

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



Projet de Fin d'Etudes
Pour l'obtention du diplôme de Master en : Génie Mécanique
Domaine : Sciences et Technologies
Filière : Génie Mécanique
Spécialité : Énergétique
Thème

Une étude comparative sur la prédiction de la production de l'énergie photovoltaïque à l'aide des méthodes d'apprentissage automatique

Présenté Par :

- 1) M.MOHAMMED KRACHAI Saïd Yacine.
- 2) Melle. BENZEGUIR Aicha Chahrazed.

Devant le jury composé de :

Dr. BELGHOUL Hakima	MCB UAT(Ain Temouchent)	Président
Dr. ELAHMAR Kadi	MCB UAT(Ain Temouchent)	Examinateur
Dr. DORBANE Abdelhakim	MCA UAT.B.B.(Ain Temouchent)	Encadrant

Année Universitaire 2024/2025

Remerciements

Nous remercions avant tout Dieu le Tout-Puissant pour la force et la persévérance qui nous ont permis de mener à bien ce travail.

Nos sincères remerciements vont à nos familles et amis pour leur soutien constant tout au long de notre parcours.

*Nous exprimons notre profonde gratitude à notre encadreur, **Dr. DORBANE Abdelhakim**, pour sa disponibilité et ses précieux conseils.*

Nous remercions également les membres du jury pour l'attention portée à notre mémoire.

Nos remerciements s'adressent aux enseignants des départements de génie Mécanique pour leur accompagnement et leurs enseignements.

Enfin, merci à toutes les personnes ayant contribué de près ou de loin à la réalisation de ce travail

Dédicace

Je dédie ce mémoire à mes chers parents qui ont été toujours à mes côtés et m'ont toujours soutenu tout au long de ces longues années d'études. En signe de reconnaissance qu'ils trouvent ici, l'expression de ma profonde gratitude pour tout ce qu'ils ont consenti d'efforts et de moyens pour me voir réussir dans mes études.

A toute ma famille

A toutes mes amies

A Dr.DORBANE Abdelhakim

A tous les enseignants de Génie Mécanique

A tout le promo de Génie Mécanique option Énergétique année

2024 /2025

Et à tous ceux qui aiment le bon travail et ne reculent pas

devant les obstacles de la vie

MOHAMMED KRACHAI

Saïd Yacine

Dédicace

*Je remercie Dieu de m'avoir donné la force et le courage de
faire ce travail humble.*

*Je tiens à exprimer mes remerciements et ma gratitude
À l'homme qui m'a appris que la patience est le chemin du
succès, mon soutien et mon modèle, mon père*

*À qui sa satisfaction était mon but et mon ambition, à qui elle
m'a donné tant de choses sans attendre des remerciements, à qui
elle m'a donné la détermination, la persévérance et la volonté, à
qui elle a laissé une véritable empreinte sur ma vie. Chère
maman*

*À mes frères, sœurs et nièces dont le visage me réjouit et dont le
rire me fait plaisir.*

*À mon professeur estimé Dorbane Abdel Hakim, qui nous a
supervisés et nous a donné tous les conseils et l'orientation.*

*À tous ceux qui m'aiment sincèrement et sincèrement, à tous
ceux dont j'ai reçu conseil et soutien... Merci.*

**BENZEGUIRE AICHA
CHAHRAZED**

Résumé

Ce projet fin d'étude (PFE) présente une étude comparative de plusieurs algorithmes d'apprentissage automatique appliqués à la prédiction de la production d'énergie photovoltaïque. À l'aide du logiciel Orange3, différents modèles comme la régression linéaire, les arbres de décision, AdaBoost et PLS ont été évalués selon des métriques de performance adaptées à la classification. L'objectif principal est d'identifier les modèles les plus efficaces pour anticiper la production d'énergie solaire à partir de données environnementales. Cette recherche met en évidence l'importance de l'intelligence artificielle dans la gestion intelligente des énergies renouvelables, tout en ouvrant des perspectives vers l'intégration de données météorologiques en temps réel et l'utilisation de techniques avancées comme l'Apprentissage automatique.

Mots clés : apprentissage automatique, intelligence artificielle, prediction, energie photovoltaïque.

Abstract

This Master's thesis presents a comparative study of various machine learning algorithms used to predict photovoltaic energy production. Using the Orange3 software, models such as linear regression, decision trees, AdaBoost, and PLS were evaluated based on classification performance metrics. The main objective is to identify the most effective models for forecasting solar energy output using environmental data. The study highlights the growing role of artificial intelligence in the smart management of renewable energy and suggests future directions, including the use of real-time weather data and advanced techniques such as Machine Learning.

Keywords: Machine learning, artificial intelligence, prediction, photovoltaic energy.

ملخص

يقدم هذا البحث دراسة مقارنة بين عدة خوارزميات للتعلم الآلي تُستخدم في التنبؤ بإنتاج الطاقة الكهروضوئية. تم تنفيذ التجارب باستخدام برنامج Orange حيث تم تقييم نماذج مثل الانحدار الخطي أشجار القرار خوارزمية AdaBoost و PLS وذلك باستخدام مقاييس أداء مخصصة للتصنيف. الهدف الرئيسي هو تحديد النماذج الأكثر كفاءة في التنبؤ بإنتاج الطاقة الشمسية بناءً على البيانات البيئية. يسلط هذا العمل الضوء على أهمية الذكاء الاصطناعي في إدارة مصادر الطاقة المتجددة ويفتح آفاقاً مستقبلية لاستخدام بيانات الأرصاد الجوية في الوقت الحقيقي وتقنيات متقدمة مثل التعلم الآلي التلقائي.

الكلمات المفتاحية: التعلم الآلي، الذكاء الاصطناعي، التنبؤ، الطاقة الكهروضوئية

Liste des Figures

Figure 1: Les différents types d'énergies renouvelables[9]	16
Figure 2 :L'énergie solaire[11].....	17
Figure 3:L'énergie hydraulique[13]	18
Figure 4:L'énergie éolienne[15].....	19
Figure 5:L'énergie biomasse[16].....	19
Figure 6:L'énergie géothermique[18]	20
Figure 7: Le fonctionnement de la cellule photovoltaïque[21]	22
Figure 8 : la cellule monocristalline[23]	23
Figure 9 :la cellule poly cristallin[25].....	24
Figure 10 : la cellule amorphe[27].....	24
Figure 11: IA dans le domaine solaire[28].....	25
Figure 12: Le Machine Learning est une sous-discipline de l'IA [31]	26
Figure 13: principe de fonctionnement de l' Apprentissage automatique [33].....	27
Figure 14: types d'apprentissage automatique [34]	27
Figure 15: schéma explicative de L' Apprentissage automatique supervisé [36]	28
Figure 16: apprentissage non supervisé [39].....	29
Figure 17 : traitement 'NEWS' à partir D' Apprentissage automatique semi-supervisé [43]	30
Figure 18: apprentissage automatique et profond [46]	31
Figure 19 : Schéma de la structure GBDD[49].....	34
Figure 20 :Diagramme de structure des méthodes proposées.[50]	35
Figure 21 : Diagramme de flux des modèles d'apprentissage automatique pour la conversion de l'irradiance en puissance [51].....	36
Figure 22: Présentation schématique de la prévision de la puissance photovoltaïque (PV) basée sur l'apprentissage profond.[53].....	37
Figure 23 : Schéma de principe des prévisions d'énergie photovoltaïque (a) prévisions directes et (b) prévisions indirectes.[54].....	38

Figure 24: Architecture de (a) cellule LSTM (b) LSTM bidirectionnel[55].....	39
Figure 25 : Classification des prévisions de puissance photovoltaïque en fonction du temps [58] ...	40
Figure 26 : Organigramme du modèle de prévision basé sur la régression à vecteur de support (SVR) proposé [61]	41
Figure 27 : Algorithme AdaBoost [67]	46
Figure 28 : régression des moindres carrés partiels (PLS)[69]	47
Figure 29 : Régression linéaire [71].....	47
Figure 30 : Organigramme de l'arbre de décision[75].....	49
Figure 31 : Interface utilisateur et exemple de flux de travail dans Orange3	54
Figure 32: Photo écran du programme de la simulation de notre étude.....	57
Figure 33 : Représentation d'un diagramme en violon avec les statistiques descriptives [87].....	58
Figure 34 : Diagramme en violon de la variable puissance_PACE	59
Figure 35 : Diagramme en violon de la distribution des vitesses moyennes du vent (km/h).....	60
Figure 36 : Diagramme en violon des directions moyennes du vent	61
Figure 37 : "Distribution de la vitesse minimale du vent en km/h"	62
Figure 38 : "Distribution de la vitesse maximale du vent en km/h"	63
Figure 39 : Diagramme de distribution des températures moyennes (°C)	64
Figure 40 : Distribution de l'Humidité Relative Moyenne en Pourcentage (%)	65
Figure 41: Distribution de la pression atmosphérique moyenne au niveau de la mer en hPa.....	66
Figure 42 : diagramme de distribution moyenne de rayonnement solaire	66
Figure 43 : Organigramme du cadre de prédiction de notre étude.....	67
Figure 44 : Résultat du plot SHAP du modèle Linear regression	69

Abréviations

AA	L'apprentissage automatique
IA	L'intelligence artificielle
AAS	Apprentissage automatique supervisé
DOY	le jour de l'année
R ²	le coefficient de détermination
GBDD	Réseau dendritique à gradient boostant
GHI	l'irradiance globale
IQR	l'écart interquartile
Kw	kilowatt
MAE	l'erreur absolue moyenne
P	la puissance générée
PLS	Les moindres carrés partiels
PV	photovoltaïque
RMSE	Erreur quadratique moyenne
SHAP	Explications additives de Shapley
W	Watts
ICEEMDAN	Décomposition en mode empirique d'ensemble complet améliorée avec bruit adaptatif
BiLSTM-ICOA	Mémoire à long terme bidirectionnelle alimentée par l'algorithme d'optimisation Crayfish
DNI	l'irradiance directe

Listes des tableaux

Tableau 1: Paramètres d'entrées	44
Tableau 2 : Paramètres de sortie	45
Tableau 3 : échantillon de données météorologiques de notre étude [86].....	56
Tableau 4 : Évaluation comparative des performances de quatre modèles.....	68

Sommaire

Chapitre I: Généralités.....	15
I.1 Introduction.....	16
I.2 Les énergies renouvelables.....	16
I.2.1 Les différents types des énergies renouvelables	17
I.2.2 Les avantages et Les inconvénients des énergies renouvelables :.....	20
I.3 Énergie solaire photovoltaïque	21
I.3.1 Définition.....	21
I.3.2 L'effet photovoltaïque.....	21
I.3.3 Principe de fonctionnement d'une cellule solaire photovoltaïque	21
I.3.4 Différent types des cellules solaires.....	23
I.4 L'intelligence artificielle dans l'énergie solaire.....	24
I.4.1 L'apprentissage automatique	25
I.5 Apprentissage profond	30
I.5.1 Différence entre Apprentissage automatique et Apprentissage profond	30
I.6 Le rôle de l' Apprentissage automatique dans l'énergie solaire photovoltaïque.....	31
I.7 Conclusion	32
Chapitre II: Recherche bibliographique.....	33
II.1 Introduction	34
II.2 Les études précédentes.....	34
II.3 Conclusion.....	42
Chapitre III: Méthodologie de La Recherche	43

III.1 Introduction.....	44
III.2 Description des paramètres d'entrée et de sortie.....	44
III.3 Présentation des différents modèles de prédiction utilisant l'IA.....	45
III.3.1 AdaBoost.....	45
III.3.2 Les moindres carrés partiels PLS	46
III.3.3 Régression linéaire	47
III.3.4 Arbres de décision	48
III.4 Défis et Considérations.....	49
III.4.1 Défis.....	49
III.4.2 Considérations.....	50
III.5 Conclusion	51
Chapitre IV: Partie Simulation.....	52
IV.1 Introduction.....	53
IV.2 Présentation Orange 3	53
IV.3 Les mesures de performances	54
IV.3.1 Mesures de classification.....	54
IV.3.2 Mesures de Regression.....	55
IV.4 Résultats et Discussions	57
IV.4.1 Présentation du programme	57
IV.5 Cadre de prédiction	66
IV.6 Interprétation du tableau	67
IV.7 Résultats du tracage du diagramme SHAP.....	68
IV.8 Conclusion	70
Références	73

Introduction Générale

L'intérêt pour les énergies renouvelables a commencé après la crise pétrolière en 1973, où l'exploitation des énergies renouvelables est l'une des questions les plus importantes de notre temps actuel [1], L'énergie renouvelable (ou ENR) désigne un ensemble de moyens de production d'énergie à partir de sources ou de ressources théoriquement illimitées, disponibles sans limite de temps ou reconfigurables plus rapidement qu'elles ne sont consommées. On parle généralement d'énergies renouvelables par opposition aux énergies fossiles, dont les réserves sont finies et non renouvelables à l'échelle du temps humain charbon, pétrole, gaz naturel, etc. En revanche, les énergies renouvelables sont produites à partir de sources telles que le soleil ou le vent, qui sont théoriquement illimitées à l'échelle du temps humain. Les énergies renouvelables sont également parfois appelées « énergies vertes » ou « énergies propres » [2]. Parmi les énergies renouvelables L'énergie solaire photovoltaïque est une possibilité de développement efficace assez rapide pour qu'elles puissent être considérées comme une puissante à l'échelle du temps humaine. L'énergie solaire photovoltaïque provient de la conversion directe de l'énergie photonique du rayonnement solaire en énergie électrique, à l'aide de capteurs constitués de matériaux sensibles aux longueurs d'onde visibles (appelés cellules photovoltaïques) [3]. On estime que l'énergie émise par le Soleil serait environ 40 000 fois plus grande. Besoins énergétiques humains sous forme de combustibles fossiles [4]. La nature fluctuante du rayonnement solaire constitue un défi pour la gestion efficace et fiable des systèmes solaires photovoltaïques L'un des principaux défis des énergies renouvelables est leur dépendance aux conditions naturelles, comme l'ensoleillement ou la force du vent. Par exemple, une journée nuageuse peut réduire considérablement la production d'énergie solaire, tandis qu'une absence de vent rend les éoliennes inefficaces. Pour surmonter cette variabilité, l'IA propose des modèles prédictifs capables d'anticiper ces fluctuations avec une grande précision[5]. Ces dernières années, les technologies d'IA sont apparues comme des outils puissants pour répondre à l'incertitude dans les systèmes d'énergie renouvelable, notamment les systèmes solaires photovoltaïques.[6]. Des études récentes montrent que les algorithmes de machine learning peuvent analyser simultanément des données historiques et des informations météorologiques en temps réel[5].

L'IA s'attaque aux défis les plus complexes du secteur des énergies renouvelables en proposant des solutions innovantes et efficaces. L'un des principaux problèmes résolus est la prévision précise de la production d'énergie. En utilisant des modèles d'apprentissage automatique, l'IA peut anticiper les fluctuations météorologiques et ajuster la production en conséquence, assurant une plus grande stabilité énergétique. En outre, l'IA a résolu le problème de

l'intégration des énergies renouvelables dans les réseaux électriques traditionnels. En améliorant le stockage et la distribution de l'énergie, il permet une meilleure gestion des pics de production et une utilisation plus efficace des ressources disponibles. Enfin, AI s'est également attaqué aux problèmes de sécurité des infrastructures énergétiques en détectant et en neutralisant les menaces potentielles avant qu'elles ne causent des dommages. Ces solutions spécifiques montrent que l'IA est un catalyseur clé pour surmonter les obstacles et propulser les énergies renouvelables dans le futur[7].

Ce mémoire s'inscrit dans cette dynamique en proposant une étude comparative entre plusieurs modèles d'apprentissage automatique nous avons utilisé le logiciel Orange3, une plateforme open-source visuelle, qui permet la création de workflows de traitement de données et de modélisation prédictive de manière simple et efficace afin de déterminer lesquels offrent les meilleures performances pour la prédiction de la production d'énergie solaire. Pour ce faire, différents modèles, tels que linear regression , les arbres de décision, ou encore des méthodes d'ensemble comme AdaBoost, PLS seront analysés et comparés à l'aide de métriques de performance adaptées.

Notre mémoire est structuré en quatre chapitres.

- Le premier chapitre introduit les énergies renouvelables, en particulier l'énergie photovoltaïque, ainsi que l'intérêt de l'apprentissage automatique pour prévoir sa production.
- Le deuxième chapitre est une revue bibliographique des travaux antérieurs utilisant des algorithmes d'apprentissage automatique pour la prédiction photovoltaïque.
- Le troisième chapitre présente la méthodologie adoptée, avec les algorithmes utilisés (AdaBoost, régression linéaire, arbre de décision, PLS) et la description des données.
- Enfin, le quatrième chapitre expose les simulations effectuées via Orange3, l'analyse des résultats, et la comparaison des modèles (AdaBoost, régression linéaire, arbre de décision, PLS) selon des indicateurs de performance.

Chapitre I: Généralités

Ce chapitre couvre les concepts généraux des différents types d'énergie renouvelable, l'énergie solaire photovoltaïque, en plus des performances physiques de la cellule photovoltaïque avec ses types, ainsi que le rôle de l'apprentissage automatique dans la prédiction de la production d'énergie photovoltaïque et l'importance de la production d'énergie photovoltaïque.

I.1 Introduction

À l'ère de la transition énergétique, les énergies renouvelables sont devenues une solution clé pour relever les défis climatiques et réduire notre dépendance aux combustibles fossiles. Parmi ces énergies propres, le solaire, l'éolien et l'hydroélectricité sont des sources importantes qui offrent un grand potentiel mais présentent des défis importants liés à leur intermittence et variabilité. Dans ce contexte, l'intelligence artificielle (IA) émerge comme un outil révolutionnaire capable d'améliorer la production, la distribution et le stockage de ces énergies vertes.

I.2 Les énergies renouvelables

L'énergie renouvelable est une énergie dérivée de ressources naturelles qui se renouvelle à un rythme supérieur à celui de sa consommation. Par exemple, la lumière du soleil et le vent sont des ressources constamment renouvelables. Les sources d'énergie renouvelables sont abondantes et présentes partout autour de nous. En fait, les combustibles fossiles (carbone, pétrole et gaz) sont des ressources non renouvelables qui prennent des centaines de millions d'années à se former. La combustion de combustibles fossiles pour produire de l'énergie produit des émissions gazeuses qui se transforment en dioxyde de carbone. La production d'énergie renouvelable génère moins d'émissions que la combustion de combustibles fossiles. Face au changement climatique, il est nécessaire de passer des énergies fossiles, qui sont actuellement à l'origine de la majorité des émissions, aux énergies renouvelables. Les énergies renouvelables restent moins chères dans la plupart des régions du pays et créent trois fois plus d'emplois que les combustibles fossiles. Représentation des différents types d'énergies renouvelables dans la Figure 1.[8]



Figure 1: Les différents types d'énergies renouvelables[9]

I.2.1 Les différents types des énergies renouvelables

Énergie solaire

L'énergie solaire est une forme d'énergie renouvelable qui provient du soleil. Lorsque le soleil brille, nous pouvons capter son énergie et produire de l'électricité, de la chaleur ou les deux. L'un des problèmes liés à l'utilisation de l'énergie solaire est son coût relativement élevé par rapport à d'autres sources d'énergie renouvelables, comme l'énergie éolienne. Une autre limitation est que l'énergie solaire dépend des conditions météorologiques, car elle ne peut être captée que si le soleil brille. Cela signifie qu'une autre source d'énergie devra être utilisée les jours nuageux ou si la demande en électricité est élevée.

Il existe plusieurs façons d'utiliser l'énergie solaire.

Énergie photovoltaïque – L'électricité est produite à partir de panneaux solaires qui captent l'énergie du soleil, c'est-à-dire des panneaux noirs recouverts de petites cellules solaires. Lorsque ses rayons sont réfléchis sur ces panneaux, l'électricité est produite et dirigée vers une batterie électrique ou directement vers le système électrique.

Énergie thermique – L'énergie solaire peut également être utilisée pour produire de la chaleur. L'eau chaude peut être produite en plaçant de grands panneaux noirs sur le toit, puis en faisant circuler l'eau à l'intérieur, réduisant ainsi le besoin d'utiliser l'électricité pour chauffer l'eau de la maison. Comme le montre Figure 2.[10]



Figure 2 :L'énergie solaire[11]

Énergie hydraulique

L'énergie hydraulique est une source d'énergie renouvelable qui exploite la force de l'eau pour produire de l'électricité. Elle est l'une des plus anciennes et des plus répandues dans le monde, contribuant de manière significative à la production d'énergie décarbonnée. On

distingue deux principales filières d'énergie hydraulique : la grande (centrale de grande et moyenne chute) et la petite hydraulique (usine au fil de l'eau). Les énergies marines (usine marémotrices, hydroliennes, houlomotrices) utilisent aussi la force motrice de l'eau pour produire de l'énergie hydraulique [12]. Comme le montre la Figure 3.



Figure 3:L'énergie hydraulique[13]

Énergie éolienne

L'énergie éolienne est une source d'énergie qui dépend du vent. Le soleil chauffe inégalement la Terre, ce qui crée des zones de températures et de pression atmosphérique différentes tout autour du globe. De ces différences de pression naissent des mouvements d'air, appelés vent. Cette énergie permet de fabriquer de l'électricité dans des éoliennes, appelées aussi aérogénérateurs, grâce à la force du vent. L'énergie éolienne, sans conteste est l'énergie propre par excellence. Sans rejet ni déchet d'aucune sorte, elle permet avec un bon coefficient de performance d'assurer une bonne indépendance énergétique. Sur terre ou off-shore, les éoliennes démontrent d'excellentes capacités de production et sont des alternatives viables pour les territoires les plus reculés dont l'accès à l'électricité est difficile [14]. Comme le montre la Figure 4.



Figure 4: L'énergie éolienne[15]

Énergie biomasse

Le terme, Biomasse, désigne l'ensemble de la matière organique d'origine végétale ou animale (bois, déchets végétaux, algues) pouvant devenir source d'énergie par combustion. La biomasse constitue parmi les premières sources d'énergies renouvelables produites en Europe devant l'énergie hydraulique, éolienne et géothermique. Cette énergie permet de fabriquer de l'électricité grâce à la chaleur dégagée par la combustion de ces matières ou du biogaz issu de la fermentation de ces matières, dans des centrales biomasses. Elle rejette des gaz à effet de serre mais qui sont toutefois absorbés naturellement par les arbres. Cependant, elle pourra procurer l'énergie la moins chère et surtout la plus locale du marché [14]. Comme le montre Figure 5.



Figure 5: L'énergie biomasse[16]

Énergie géothermique

L'énergie géothermique, dérivée de la chaleur naturelle stockée sous la surface de la Terre, est exploitée pour produire de la chaleur et de l'électricité renouvelables. Il est extrait en forant des puits et en distribuant des fluides dans des réservoirs souterrains, avec des applications variant en fonction de la température, de la profondeur et des propriétés de la roche. Il existe trois types de centrales électriques : les centrales à vapeur sèche (utilisant directement la vapeur naturelle, comme en Italie depuis 1904), les centrales à cycle instantané (convertissant l'eau chaude en vapeur avant de la réinjecter, ce qui est le plus courant), et les centrales à cycle binaire (où la chaleur est transférée à un autre fluide pour produire de l'électricité). Cette énergie durable s'adapte à différentes conditions géologiques. [17]. Comme le montre Figure 6.



Figure 6:L'énergie géothermique[18]

I.2.2 Les avantages et Les inconvénients des énergies renouvelables :

Les avantages

- Les énergies renouvelables sont disponibles naturellement.
- Elles ne produisent pas de gaz à effet de serre.
- Les conséquences des accidents, dus à ces énergies, ne sont pas importantes et ils sont faciles à résoudre par rapport aux énergies conventionnelles (centrale nucléaire ou pétrolière).
- Le coût d'une énergie renouvelable est principalement fixé par l'investissement industriel.[14]

Les inconvénients

- L'intermittence des sources des énergies renouvelables (à l'exception de l'énergie géothermique).
- Les énergies renouvelables demandent de l'espace et ont un impact sur les usages des territoires.
- Une énergie renouvelable n'est pas nécessairement propre, et inversement (l'énergie issue de la combustion de la biomasse est propre à condition que la production ne soit pas excessive et permette à la flore de réabsorber tout le dioxyde de carbone dégagée).
- La construction des installations influence l'aspect visuel du paysage.

I.3 Énergie solaire photovoltaïque

I.3.1 Définition

L'énergie solaire photovoltaïque est l'une des énergies renouvelables et la plus utilisée. Elle est transmise à la terre à travers l'espace sous forme de photons et de rayonnement électromagnétique. Elle consiste à convertir directement le rayonnement solaire en électricité. On utilise des convertisseurs photovoltaïques ou cellules photovoltaïques qui représentent l'élément de base dans la conversion photovoltaïque [19].

I.3.2 L'effet photovoltaïque

L'effet photovoltaïque se manifeste quand un photon est absorbé dans un matériau composé de semi-conducteurs dopés p (positif) et n (négatif), dénommé comme jonction p-n (ou n-p). Sous l'effet de ce dopage, un champ électrique est présent dans le matériau de manière permanente (comme un aimant possède un champ magnétique permanent). Quand un photon incident (grain de lumière) interagit avec les électrons du matériau, il cède son énergie ($h\nu$) à l'électron qui se retrouve libéré de sa bande de valence et subit donc le champ électrique intrinsèque. Sous l'effet de ce champ, l'électron migre vers la face supérieure laissant place à un trou qui migre en direction inverse. Des électrodes placées sur les faces supérieure et inférieure permettent de récolter les électrons et de leur faire réaliser un travail électrique pour rejoindre le trou de la face antérieure, ainsi se crée un courant électrique.[19]

I.3.3 Principe de fonctionnement d'une cellule solaire photovoltaïque

Le fonctionnement de la Cellule photovoltaïque est fondé sur les propriétés de semi-conducteurs qui, percutés par les photons de la lumière, mettent en mouvement un flux d'électrons. Les photons sont des particules élémentaires qui transportent l'énergie solaire à 300 000 km/s et qu'Albert Einstein appelait dans les années 1920 les « grains de lumière ».

Lorsqu'ils frappent un élément semi-conducteur comme le silicium ils arrachent des électrons à ses atomes. Ces électrons se mettent en mouvement, de façon désordonnée, à la recherche d'autres « trous » où se repositionner.

Mais pour qu'il y ait un courant électrique, il faut que ces mouvements d'électrons aillent tous dans le même sens. Pour les y aider, on va associer deux types de silicium. La face exposée au soleil est « dopée » avec des atomes de phosphore qui comportent plus d'électrons que le silicium, l'autre face est dopée avec des atomes de bore qui comportent moins d'électrons. Cette double face devient une sorte de pile : le côté très chargé en électrons devient la borne négative (N), le côté avec moins d'électrons devient la borne positive (P). Entre les deux il se crée un champ électrique.

Quand les photons viennent exciter les électrons, ceux-ci vont migrer vers la zone N grâce au champ électrique, tandis que les « trous » vont vers la zone P. Les électrons sont récupérés par des contacts électriques déposés à la surface des deux zones avant d'aller dans le circuit extérieur sous forme d'énergie électrique. Un courant continu se crée. Une couche antireflet permet d'éviter que trop de photons se perdent en étant réfléchis par la surface [20]. qui sont représenté ci-dessous dans la Figure 7.

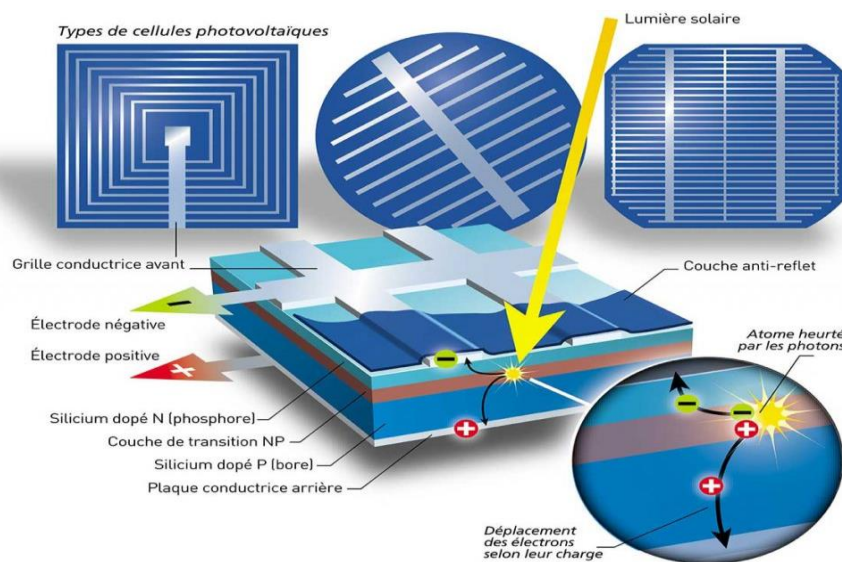


Figure 7: Le fonctionnement de la cellule photovoltaïque[21]

I.3.4 Différent types des cellules solaires

I.3.4.1 Cellules monocristallines

On reconnaît les cellules monocristallines grâce à leur couleur très foncée. Il s'agit d'un seul cristal de silicium. Au sein d'un même module solaire, tous les cristaux de silicium sont orientés dans le même sens. Le rendement de ce matériau est supérieur à celui du silicium polycristallin, mais il est également vendu plus cher, car sa fabrication est plus délicate [22]. Comme le montre Figure 8.



Figure 8 : la cellule monocristalline[23]

I.3.4.2 solaires en silicium poly cristallin

- Les cellules solaires en silicium poly cristallin sont fabriquées à partir de lingots de silicium formés par refroidissement rapide. Ces lingots présentent des cristaux de différentes tailles, ce qui peut entraîner des limites de rendement et d'efficacité légèrement inférieures par rapport au silicium monocristallin.
- Cependant, les cellules solaires poly cristallines sont moins coûteuses à produire que les monocristallines en raison de leur processus de fabrication moins complexe [24]. Comme le montre Figure 9.



Figure 9 :la cellule poly cristallin[25]

I.3.4.3 Cellules solaire en silicium amorphe

La cellule photovoltaïque au silicium amorphe est composée d'une couche fine de silicium, bien plus fine que les monocristallines ou les poly cristallines. On la trouve essentiellement pour alimenter les appareils de faible puissance, comme les montres solaires, les éclairages de jardin ou encore les calculatrices solaires. Le silicium amorphe est obtenu à partir de gaz de silicium. Ce gaz est vaporisé sur un support, en verre, en plastique souple ou en métal, grâce à un procédé de projection sous vide. Ces cellules photovoltaïques sont gris foncé [26]. Comme le montre Figure 10.



Figure 10 : la cellule amorphe[27]

I.4 L'intelligence artificielle dans l'énergie solaire

L'intelligence artificielle est un outil puissant pour l'industrie de l'énergie solaire. Et son influence ne cesse de croître. L'avenir promet un dynamisme où l'intelligence artificielle jouera

un rôle central, non seulement comme accessoire, mais comme moteur de l'innovation et de la responsabilité environnementale.

Cette vision suggère que l'IA deviendra un pilier essentiel, propulsant les progrès technologiques et favorisant une conscience écologique accrue. Imaginer un tel avenir inspire à anticiper des avancées majeures, où l'IA sera au cœur des efforts visant à façonner un monde plus durable et innovant

Dans un monde en quête de solutions énergétiques plus propres, l'intégration de l'intelligence artificielle par l'industrie de l'énergie solaire se profile comme une avancée majeure. Cette convergence ouvre la voie à un environnement énergétique plus intelligent, flexible et durable. Ainsi, la combinaison de l'IA et de l'énergie solaire promet des perspectives prometteuses pour répondre aux défis actuels de durabilité énergétique. La Figure 11 est une photo de l'IA dans le domaine solaire[28]



Figure 11: IA dans le domaine solaire[28]

I.4.1 L'apprentissage automatique

I.4.1.1 Définition

Les systèmes d'apprentissage automatique apprennent automatiquement des programmes à partir de données. C'est souvent une alternative très attrayante à leur construction manuelle, et au cours de la dernière décennie, l'utilisation de l'apprentissage automatique s'est rapidement répandue dans l'informatique et au-delà. L'apprentissage automatique est utilisé dans la recherche sur le Web, les filtres anti-spam, les systèmes de recommandation, le placement

d'annonces, la notation de crédit, la détection de fraude, le tradings d'actions, la conception de médicaments et de nombreuses autres applications. [29]

D'autre part L'apprentissage automatique est un sous-ensemble de l'intelligence artificielle et peut être résumé comme le domaine d'étude qui donne aux ordinateurs la capacité d'apprendre sans être explicitement programmés. Les récents progrès réalisés grâce à des quantités massives de données facilement disponibles et à une puissance de calcul inégalée qui devient moins chère chaque année ont rendu les perspectives de l'apprentissage automatique et de l'IA en général plus grandes que jamais auparavant. [30]. La Figure 12 montre que L'AA est un sous-domaine d'IA

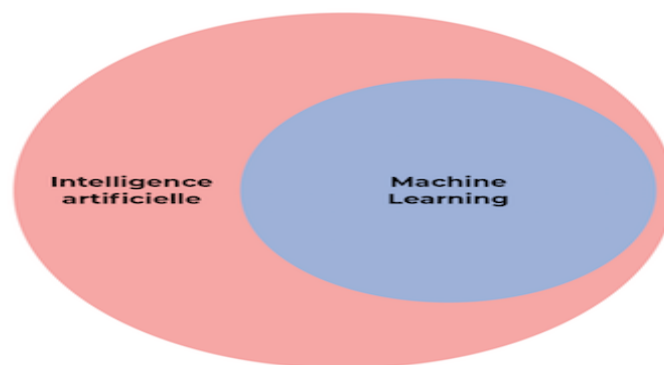


Figure 12: Le Apprentissage automatique est une sous-discipline de l'IA [31]

I.4.1.2 Principe de fonctionnement

L'apprentissage automatique est à la fois simple et complexe.

À la base, la méthode utilise simplement des algorithmes - essentiellement des listes de règles - ajustés et affinés en utilisant des ensembles de données antérieures pour faire des prédictions et des catégorisations lorsqu'ils sont confrontés à de nouvelles données. Par exemple, un algorithme d'apprentissage automatique peut être « entraîné » sur un ensemble de données composé de milliers d'images de fleurs étiquetées avec chacun de leurs différents types de fleurs afin qu'il puisse ensuite identifier correctement une fleur dans une nouvelle photographie sur la base des caractéristiques de différenciation qu'il a apprises à partir d'autres images.

Toutefois, pour que ces algorithmes fonctionnent efficacement, ils doivent généralement être affinés à de nombreuses reprises jusqu'à ce qu'ils accumulent une liste complète d'instructions leur permettant de fonctionner correctement. Les algorithmes qui ont été suffisamment entraînés finissent par devenir des « modèles d'apprentissage automatique », qui sont essentiellement des

algorithmes qui ont été entraînés à effectuer des tâches spécifiques telles que trier des images, prédire les prix de l'immobilier ou faire des coups d'échecs. Dans certains cas, les algorithmes sont superposés pour créer des réseaux complexes qui leur permettent d'effectuer des tâches de plus en plus complexes et nuancées, comme la génération de texte et l'alimentation de chatbots, grâce à une méthode connue sous le nom d'« apprentissage profond » (ou deep learning).

Par conséquent, bien que les principes généraux qui sous-tendent l'apprentissage automatique soient relativement simples, les modèles qui sont produits à la fin du processus peuvent être très élaborés et complexes. [32]. La Figure 13 est une illustration de principe de fonctionnement D'AA

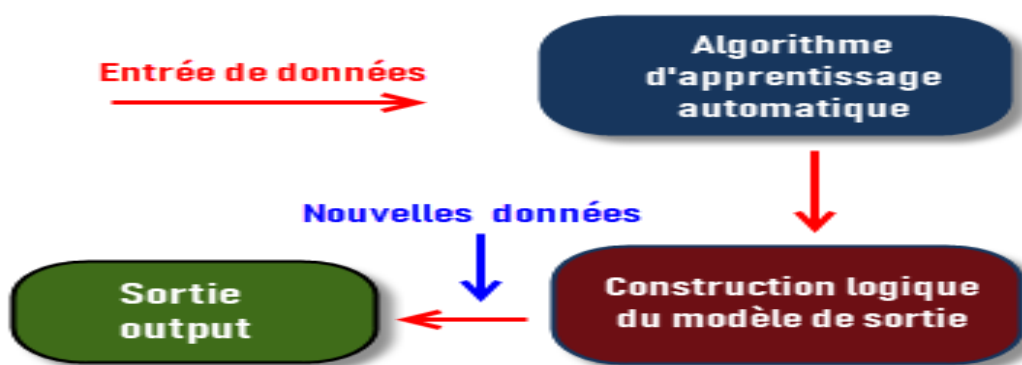


Figure 13: principe de fonctionnement de l' Apprentissage automatique [33]

I.4.1.3 Types d'apprentissage automatique

Il existe plusieurs types d'apprentissage automatique, mais ils peuvent être largement classés en quatre catégories principales. Comme l'explique la Figure 14

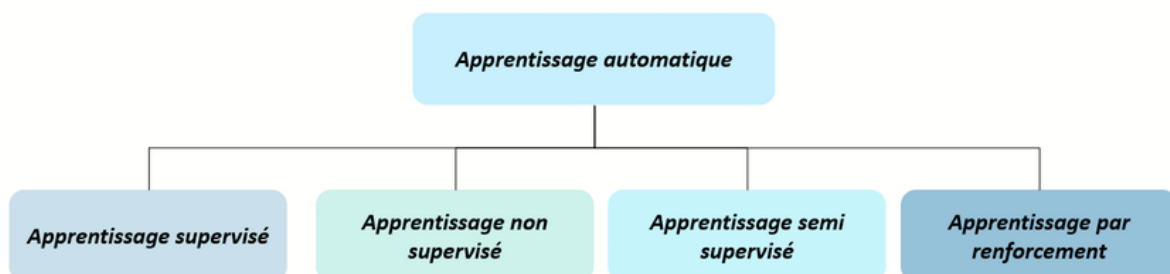


Figure 14: types d'apprentissage automatique [34]

a) Apprentissage supervisé

L'apprentissage supervisé est une technique d'apprentissage automatique qui utilise des ensembles de données étiquetés pour former des modèles d'algorithmes d'intelligence artificielle afin d'identifier les modèles et les relations sous-jacents entre les caractéristiques d'entrée et les sorties. L'objectif du processus d'apprentissage est de créer un modèle capable de prédire des sorties correctes sur de nouvelles données du monde réel. [17]. La Figure 15 représente un schéma d'AAS.

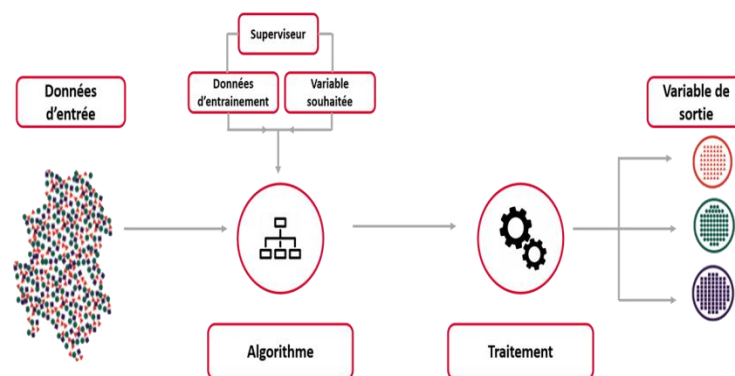


Figure 15: schéma explicatif de L' Apprentissage automatique supervisé [36]

Des algorithmes d'apprentissage supervisé sont utilisés lorsque la sortie est classée ou étiquetée. Ces algorithmes apprennent des données passées qui sont entrées, appelées données d'entraînement, exécutent leur analyse et utilisent cette analyse pour prédire les événements futurs de toute nouvelle donnée dans les classifications connues. La prédiction précise des données de test nécessite des données volumineuses pour avoir une compréhension suffisante des modèles. L'algorithme peut être entraîné davantage en comparant les résultats d'apprentissage aux résultats réels et en utilisant les erreurs pour la modification des algorithmes. [37]

b) Apprentissage non supervisé

Dans le cas de l'apprentissage non supervisé ou apprentissage des caractéristiques, l'algorithme se charge de déterminer seul la structure des données d'entrée. Aucune des données n'est annotée et la machine s'applique à explorer les données afin d'identifier des patterns. Cette

approche permet d'identifier des relations qui peuvent facilement échapper à l'attention des humains.

L'algorithme de reconnaissance faciale de Facebook est un exemple de Machine Learning non supervisé [38]. La Figure 16 représente L'AA non supervisé

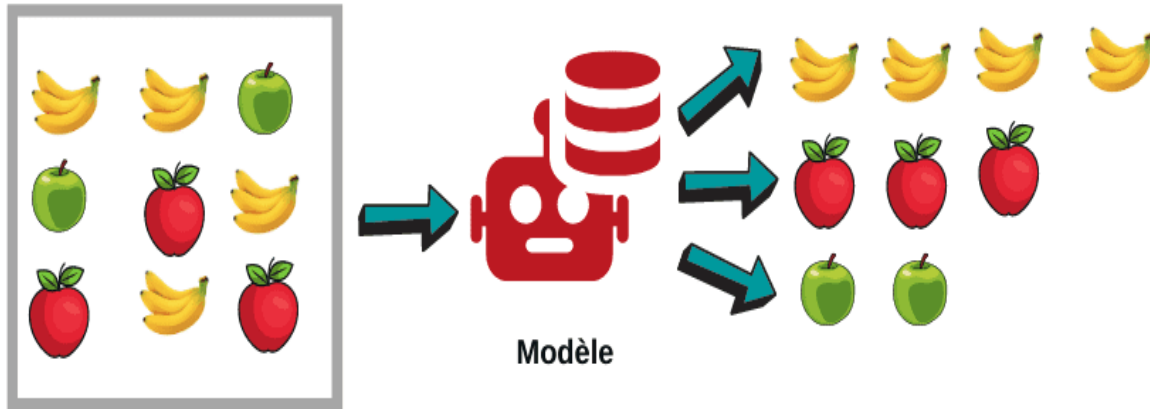


Figure 16: apprentissage non supervisé [39]

c) Apprentissage par renforcement

Dans ce type d'apprentissage, vous commencez avec un agent {algorithme} qui doit choisir parmi une liste d'actions. Puis, pour l'action choisie, il recevra, de l'environnement, un retour qui pourrait provenir d'un humain dans certaines situations ou de tout autre algorithme. Il s'agit d'une récompense pour un bon choix ou d'une pénalité pour une mauvaise action. L'agent (l'algorithme) apprend quelle stratégie (ou choix d'actions) maximise le cumul de récompenses.

Ce type d'apprentissage est souvent utilisé dans le cadre de la robotique, de la théorie des jeux et des véhicules autonomes. [40]

d) Apprentissage semi-supervisé

L'apprentissage semi-supervisé est une branche du machine learning qui combine l'apprentissage supervisé et non supervisé, Les problèmes pour lesquels vous avez une grande quantité de données d'entrée (X) et que seules certaines données sont étiquetées (Y) sont appelés problèmes d'apprentissage semi-supervisés. Par conséquent, ces problèmes se situent entre l'apprentissage supervisé et l'apprentissage non supervisé. pour entraîner des modèles d'intelligence artificielle (IA) sur des tâches de classification et de régression.[41], [42]. La Figure 17 est un exemple de traitement 'NEWS' à partir D' Apprentissage automatique semi-supervisé

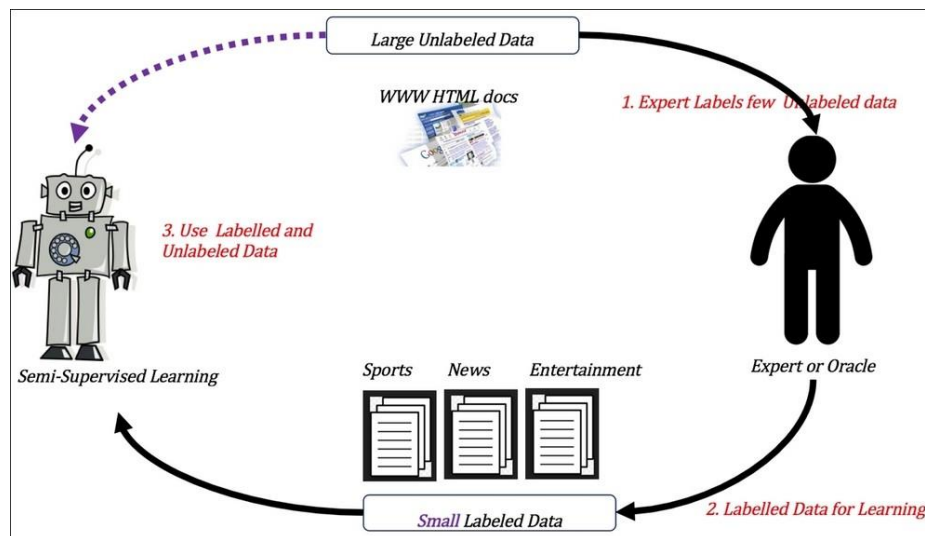


Figure 17 : traitement ‘NEWS’ à partir D’ Apprentissage automatique semi-supervisé [43]

I.5 Apprentissage profond

Le Apprentissage profond, ou apprentissage profond, est un sous-ensemble du Machine Learning, ou apprentissage automatique, basé sur des réseaux neuronaux artificiels. Le processus d’apprentissage est qualifié de profond parce que la structure des réseaux neuronaux artificiels se compose de plusieurs couches d’entrée, de sortie et masquées. Chaque couche contient des unités qui transforment les données d’entrée en informations que la couche suivante peut utiliser une tâche prédictive spécifique. Grâce à cette structure, une machine est capable d’apprendre au travers de son propre traitement de données. [44]

I.5.1 Déférence entre Apprentissage automatique et Apprentissage profond

Apprentissage automatique et Apprentissage profond sont devenus des termes extrêmement utilisés dans le cadre de nos activités, avec des applications toujours plus nombreuses. Lorsque l’on parle de Deep Learning, nous parlons d’algorithmes qui peuvent même imiter les actions du cerveau humain grâce à des réseaux de neurones d’où le terme d’Intelligence Artificielle. Il est souvent expliqué que la différence entre AA et Apprentissage profond réside dans le fait que les algorithmes de Machine Learning vont traiter des données quantitatives et structurées (des valeurs numériques), lorsque ceux de Deep Learning traiteront des données non-structurées, comme le son, le texte, l’image. [45]. La Figure 18 montre la différence entre L’ Apprentissage automatique et L’ Apprentissage profond.

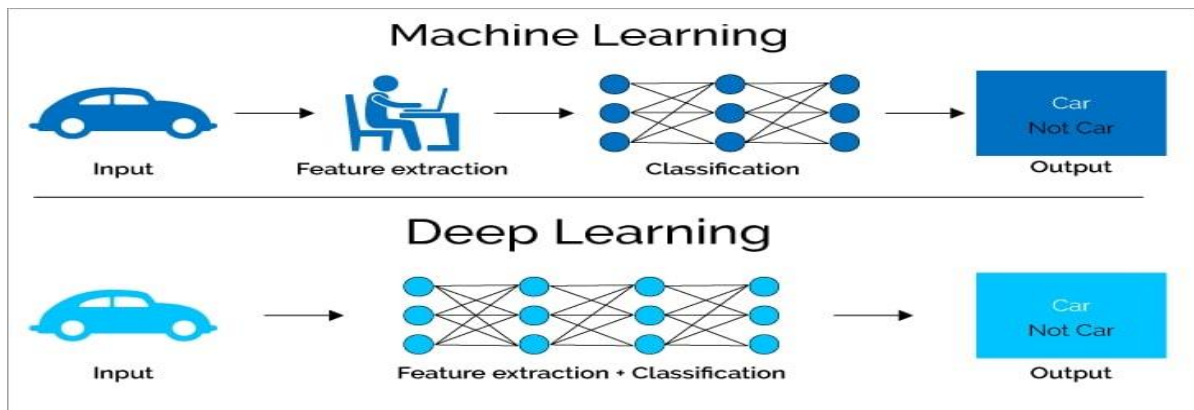


Figure 18: apprentissage automatique et profond [46]

I.6 Le rôle de L'apprentissage automatique dans l'énergie solaire photovoltaïque

L'apprentissage automatique (AA) joue un rôle de plus en plus important dans l'augmentation de l'efficacité des systèmes photovoltaïques. En utilisant les technologies d'AA, divers aspects de l'exploitation et de la maintenance des systèmes d'énergie solaire peuvent être améliorés, ce qui conduit à une augmentation de la production d'énergie et de l'efficacité des coûts.[47]

Analyse de données et reconnaissance de formes Le principal avantage de l'AA dans les systèmes photovoltaïques est la capacité d'analyser de grandes quantités de données. Les systèmes d'AA peuvent analyser les prévisions météorologiques, les données historiques de productivité et la surveillance en temps réel pour calculer le mode de fonctionnement optimal pour le système solaire. Cela permet d'ajuster l'inclinaison des panneaux solaires ou de modifier leur orientation pour obtenir un rendement maximal.

L'AA peut également reconnaître des modèles que les opérateurs humains pourraient manquer. Cela inclut la détection d'ombrage grâce à des algorithmes de traitement d'image qui peuvent ajuster l'orientation des panneaux solaires en conséquence pour obtenir des performances maximales. [47]

Gestion efficace de l'énergie Les systèmes de gestion de l'énergie basés sur l'AA optimisent non seulement le stockage, mais également la consommation d'énergie solaire photovoltaïque. L'AA analyse les prix de l'énergie et les habitudes de consommation, puis décide quand il est judicieux d'utiliser l'électricité pré-stockée ou de l'injecter directement dans le réseau. Cela contribue à l'efficacité des coûts et à la stabilité du réseau. [47]

Développement des réseaux solaires intelligents L'essor des réseaux solaires intelligents repose sur la capacité de l'IA à analyser d'immenses quantités de données issues des panneaux photovoltaïques. En intégrant ces informations à d'autres sources, telles que les prévisions météorologiques ou les habitudes de consommation énergétique, il devient possible d'affiner les stratégies de gestion et d'accroître significativement l'efficacité énergétique globale. Par exemple, lors d'une journée nuageuse, un réseau intelligent pourrait rediriger automatiquement l'énergie stockée vers les zones déficitaires, assurant ainsi une alimentation continue. [48]

I.7 Conclusion

Les énergies renouvelables, en particulier l'énergie photovoltaïque, jouent un rôle central dans la transition énergétique mondiale. La production d'énergie solaire est devenue un pilier central des systèmes énergétiques durables, grâce à son abondance et à son faible impact environnemental. L'intégration de l'apprentissage automatique dans la gestion des installations photovoltaïques est une étape majeure pour assurer une production stable et améliorée. Ces innovations technologiques améliorent non seulement la fiabilité des énergies renouvelables, mais accélèrent également leur adoption généralisée, contribuant ainsi à un avenir énergétique plus propre et plus durable.

Chapitre II: Recherche bibliographique

Dans ce chapitre, une revue des articles de recherche récents sera fournie, entrée sur des techniques innovantes qui ont été développées pour améliorer la prédiction à court, moyen et long terme de la production d'énergie solaire photovoltaïque. Des scientifiques comme Wang et al , ainsi que Li et al., dans lesquels chaque chercheur a adopté des stratégies différentes, par exemple des réseaux neuronaux avancés, les algorithmes d'optimisation ainsi que des modèles physiques, afin de réduire la marge d'erreur en prédiction et d'améliorer la fiabilité des systèmes photovoltaïques.

II.1 Introduction

Dans le secteur de la prédiction de la production d'énergie photovoltaïque, de nouvelles approches et méthodes doivent être employées pour augmenter la précision des modèles de prédiction afin de rendre possible l'intégration appropriée de l'énergie solaire dans le réseau électrique. Un certain nombre de stratégies ont été suggérées afin d'affiner ces prédictions, notamment des modèles mixtes qui sont une combinaison d'apprentissage automatique et de modèles physiques dynamiques.

II.2 Les études précédentes

D'abord, Wang et al. ont développé une étude dans le but d'améliorer la précision de la prévision à ultra-court terme de la production photovoltaïque en construisant un modèle hybride fondé sur un réseau dendritique avec gradient boosting (GBDD - Gradient Boosting Dendritic Network). Ils ont appliqué une stratégie de gradient boosting pour ajuster les erreurs de prévision en fonction des conditions météorologiques et ont incorporé un réseau dendritique comme modèle de base pour améliorer la robustesse et l'interprétabilité des prévisions. Les résultats obtenus indiquent que le modèle GBDD a diminué le MAE de 38,85 %, le MAPE de 32,63 % et le RMSE de 30,89 % par rapport aux méthodes traditionnelles. L'étude conclut que le modèle GBDD constitue une amélioration considérable des prévisions à ultra-court terme de l'énergie photovoltaïque et peut être utilisé comme outil d'optimisation pour d'autres modèles de prévision en exploitant le lien entre les erreurs de prévision et les conditions météorologiques. La Figure 19 exprime un schéma de GBDD [49]

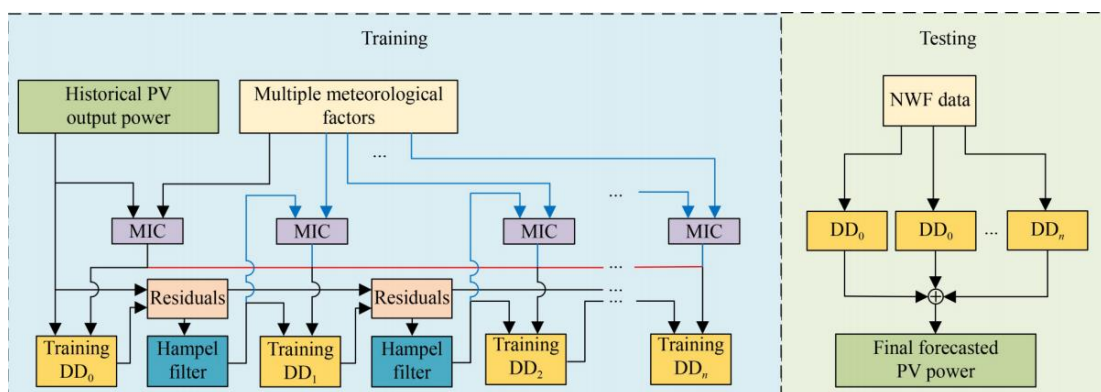


Figure 19 : Schéma de la structure GBDD[49]

De plus, Jiang et al. ont mené une étude prospective pour rendre plus précise la prédiction de la production photovoltaïque en utilisant des techniques basées sur des séries temporelles et des recherches similaires. À cette fin, ils ont utilisé le modèle BiLSTM-ICOA (Bidirectionnel Long Short-Term Memory optimisé par l'algorithme d'optimisation des écrevisses) pour la prédiction à court terme et FastDTW-HNSW (Fast Dynamic Time Warping with Hierarchical Navigable Small World) pour la prédiction à moyen et long terme. Ils ont également utilisé ICEEMDAN (Improved Complete Ensemble Empirical Mode Decomposition with Adaptive Noise) pour le prétraitement des données de séries temporelles. Les résultats obtenus indiquent que BiLSTM-ICOA avait une valeur RMSE de 0,02081 kW et une valeur R² de 99,898 %, ce qui est excellent pour la prévision à court terme. Cependant, FastDTW-HNSW avait une valeur RMSE de 0,01645 kW et une valeur R² de 99,923 %, offrant de meilleures performances dans les prévisions à moyen et long terme. L'étude conclut que l'approche hybride fusionnant ces méthodes a le potentiel d'améliorer considérablement la précision de la prédiction photovoltaïque, permettant ainsi une meilleure gestion du réseau énergétique [50]. La Figure 20 montre le Diagramme de structure des méthodes proposées.

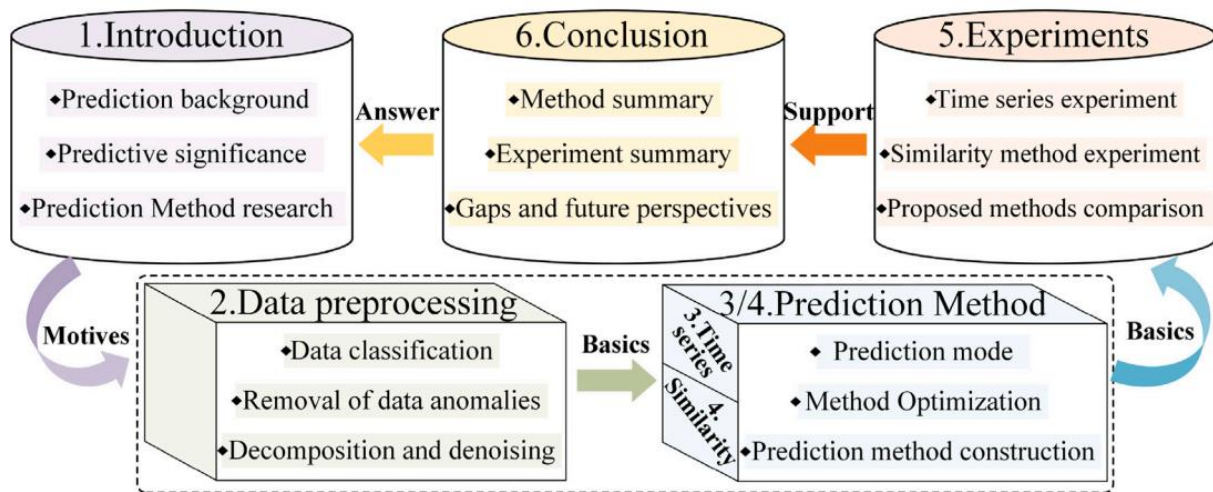


Figure 20 :Diagramme de structure des méthodes proposées.[50]

En addition, Li et al ont mené des recherches pour améliorer la précision de la prévision de puissance des systèmes photovoltaïques en utilisant l'aide d'un modèle physique dynamique (PVPro) de conversion de l'irradiance en puissance. Ils ont mis en œuvre une approche qui permet d'estimer les paramètres du modèle à partir des données de production actuelles, indépendamment des mesures de la courbe I-V. Pour comparer les performances, PVPro a été comparé à des modèles de persistance intelligents, à des modèles nominaux physiques et à un certain nombre d'approches d'apprentissage automatique pour la prévision de puissance à un

jour. D'après les résultats obtenus, il est évident que PVPro atteint une erreur absolue moyenne normalisée (nMAE) de 1,4 % pour quatre sites photovoltaïques, réduisant l'erreur de 17,6 % par rapport aux meilleures méthodes disponibles. L'étude révèle que PVPro n'est pas sensible aux variations saisonnières et météorologiques et peut même être utilisé avec un historique minimal de données, le rendant ainsi adapté aux nouveaux systèmes photovoltaïques. La Figure 21 illustre le processus de conversion de l'irradiance en puissance à l'aide de différents modèles de L'AA [51]

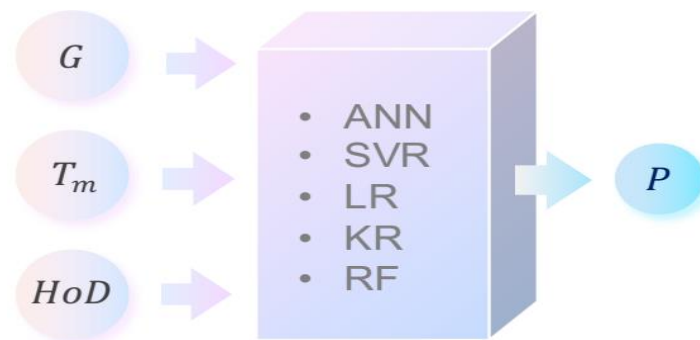


Figure 21 : Diagramme de flux des modèles d'apprentissage automatique pour la conversion de l'irradiance en puissance [51]

Ensuite, Antonanzas et al. [52] ont conduit une revue sur les méthodes de prévision de la production d'énergie photovoltaïque (PV). Les auteurs ont analysé et classé les méthodes de prévision en fonction de leur horizon temporel (court, moyen et long terme) et des techniques utilisées, telles que les approches physiques, statistiques et d'apprentissage automatique. Ils ont également examiné les facteurs influençant la précision des prévisions, comme les conditions météorologiques et la qualité des données. Les résultats montrent que les méthodes hybrides, combinant plusieurs approches, offrent souvent les meilleures performances. Les auteurs concluent que la prévision de la production PV reste un domaine en évolution, nécessitant des améliorations continues pour intégrer des données plus précises et des techniques avancées afin de répondre aux besoins croissants du secteur de l'énergie solaire. Puis, Dairi et al. ont conduit une étude sur la Prévision à court terme de la production d'énergie solaire photovoltaïque à l'aide d'une approche d'apprentissage profond pilotée par un encodeur automatique variationnel, leur étude vise à améliorer la précision de la prévision à court terme de la production d'énergie solaire photovoltaïque (PV) en utilisant un modèle basé sur les autoencodeurs variationnels (VAE). Les auteurs ont comparé les performances de huit modèles d'apprentissage profond, dont les réseaux de neurones récurrents (RNN), les réseaux de

mémoire à long terme (LSTM), les unités récurrentes à grille (GRU), les autoencodeurs empilés (SAE), les machines de Boltzmann restreintes (RBM) et les VAE, ainsi que deux méthodes d'apprentissage machine classiques (régression logistique et régression par vecteurs de support). Les données provenaient de deux centrales PV connectées au réseau, l'une aux États-Unis et l'autre en Algérie. Les résultats ont montré que le modèle VAE surpassait les autres méthodes en termes de précision de prévision, avec des valeurs plus élevées de R^2 et des erreurs plus faibles (RMSE, MAE). Les auteurs ont conclu que les modèles d'apprentissage profond, en particulier le VAE, sont efficaces pour capturer les caractéristiques linéaires et non linéaires des données de production d'énergie solaire, ce qui en fait des outils puissants pour la gestion des systèmes PV dans les réseaux électriques intelligents.[53] la Figure 22 est une vue schématique d'une prévision photovoltaïque (PV) basée sur l'apprentissage profond.

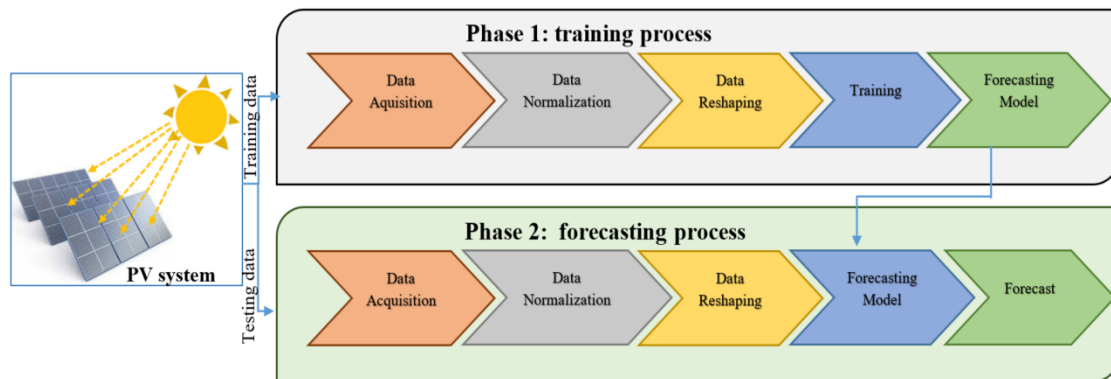


Figure 22: Présentation schématique de la prévision de la puissance photovoltaïque (PV) basée sur l'apprentissage profond.[53]

De plus, Gupta et al. ont réalisé une revue des techniques de prévision de la puissance photovoltaïque (PV) basées sur des modèles pilotés par les données, en mettant l'accent sur les méthodes d'apprentissage automatique (ML). L'étude vise à comparer les approches de prévision directe et indirecte de la puissance PV, en examinant les modèles de prévision du rayonnement solaire, les modèles d'estimation de l'irradiance sur le plan du tableau (POA) et les modèles de performance des systèmes PV. Les auteurs ont utilisé une analyse approfondie de la littérature pour classer et comparer les différentes méthodes, en mettant en avant les avantages des modèles hybrides et d'ensemble, ainsi que les techniques d'apprentissage profond (deep learning) pour les prévisions à court terme. Les résultats montrent que les modèles d'apprentissage profond et les modèles hybrides surpassent souvent les méthodes conventionnelles en termes de précision. Les auteurs concluent qu'il n'existe pas de modèle universel pour la prévision de la puissance PV, mais que les modèles hybrides et d'ensemble offrent généralement de meilleures performances, en particulier pour les prévisions à court

terme. Ils soulignent également l'importance de l'optimisation des hyperparamètres et de la préparation des données pour améliorer la précision des prévisions.[54] la Figure 23 représente un diagramme des prévisions PV (a) prévisions directes et (b) prévisions indirectes.

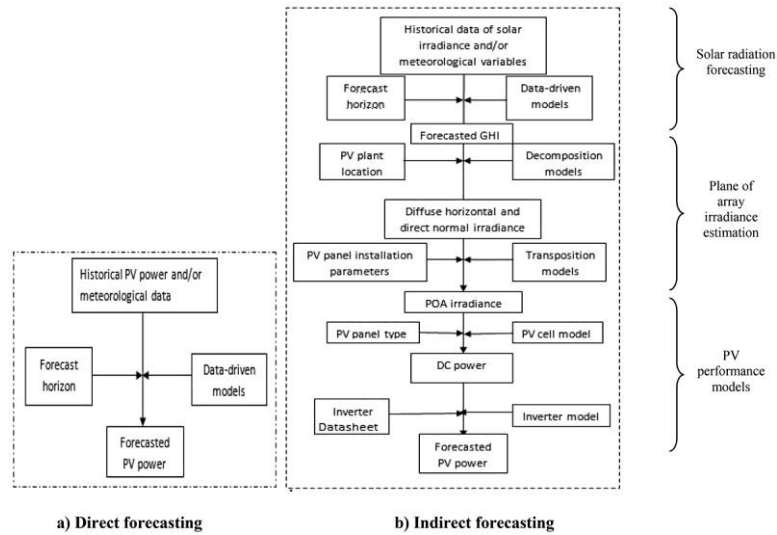


Figure 23 : Schéma de principe des prévisions d'énergie photovoltaïque (a) prévisions directes et (b) prévisions indirectes.[54]

Puis, Sharadga et al. ont étudié la prévision des séries temporelles de la production d'énergie solaire pour les grandes centrales photovoltaïques. Les auteurs ont utilisé des méthodes de régression linéaire, de réseaux de neurones artificiels (ANN) et de machines à vecteurs de support (SVM) pour prédire la génération d'énergie. Ils ont obtenu des résultats montrant que les modèles hybrides, combinant plusieurs approches, offraient une meilleure précision par rapport aux méthodes individuelles. Ils ont conclu que l'intégration de données météorologiques et de caractéristiques spécifiques aux centrales photovoltaïques améliore significativement la performance des modèles de prévision [55]. la Figure 24 représente l'architecture de (a) une cellule LSTM (b) une cellule LSTM bidirectionnelle.

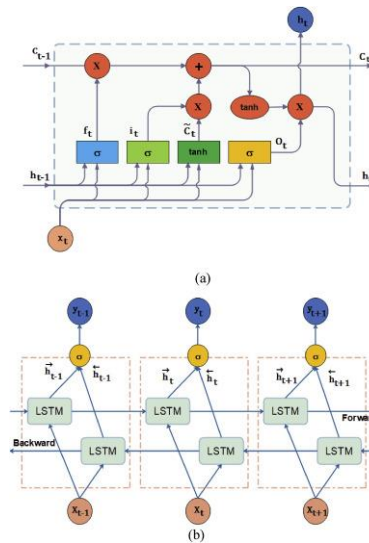


Figure 24: Architecture de (a) cellule LSTM (b) LSTM bidirectionnel[55]

En outre, Harrou et al. [56], [57] ont étudié le monitoring des systèmes photovoltaïques en utilisant des méthodes d'apprentissage basées sur le noyau améliorées. Les auteurs ont utilisé des techniques avancées de détection d'anomalies et de prédiction de performance pour surveiller l'efficacité des systèmes photovoltaïques. Ils ont obtenu des résultats significatifs en termes de précision et de rapidité dans la détection des défauts et l'évaluation des performances. Les auteurs ont conclu que ces méthodes améliorées offrent une solution robuste et efficace pour la surveillance en temps réel des installations photovoltaïques, contribuant ainsi à optimiser leur rendement et à réduire les coûts de maintenance.

Quant à Akhter et al. Ont réalisé une revue approfondie des méthodes de prévision de la production d'énergie photovoltaïque en se concentrant sur les techniques d'apprentissage automatique et les métaheuristiques. Le but de cette étude était d'analyser et de comparer les différentes approches utilisées pour améliorer la précision des prévisions de la production d'énergie solaire. Les auteurs ont examiné les méthodes d'apprentissage automatique, telles que les réseaux de neurones artificiels et les machines à vecteurs de support, ainsi que les techniques métaheuristiques comme les algorithmes génétiques et l'optimisation par essaims particulaires. Les résultats ont montré que ces méthodes offrent des performances supérieures par rapport aux approches traditionnelles, avec une capacité accrue à gérer les incertitudes et les variations météorologiques. Ils ont conclu que l'intégration de l'apprentissage automatique et des métaheuristiques est essentielle pour améliorer la précision des prévisions et optimiser la

gestion des systèmes photovoltaïques [58]. La Figure 25 représente la classification des projections PV en fonction du temps.

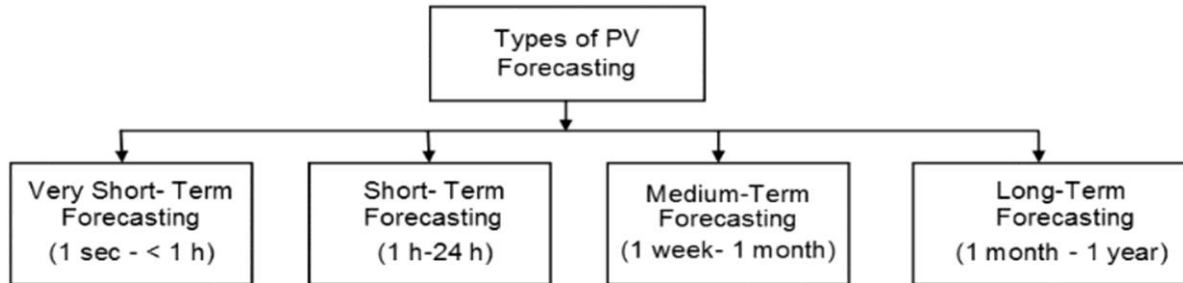


Figure 25 : Classification des prévisions de puissance photovoltaïque en fonction du temps [59]

Par ailleurs, Banik et al. ont proposé une approche innovante pour améliorer la précision des prévisions de la production d'énergie solaire photovoltaïque (PV) en combinant deux méthodes d'apprentissage automatique : la forêt aléatoire (RF) et CatBoost. Le but de cette étude était de développer un modèle robuste et complémentaire pour prédire la production d'énergie solaire en exploitant les avantages des deux algorithmes. Les auteurs ont utilisé un modèle d'ensemble RF-CatBoost, qui intègre les forces de la forêt aléatoire (pour la gestion des données complexes) et de CatBoost (pour la gestion des données catégorielles et les performances de prédiction). Les résultats ont montré que cette approche améliore significativement la précision des prévisions par rapport aux méthodes individuelles, avec une réduction des erreurs et une meilleure adaptabilité aux variations des conditions météorologiques. Ils ont conclu que l'approche RF-CatBoost est une solution efficace et robuste pour la prévision de la production d'énergie solaire, contribuant ainsi à une meilleure gestion des systèmes PV et à une intégration optimale des énergies renouvelables.[60]

En outre, Das et al. ont développé un modèle basé sur les machines à vecteurs de support pour la régression (SVR) afin de prévoir la production d'énergie photovoltaïque (PV) sous différentes conditions météorologiques. Le but de cette étude était de créer un modèle précis et fiable pour prédire la production d'énergie solaire en tenant compte des variations météorologiques. Les auteurs ont utilisé la méthode SVR, une technique d'apprentissage automatique, pour traiter les données non linéaires et capturer les relations complexes entre les variables météorologiques et la production d'énergie PV. Les résultats ont montré que le modèle SVR offre une grande précision dans les prévisions, même sous des conditions météorologiques changeantes, avec des erreurs de prévision réduites par rapport aux méthodes

traditionnelles. Ils ont conclu que le modèle SVR est une approche efficace pour la prévision de la production d'énergie PV, ce qui peut contribuer à une meilleure gestion des systèmes solaires et à une intégration optimale des énergies renouvelables dans les réseaux électriques [61]. La Figure 26 est un organigramme d'une régression vectorielle de soutien (SVR) proposée - modèle de prévision basé.

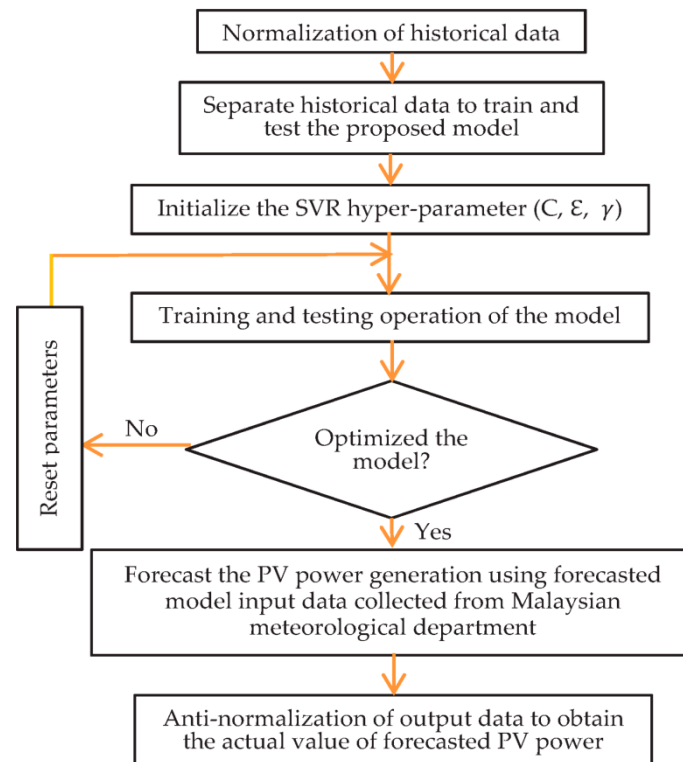


Figure 26 : Organigramme du modèle de prévision basé sur la régression à vecteur de support (SVR) proposé [62]

Ainsi que, Ahmad et al. ont exploré l'utilisation de méthodes d'ensemble basées sur les arbres de décision pour prédire la production d'énergie photovoltaïque (PV) et les ont comparées à la régression par machines à vecteurs de support (SVR). Le but de cette étude était d'évaluer l'efficacité des méthodes d'ensemble, telles que la forêt aléatoire et le gradient boosting, par rapport à la SVR pour la prévision de la production d'énergie solaire. Les auteurs ont utilisé des données historiques de production d'énergie PV et des variables météorologiques pour entraîner et comparer les modèles. Les résultats ont montré que les méthodes d'ensemble basées sur les arbres, en particulier le gradient boosting, offrent une précision supérieure et des erreurs de prévision plus faibles que la SVR. Ils ont conclu que les méthodes d'ensemble basées sur les arbres sont plus adaptées pour la prévision de la production d'énergie PV, en raison de

leur capacité à gérer des données complexes et non linéaires, et qu'elles constituent une approche prometteuse pour optimiser la gestion des systèmes solaires.[63]

II.3 Conclusion

L'amélioration des modèles de prévision de la production photovoltaïque est une question clé pour une meilleure intégration de l'énergie solaire dans les réseaux électriques. Pour atteindre cet objectif, l'utilisation d'approches novatrices est essentielle. Les solutions prometteuses comprennent des modèles hybrides, qui combinent des techniques d'apprentissage automatique avec des modèles physiques dynamiques, fournissant des prédictions plus précises et robustes. Ces développements méthodologiques ouvrent la voie à une gestion plus efficace des énergies renouvelables et contribuent à la transition vers un système énergétique durable.

Chapitre III: Méthodologie de La Recherche

Dans ce chapitre, nous décrivons chaque paramètre d'entrée et de sortie, en soulignant son rôle et son importance dans notre étude. Nous avons ensuite abordé la sélection des modèles de machine learning, en justifiant nos choix selon les critères de performance, de complexité et d'ajustement avec les données disponibles. Enfin, nous identifions les défis et les considérations rencontrées, tels que la qualité des données, le surapprentissage ou les limitations de calcul, tout en proposant des moyens pour les surmonter. Cette analyse approfondie fournit une base méthodologique solide pour nos travaux ultérieurs.

III.1 Introduction

Un modèle d'apprentissage automatique est un programme qui peut trouver des modèles ou prendre des décisions à partir de tout un ensemble de données. Il existe de nombreux modèles d'apprentissage automatique. Presque tous sont basés sur un algorithme d'apprentissage automatique. Les algorithmes les plus couramment utilisés pour la classification et la régression appartiennent à l'apprentissage machine supervisé, tandis que les algorithmes de clustering sont principalement utilisés dans des scénarios d'apprentissage machine non supervisés.[64]

III.2 Description des paramètres d'entrée et de sortie

Les paramètres d'entrée utilisés dans nos modèles de prédiction incluent des variables météorologiques, temporelles et liées au système photovoltaïque. Les données météorologiques englobent l'irradiance globale (GHI) et directe (DNI) mesurées en W/m^2 , la température ambiante et celle des modules en $^{\circ}C$, la vitesse du vent en m/s , ainsi que l'humidité relative en pourcentage. Les paramètres temporels comprennent l'heure et la date de la mesure, ainsi que le jour de l'année (DOY). Enfin, un paramètre propre au système PV est l'inclinaison des panneaux, exprimée en degrés. Le paramètre de sortie principal est la puissance générée (P), correspondant à l'énergie électrique effectivement produite par l'installation photovoltaïque, exprimée en watts (W) ou kilowatts (kW). Le Tableau 1 et le **Error! Reference source not found.** représente les Paramètres d'entrée et de sortie, respectivement.

Tableau 1: Paramètres d'entrées

Nom du paramètre	Description	Unité	Type
Irradiance globale (GHI)	Rayonnement solaire total sur une surface horizontale	W/m^2	Météorologique
Irradiance directe (DNI)	Composante directe du rayonnement solaire	W/m^2	Météorologique
Température ambiante	Température de l'air à proximité du champ PV	$^{\circ}C$	Météorologique
Température du module	Température des panneaux solaires (si mesurée)	$^{\circ}C$	Météorologique
Vitesse du vent	Vitesse du vent local, influençant le refroidissement des panneaux	m/s	Météorologique
Humidité relative	Quantité d'humidité présente dans l'air	%	Météorologique
Heure	Heure locale de la mesure	hh:mm	Temporelle

Date	Date de la mesure	jj-mm-aaaa	Temporelle
Jour de l'année (DOY)	Numéro du jour dans l'année (1 à 365/366)	Entier	Temporelle
Inclinaison des panneaux	Angle d'inclinaison des panneaux PV par rapport à l'horizontale	°	Système PV

Tableau 2 : Paramètres de sortie

Nom du paramètre	Description	Unité	Type
Puissance générée (P)	Puissance électrique réellement produite à un instant donné	W ou kW	Cible principale

III.3 Présentation des différents modèles de prédiction utilisant l'IA

III.3.1 AdaBoost

L'algorithme AdaBoost, introduit par Freund et Schapire en 1997, a révolutionné la modélisation d'ensemble. Depuis son apparition, AdaBoost est devenu une technique largement adoptée pour relever les défis de la classification binaire. Ce puissant algorithme améliore la précision des prédictions en transformant une multitude d'apprenants faibles en apprenants robustes et performants.[65]

AdaBoost est un modèle d'apprentissage automatique d'ensemble qui crée une séquence d'arbres de décision pondérés, généralement à partir d'arbres peu profonds (souvent de simples souches à un seul niveau). Chaque arbre est entraîné sur l'ensemble du jeu de données, mais avec des pondérations d'échantillon adaptatives qui accordent plus d'importance aux exemples précédemment mal classés. Pour les tâches de classification, AdaBoost combine les arbres grâce à un système de vote pondéré, où les arbres les plus performants ont plus d'influence sur la décision finale. La force du modèle réside dans son processus d'apprentissage adaptatif : si chaque arbre simple peut être un « apprenant faible » dont les performances sont légèrement supérieures à celles d'une estimation aléatoire, la combinaison pondérée des arbres crée un « apprenant fort » qui se concentre progressivement sur les erreurs et les corrige.[66]

AdaBoost est l'un des meilleurs algorithmes de boosting. Il repose sur une base théorique solide et a rencontré un franc succès dans des applications pratiques[67].la Figure 27 représente l'Algorithme AdaBoost [68]

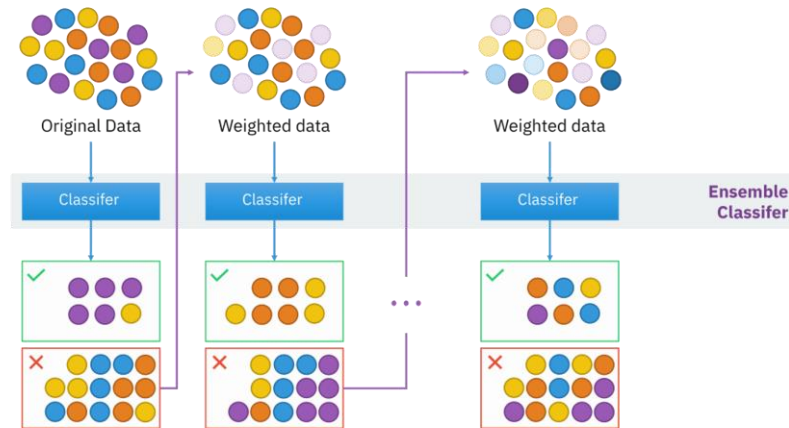


Figure 27 : Algorithme AdaBoost [68]

III.3.2 Les moindres carrés partiels PLS

L'analyse de régression est une méthode statistique utilisée pour modéliser la relation entre une variable dépendante, Y , et une ou plusieurs variables indépendantes, généralement représentées par X . Aujourd'hui, plusieurs algorithmes de régression peuvent être utilisés pour modéliser la relation entre la réponse Y et les variables prédictives. Parmi tous les algorithmes de régression possibles, un en particulier, les moindres carrés partiels (PLS), mérite l'attention car, dans cet algorithme, les paramètres du modèle restent les mêmes lorsque de nouveaux échantillons d'entraînement sont échantillonnés à partir de la population totale. Cela signifie que PLS crée des modèles qui correspondent plus étroitement aux observations données, vous offrant une vue strictement réaliste de l'ensemble de données et des informations contenues dans votre ensemble de données que le surajustement de la variable indépendante X pour vous fournir la meilleure prédiction possible (torture des données). Bien entendu, disposer de prévisions plus précises et plus exactes est l'objectif de tous les modèles d'apprentissage automatique et de régression. Cependant, la plupart du temps, les utilisateurs/scientifiques ont besoin de comprendre les données pour prendre des décisions rationnelles. Dans ce contexte, PLS possède des outils de diagnostic intéressants qui permettent au scientifique de comprendre ce qui se passe dans son modèle, de savoir pourquoi il peut échouer et de prendre une décision consciente à ce sujet.[69]

possible de prévoir une valeur de réponse à partir de la valeur du prédicteur avec une précision améliorée.[73]

La régression linéaire multiple examine les relations linéaires entre une réponse continue et deux prédicteurs ou plus. Si le nombre des prédicteurs est important, avant d'ajuster un modèle de régression avec tous les prédicteurs, vous devez utiliser la régression pas à pas ou celle sur les meilleurs sous-ensembles pour sélectionner le modèle en éliminant les prédicteurs non associés aux réponses.[73]

Dans la régression sur les moindres carrés, l'équation estimée correspond à l'équation qui minimise la somme des distances quadratiques entre les points de données de l'échantillon et les valeurs prévues par l'équation.[73]

La régression linéaire un concept relativement facile à comprendre et à appliquer. Le modèle de régression linéaire simple qui en résulte est une équation simple qui montre l'influence d'une variable sur une autre. Cela facilite l'explication et la fiabilité des résultats par rapport à des modèles plus complexes. La régression linéaire permet de prédire des valeurs futures à partir de données existantes. Elle sert de base à de nombreuses autres méthodes de science des données et d'apprentissage automatique. Même les algorithmes complexes s'appuient souvent sur la régression linéaire comme point de départ ou à des fins de comparaison. La régression linéaire peut être utilisée dans divers domaines, de la finance et de l'économie aux sciences et aux sciences sociales. C'est un outil polyvalent permettant de découvrir les relations entre les variables dans de nombreux scénarios réels.[74]

III.3.4 Arbres de décision

Les arbres sont une métaphore courante dans la vie quotidienne. Les arbres symbolisent souvent la croissance, car ils sont constitués d'une combinaison de racines, de tronc, de branches et de feuilles. Dans l'apprentissage automatique, un arbre de décision est un algorithme capable de créer des modèles de classification et de régression. Un arbre de décision est ainsi nommé parce qu'il commence à la racine, comme un arbre inversé, et se ramifie pour montrer différents résultats. Étant donné que l'apprentissage automatique est basé sur la résolution de problèmes, les arbres de décision nous aident à visualiser ces modèles et à affiner la manière dont ils sont formés. Un arbre de décision est un algorithme d'apprentissage supervisé utilisé dans la modélisation de classification et de régression. La régression est une technique utilisée dans la modélisation prédictive, donc ces arbres sont utilisés soit pour classer les données, soit pour prédire ce qui se passera ensuite. Les arbres de décision sont similaires

aux organigrammes. Ils commencent à un nœud racine avec une question spécifique à partir des données, menant à des branches avec des réponses possibles. Les branches mènent ensuite à des nœuds de décision (nœuds internes), qui posent d'autres questions, conduisant à d'autres conclusions. Cela continue jusqu'à ce que les données atteignent un nœud feuille (ou « feuille ») et se terminent. Dans l'apprentissage automatique, vous pouvez choisir entre quatre principaux algorithmes de formation : l'apprentissage supervisé, l'apprentissage non supervisé, l'apprentissage par renforcement et l'apprentissage semi-supervisé. Un arbre de décision nous aide à visualiser comment un algorithme d'apprentissage supervisé parvient à des résultats spécifiques[75]. la Figure 30 représente Organigramme de l'arbre de décision[76]

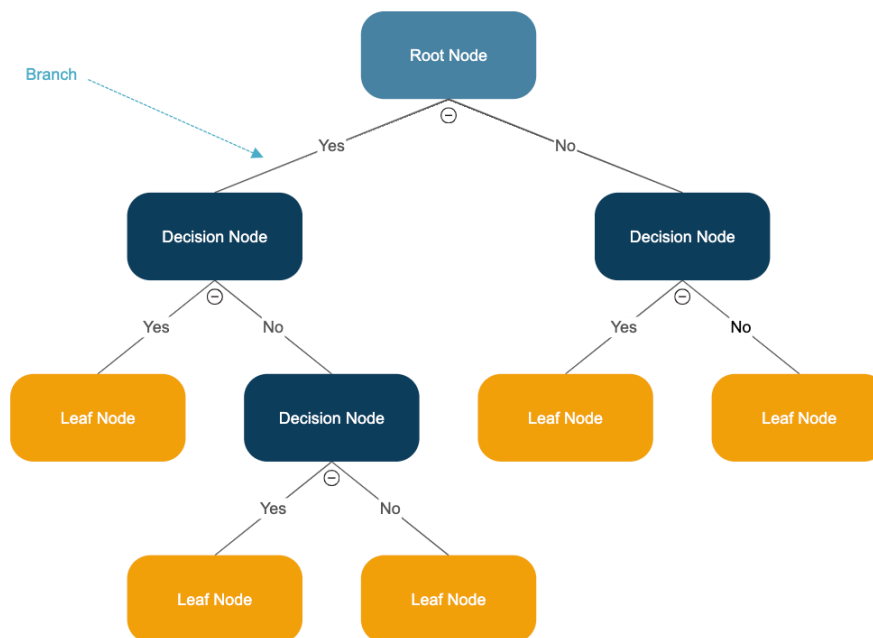


Figure 30 : Organigramme de l'arbre de décision[76]

III.4 Défis et Considérations

III.4.1 Défis

Les systèmes de machine Learning s'appuient fortement sur de grands volumes de données de haute qualité pour produire des résultats précis et fiables. Une mauvaise qualité des données, des incohérences ou des informations manquantes peuvent entraîner des prédictions erronées et une efficacité réduite. De plus, les organisations peuvent avoir des difficultés à collecter suffisamment de données pour entraîner des modèles robustes, en particulier dans les secteurs

où les informations propriétaires ou les préoccupations en matière de confidentialité limitent l'accessibilité.[77]

Le sous-ajustement se produit lorsqu'un modèle ne parvient pas à capturer la structure sous-jacente des données, ce qui entraîne de mauvaises performances sur les ensembles d'entraînement et de validation. Cela peut être dû à une architecture de modèle trop simpliste ou à un entraînement insuffisant. Le surajustement se produit lorsqu'un modèle apprend à fonctionner correctement sur les données d'entraînement, mais ne généralise pas correctement aux données non vues. Cela peut être dû à un nombre excessif de paramètres, à un entraînement trop long ou à un manque de données d'entraînement.[78]

De nombreux modèles d'apprentissage automatique fonctionnent comme des « boîtes noires », ce qui complique la compréhension du raisonnement qui sous-tend leurs décisions. L'interopérabilité et l'applicabilité sont cruciales dans les applications où des vies humaines ou des décisions financières importantes sont en jeu. Comprendre pourquoi un modèle a pris une décision spécifique peut être essentiel pour la transparence, la responsabilité et la confiance.[79]

Le coût de calcul est un autre défi dans cette intégration. Les algorithmes d'apprentissage automatique, en particulier ceux impliquant de grands ensembles de données ou des modèles complexes, nécessitent des ressources de calcul importantes. Vous devez tenir compte du compromis entre les performances du modèle et les dépenses de calcul. Parfois, des méthodes statistiques plus simples peuvent être plus pratiques, même si elles sont moins sophistiquées.[80]

III.4.2 Considérations

Choisir le modèle de ML approprié pour une tâche donnée est une décision cruciale qui impacte significativement la réussite du projet. Pour ce faire, plusieurs aspects doivent être pris en compte. La taille et la qualité des données disponibles jouent un rôle crucial dans le choix du modèle. Pour les petits ensembles de données, des modèles plus simples peuvent être plus adaptés pour éviter le surapprentissage, tandis que des ensembles de données plus volumineux peuvent prendre en charge des modèles plus complexes. La complexité du modèle doit être adaptée à la tâche spécifique et aux ressources de calcul disponibles. Les modèles simples sont généralement plus faciles à interpréter et moins coûteux en calcul, tandis que les modèles plus complexes peuvent atteindre une plus grande précision, mais nécessitent davantage de données et de ressources de calcul. L'interprétabilité du modèle est essentielle pour comprendre son

processus décisionnel et établir la confiance dans ses résultats. Les modèles interprétables sont plus transparents et plus faciles à déboguer, ce qui les rend préférables pour les applications où l'explicabilité est cruciale. Les ressources de calcul disponibles pour l'entraînement et le déploiement du modèle doivent être prises en compte. Si les ressources sont limitées, des modèles plus simples ou des stratégies de calcul distribué peuvent être nécessaires. L'expertise de l'équipe qui développe et déploie le modèle est également un facteur important. Si l'expertise du domaine est limitée, des modèles plus simples ou des solutions standard peuvent être plus appropriés. Le choix des indicateurs d'évaluation dépend de la tâche spécifique et du résultat souhaité. Les indicateurs courants incluent l'exactitude, la précision et le rappel. Les bénéfices potentiels d'une amélioration des performances doivent être mis en balance avec les coûts de formation, de déploiement et de maintenance du modèle. Une analyse coûts-avantages peut aider à déterminer si le modèle mérite d'être adopté.[81]

III.5 Conclusion

L'apprentissage automatique repose sur la capacité des algorithmes à détecter les modèles et les associations dans de grands ensembles de données. Par conséquent, ces algorithmes peuvent améliorer leurs performances au fil du temps à mesure qu'ils sont exposés à davantage de données. L'importance du machine learning dans la technologie est indéniable : il permet d'automatiser des tâches complexes, améliore la précision des prédictions et améliore les processus opérationnels.[82]

Chapitre IV: Partie Simulation

Dans ce chapitre, la partie simulation est consacrée à la définition des modèles et aux métriques d'évaluation utilisées pour analyser la performance de l'apprentissage automatique dans le contexte de la prédiction de l'énergie photovoltaïque. Nous évaluons plusieurs approches d'apprentissage supervisé, notamment AdaBoost, la régression par moindres carrés partiels (PLS), les arbres de décision, ainsi que la régression linéaire. L'objectif est de prédire avec précision la production d'énergie photovoltaïque à partir de données météorologiques et environnementales. Pour mesurer la performance de ces modèles, nous utilisons des métriques telles que l'erreur quadratique moyenne (RMSE), le coefficient de détermination R^2 , et l'erreur absolue moyenne (MAE). Une discussion critique est menée autour des résultats obtenus, mettant en évidence les points forts et les limitations de chaque méthode. Ce focus sur des techniques spécifiques d'apprentissage automatique permet d'approfondir leur impact dans l'optimisation de la gestion énergétique, et ouvre des perspectives pour des recherches futures visant à améliorer la précision des prévisions PV.

IV.1 Introduction

L'énergie photovoltaïque joue un rôle central dans la transition vers des systèmes énergétiques durables, grâce à son abondance et à son faible impact environnemental. Cependant, son caractère intermittent et dépendant des conditions météorologiques pose des défis majeurs pour son intégration optimale dans les réseaux électriques. Dans ce contexte, des programmes innovants comme Orange cherchent à développer des solutions intelligentes pour améliorer la prévision, la gestion et l'efficacité de cette énergie renouvelable.

IV.2 Présentation Orange 3

Orange 3 est un logiciel de visualisation et d'analyse de données qui permet aux utilisateurs d'explorer des ensembles de données, de créer des modèles prédictifs et de visualiser les résultats de manière intuitive. Développée à l'origine à l'Université de Ljubljana. Contrairement à d'autres outils, Orange se base sur une approche "no-code" ou "low-code", ce qui signifie qu'il permet aux utilisateurs d'effectuer des analyses complexes sans avoir besoin de connaissances approfondies en programmation. Grâce à une interface glisser-déposer, il est possible de construire des workflows analytiques en intégrant divers composants appelés "widgets"[83] :

- Fichier : importer un fichier CSV ou Excel.
- Data Table : voir les données.
- Select Columns : choisir les colonnes à utiliser.
- Logistic Regression, forêt aléatoire, arbre de décision : appliquer des algorithmes.
- Test & Score : évaluer la performance. La Figure 31 représente l'interface Orange3 et exemple d'utilisation

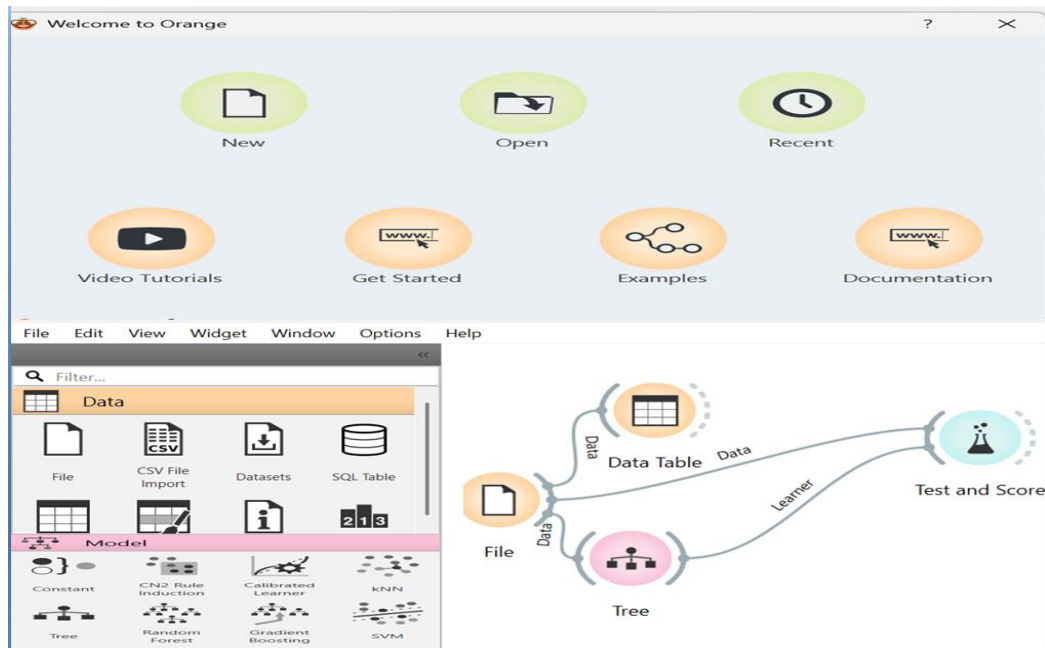


Figure 31 : Interface utilisateur et exemple de flux de travail dans Orange3

IV.3 Les mesures de performances

L'évaluation des performances est la mesure quantitative des performances d'un modèle entraîné selon des indicateurs d'évaluation spécifiques en apprentissage automatique. Ces informations permettent ensuite de déterminer si un modèle est prêt à passer à l'étape suivante de test, à être déployé à plus grande échelle ou s'il nécessite un entraînement supplémentaire ou un recyclage [84]. Les mesure de performances divisent en deux catégories :

IV.3.1 Mesures de classification

Dans un cadre de classification, la manière d'évaluer les performances d'un modèle est différente. Nous allons présenter trois des principales métriques de classification : l'exactitude, la précision et le rappel avec F1 score [85].

Exactitude (Accuracy) La précision mesure la fréquence à laquelle un modèle prédit correctement le résultat par rapport au nombre total de prédictions.

$$\text{Exactitude} = \frac{\text{vrais positifs} + \text{vrais négatifs}}{\text{total des prédictions}} \quad (1)$$

Précision est une mesure qui mesure la proportion de vrais positifs (prédictions correctes) par rapport au total des prédictions positives du modèle.

$$\text{Précision} = \frac{\text{vrais positifs}}{\text{vrais positifs} + \text{faux positifs}} \quad (2)$$

Rappel également connu sous le nom de sensibilité ou de taux de vrais positifs, est une mesure qui mesure la proportion de vrais positifs correctement identifiés par un modèle.

$$Rappel = \frac{\text{vrais positifs}}{\text{vrais positifs} + \text{faux négatifs}} \quad (3)$$

Score F1 est la moyenne harmonique des scores de précision et de rappel. La formule ci-dessous illustre ce concept.

$$F1 = \frac{2(\text{Précision} \times \text{Rappel})}{(\text{Précision} + \text{Rappel})} \quad (4)$$

IV.3.2 Mesures de Regression

Les tâches de régression impliquent la prédiction de valeurs numériques continues. Les indicateurs de performance des modèles de régression quantifient la capacité du modèle à prédire avec précision la variable cible [86]. Voici quelques indicateurs de régression courants

- **Erreur absolue moyenne (MAE)** L'EMA mesure la différence absolue moyenne entre les valeurs prédites et les valeurs réelles.

$$MAE = \frac{1}{n} \sum_{i=1}^n |O_i - P_i| \quad (5)$$

où :

- (n) est le nombre total d'observations,
 - (O_i) est la valeur réelle pour l'observation
 - (P_i) est la valeur prédite pour l'observation
- **Erreur quadratique moyenne (MSE)** L'MSE mesure la différence quadratique moyenne entre les valeurs prévues et les valeurs réelles.

$$MAE = \frac{1}{n} \sum_{i=1}^n (O_i - P_i)^2 \quad (6)$$

- **Erreur quadratique moyenne (RMSE)** La RMSE est la racine carrée de l'erreur quadratique moyenne (MSE), fournissant une mesure de l'erreur de prédiction dans les mêmes unités que la variable cible.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (O_i - P_i)^2} \quad (7)$$

- **Coefficient de détermination (R²)** Le R² quantifie la proportion de la variance de la variable cible expliquée par le modèle. Sa valeur varie de 0 à 1, les valeurs élevées indiquant un meilleur ajustement du modèle.

$$R^2 = 1 - \frac{\sum_{i=1}^n (P_i - O_i)^2}{\sum_{i=1}^n (P_i - O'_i)^2} \quad (8)$$

$$O'_i = \text{moyenne des valeurs observées} = \frac{1}{n} \sum_{i=1}^n O_i$$

IV.4 Données Météorologiques

Les données Météorologiques sont des informations numériques ou descriptives représentant les éléments du temps tels que : la température, l'humidité, le rayonnement solaire ... Elles sont utilisées pour la prévision météorologiques ou pour des études climatiques. Tableau 3 illustre un échantillon des données météorologiques de notre étude.

Orientation du vent (°)	v.vent moy (km/h)	v.vent Min (km/h)	v.vent Max (km/h)	Température Moy (°)	Humidité Moy (hpa)	Irradiance solaire Moy (w/m ²)	Pression au niveau de la mer (hpa)	Puissance (watts)
260	7,56	6,12	9,72	11,7	41,3	0,027008	1020,42	0
197	2,88	2,16	3,24	10,2	85,4	66,157264	1016,42	88
323	9,72	7,56	11,88	27,2	12,6	68,27	1011,42	11800
277	12,24	8,28	17,64	28,1	12,8	337,29248	1011,22	55036

Tableau 3 : échantillon de données météorologiques de notre étude [87]

IV.5 Résultats et Discussions

IV.5.1 Présentation du programme

La Figure 32 présente une photo écran du fichier exécutable de notre étude, depuis cette figure, on voit bien la structure de notre programme, contenant les quatre modèles de classification, ainsi que la partie violon, , explain model [on parlera de ça dans une partie qui suit], et finalement Test and Score, cette partie de notre programme nous donnera les résultats de l'évaluation de chaque modèle.

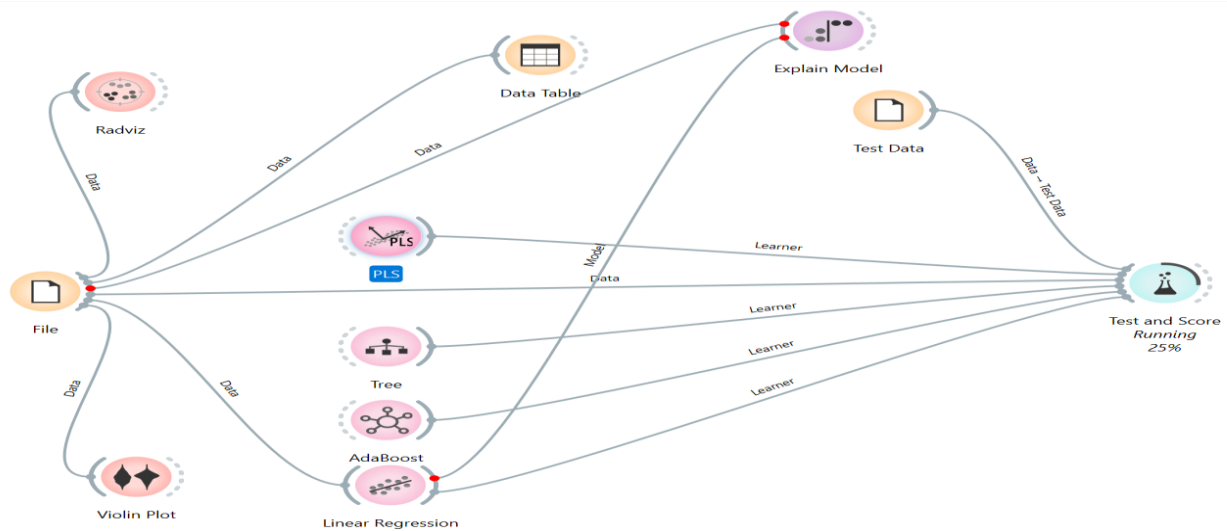


Figure 32: Photo écran du programme de la simulation de notre étude

IV.5.1.1 Diagramme en Violon

Un graphique en violon permet de représenter graphiquement la distribution des données métriques . Il permet de visualiser la distribution d'un ensemble de données, en affichant non seulement la moyenne et la variance , mais aussi l'ensemble des données. Similaire au diagramme en boîte à moustaches , le graphique en violon affiche également la distribution de probabilité de la variable correspondante. Il s'agit donc d'une combinaison d'un diagramme en boîte à moustaches et d'un diagramme de densité du noyau.

Le diagramme en violon présente une forme similaire à celle d'un diagramme en boîte traditionnel : un rectangle représente l'écart interquartile (IQR) et des lignes représentent la médiane et les quartiles supérieur et inférieur [88]

La Figure 33 est une représentation d'un diagramme en violon avec les statistiques descriptives.

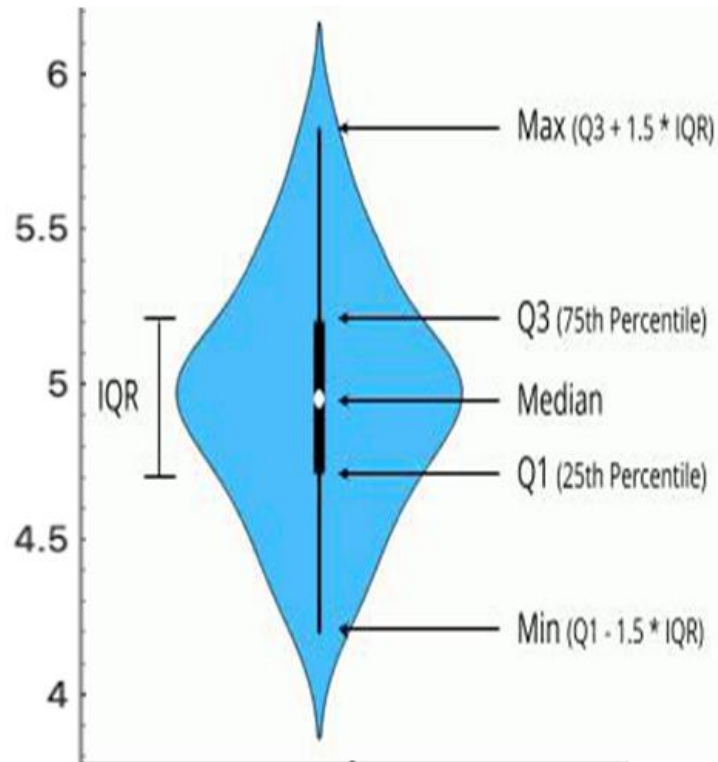


Figure 33 : Représentation d'un diagramme en violon avec les statistiques descriptives [89]

IV.5.1.2 Analyse de la base de données

a) Puissance

La Figure 34 montre la distribution de la variable Puissance_PACE. En fait, il s'agit d'un box-plot avec estimation de densité faite comme un violon. Ainsi, il est possible d'observer non seulement les indicateurs descriptifs, mais aussi la distribution elle-même. Plusieurs densités se constituent autour de zéro, alors que les pics de densité excellent environné par 40 000–50 000 et par 90 000-100 000. Pareilles tendances suggèrent un modèle de l'ensemble. Le box-plot central comprend la médiane; l'intervalle interquartile (boîte noire); la dispersion des données (lignes fines). Le fait qu'il n'y ait pas de groupes sur l'axe vertical signale qu'on regarde le plein ensemble des données. Ce graphique révèle une distribution asymétrique, possiblement influencée par plusieurs sous-populations.

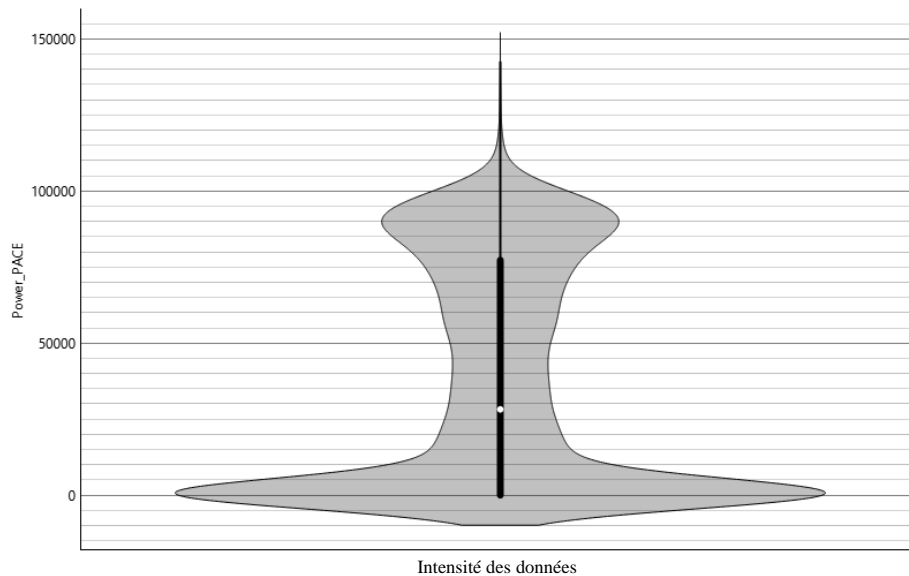


Figure 34 : Diagramme en violon de la variable Puissance

b) Vitesse moyenne du vent

La Figure 35 présentée est un diagramme en violon illustrant la distribution de la variable Vitesse moyenne du vent (km/h), correspondant à la vitesse moyenne du vent en kilomètres par heure. L'axe vertical s'étend de 0 à 40 km/h, couvrant ainsi l'ensemble des valeurs observées pour cette variable. La forme grise du violon reflète la densité de distribution des données : plus la largeur est importante à un certain niveau, plus les valeurs sont fréquentes autour de cette vitesse. On observe que la distribution est asymétrique, avec une queue allongée vers le haut. Cela indique que, bien que la plupart des vitesses moyennes soient relativement faibles, entre 2 et 10 km/h, quelques valeurs élevées peuvent atteindre jusqu'à 40 km/h, bien qu'elles soient rares. Le cœur de la distribution, plus épais, se situe entre 5 et 10 km/h, ce qui montre une forte concentration des données dans cette plage. Au centre du violon, un boxplot est superposé. Le point blanc représente la médiane, soit la valeur centrale de la distribution. Le rectangle noir correspond à l'intervalle interquartile (IQR), englobant les 50 % des données les plus centrales, tandis que les lignes fines s'étendent jusqu'aux limites des valeurs considérées comme normales (hors valeurs aberrantes). Dans l'ensemble, ce graphique permet de visualiser efficacement la répartition des vitesses moyennes du vent, mettant en évidence une concentration autour de faibles vitesses, une variabilité modérée et la présence occasionnelle de vitesses plus élevées.

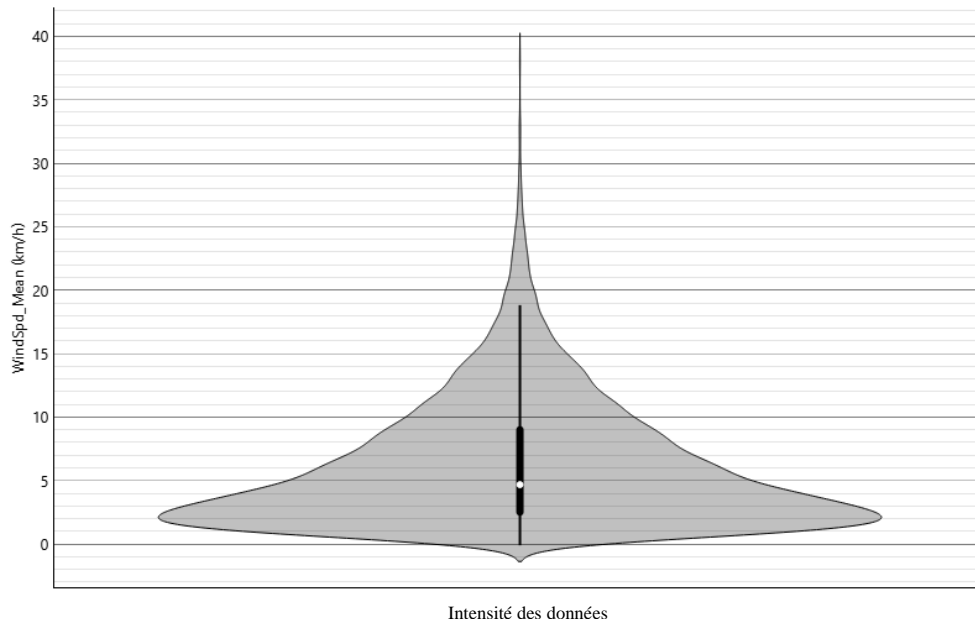


Figure 35 : Diagramme en violon de la distribution des vitesses moyennes du vent (km/h)

c) Moyenne de la direction du vent (deg)

La Figure 36 illustre la distribution des directions moyennes du vent à l'aide d'un diagramme en violon, où les mesures sont exprimées en degrés. L'axe vertical s'étend de 0° à 360° , couvrant ainsi l'ensemble des orientations possibles du vent sur un cercle. La forme du violon met en évidence la densité des données : plus la forme est large à un certain niveau, plus les observations sont concentrées autour de cette direction. Le graphique révèle plusieurs zones d'élargissement, notamment autour de 45° , 150° , 260° et 350° , indiquant une distribution multimodale, ce qui signifie que le vent souffle fréquemment dans ces directions spécifiques. Au centre du graphique, un point blanc représente vraisemblablement la médiane, tandis qu'une boîte noire verticale illustre l'intervalle interquartile (IQR), mettant en évidence la concentration des données autour des valeurs centrales. Des points noirs dispersés sur la hauteur du graphique correspondent aux valeurs individuelles mesurées. Ce type de visualisation permet de saisir en un coup d'œil la tendance centrale, la variabilité et les zones de concentration dans la distribution des directions moyennes du vent, offrant ainsi une compréhension approfondie de son comportement directionnel.

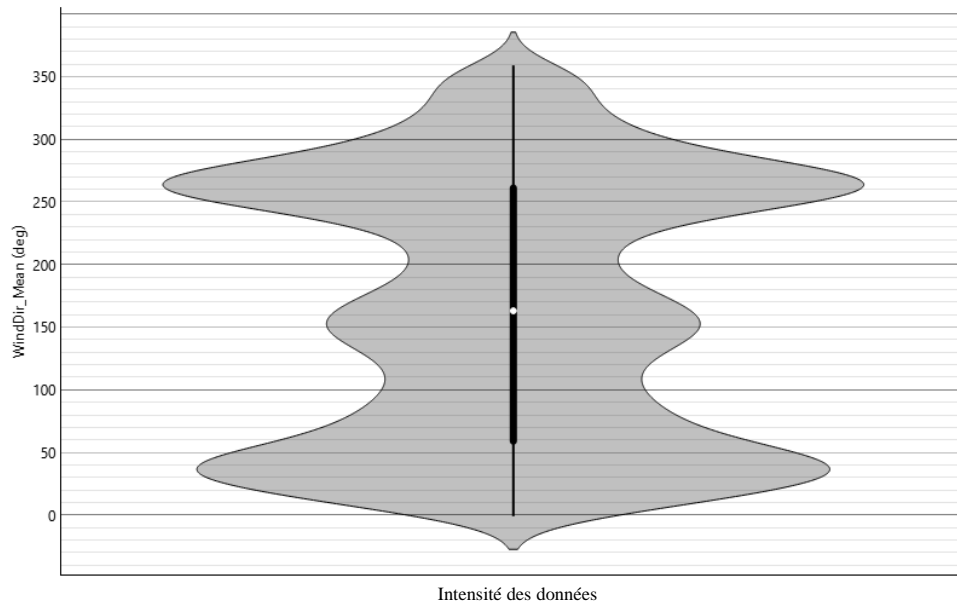


Figure 36 : Diagramme en violon des directions moyennes du vent

d) Vitesse du vent min.

La Figure 37 présente un diagramme en violon représentant la distribution de la variable Vitesse du vent min. (km/h), c'est-à-dire la vitesse minimale du vent exprimée en kilomètres par heure. L'axe vertical (Y) indique les valeurs de cette vitesse, variant de 0 à 40 km/h. La forme du violon, en gris, visualise la densité de probabilité : plus le violon est large à une certaine hauteur, plus les données sont concentrées autour de cette valeur. Un boxplot est superposé au centre du violon, offrant des repères statistiques : le point blanc représente la médiane, le rectangle noir indique l'intervalle interquartile (allant du premier au troisième quartile), et les lignes verticales noires délimitent l'étendue des valeurs normales, excluant les valeurs extrêmes. L'analyse du graphique montre que la majorité des vitesses minimales du vent se situe entre 2 et 5 km/h, ce qui en fait la plage la plus représentative. Toutefois, une queue allongée vers les valeurs élevées, jusqu'à 40 km/h, signale la présence de quelques vitesses minimales plus élevées, bien que celles-ci soient rares. Cette configuration traduit une distribution asymétrique, avec une asymétrie positive, puisque la queue est plus prononcée du côté des grandes valeurs. Ainsi, ce diagramme en violon permet de visualiser clairement la concentration des données, leur dispersion, ainsi que la présence de valeurs atypiques dans les vitesses minimales du vent.

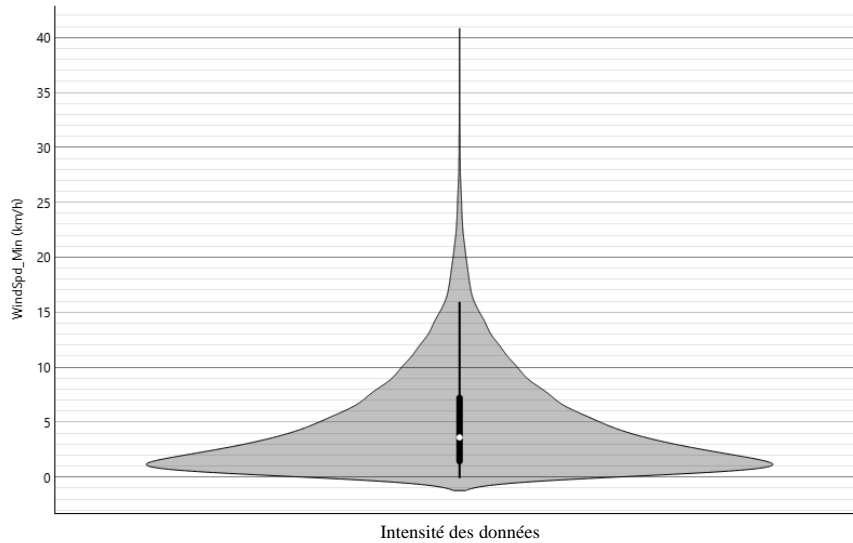


Figure 37 : "Distribution de la vitesse minimale du vent Vitesse du vent min.en km/h"

e) Vitesse du vent max

La Figure 38 illustre la distribution des vitesses maximales du vent (exprimées en km/h) à l'aide d'un diagramme en violon. L'axe vertical représente les différentes vitesses enregistrées, tandis que la forme grise du violon indique la fréquence des observations : les zones les plus larges correspondent aux vitesses les plus fréquentes, tandis que les parties étroites signalent des valeurs plus rares. Un boxplot est intégré au centre du violon, offrant des repères statistiques essentiels : le point blanc représente la médiane, la barre noire montre l'intervalle interquartile (du premier au troisième quartile), et les lignes fines illustrent l'étendue des valeurs considérées comme normales. On observe une traîne vers le haut, signe de la présence de quelques vitesses très élevées, allant jusqu'à 50 km/h, bien que ces valeurs soient peu fréquentes. En revanche, la majorité des vitesses maximales se situe entre 5 et 15 km/h, ce qui constitue la plage la plus représentative. Ainsi, la distribution n'est pas symétrique et présente une asymétrie positive, avec une tendance tirée vers la droite en raison de ces valeurs extrêmes. Ce violon plot offre donc une représentation visuelle complète, permettant de saisir d'un coup d'œil la concentration des vitesses, leur étendue et leur distribution globale.

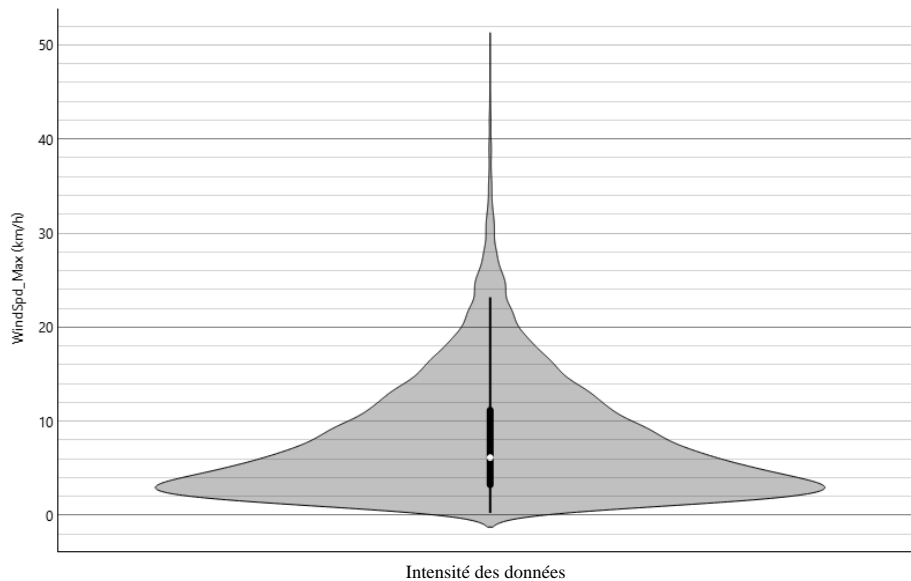


Figure 38 : "Distribution de la vitesse maximale du vent en km/h"

f) Temp_Moyenne (deg)

La Figure 39 présente un diagramme en violon représentant la distribution de la variable Temp_Moyenne (deg), soit la température moyenne exprimée en degrés Celsius. Ce graphique permet de visualiser à la fois la tendance centrale, la dispersion et la forme de la distribution des températures. L'étendue des données s'étale d'environ 5 à 30 degrés, avec une forte concentration des valeurs entre 18° et 22°, comme l'indique l'élargissement du violon dans cette plage. Cette concentration suggère que les trois quarts des températures observées se situent dans cet intervalle. La distribution présente une légère asymétrie : elle est davantage étalée au-dessus de 22° qu'en dessous de 18°, ce qui indique une présence plus marquée de températures élevées. La médiane, représentée par une ligne noire horizontale au centre du violon, se situe autour de 18°. En résumé, ce violon plot offre une représentation visuelle riche et détaillée de la distribution des températures moyennes, soulignant clairement la concentration des valeurs autour de la médiane ainsi que leur dispersion. C'est un outil efficace pour appréhender non seulement la tendance centrale, mais aussi la densité et la variabilité des données.

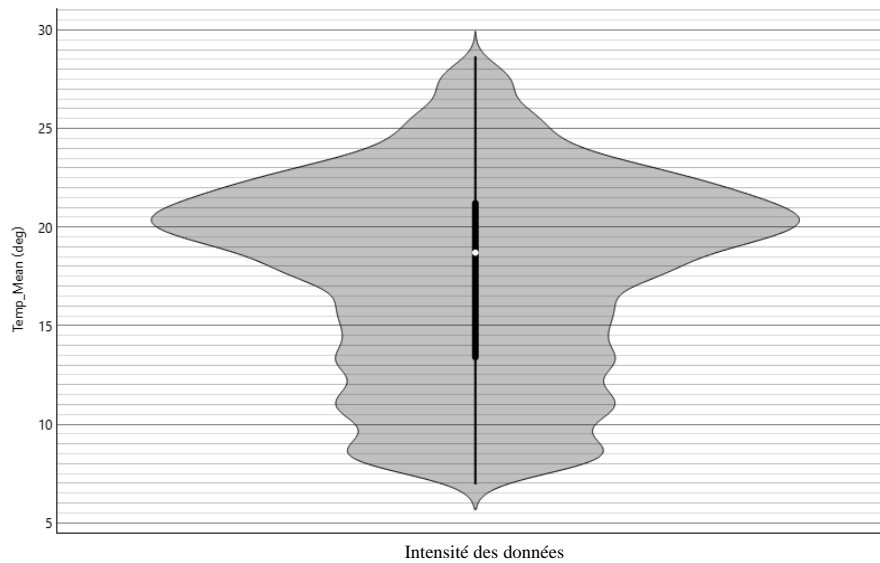


Figure 39 : Diagramme de distribution des températures moyennes (°C)

g) Humidité Relative_Moyenne

La Figure 40 présente un diagramme en violon illustrant la distribution de la variable "RH_Moyenne (%)", correspondant au taux d'humidité moyen. Ce type de graphique permet de visualiser d'un seul coup d'œil la forme de la distribution, sa tendance centrale ainsi que sa dispersion. La largeur du violon reflète la concentration des données à chaque niveau : on observe une concentration notable des valeurs autour de 50 à 60 %, ce qui en fait la plage la plus fréquente. La médiane est représentée par une ligne noire horizontale, tandis que la moyenne est indiquée par un point noir. La distribution montre une légère asymétrie, avec un étalement plus marqué vers la droite, ce qui suggère une tendance à des valeurs plus élevées dans certains cas. De petits graphiques internes situés sur les violons permettent également d'apprécier la dispersion des valeurs autour de la médiane, offrant un aperçu des valeurs extrêmes et de la variabilité au sein de l'échantillon. En somme, ce diagramme en violon fournit une représentation visuelle complète de la distribution des données d'humidité relative, mettant en évidence leur concentration, leur asymétrie et leur plage de variation. Il constitue ainsi une alternative à la fois esthétique et informative aux histogrammes ou aux boxplots traditionnels

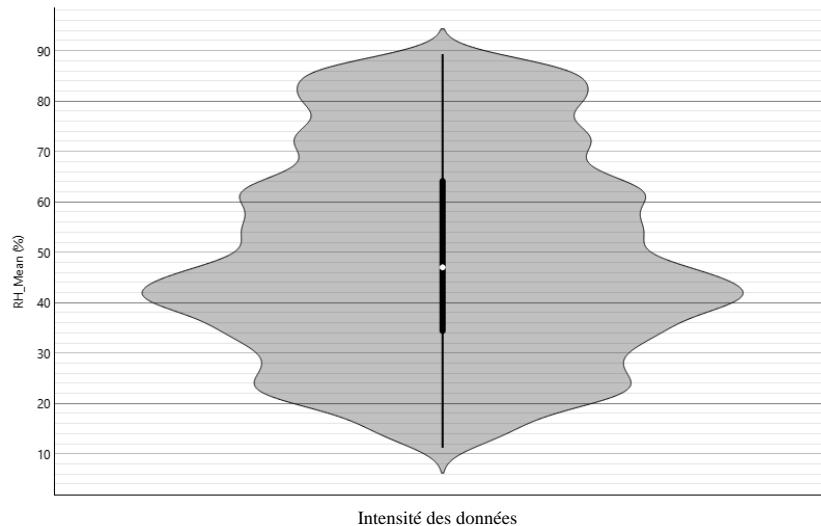


Figure 40 : Distribution de l'Humidité Relative Moyenne (RH_Moyenne) en Pourcentage (%)

h) Pression au niveau de la mer moyenne en hpa (MSLP)

La Figure 41 est un graphique en forme de violon montrant la distribution de la pression atmosphérique prévue au niveau de la mer (Pression au niveau de la mer, MSLP) en hPa. Sa forme grise représente la densité des observations : plus c'est épais, plus il s'agit de valeurs fréquentes. Au centre se trouve un boxplot vertical en noir qui illustre les statistiques principales. Le point blanc représente la médiane, tandis que la zone sombre montre les 1er et 3ème quartiles et les lignes minces montrent les valeurs extrêmes. La pression la plus fréquemment observée est entre 1018 et 1022 hPa, qui correspond à la partie la plus épaisse du violon. La lecture visuelle du graphique suggère qu'il y a de façon surprenante – deux principales tendances de données originales. La plupart des pressions se trouvent entre 1017 et 1022 hPa, ce qui est proche de la pression atmosphérique normale au niveau de la mer.

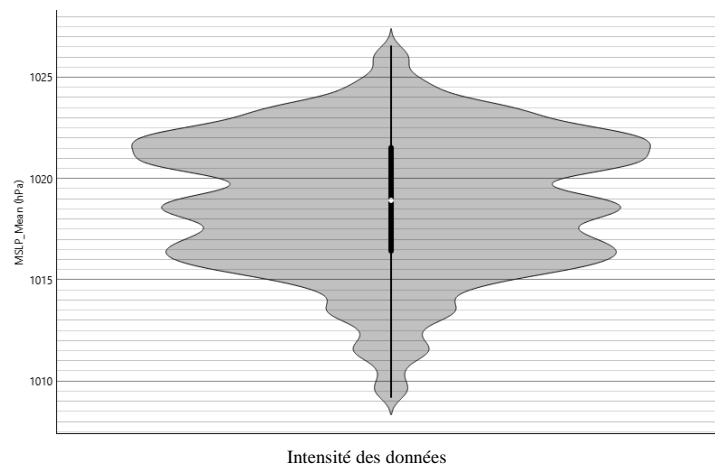


Figure 41: Distribution de la pression atmosphérique moyenne au niveau de la mer en hPa

i) Irradiance moyenne

La Figure 42 présente un diagramme en violon illustrant la distribution du rayonnement solaire moyen, exprimé en W/m^2 . Sur l'axe vertical (Y), les valeurs du rayonnement solaire varient de 0 jusqu'à plus de 1000 W/m^2 . L'axe horizontal ne correspond pas à une variable précise, mais permet de visualiser la densité de distribution des données. La zone grise du graphique représente cette densité : plus elle est large à un certain niveau, plus la concentration de valeurs est élevée à ce niveau. Les points noirs indiquent les valeurs individuelles du rayonnement solaire, similaires à un diagramme en points, tandis que le point blanc central marque la médiane des observations. La forme générale du diagramme reflète une distribution bimodale, caractérisée par un premier pic de densité à de très faibles valeurs et un second situé autour de 600 à 700 W/m^2 . La médiane, relativement basse (entre 200 et 300 W/m^2), suggère que la majorité des observations sont faibles, malgré la présence d'un groupe important de fortes valeurs.

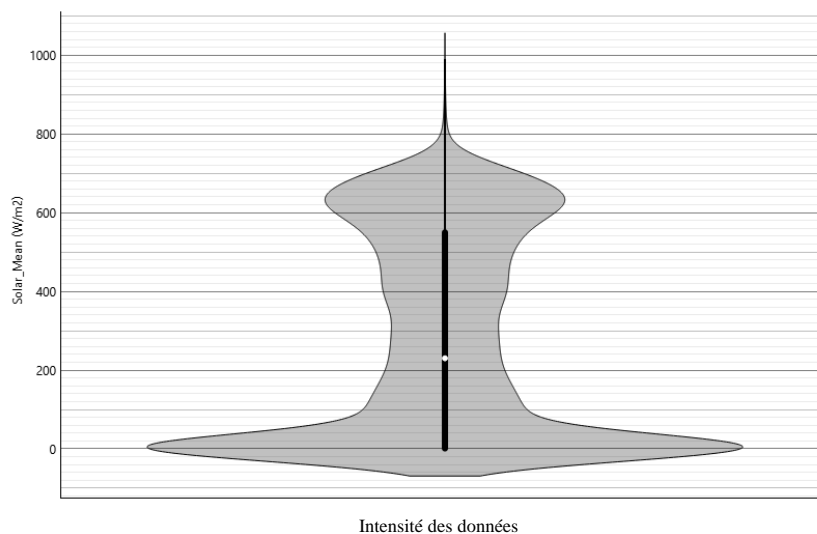


Figure 42 : diagramme de distribution moyenne de rayonnement solaire

IV.6 Cadre de prédiction

La Figure 43 représente le cadre de prédiction de notre étude basé sur des données photovoltaïque. Ce processus de prédiction est bien structuré en plusieurs étapes fondamentales. Il commence par l'acquisition de données solaires, suivies de la préparation et de la partition en deux ensembles : 80 % sont destinés à l'entraînement des modèles prédictifs et 20 % à leur évaluation. Ensuite, les modèles issus de la formation sont entraînés et utilisés

pour produire des prédictions du modèle. Les valeurs de prédictions sont comparées aux valeurs réelles mesurées pour déterminer la pertinence du modèle.

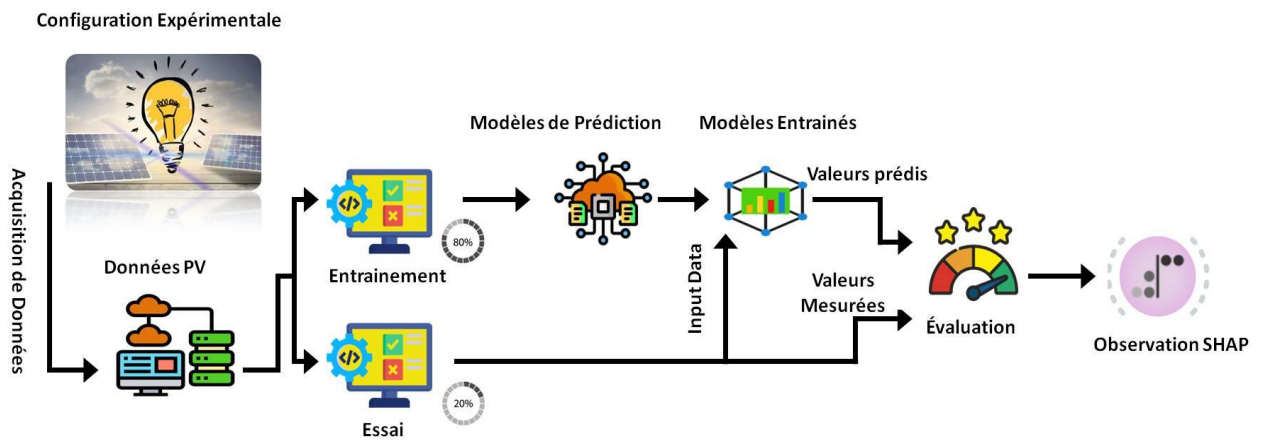


Figure 43 : Organigramme du cadre de prédiction de notre étude

IV.7 Interprétation du tableau

Tableau 4 compare les performances de quatre modèles de régression à l'aide de plusieurs métriques : MSE, RMSE, MAE et R^2 . L'analyse des performances des quatre modèles (Tree, AdaBoost, Régression Linéaire et PLS) montre clairement que la régression linéaire est le modèle le plus performant. En effet, elle présente les plus faibles valeurs d'erreurs avec un MSE de 3 899 6047.557, un RMSE de 6 244.682, un MAE de 5 149.689. De plus, son coefficient de détermination R^2 atteint 0.975, indiquant qu'elle explique 97.5 % de la variance des données, ce qui est excellent.

Le modèle AdaBoost suit avec des résultats relativement bons, meilleurs que ceux de l'arbre de décision, avec un R^2 de 0.894 et des erreurs modérées. Il constitue donc une alternative valable, notamment si la relation entre les variables n'est pas parfaitement linéaire. En revanche, le modèle "Tree" présente des performances inférieures, affichant les plus grandes erreurs quadratiques et absolues (MSE et RMSE élevés), ce qui limite sa fiabilité malgré un R^2 correct de 0.866. Enfin, le modèle PLS montre également des erreurs significatives, en particulier un MAE malgré un R^2 de 0.880, ce qui laisse penser qu'il s'adapte mal aux données dans ce cas précis.

En conclusion, la régression linéaire est le modèle à privilégier ici en raison de sa précision et de sa capacité explicative supérieure, suivie éventuellement d'AdaBoost. Les modèles Arbre de décision et PLS semblent moins adaptés aux données de ce cas d'étude.

<i>Modèle</i>	<i>MSE</i>	<i>RMSE</i>	<i>MAE</i>	<i>R²</i>
Arbre de décision	2 127 495 59.078	14 585.937	8 071.833	0.866
AdaBoost	1 686 912 74.377	12 988.121	7 242.852	0.894
Linear Regression	38 996 047.557	6 244.682	5 149.689	0.975
PLS	1 908 715 83.284	13 815.628	10 550.116	0.880

Tableau 4 : Évaluation comparative des performances de quatre modèles

IV.8 Résultats du tracage du diagramme SHAP

Les valeurs SHAP (Explications additives de Shapley) permettent d'expliquer les résultats de tout modèle d'apprentissage automatique. En apprentissage automatique, chaque caractéristique se voit attribuer une valeur d'importance représentant sa contribution aux résultats du modèle. Elles montrent comment chaque fonctionnalité affecte chaque prédiction finale, l'importance de chaque fonctionnalité par rapport aux autres et la dépendance du modèle à l'interaction entre les fonctionnalités.[90]

Les valeurs SHAP sont indépendantes du modèle, ce qui signifie qu'elles peuvent être utilisées pour interpréter n'importe quel modèle d'apprentissage automatique, notamment :

- Régression linéaire
- Arbres de décision
- Adaboost
- PLS

La Figure 44 présentée est un graphique d'interprétabilité de type SHAP (Explications additives de Shapley), illustrant l'impact de différentes variables explicatives sur la sortie d'un modèle d'apprentissage automatique (linear regression). Chaque ligne correspond à une variable (ou "feature") utilisée dans le modèle, et l'axe horizontal représente l'impact de cette variable sur la prédiction (positif ou négatif). Les points colorés le long de chaque ligne représentent des observations individuelles, où la couleur indique la valeur de la variable (du

bleu pour des valeurs faibles au rouge pour des valeurs élevées), et leur position sur l'axe horizontal montre comment cette valeur a influencé la prédiction du modèle (vers la gauche = effet négatif, vers la droite = effet positif). On observe que la variable Solar_Mean (W/m^2) a le plus grand impact sur les prédictions, avec une répartition claire entre des valeurs faibles (bleues) qui réduisent la sortie du modèle, et des valeurs élevées (rouges) qui l'augmentent fortement. Ensuite, des variables comme Temp_Moyenne (deg), RH_Moyenne (%), et Vitesse du vent min. (km/h) montrent également une certaine influence, bien que moins marquée. Les autres variables, telles que WindDir_Mean (deg) ou Vitesse du vent max. (km/h), ont un impact plus faible sur la sortie du modèle.

Ce graphique permet donc de visualiser non seulement l'importance relative des variables pour le modèle, mais aussi la direction et l'intensité de leur contribution aux prédictions, tout en tenant compte des valeurs réelles observées pour chaque variable.

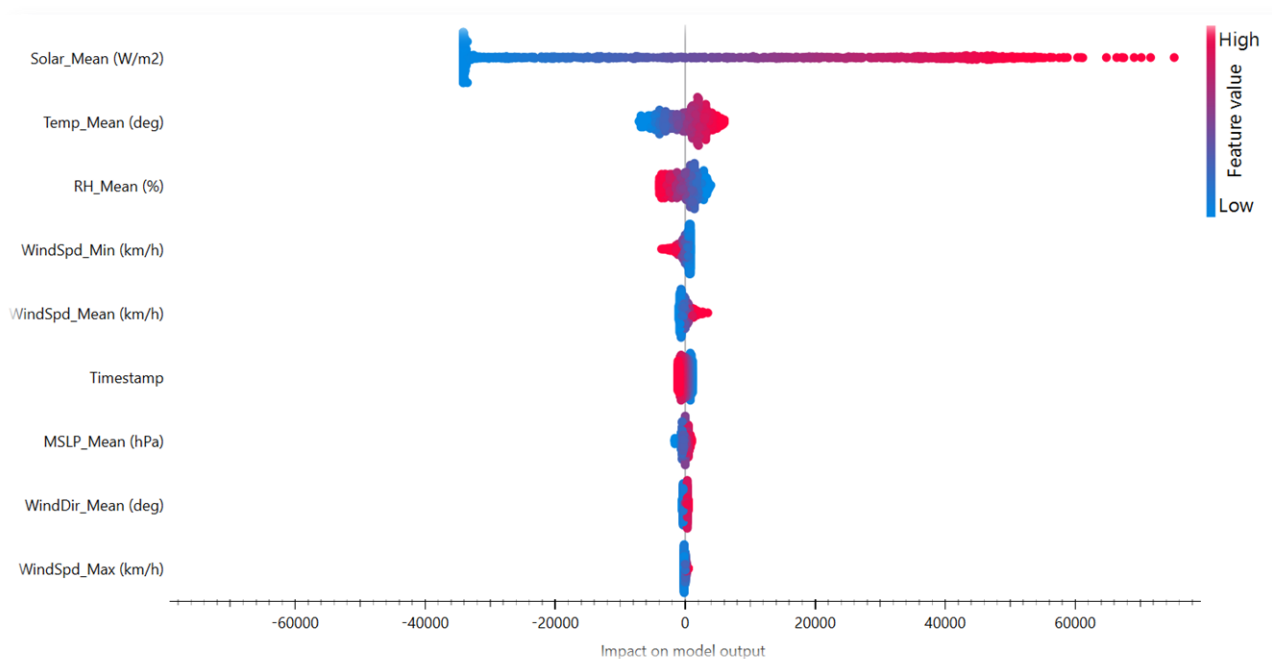


Figure 44 : Résultat du plot SHAP du modèle linear regression

IV.9 Conclusion

Cette étude a démontré l'exploration du potentiel des méthodes d'apprentissage automatique pour prédire la génération de PV à travers une approche pratique via le logiciel Orange3. Cette plate-forme a facilité la mise en place, la visualisation et la comparaison de différents modèles prédictifs tout en fournissant une interface accessible pour analyser des données complexes. L'étude comparative a identifié les forces et les limites de chaque algorithme utilisé. Certains modèles se sont caractérisés par leur capacité à faire des prédictions plus précises, tandis que d'autres ont montré une plus grande robustesse face à la variabilité des données. Ces résultats soulignent l'importance de sélectionner des algorithmes en fonction du contexte de l'application.

.

Conclusion Générale

Conclusion Générale

Cette étude complète sur la prévision de la production d'énergie photovoltaïque à l'aide de l'apprentissage automatique contient plusieurs conclusions clés. Le cadre théorique a d'abord mis en évidence le rôle central de l'énergie solaire dans la transition énergétique et l'importance cruciale des technologies d'IA pour surmonter les défis des pannes de courant. L'analyse bibliographique a révélé l'évolution des méthodes, des modèles physiques traditionnels aux architectures hybrides combinant réseaux de neurones et algorithmes d'optimisation. La partie méthodologique a démontré l'importance des quatre modèles sélectionnés (AdaBoost, PLS, arbres de décision et régression linéaire) en identifiant leurs spécificités. Enfin, les simulations ont démontré que cette étude démontrait le grand potentiel de l'intégration de l'énergie solaire photovoltaïque dans le programme d'Orange, ouvrant ainsi de nouveaux horizons pour une gestion de l'énergie plus durable et plus intelligente. Les progrès technologiques, notamment les modèles de prévision hybrides et les solutions numériques innovantes, améliorent considérablement la production et la distribution de cette énergie renouvelable.

Avec son infrastructure technologique et son expertise en transformation digitale, Orange est un acteur clé pour faciliter cette intégration. Les solutions développées dans ce cadre contribuent non seulement à améliorer l'efficacité énergétique, mais également à accroître la stabilité du réseau et à accélérer la transition vers des systèmes plus propres.

Cette thèse ouvre la voie à des recherches futures, notamment l'intégration de données météorologiques en temps réel, l'amélioration des modèles à l'aide de techniques d'apprentissage automatique ou l'exploration de méthodes avancées telles que les réseaux neuronaux profonds. L'apprentissage automatique, soutenu par des outils accessibles comme Orange3, est un élément clé pour une transition énergétique plus intelligente et plus efficace et pour accélérer la transition vers les énergies renouvelables.

Cette synergie entre les technologies photovoltaïques et numériques est un parfait exemple de la manière dont l'innovation peut répondre aux défis énergétiques actuels. Cela ouvre la porte à de futures recherches et applications qui pourraient changer de manière permanente la façon dont nous produisons et consommons l'énergie.

Références Bibliographiques

- [1] Z. Hamza, « Utilisation de la méthode du plan d'expériences pour un choix optimal des paramètres géométriques d'une cheminée solaire. ».
- [2] « Énergies renouvelables : définition, exemples, avantages et limites », Youmatter. Consulté le: 6 février 2025. [En ligne]. Disponible sur: <https://youmatter.world/fr/definition/energies-renouvelables-definition/>
- [3] N. Benabbas, N. E. H. Chiha, et A. Benyahia, « DIMENSIONNEMENT D'UN SYSTEME PHOTOVOLTAÏQUE AVEC STOCKAGE POUR ALIMENTER UN SITE ISOLE (APPLIQUEE SUR L'ECOLE PRIMAIRE BELGHAZALI MOHAMMED-NAAMA-) », CENTRE UNIVERSITAIRE SALHI AHMED - NAAMA, 2020. Consulté le: 6 février 2025. [En ligne]. Disponible sur: [file:///C:/Users/Splendid/Desktop/Memoire-finale-2020-1-1%20\(1\).pdf](file:///C:/Users/Splendid/Desktop/Memoire-finale-2020-1-1%20(1).pdf)
- [4] « ref10x10_3.pdf ». Consulté le: 6 février 2025. [En ligne]. Disponible sur: https://www.ctc-n.org/sites/www.ctc-n.org/files/UNFCCC_docs/ref10x10_3.pdf
- [5] W. Mchara, « L'IA au service des énergies renouvelables : prédire pour mieux gérer », Globaldev Blog. Consulté le: 12 mai 2025. [En ligne]. Disponible sur: <https://globaldev.blog/fr/lia-au-service-des-energies-renouvelables-predire-pour-mieux-gerer/>
- [6] G. Narvaez Morales, « Intelligence artificielle pour les systèmes d'énergie renouvelable », These de doctorat, Université de Toulouse (2023-....), 2024. Consulté le: 6 février 2025. [En ligne]. Disponible sur: <https://theses.fr/2024TLSES057>
- [7] « Cas d'usage de l'IA dans le secteur : Énergies renouvelables ».
- [8] U. Nations, « Les énergies renouvelables : qu'est-ce que c'est ? | Nations Unies », United Nations. Consulté le: 6 février 2025. [En ligne]. Disponible sur: <https://www.un.org/fr/climatechange/what-is-renewable-energy>
- [9] M. Novel, « Quelle est la meilleure énergie renouvelable ? La réponse va vous surprendre », GreenWorks. Consulté le: 7 février 2025. [En ligne]. Disponible sur: <https://www.greenworks.lu/energie/quelle-est-la-meilleure-energie-renouvelable-la-reponse-va-vous-surprendre/594566/>
- [10] C. Gouvernement du Nouveau-Brunswick, « Sources d'énergie renouvelable - Énergie et Mines ». Consulté le: 6 février 2025. [En ligne]. Disponible sur: https://www2.gnb.ca/content/gnb/fr/ministeres/der/energie/content/renewable.html?__cf_chl_tk=BSegLsw8mts5I3jhmz5K7UXRoUKZR02ge4IXNSj9rPM-1738256286-1.0.1.1-Hvh6vn70h_iRS6A6eVBzT.YjCmgYXtZbI1w9W7vZL.w
- [11] J. Kitto, « THE SPIRITUAL SYMBOLISM OF SOLAR ENERGY », The Daily Guardian. Consulté le: 7 février 2025. [En ligne]. Disponible sur: <https://theguardian.com/>
- [12] « Hydraulique | EDF FR ». Consulté le: 6 février 2025. [En ligne]. Disponible sur: <https://www.edf.fr/groupe-edf/comprendre/production/hydraulique>
- [13] « La Quotidienne Actu-Environnement ». Consulté le: 17 février 2025. [En ligne]. Disponible sur: https://www.actu-environnement.com/ae/newsletter/newsletter_quotidienne.php?id=1943

- [14] I. BENGOURAIN, « Représentations des systèmes hybrides à énergie renouvelables », Université Aboubakr Belkaïd- Tlemcen Faculté de TECHNOLOGIE, 2022. [En ligne]. Disponible sur: [file:///C:/Users/Splendid/Downloads/Ms.Gm.Bengourain%20\(1\).pdf](file:///C:/Users/Splendid/Downloads/Ms.Gm.Bengourain%20(1).pdf)
- [15] Stéphanie, « Combien coûte l'énergie éolienne ? », ECO Delta. Consulté le: 7 février 2025. [En ligne]. Disponible sur: <https://ecodelta.fr/combien-coute-lenergie-eolienne/>
- [16] « Chancen und Herausforderungen von Biomasse ». Consulté le: 17 février 2025. [En ligne]. Disponible sur: <https://www.bakertilly.de/beitrag/neue-perspektiven-fuer-biomasse-chancen-und-herausforderungen-in-der-energiepolitik.html>
- [17] R. naturelles Canada, « Énergie géothermique ». Consulté le: 6 février 2025. [En ligne]. Disponible sur: <https://ressources-naturelles.canada.ca/nos-ressources-naturelles/sources-denergie-reseau-distribution/energies-renouvelables/energie-geothermique/25791>
- [18] « Géothermie : définition, procédés d'exploitation, chiffres clés ». Consulté le: 7 février 2025. [En ligne]. Disponible sur: <https://www.connaissancedesenergies.org/fiche-pedagogique/geothermie>
- [19] Z. Belhait et A. Kenouni, « L'énergie photovoltaïque », -Annaba- Mokhtar Badji Université, 2021. [En ligne]. Disponible sur: <file:///C:/Users/Splendid/Desktop/memoire%20sur.pdf>
- [20] « La cellule photovoltaïque : comment ça marche ? » Consulté le: 7 février 2025. [En ligne]. Disponible sur: <https://www.planete-energies.com/fr/media/article/cellule-photovoltaique-comment-ca-marche>
- [21] « Comment fonctionne un panneau solaire ». Consulté le: 7 février 2025. [En ligne]. Disponible sur: <https://www.manomano.fr/conseil/comment-fonctionne-un-panneau-solaire-5973>
- [22] « Cellule photovoltaïque : fonctionnement, types, rendement ». Consulté le: 7 février 2025. [En ligne]. Disponible sur: <https://mypower.engie.fr/conseils/panneaux-solaires/panneau-solaire-photovoltaique/types-cellules-solaires.html>
- [23] « Fichier:Monocristallin.png — Solarpedia ». Consulté le: 7 février 2025. [En ligne]. Disponible sur: <https://fr.solarpedia.net/wiki/index.php?title=Fichier:Monocristallin.png>
- [24] « شعشوعة حبيب (1).pdf ». Consulté le: 7 février 2025. [En ligne]. Disponible sur: <https://ds.univ-oran2.dz:8443/bitstream/123456789/8651/1/%D8%B4%D8%B9%D8%B4%D9%88%D8%B9%D8%A9%20%D8%AD%D8%A8%D9%8A%D8%A8%20%281%29.pdf>
- [25] « Cellule au silicium polycristallin — Solarpedia ». Consulté le: 7 février 2025. [En ligne]. Disponible sur: https://fr.solarpedia.net/wiki/index.php?title=Cellule_au_silicium_polycristallin
- [26] S. ZAÏRI et S. BOUBIADA, « Etude et Dimensionnement d'une centrale Photovoltaïque », Université Mohamed Khider de Biskra, 2020.
- [27] « Figure 1.7: Cellule en silicium amorphe[6] », ResearchGate. Consulté le: 7 février 2025. [En ligne]. Disponible sur: https://www.researchgate.net/figure/Cellule-en-silicium-amorphe6_fig12_301487491
- [28] « Énergie solaire : IA et machine learning dans l'évolution technologique », France renouvelables. Consulté le: 7 février 2025. [En ligne]. Disponible sur: <https://www.france-renouvelables.fr/guide-energie-solaire/energie-solaire-ia-machine-learning-evolution-technologique/>
- [29] P. Domingos, « A few useful things to know about machine learning », *Commun. ACM*, vol. 55, n° 10, p. 78-87, oct. 2012, doi: 10.1145/2347736.2347755.
- [30] « On-Machine-Learning.pdf ». Consulté le: 5 février 2025. [En ligne]. Disponible sur: <https://blas.com/wp-content/uploads/2019/07/On-Machine-Learning.pdf>

- [31] « Initiez-vous aux fondamentaux du Machine Learning ou de l'apprentissage automatique ». Consulté le: 7 février 2025. [En ligne]. Disponible sur: <https://openclassrooms.com/fr/courses/6417031-objectif-ia-initiez-vous-a-lintelligence-artificielle/6823501-initiez-vous-aux-fondamentaux-du-machine-learning-ou-de-lapprentissage-automatique>
- [32] « Qu'est-ce que l'apprentissage automatique ? Définition, types et exemples », Coursera. Consulté le: 5 février 2025. [En ligne]. Disponible sur: https://www.coursera.org/fr-FR/articles/what-is-machine-learning?utm_source=link&utm_medium=page_share&utm_content=article&utm_campaign=sharing_cta
- [33] PythonTech, « Introduction à l'apprentissage automatique (Machine Learning) ». Consulté le: 7 février 2025. [En ligne]. Disponible sur: <https://www.tresfacile.net/introduction-a-lapprentissage-automatique-machine-learning/>
- [34] M. Naila, « Conception et développement des méthodes de prédiction de la durée de séjour hospitalier centrées sur des techniques de “machine learning” », 2022.
- [35] « What Is Supervised Learning? | IBM ». Consulté le: 5 février 2025. [En ligne]. Disponible sur: <https://www.ibm.com/think/topics/supervised-learning>
- [36] « DJEMMAM_AYET EL NOUR_F5.pdf ». Consulté le: 7 février 2025. [En ligne]. Disponible sur: https://dspace.univ-guelma.dz/jspui/bitstream/123456789/14976/1/DJEMMAM_AYET%20EL%20NOUR_F5.pdf
- [37] « Types d'apprentissage automatique - Différentes méthodes et types de modèles | Big Data 2025 », Education-WIKI.com. Consulté le: 5 février 2025. [En ligne]. Disponible sur: <https://fr.education-wiki.com/1309569-types-of-machine-learning>
- [38] « Les principaux algorithmes de Machine Learning ». Consulté le: 5 février 2025. [En ligne]. Disponible sur: <https://www.jedha.co/formation-ia/algorithmes-machine-learning>
- [39] « Modern artificial intelligence : What is unsupervised learning? - Phedone ». Consulté le: 7 février 2025. [En ligne]. Disponible sur: <https://phedone.com/en/blog/artificial-intelligence-what-is-the-unsupervised-learning/>
- [40] « Les 3 étapes essentielles de l'apprentissage automatique (Machine Learning) », Spiria. Consulté le: 5 février 2025. [En ligne]. Disponible sur: <https://www.spiria.com/fr/blogue/intelligence-artificielle/3-etapes-essentielles-apprentissage-automatique-machine-learning/>
- [41] Z. ISMAILI, « Apprentissage Supervisé Vs. Non Supervisé », BrightCape. Consulté le: 5 février 2025. [En ligne]. Disponible sur: <https://brightcape.co/apprentissage-supervise-vs-non-supervise/>
- [42] « Qu'est-ce que l'apprentissage semi-supervisé ? | IBM ». Consulté le: 5 février 2025. [En ligne]. Disponible sur: <https://www.ibm.com/fr-fr/topics/semi-supervised-learning>
- [43] « Semi-supervised learning - Mastering ». Consulté le: 7 février 2025. [En ligne]. Disponible sur: https://www.jimi2mall.com/?product_id=106839404_38
- [44] Igayhardt, « Apprentissage profond et apprentissage automatique - Azure Machine Learning ». Consulté le: 5 février 2025. [En ligne]. Disponible sur: <https://learn.microsoft.com/fr-fr/azure/machine-learning/concept-deep-learning-vs-machine-learning?view=azureml-api-2>
- [45] « Machine Learning vs Deep Learning : quelles différences ? » Consulté le: 5 février 2025. [En ligne]. Disponible sur: <https://www.jedha.co/formation-ia/vraie-difference-machine-learning-deep-learning>
- [46] M. B. HENDA, « Deep Learning, Machine Learning », Site personnel de Mokhtar BEN HENDA. Consulté le: 7 février 2025. [En ligne]. Disponible sur: <http://www.benhenda.com/fra/?p=1802>

- [47] K. Wolfenstein, « Comment l'intelligence artificielle ☐ peut augmenter l'efficacité des systèmes photovoltaïques ☀☑ ». Consulté le: 5 février 2025. [En ligne]. Disponible sur: <https://xpert.digital/ar/كفاءة-كي-الكهروضوئية/>
- [48] « France renouvelables - Système électrique et énergies renouvelables », France renouvelables. Consulté le: 5 février 2025. [En ligne]. Disponible sur: <https://www.france-renouvelables.fr/>
- [49] C. Wang, M. Li, Y. Cao, et T. Lu, « Gradient boosting dendritic network for ultra-short-term PV power prediction », *Front. Energy*, vol. 18, n° 6, p. 785-798, déc. 2024, doi: 10.1007/s11708-024-0915-y.
- [50] M. Jiang *et al.*, « Research on time-series based and similarity search based methods for PV power prediction », *Energy Convers. Manag.*, vol. 308, p. 118391, mai 2024, doi: 10.1016/j.enconman.2024.118391.
- [51] B. Li, X. Chen, et A. Jain, « Enhancing Power Prediction of Photovoltaic Systems: Leveraging Dynamic Physical Model for Irradiance-to-Power Conversion ».
- [52] J. Antonanzas, N. Osorio, R. Escobar, R. Urraca, F. J. Martinez-de-Pison, et F. Antonanzas-Torres, « Review of photovoltaic power forecasting », *Sol. Energy*, vol. 136, p. 78-111, oct. 2016, doi: 10.1016/j.solener.2016.06.069.
- [53] A. Dairi, F. Harrou, Y. Sun, et S. Khadraoui, « Short-Term Forecasting of Photovoltaic Solar Power Production Using Variational Auto-Encoder Driven Deep Learning Approach », *Appl. Sci.*, vol. 10, n° 23, Art. n° 23, janv. 2020, doi: 10.3390/app10238400.
- [54] P. Gupta et R. and Singh, « PV power forecasting based on data-driven models: a review », *Int. J. Sustain. Eng.*, vol. 14, n° 6, p. 1733-1755, nov. 2021, doi: 10.1080/19397038.2021.1986590.
- [55] H. Sharadga, S. Hajimirza, et R. S. Balog, « Time series forecasting of solar power generation for large-scale photovoltaic plants », *Renew. Energy*, vol. 150, p. 797-807, mai 2020, doi: 10.1016/j.renene.2019.12.131.
- [56] « Monitoring of Photovoltaic Systems Using Improved Kernel-Based Learning Schemes ». Consulté le: 9 mai 2025. [En ligne]. Disponible sur: <https://ieeexplore.ieee.org/abstract/document/9363277>
- [57] F. Harrou, A. Dorbane, et Y. Sun, « Enhancing photovoltaic system performance using advanced gradient boosting methods », in *2024 International Conference on Control, Automation and Diagnosis (ICCAD)*, mai 2024, p. 1-6. doi: 10.1109/ICCAD60883.2024.10553919.
- [58] « Review on forecasting of photovoltaic power generation based on machine learning and metaheuristic techniques - Akhter - 2019 - IET Renewable Power Generation - Wiley Online Library ». Consulté le: 10 février 2025. [En ligne]. Disponible sur: <https://ietresearch.onlinelibrary.wiley.com/doi/full/10.1049/iet-rpg.2018.5649>
- [59] M. N. Akhter, S. Mekhilef, H. Mokhlis, et N. Mohamed Shah, « Review on forecasting of photovoltaic power generation based on machine learning and metaheuristic techniques », *IET Renew. Power Gener.*, vol. 13, n° 7, p. 1009-1023, mai 2019, doi: 10.1049/iet-rpg.2018.5649.
- [60] R. Banik et A. Biswas, « Improving Solar PV Prediction Performance with RF-CatBoost Ensemble: A Robust and Complementary Approach », *Renew. Energy Focus*, vol. 46, p. 207-221, sept. 2023, doi: 10.1016/j.ref.2023.06.009.
- [61] « SVR-Based Model to Forecast PV Power Generation under Different Weather Conditions ». Consulté le: 10 février 2025. [En ligne]. Disponible sur: <https://www.mdpi.com/1996-1073/10/7/876>
- [62] U. K. Das *et al.*, « SVR-Based Model to Forecast PV Power Generation under Different Weather Conditions », *Energies*, vol. 10, n° 7, Art. n° 7, juill. 2017, doi: 10.3390/en10070876.

- [63] M. W. Ahmad, M. Mourshed, et Y. Rezgui, « Tree-based ensemble methods for predicting PV power generation and their comparison with support vector regression », *Energy*, vol. 164, p. 465-474, déc. 2018, doi: 10.1016/j.energy.2018.08.207.
- [64] « Qu'est-ce qu'un modèle de machine learning ? », Databricks. Consulté le: 11 mai 2025. [En ligne]. Disponible sur: <https://www.databricks.com/fr/glossary/machine-learning-models>
- [65] Anshul, « Guide on AdaBoost Algorithm », Analytics Vidhya. Consulté le: 23 avril 2025. [En ligne]. Disponible sur: <https://www.analyticsvidhya.com/blog/2021/09/adaboost-algorithm-a-complete-guide-for-beginners/>
- [66] S. Baladram, « AdaBoost Classifier, Explained: A Visual Guide with Code Examples », TDS Archive. Consulté le: 23 avril 2025. [En ligne]. Disponible sur: <https://medium.com/data-science/adaboost-classifier-explained-a-visual-guide-with-code-examples-fc0f25326d7b>
- [67] Y. Cao, Q.-G. Miao, J.-C. Liu, et L. Gao, « Advance and Prospects of AdaBoost Algorithm », *Acta Autom. Sin.*, vol. 39, n° 6, p. 745-758, juin 2013, doi: 10.1016/S1874-1029(13)60052-X.
- [68] « AdaBoost Algorithm », AlmaBetter. Consulté le: 12 avril 2025. [En ligne]. Disponible sur: <https://www.almabetter.com/bytes/tutorials/data-science/adaboost-algorithm>
- [69] G. M. Randazzo, « Why do Partial Least Squares (PLS) Still Deserve Attention Today? », Medium. Consulté le: 20 avril 2025. [En ligne]. Disponible sur: <https://medium.com/@gmrando/why-do-partial-least-squares-pls-still-deserve-attention-today-77f939db2eea>
- [70] « Régression par les moindres carrés partiels (PLS) », XLSTAT, Your data analysis solution. Consulté le: 24 avril 2025. [En ligne]. Disponible sur: <https://www.xlstat.com/fr/solutions/fonctionnalites/regression-par-moindres-carres-partiels-pls>
- [71] « What is Linear Regression?- Spiceworks - Spiceworks », Spiceworks Inc. Consulté le: 11 avril 2025. [En ligne]. Disponible sur: <https://www.spiceworks.com/tech/artificial-intelligence/articles/what-is-linear-regression/>
- [72] « t-Test, khi-deux, ANOVA, Régression, Corrélation... » Consulté le: 12 avril 2025. [En ligne]. Disponible sur: <https://datatab.fr/tutorial/linear-regression>
- [73] « Types d'analyses de régression ». Consulté le: 23 avril 2025. [En ligne]. Disponible sur: <https://support.minitab.com/fr-fr/minitab/help-and-how-to/statistical-modeling/regression/supporting-topics/basics/types-of-regression-analyses/>
- [74] KAVITA, « Linear Regression in Machine Learning », Analytics Vidhya. Consulté le: 23 avril 2025. [En ligne]. Disponible sur: <https://www.analyticsvidhya.com/blog/2021/10/everything-you-need-to-know-about-linear-regression/>
- [75] « Decision Trees in Machine Learning: Two Types (+ Examples) », Coursera. Consulté le: 11 avril 2025. [En ligne]. Disponible sur: <https://www.coursera.org/articles/decision-tree-machine-learning>
- [76] « Decision Tree Data Mining: All You Need to Know in 2025 », upGrad blog. Consulté le: 23 avril 2025. [En ligne]. Disponible sur: <https://www.upgrad.com/blog/what-is-decision-tree-in-data-mining/>
- [77] « 13 benefits and challenges of machine learning | Machine learning benefits and challenges », Lumenalta. Consulté le: 23 avril 2025. [En ligne]. Disponible sur: <https://lumenalta.com/insights/13-benefits-and-challenges-of-machine-learning>
- [78] L. and Found, « Common Problems When Training Deep Learning Models and How to overcome Them », Medium. Consulté le: 23 avril 2025. [En ligne]. Disponible sur:

- <https://medium.com/@lostandfound2654/common-problems-when-training-deep-learning-models-and-how-to-overcome-them-e37d0ac0a13b>
- [79] S. Devshatwar, « Challenges and Limitations of Machine Learning: What to Consider Before Implementation », Medium. Consulté le: 23 avril 2025. [En ligne]. Disponible sur: <https://medium.com/@shruti2402devshatwar/challenges-and-limitations-of-machine-learning-what-to-consider-before-implementation-d2c0af137647>
- [80] « Quels sont les défis auxquels vous êtes confrontés lors de l'intégration de l'apprentissage automatique dans les statistiques ? » Consulté le: 24 avril 2025. [En ligne]. Disponible sur: <https://fr.linkedin.com/advice/0/what-challenges-do-you-face-when-integrating-machine-xlbsf?lang=fr>
- [81] « Understanding ML Models: Types, Challenges, & Considerations ». Consulté le: 23 avril 2025. [En ligne]. Disponible sur: <https://www.markovml.com/blog/ml-models>
- [82] « Le machine learning : fondements, applications et perspectives futures en intelligence artificielle », <https://www.crossdata.tech/>. Consulté le: 11 mai 2025. [En ligne]. Disponible sur: <https://www.crossdata.tech/le-machine-learning-fondements-applications-et-perspectives-futures-en-intelligence-artificielle/>
- [83] L. Rédaction, « Orange 3 : Découvrez les nouveautés et fonctionnalités clés », Umvie. Consulté le: 22 mars 2025. [En ligne]. Disponible sur: <https://umvie.com/orange-3-decouvrez-les-nouveautes-et-fonctionnalites-cles/>
- [84] « What is model performance evaluation? | Fiddler AI ». Consulté le: 19 avril 2025. [En ligne]. Disponible sur: <https://www.fiddler.ai/model-evaluation-in-model-monitoring/what-is-model-performance-evaluation>
- [85] L. Molina, « Métriques en Machine Learning : Tout ce qu'il faut savoir », Formation Data Science | DataScientest.com. Consulté le: 20 avril 2025. [En ligne]. Disponible sur: <https://datascientest.com/metriques-en-machine-learning>
- [86] N. V. Otten, « Top 9 Performance Metrics In Machine Learning & How To Use Them », Spot Intelligence. Consulté le: 19 avril 2025. [En ligne]. Disponible sur: <https://spotintelligence.com/2024/03/12/performance-metrics-in-machine-learning/>
- [87] « PVdata_Power_PACE ».
- [88] « t-Test, Chi-Square, ANOVA, Regression, Correlation... » Consulté le: 10 avril 2025. [En ligne]. Disponible sur: <https://datatab.net/statistics-calculator/charts/violin-plot>
- [89] B. L. Ljubljana University of, « Orange Data Mining - undefined », Orange Data Mining. Consulté le: 10 avril 2025. [En ligne]. Disponible sur: <https://orangedatamining.com>
- [90] « An Introduction to SHAP Values and Machine Learning Interpretability ». Consulté le: 13 avril 2025. [En ligne]. Disponible sur: https://www.datacamp.com/tutorial/introduction-to-shap-values-machine-learning-interpretability?utm_source=google&utm_medium=paid_search&utm_campaignid=19589720821&utm_adgroupid=157156374951&utm_device=c&utm_keyword=&utm_matchtype=&utm_network=g&utm_adposition=&utm_creative=733936221293&utm_targetid=dsa-2218886984380&utm_loc_interest_ms=&utm_loc_physical_ms=9069683&utm_content=dsa~tofu~tutorial~machine-learning&accountid=9624585688&utm_campaign=230119_1-sea~dsa~tofu_2-b2c_3-emea_4-prc_5-na_6-na_7-le_8-pdsh-go_9-nb-e_10-na_11-na&gad_source=1&gclid=CjwKCAjwwe2_BhBEEiwAM1I7sdfE48AHU1284yvIBfjFs6Um_ThMFQasxhOCM_QyjsiEq6hi9XsxoCezoQAvD_BwE