Mengxue, Fang Lei, Zhang Chunyan, Zeng Zhaoliang, Solar Radiation Prediction Using Different Machine Learning Algorithms and Implications for Extreme Climate Events, Frontiers in Earth Science vol 9, 2021.
- [32] <https://machinelearningmastery.com/linear-regression-for-machine-learning/>. Available online in 2019. Accessed on 06 April 2024.
- [33] https://en.wikipedia.org/wiki/Linear_regression. Available online in April 2024. Accessed on April 06,2024.
- [34] Freedman David A. (2009). Statistical Models: Theory and Practice. Cambridge University Press. p. 26. A simple regression equation has on the right hand side an intercept and an explanatory variable with a slope coefficient.
- [35] Rencher, Alvin C.; Christensen, William F. (2012), "Chapter 10, Multivariate regression–section 10.1, Introduction", Methods of Multivariate Analysis, Wiley Series in Probability and Statistics, 709 (3rd ed.), John Wiley & Sons, p. 19, ISBN 9781118391679.
- [36] Hilary L. Seal (1967). "The historical development of the Gauss linear model". Biometrika. 54 (1/2): 1–24. doi:10.1093/biomet/54.1-2.1. JSTOR 2333849.
- [37] Kamiński, B.; Jakubczyk, M.; Szufel, P. (2017). "A framework for sensitivity analysis of decision trees". Central European Journal of Operations Research. 26 (1): 135–159. doi:10.1007/s10100-017-0479-6. PMC 5767274. PMID 29375266.

- [38] Quinlan, J. R. (1987). "Simplifying decision trees". *International Journal of Man-Machine Studies*. 27 (3): 221–234. CiteSeerX 10.1.1.18.4267. doi:10.1016/S0020-7373(87)80053-6.
- [39] https://www.researchgate.net/figure/An-automatically-generated-expression-tree-for-the-question-What-is-the-entropy-of-the_fig4_353052827
- [40] <https://asana.com/fr/resources/decision-tree-analysis>. Available online in February 2024, Accessed on April 06,2024.
- [41] Quinlan, R., 1983. Learning efficient classification procedures. In: Michalski, Carbonell, Mitchell (Eds.), *Machine Learning: an artificial intelligence approach*. Morgan Kaufmann, pp. 463–482. https://doi.org/10.1007/978-3-662-12405-5_15.
- [42] Quinlan, J. R. (1999). Simplifying decision trees. *Int. J. Hum. Comput. Stud.* 51,497–510. Doi : 10.1006/ijhc.1987.0321.
- [43] <https://konfuzio.com/xmlrpc.php> Guide du développeur Random Forest : 5 façons de l'implémenter en Python. Available online in march 2023, Accessed on April 07,2024.
- [44] Ho, Tin Kam, 1995. Random Decision Forests (PDF). In: *Proceedings of the 3rd International Conference on Document Analysis and Recognition*, Montreal, QC,14–16 August 1995, pp. 278–282.
- [45] Min, Yoon, Yeboon, Yun, Hirotaka, Nakayama, 2003. A role of total margin in support vector machines. In: *Proceedings of the International Joint Conference on Neural Networks*. IEEE, p. 2003.
- [46] <https://scikit-learn.org/stable/modules/svm.html> . Accessed on 03 May 2024.
- [47] Breiman, Leo. 2001. "Random Forests." *Machine Learning* 45 (1): 5–32.
- [48] By Larhmam-Ownwork, CCBYSA 4.0, [https:// commons.wikimedia.org/w/index.php?curid = 73710028](https://commons.wikimedia.org/w/index.php?curid=73710028)].
- [49] Bayrakçı, H.C., Demircan, C., Keçebas, A., 2018. The development of empirical models for estimating global solar radiation on horizontal surface: a case study. *Renew. Sust.Energ. Rev.* 81, 2771–2782.
- [50] Huang, L., Kang, J., Wan, M., Fang, L., Zhang, C., Zeng, Z., 2001. Solar radiation prediction using different machine Learning algorithms and implications for extreme climate events. *Front. Earth Sci.* 9, 596860 <https://doi.org/10.3389/feart.2021.596860>.

- [51] Fan, J., Wang, X., Wu, L., Zhang, F., Bai, H., Lu, X., Xiang, Y., 2018c. Nouveaux modèles combinés pour estimer le rayonnement solaire global quotidien basé sur la durée d'ensoleillement dans les régions humides: une étude de cas dans le sud de la Chine. *Convertisseurs d'énergie. Gérer.* 156, 618625.
- [52] Rehman, S., 1998. Rayonnement solaire sur l'Arabie saoudite et comparaisons avec des modèles empiriques. *Énergie* 23 (12), 1077-1082.
- [53] Moore, D. S., Notz, W. I, & Flinger, M. A. (2013). *The basic practice of statistics* (6th ed.). New York, NY: W. H. Freeman and Company. Page (138)