

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE MINISTRE DE L'ENSEIGNEMENT
SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE

Université Echahid Hamma Lakhdar - El Oued



MASTER ACADÉMIQUE

MÉMOIRE



Faculté : Science Exacte

Département : Informatique

Spécialisation : Intelligence artificielle et science des données

Intitulé :

**Application des techniques d'intelligence
artificielle à l'analyse et la prédiction des
données climatiques : Cas de El-Oued**

Mémoire présenté en vue de l'obtention du diplôme de Master 2

Présenté par :

Guezei Tarek
Benali Mohamed El Fadhel

Encadrement:

Encadreur: Dr. Kara Mostefa
Co-encadreur: Dr.Boucherit Ammar

Soutenu le : 06/06/2026

Composés du jury

Mme. Belila Khaoula	Presidente	MCA	Université d'El-Oued
Mr. Mostefa Kara	Encadreur	MCB	Université d'El-Oued
Mr. Gherbi Kaddour	Examineur	MCA	Université d'El-Oued

Année Universitaire : 2025/2026

Dédicace

Je dédie ce modeste travail :

À mes chers parents,

pour leur amour, leur soutien, leurs sacrifices et leurs encouragements tout au long de mon parcours.

À ma famille (Ma femme et mes enfants),

pour leur présence, leur patience et leurs conseils précieux.

À mes enseignants et encadrateurs,

pour leur accompagnement, leurs orientations et les connaissances qu'ils nous ont transmises.

À mes amis et collègues,

pour leur soutien moral et les moments de motivation et d'entraide.

Enfin, à toute personne ayant contribué de près ou de loin à la réalisation de ce travail.

Tarek Guezei

Dédicace

Avec un immense plaisir, je dédie ce travail :

À mes très chers parents,
qui ont toujours été une source de soutien, de confiance et
d'encouragement. Que Dieu les protège et les récompense
pour tous leurs sacrifices.

À mes frères et sœurs,
pour leur affection, leur aide et leur présence dans les
moments difficiles.

À toute ma famille,
qui m'a soutenu durant tout mon parcours universitaire.

À mes enseignants et mon encadreur,
pour leurs précieux conseils, leur patience et leur
accompagnement scientifique.

À mes amis et collègues,
avec qui j'ai partagé de nombreux moments d'apprentissage
et de motivation.

Et à toutes les personnes qui ont contribué, de près ou de
loin, à la réalisation de ce mémoire.

Mohammed El Fadhel BenAli

Remerciements

Avant tout, nous remercions Dieu le Tout-Puissant de nous avoir donné la force, la patience et le courage pour réaliser ce travail.

Nous adressons nos sincères remerciements à notre encadreur Docteur Kara Mostefa pour son accompagnement, ses conseils précieux, sa disponibilité et son soutien tout au long de la réalisation de ce mémoire.

Nous remercions également tous les enseignants du département informatique pour la qualité de leur enseignement et les connaissances qu'ils nous ont transmises durant notre parcours universitaire.

Nos remerciements s'adressent aussi aux membres du jury qui ont accepté d'évaluer ce travail.

Nous exprimons notre profonde gratitude à nos parents et à nos familles pour leurs encouragements, leur patience et leur soutien moral tout au long de nos études.

Enfin, nous remercions toutes les personnes qui ont contribué de près ou de loin à la réalisation de ce travail.

Résumé

Le changement climatique constitue aujourd'hui l'un des défis environnementaux majeurs, particulièrement dans les régions arides où les écosystèmes et les activités humaines sont fortement sensibles aux variations climatiques. La région d'El-Oued, caractérisée par un climat désertique marqué par des températures élevées, une faible pluviométrie et une forte évaporation, représente un cas d'étude pertinent pour analyser les impacts du changement climatique à l'échelle locale.

L'objectif principal de ce travail est d'appliquer des techniques d'intelligence artificielle (IA) afin d'analyser les données climatiques historiques et de prédire les tendances climatiques futures dans cette région aride. L'étude repose sur l'exploitation de séries temporelles climatiques comprenant plusieurs variables météorologiques telles que la température, les précipitations, l'humidité et la vitesse du vent.

Après une phase de prétraitement, de nettoyage et d'analyse exploratoire des données, une approche basée sur le modèle Transformer est mise en œuvre pour la modélisation prédictive. Ce choix repose sur la capacité de cette architecture à capturer efficacement les dépendances temporelles à court et à long terme au sein des séries climatiques. Les performances du Transformer sont ensuite évaluées et comparées à celles d'autres modèles d'apprentissage automatique et d'apprentissage profond afin d'identifier l'approche la plus performante pour la prédiction climatique.

Mots-clés : Intelligence artificielle, changement climatique, région aride, prédiction climatique, apprentissage automatique, modèle Transformer, El-Oued.

Abstract

Climate change is currently one of the major environmental challenges, particularly in arid regions where ecosystems and human activities are highly sensitive to climate variations. The El-Oued region, characterized by a desert climate with high temperatures, low rainfall, and intense evaporation, represents a relevant case study for analyzing the impacts of climate change at the local scale.

The main objective of this work is to apply Artificial Intelligence (AI) techniques to analyze historical climate data and predict future climate trends in this arid region. The study is based on the exploitation of climate time series containing several meteorological variables, including temperature, precipitation, humidity, and wind speed.

After a phase of data preprocessing, cleaning, and exploratory data analysis, a predictive modeling approach based on the Transformer model is implemented. This choice is motivated by the ability of this architecture to effectively capture both short-term and long-term temporal dependencies within climate time series. The performance of the Transformer model is then evaluated and compared with other machine learning and deep learning models in order to identify the most effective approach for climate prediction.

Keywords: Artificial Intelligence, Climate Change, Arid Region, Climate Prediction, Machine Learning, Transformer Model, El-Oued.

الملخص

يُعدّ التغير المناخي في الوقت الحاضر أحد أبرز التحديات البيئية العالمية، لا سيما في المناطق الجافة التي تتأثر فيها النظم البيئية والأنشطة البشرية بشكل كبير بالتقلبات المناخية. وتمثل منطقة الوادي، التي تتميز بمناخ صحراوي يتسم بارتفاع درجات الحرارة، وقلة التساقطات المطرية، وارتفاع معدلات التبخر، نموذجًا مناسبًا لدراسة آثار التغير المناخي على المستوى المحلي.

يهدف هذا العمل أساسًا إلى تطبيق تقنيات الذكاء الاصطناعي لتحليل البيانات المناخية التاريخية والتنبؤ بالاتجاهات المناخية المستقبلية في هذه المنطقة الجافة. وتعتمد الدراسة على استغلال السلاسل الزمنية المناخية التي تتضمن عدة متغيرات أرضادية، من بينها درجة الحرارة، والتساقطات المطرية، والرطوبة، وسرعة الرياح.

بعد مرحلة المعالجة المسبقة للبيانات وتنظيفها وإجراء التحليل الاستكشافي لها، تم اعتماد مقارنة تنبؤية قائمة على نموذج Transformer. ويستند اختبار هذا النموذج إلى قدرته على التقاط العلاقات الزمنية قصيرة وطويلة المدى داخل السلاسل الزمنية المناخية بكفاءة عالية. بعد ذلك، تم تقييم أداء نموذج Transformer ومقارنته بنماذج أخرى من التعلم الآلي والتعلم العميق بهدف تحديد النهج الأكثر فعالية في التنبؤ بالمناخ.

الكلمات المفتاحية: الذكاء الاصطناعي، التغير المناخي، المناطق الجافة، التنبؤ المناخي، التعلم الآلي، نموذج Transformer، الوادي.

Table des matières

Introduction générale.....	12
Chapitre I : Intelligence Artificielle et Climatologie	13
1. Introduction.....	14
2. Définition de l'Intelligence Artificielle (IA)	14
3. Définition apprentissage automatique	15
4. Définition d'Apprentissage profond	15
5. Différence entre IA, apprentissage automatique et l'apprentissage profond	16
6. Applications de l'Intelligence Artificielle en climatologie	17
a) Préviation météorologique et climatique	17
b) Détection du changement climatique	17
c) Analyse des événements climatiques extrêmes	18
d) Gestion des ressources en eau	18
e) Agriculture intelligente (Smart Agriculture)	18
f) Aide à la prise de décision environnementale	18
7. Caractéristiques climatiques de la région d'El-Oued	19
8. Analyse des séries temporelles climatiques.....	21
a) Définition	21
b) Objectifs de l'analyse des séries temporelles climatiques.....	21
c) Méthodes d'analyse des séries temporelles.....	21
- Analyse graphique	21
- Analyse statistique	21
- Analyse de la décomposition	21
d) Applications dans l'étude du climat	22
e) Importance pour la prédiction climatique.....	22
9. Comparaison entre RNN, LSTM et Transformer.....	22
10. Etat de l'art	24
10.1 Approches statistiques traditionnelles	24
10.2 Méthodes de Machine Learning	24
10.3 Apprentissage profond pour les séries temporelles climatiques.....	25
10.4 Les architectures Transformer	25
10.5 Applications récentes du Transformer à la prévision climatique	26
10.6 Synthèse comparative	26
11. Justification du modèle	27

12.	Conclusion	28
CHAPITRE II : Collecte et prétraitement des données climatiques		29
1.	Introduction.....	30
2.	Sources des données climatiques	30
a)	Collecte des données climatiques depuis Ogimet	31
b)	Méthode d'Extraction des Données.	32
c)	Exemple d'Extraction des données depuis Ogimet	32
3.	Prétraitement des données climatiques	33
a)	Nettoyage des données	33
1.	Gestion des valeurs manquantes.....	33
2.	Détection des valeurs aberrantes (Outliers)	34
b)	Transformation des données.....	35
c)	Normalisation des données.....	35
4.	Ingénierie des fonctionnalités.....	36
5.	Structuration pour les modèles d'apprentissage profond	37
6.	Importance du prétraitement.....	38
7.	Analyse exploratoire des données (EDA)	38
a)	Analyse descriptive des variables.....	38
b)	Analyse des distributions.....	39
c)	Analyse des séries temporelles	39
d)	Analyse des corrélations.....	39
e)	Détection des anomalies	40
8.	Visualisation des données.....	40
9.	Conclusion.....	42
CHAPITRE III : Intelligence artificielle et modèle Transformer appliquer aux séries temporelles climatiques		43
1.	Introduction.....	44
2.	Intelligence artificielle et séries temporelles	44
a)	Définition des séries temporelles	44
b)	Apport de l'intelligence artificielle	46
3.	Limites des méthodes traditionnelles	47
4.	L'apprentissage profond pour les séries temporelles	49
5.	Limites des modèles RNN et LSTM	50
a)	Difficulté à capturer les dépendances à long terme	51

6.	Transition vers le Transformer	52
7.	Le modèle transformer	53
a)	Principe de fonctionnement.....	53
b)	Mécanisme d'attention (explication détaillée)	54
a)	Principe général de l'attention	54
b)	Formulation mathématique de l'attention	55
c)	Self-Attention, Multi-Head Attention et application aux séries temporelles	56
d)	Architecture du Transformer	57
8.	Adaptation aux séries temporelles	59
9.	Application du modèle Transformer aux séries temporelles climatiques.....	59
a)	Préparation des données	59
b)	Transformation en séquences temporelles.....	60
c)	Application du Transformer	60
d)	Fonction de perte et optimisation	61
e)	Exemple d'application (El-Oued)	61
f)	Résultats attendus.....	61
10.	Avantages de l'approche	61
11.	Conclusion.....	61
CHAPITRE IV : L'implémentation pratique du modèle pour la prédiction climatique		63
1.	Introduction.....	64
2.	Environnement de développement	64
3.	Description des données utilisées	65
4.	Préparation des données	65
a)	Extraction des données :.....	65
b)	Décodage et Nettoyage des données :	66
c)	Normalisations des données :	71
5.	Apprentissage profond	73
6.	Prédiction.....	74
7.	Résultats comparatifs entre différentes régions.....	76
8.	Conclusion.....	77
Conclusion générale		79
Bibliographie.....		80

Liste des Figures :

Figure 1 : Relation entre IA, Apprentissage automatique, Apprentissage profond et la prédiction climatique	16
Figure 2: Grand Erg Oriental avec El Oued en haut, au centre	19
Figure 3: Données climatiques mensuelles de la région d'El Oued (températures et précipitations) ..	20
Figure 4: Diagramme climatique de la région d'El Oued montrant les températures minimales, maximales et les précipitations mensuelles.....	20
Figure 5: Principe de la fenêtre glissante (sliding window)	37
Figure 6: Évolution de la température dans la région d'El Oued(2023).....	41
Figure 7: Distribution des précipitations dans la région d'El Oued (2023)	41
Figure 8: Corrélation entre variables climatiques.....	42
Figure 9: Évolution typique de la température à El Oued	45
Figure 10: Schéma simplifié d'un modèle d'intelligence artificielle pour la prédiction climatique	47
Figure 11: Architecture générale du modèle Transformer	58

Liste des Tableaux :

Tableau 1: Différences fondamentales entre IA, apprentissage automatique et l'apprentissage profond	16
Tableau 2: Applications de l'IA en climatologie	18
Tableau 3: Comparaison entre les modèles RNN, LSTM et Transformer appliqués aux séries climatiques	23
Tableau 4: Exemple de caractéristiques climatiques.....	31
Tableau 5: Exemple d'extraction des données climatiques de la station d'El-Oued (WMO : 60559) –	33
Tableau 6: Exemple des valeurs manquantes.....	34
Tableau 7: Résultat après les traitements.....	35
Tableau 8: Observation journalières de plusieurs caractéristiques.....	46
Tableau 9: Exemple de données météorologiques utilisées pour la prédiction de la température	47
Tableau 10: Exemple basé sur des données climatiques simples.....	48
Tableau 11: Exemple des valeurs aberrantes	48
Tableau 12: Exemple de série temporelle des températures	49
Tableau 13: Résultat des entrées de série temporelle des températures	50
Tableau 14: Exemple de prédiction après l'entraînement	50
Tableau 15: Exemple une série climatique contenant trois observations.....	56
Tableau 16: Poids d'attention pour les observations.....	56
Tableau 17: Estimation de la prochaine valeur	59
Tableau 18: Prédiction des valeurs de sortie	60
Tableau 19: Extrait des observations météorologiques brutes De la station météorologique 60559 ...	66
Tableau 20: Description des caractéristiques météorologiques utilisées.....	67
Tableau 21: Résumé de l'extraction des données SYNOP	67
Tableau 22: Exemple des données obtenues après décodage.....	69
Tableau 23: Résultat obtenue après l'exécution du code	73
Tableau 24: Comparaison des températures réelles et prédites par le modèle Transformer dans différentes régions algériennes.....	76

Introduction générale

Le changement climatique représente aujourd'hui un défi majeur, particulièrement dans les régions arides telles qu'El-Oued, où les conditions climatiques sont caractérisées par des températures élevées, une faible pluviométrie et une forte évaporation. Dans ce contexte, l'intelligence artificielle apparaît comme une solution prometteuse pour analyser les données climatiques et améliorer la précision des prévisions.

L'objectif de ce mémoire est d'appliquer des techniques d'intelligence artificielle, notamment le modèle Transformer, à l'analyse et à la prédiction des séries temporelles climatiques de la région d'El-Oued.

Le premier chapitre présente les concepts fondamentaux de l'intelligence artificielle, de l'apprentissage automatique, de l'apprentissage profond ainsi que leur application à la climatologie. Une comparaison entre les modèles RNN, LSTM et Transformer permet de justifier le choix de ce dernier.

Le deuxième chapitre est consacré à la collecte des données climatiques à partir de la plateforme Ogimet, ainsi qu'aux différentes étapes de prétraitement, incluant le nettoyage, la normalisation et l'analyse exploratoire des données.

Le troisième chapitre présente le modèle Transformer et son adaptation aux séries temporelles climatiques, en expliquant son architecture et son mécanisme d'attention permettant de capturer les dépendances temporelles complexes.

Enfin, le quatrième chapitre porte sur l'implémentation du modèle, l'évaluation de ses performances et l'analyse des résultats obtenus. Les prédictions réalisées démontrent l'efficacité du Transformer pour la prévision climatique dans la région d'El-Oued.

Cette étude met ainsi en évidence l'intérêt des techniques d'intelligence artificielle dans l'analyse et la prévision des phénomènes climatiques, offrant des perspectives prometteuses pour la gestion et l'adaptation aux changements climatiques dans les régions arides.

Chapitre I : Intelligence Artificielle et Climatologie

1. Introduction

Dans un contexte marqué par les changements climatiques croissants et leurs impacts significatifs sur les écosystèmes et les activités humaines, l'analyse et la prédiction des phénomènes climatiques deviennent des enjeux majeurs. Les méthodes traditionnelles, bien qu'utiles, montrent certaines limites face à la complexité, à la variabilité et au caractère non linéaire des données climatiques, en particulier dans les régions arides. (Chennouf, S. 2016)

Dans ce cadre, l'intelligence artificielle (IA) s'impose comme une approche innovante et performante, capable de traiter de grandes quantités de données et d'extraire des modèles complexes. À travers ses sous-domaines, notamment l'apprentissage automatique (Machine Learning) et l'apprentissage profond (Deep Learning), elle offre des outils puissants pour l'analyse des séries temporelles et la prédiction climatique.

Ce chapitre a pour objectif de présenter les concepts fondamentaux de l'intelligence artificielle et leur application dans le domaine de la climatologie. Il aborde successivement les définitions de l'IA, l'apprentissage automatique et l'apprentissage profond, leurs différences, ainsi que leurs principales applications en climatologie. Une attention particulière est également portée à l'analyse des séries temporelles climatiques et aux modèles utilisés pour la prédiction, notamment les réseaux de neurones récurrents (RNN), les modèles LSTM et les architectures Transformer.

Enfin, ce chapitre met en évidence les caractéristiques climatiques de la région d'El-Oued, qui constitue le cas d'étude de ce mémoire, afin de justifier l'utilisation des techniques d'intelligence artificielle dans un contexte aride et complexe.

2. Définition de l'Intelligence Artificielle (IA)

L'intelligence artificielle (IA) est un domaine de l'informatique qui vise à concevoir des systèmes capables de réaliser des tâches nécessitant normalement l'intelligence humaine, telles que l'apprentissage, le raisonnement, la prise de décision et la résolution de problèmes. Elle repose sur le développement d'algorithmes capables d'analyser des données, d'identifier des modèles et d'améliorer leurs performances automatiquement à partir de l'expérience (BERBIB, N., & Abderrazak, E. L. 2026).

Dans le contexte scientifique, l'IA permet de traiter de grandes quantités de données complexes afin d'extraire des informations utiles et de produire des prédictions fiables.

Appliquée aux sciences climatiques, elle facilite l'analyse des variables météorologiques et la modélisation des phénomènes climatiques non linéaires difficiles à représenter par les méthodes statistiques classiques (Guia et al. 2024).

Dans cette étude, l'intelligence artificielle est utilisée pour analyser les données climatiques historiques de la wilaya d'El-Oued et prédire l'évolution future des paramètres climatiques tels que la température et les précipitations (Sonhay, K. N. 2022).

3. Définition apprentissage automatique

L'apprentissage automatique est une sous-branche de l'intelligence artificielle qui permet aux systèmes informatiques d'apprendre automatiquement à partir des données sans être explicitement programmés pour chaque tâche. Les modèles apprennent des relations entre les variables d'entrée et les résultats attendus afin de réaliser des prédictions sur de nouvelles données.

Les principales catégories d'apprentissage automatique sont l'apprentissage supervisé, l'apprentissage non supervisé et l'apprentissage par renforcement.

Dans les études climatiques, l'apprentissage automatique est largement utilisé pour l'analyse des séries temporelles et la prévision météorologique.

4. Définition d'Apprentissage profond

L'apprentissage profond, ou apprentissage profond, est une branche avancée d'apprentissage automatique basée sur les réseaux de neurones artificiels comportant plusieurs couches cachées. Ces réseaux imitent le fonctionnement du cerveau humain en apprenant progressivement des représentations complexes des données (Bashardoustjoubjarkouli, F. 2025).

L'apprentissage profond est particulièrement efficace pour l'analyse de grandes bases de données, la détection de relations non linéaires complexes et la prédiction des séries temporelles.

Parmi les modèles les plus utilisés on trouve les réseaux de neurones artificiels (RNN), réseaux LSTM (Long Short-Term Memory), réseaux convolutifs (CNN) et les réseaux Transformer.

Dans le cadre de cette mémoire, les modèles d'apprentissage profond permettent d'améliorer la précision de la prédiction climatique en capturant les dépendances temporelles entre les variables météorologiques.

5. Différence entre IA, apprentissage automatique et l'apprentissage profond

Il est essentiel de distinguer clairement les notions d'intelligence artificielle (IA), d'apprentissage automatique et d'apprentissage profond qui sont souvent utilisées de manière interchangeable mais présentent des différences fondamentales (DE MATTEIS, L., Janny, S., Nathan, S., & Shu-Quartier, W. 2022).

Tableau 1: Différences fondamentales entre IA, apprentissage automatique et l'apprentissage profond

Concept	Définition	Rôle dans le mémoire
Intelligence Artificielle	Domaine global simulant l'intelligence humaine	Cadre général
Apprentissage automatique	Apprentissage à partir des données	Analyse climatique
Apprentissage profond	Réseaux neuronaux profonds	Prédiction avancée

Ainsi, l'intelligence artificielle, à travers l'apprentissage automatique et l'apprentissage profond, constitue un outil puissant pour l'analyse et la prédiction du changement climatique, notamment dans les régions arides où les phénomènes climatiques présentent une forte complexité et variabilité.

Pour mieux comprendre l'articulation entre l'intelligence artificielle, l'apprentissage automatique et l'apprentissage profond, ainsi que leur contribution à la prédiction climatique, le schéma suivant est présenté.

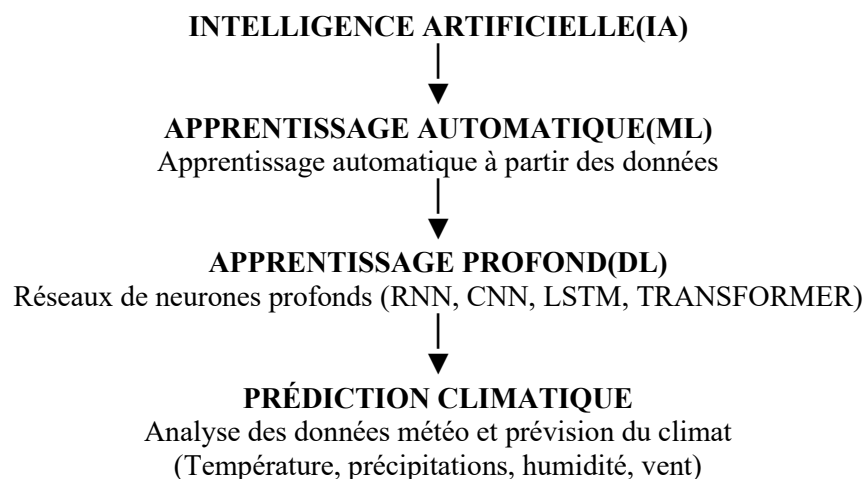


Figure 1 : Relation entre IA, Apprentissage automatique, Apprentissage profond et la prédiction climatique

Le schéma présenté illustre l'enchaînement logique des différentes approches de l'intelligence artificielle appliquées à la climatologie. Dans ce contexte, l'Intelligence Artificielle constitue le cadre global englobant les techniques d'analyse et de modélisation des données climatiques.

L'apprentissage automatique est utilisé pour exploiter les données météorologiques historiques (température, précipitations, humidité, vent) afin d'identifier des relations et des tendances cachées. Ces modèles permettent d'effectuer des analyses statistiques et de construire des modèles prédictifs de base.

L'apprentissage profond, permet d'aller plus loin en modélisant des relations complexes et non linéaires grâce à des réseaux de neurones profonds tels que les RNN, CNN, LSTM et TRANSFORMER. Ces modèles sont particulièrement adaptés aux séries temporelles climatiques, caractérisées par une forte variabilité.

Ainsi, l'intégration de ces techniques permet d'améliorer significativement la précision des prédictions climatiques, notamment dans les régions arides comme notre région (El Oued), où les conditions climatiques sont instables et difficiles à modéliser (Zeghina, A. O. 2024).

6. Applications de l'Intelligence Artificielle en climatologie

L'intelligence artificielle joue un rôle essentiel dans la climatologie grâce à sa capacité à analyser de grandes quantités de données environnementales et à modéliser des phénomènes climatiques complexes (Balogh, B. 2022).

a) Prévision météorologique et climatique

L'une des applications majeures de l'IA consiste à améliorer les modèles de prévision du climat et de la météo. Les algorithmes d'apprentissage automatique analysent les données historiques telles que la température, les précipitations, l'humidité, la pression atmosphérique et la vitesse du vent.

b) Détection du changement climatique

L'IA permet d'identifier automatiquement les tendances climatiques à long terme en analysant plusieurs décennies de données météorologiques, elle aide à détecter l'augmentation progressive des températures, analyse les anomalies climatiques et mesure l'évolution du réchauffement régional.

Cette capacité est essentielle pour comprendre l'impact du changement climatique dans les régions arides comme la wilaya d'El-Oued.

c) Analyse des événements climatiques extrêmes

Les techniques d'intelligence artificielle sont utilisées pour prévoir et analyser les vagues de chaleur, les sécheresses, les tempêtes de sable et les pluies intenses et inondations soudaines.

La prévision anticipée de ces événements extrêmes contribue significativement à la réduction des risques pour les populations ainsi que pour l'économie locale.

d) Gestion des ressources en eau

Dans les zones arides, l'IA contribue à une meilleure gestion hydrique grâce à la prévision des périodes de sécheresse, l'optimisation de l'irrigation et l'estimation de la disponibilité future des ressources en eau.

Ces applications sont particulièrement importantes pour l'agriculture saharienne.

e) Agriculture intelligente (Smart Agriculture)

L'IA combine les données climatiques avec les données agricoles afin de déterminer les périodes optimales de plantation, de prévoir les rendements agricoles et de réduire les pertes liées aux conditions climatiques.

Les agriculteurs peuvent ainsi prendre des décisions basées sur des prévisions scientifiques fiables.

f) Aide à la prise de décision environnementale

Les modèles d'IA fournissent des outils pour soutenir les politiques climatiques, planifier l'adaptation au changement climatique et améliorer la gestion territoriale.

Tableau 2: Applications de l'IA en climatologie

Application	Objectif	Méthode IA
Prévision climatique	Température, précipitations	LSTM, CNN, TRANSFORMER
Détection du changement climatique	Identifier tendances	Apprentissage automatique supervisé
Événements extrêmes	Vagues de chaleur, sécheresses	Apprentissage profond, séries temporelles
Gestion de l'eau	Optimisation irrigation	Réseaux neuronaux, ML
Agriculture intelligente	Rendement agricole	Apprentissage profond, Méga données
Politique climatique	Aide à la décision	Analyse prédictive

7. Caractéristiques climatiques de la région d'El-Oued

La région d'El-Oued, située dans le sud-est de l'Algérie, se caractérise par un climat désertique aride, typique du Sahara. Elle est marquée par des températures élevées durant la majeure partie de l'année, pouvant dépasser 45°C en été, ainsi que par une faible pluviométrie annuelle, généralement inférieure à 100 mm.

Les précipitations y sont rares, irrégulières et souvent concentrées sur de courtes périodes, ce qui accentue les phénomènes de sécheresse. L'évaporation est très élevée en raison de l'intensité du rayonnement solaire et des vents fréquents, contribuant à un déficit hydrique important.

Par ailleurs, la région connaît des variations thermiques significatives entre le jour et la nuit, ainsi qu'entre les saisons. Les vents de sable et les tempêtes (sirocco) sont également fréquents, influençant les conditions climatiques locales et les activités humaines, notamment agricoles.

Ces caractéristiques rendent la région d'El-Oued particulièrement vulnérable aux effets du changement climatique, ce qui justifie l'intérêt de recourir à des modèles d'intelligence artificielle pour analyser et prévoir l'évolution des paramètres climatiques dans cette zone (BOUKHEZNA, B. 2008).

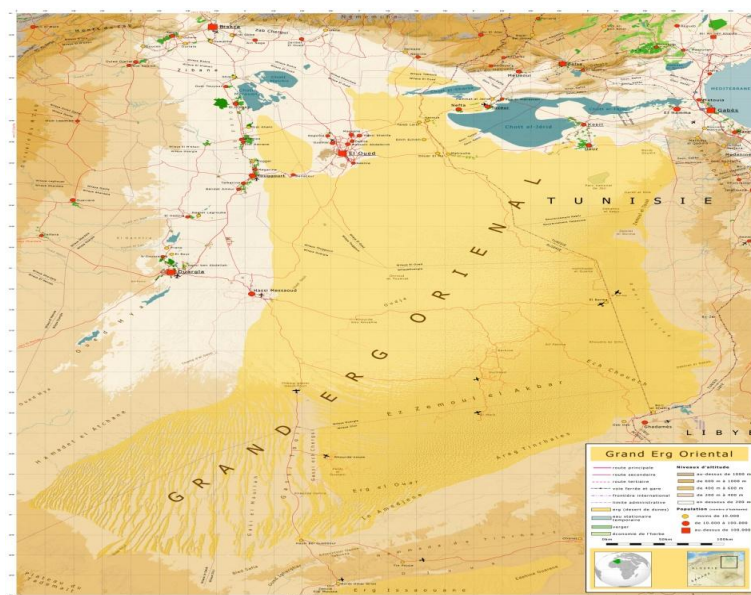


Figure 2: Grand Erg Oriental avec El Oued en haut, au centre

Le climat à El Oued, est désertique et très sec. La classification de Köppen est de type BWh. La température moyenne est de 21.8 °C et la moyenne des précipitations annuelles ne dépasse pas 100 mm.¹

¹ • Office National de la Météorologie (ONM). (2022). *Données climatologiques de l'Algérie*.

Mois	jan.	fév.	mars	avril	mai	juin	juil.	août	sep.	oct.	nov.	déc.	année
Température maximale moyenne (°C)	17	19,6	23	28	33	37,5	41,5	40,4	36	29	22	17	
Température moyenne (°C)	10	13	16	20,7	25	30,2	33,4	32,6	29	22	15,9	11	21,8
Température minimale moyenne (°C)	4	6	9	13	18	22,9	25	24	22	16	9	5	
Précipitations (mm)	8	8	11	4	7	2	0	1	5	8	11	8	74

Figure 3: Données climatiques mensuelles de la région d'El Oued (températures et précipitations)

L'analyse des données climatiques mensuelles d'El Oued met en évidence un régime thermique fortement contrasté et une pluviométrie très faible. Les températures maximales moyennes atteignent leur pic durant l'été, notamment en juillet (**41,5 °C**) et en août (**40,4 °C**), tandis que les températures minimales moyennes restent relativement douces en hiver, avec **4 °C en janvier**. La température moyenne annuelle est estimée à **21,8 °C**, confirmant la dominance de conditions chaudes. En ce qui concerne les précipitations, celles-ci sont extrêmement faibles et irrégulières, avec un cumul annuel de seulement **74 mm**, et une quasi-absence de pluie en été (0 mm en juillet). Ces caractéristiques confirment le classement d'El Oued dans la catégorie climatique désertique chaud (**BWh**) selon Köppen (Office National de la Météorologie, 2022).

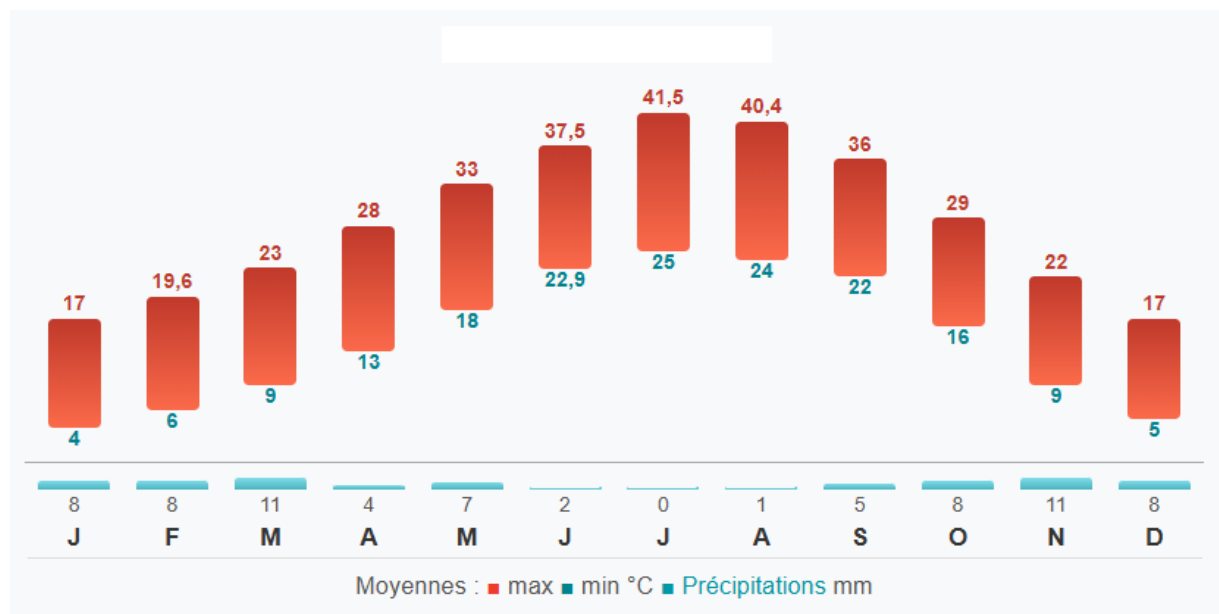


Figure 4: Diagramme climatique de la région d'El Oued montrant les températures minimales, maximales et les précipitations mensuelles

8. Analyse des séries temporelles climatiques

a) Définition

L'analyse des séries temporelles climatiques consiste à étudier l'évolution des variables météorologiques au cours du temps afin d'identifier des tendances, des variations saisonnières et des anomalies climatiques. Ces séries sont généralement composées d'observations collectées à intervalles réguliers, telles que les données journalières, mensuelles ou annuelles de température, de précipitations, d'humidité et de vitesse du vent.

L'objectif principal de cette analyse est de comprendre la dynamique du climat et de détecter les changements climatiques à long terme (Haddad, M. 2017).

b) Objectifs de l'analyse des séries temporelles climatiques

L'analyse des séries temporelles climatiques permet d'identifier les tendances climatiques à long terme, d'analyser les variations saisonnières des paramètres météorologiques, de détecter les anomalies ou les événements climatiques extrêmes, de comprendre les relations entre différentes variables climatiques, et d'améliorer la précision des modèles de prévision climatique.

Cette analyse constitue une étape essentielle avant l'application des modèles d'intelligence artificielle.

c) Méthodes d'analyse des séries temporelles

Plusieurs méthodes sont utilisées pour analyser les séries temporelles climatiques :

- Analyse graphique

La visualisation des données permet d'observer l'évolution des variables climatiques dans le temps. Les Figures de séries temporelles facilitent l'identification des tendances et des variations saisonnières.

- Analyse statistique

Les méthodes statistiques permettent de calculer plusieurs indicateurs importants (la moyenne, la variance, l'écart type, la corrélation entre les variables climatiques), ces indicateurs aident à mieux comprendre la structure des données climatiques.

- Analyse de la décomposition

La décomposition des séries temporelles consiste à séparer la série en plusieurs composantes (la tendance, la saisonnalité, le cycle, les fluctuations aléatoires), cette méthode permet d'analyser chaque composante séparément (Mélard, G. 2006).

d) Applications dans l'étude du climat

Dans le domaine de la climatologie, l'analyse des séries temporelles permet d'étudier l'évolution des températures au cours des années, d'analyser les variations des précipitations, de détecter les périodes de sécheresse ou de fortes pluies, et de prévoir les conditions climatiques futures.

Dans cette étude, les séries temporelles climatiques de la wilaya d'El-Oued sont analysées afin d'identifier les tendances climatiques et de développer des modèles de prédiction basés sur l'intelligence artificielle.

e) Importance pour la prédiction climatique

L'analyse des séries temporelles constitue une étape essentielle dans le développement des modèles de prévision climatique. Elle permet non seulement de structurer et de préparer les données, mais aussi d'identifier les dynamiques temporelles et les relations existantes entre les différentes variables climatiques.

Dans ce contexte, les modèles d'intelligence artificielle jouent un rôle central. En particulier, les réseaux de neurones récurrents (RNN) et les modèles à mémoire à long et court terme (LSTM) se distinguent par leur capacité à modéliser des dépendances temporelles complexes. Plus récemment, les modèles basés sur l'architecture Transformers ont montré des performances remarquables dans l'analyse des séries temporelles. Grâce au mécanisme d'attention, ces modèles peuvent capturer efficacement les relations à long terme entre les observations sans dépendre d'un traitement séquentiel strict, ce qui améliore leur capacité à traiter de grandes quantités de données climatiques.

Ainsi, ces différentes approches permettent de mieux comprendre les interactions temporelles et d'améliorer la précision des modèles de prévision climatique.

9. Comparaison entre RNN, LSTM et Transformer

Dans le domaine de l'analyse des séries temporelles climatiques, plusieurs modèles d'intelligence artificielle sont utilisés, notamment les RNN, les LSTM et les Transformer chacun présentant des caractéristiques spécifiques.

Les réseaux de neurones récurrents (RNN) sont conçus pour traiter des données séquentielles en tenant compte des dépendances temporelles. Cependant, ils présentent des limites lorsqu'il s'agit de capturer des dépendances à long terme, en raison du problème de disparition du gradient.

Pour pallier cette limitation, les modèles LSTM ont été introduits. Grâce à leur architecture basée sur des cellules de mémoire et des mécanismes de portes (gâtes), ils permettent de mieux conserver l'information sur de longues périodes, ce qui les rend plus performants pour les séries temporelles complexes, notamment en climatologie.

Plus récemment, les modèles basés sur l'architecture Transformer ont révolutionné l'analyse des séries temporelles. Contrairement aux RNN et LSTM, ils ne reposent pas sur un traitement séquentiel, mais utilisent un mécanisme d'attention (self-attention) qui permet de capturer efficacement les dépendances à court et à long terme en parallèle. Cette capacité améliore significativement les performances sur de grands ensembles de données climatiques et réduit le temps de calcul.

Ainsi, le choix du modèle dépend de la nature des données et des objectifs de l'étude : les RNN pour des structures simples, les LSTM pour des dépendances temporelles longues, et les Transformer pour des analyses complexes à grande échelle (Ali, Z. 2024).

Tableau 3: Comparaison entre les modèles RNN, LSTM et Transformer appliqués aux séries climatiques

Modèle	Principe	Avantages	Limites	Utilisation en climatologie
RNN (réseaux de neurones récurrents)	Traitement séquentiel des données avec prise en compte des états précédents	Simple à implémenter, adapté aux séries temporelles courtes	Problème de disparition/explosion du gradient, difficulté à capturer les dépendances à long terme	Analyse de séries climatiques simples et prévisions à court terme
LSTM (Long Short-Term Memory)	Architecture basée sur des cellules de mémoire avec mécanismes de portes (input, forget, output)	Capacité à capturer les dépendances à long terme, meilleure stabilité que les RNN	Complexité plus élevée, temps de calcul plus important	Prévision climatique avec dépendances temporelles complexes (température, précipitations)
Transformer	Basé sur le mécanisme d'attention (self-attention) avec traitement parallèle des séquences	Capture efficace des dépendances globales, haute performance sur de grands volumes de données	Nécessite des ressources de calcul importantes et de grandes quantités de données	Modélisation avancée et prévision climatique à grande échelle

10. Etat de l'art

La prévision climatique constitue un domaine de recherche majeur en raison de son importance dans de nombreux secteurs tels que l'agriculture, la gestion des ressources en eau, l'énergie, la santé publique et la prévention des catastrophes naturelles. L'augmentation des phénomènes météorologiques extrêmes liée au changement climatique a renforcé la nécessité de développer des modèles de prévision de plus en plus précis et fiables. Au fil des années, les méthodes utilisées pour l'analyse et la prédiction des données climatiques ont considérablement évolué, passant des approches statistiques traditionnelles aux techniques avancées d'intelligence artificielle et d'apprentissage profond.

10.1 Approches statistiques traditionnelles

Les premières méthodes de prévision météorologique reposaient principalement sur des techniques statistiques permettant d'identifier les tendances et les comportements des séries temporelles climatiques. Parmi les modèles les plus utilisés figurent la régression linéaire et les modèles ARIMA (AutoRegressive Integrated Moving Average) (Box et al., 2015).

La régression linéaire permet d'établir une relation mathématique entre une variable climatique à prédire et plusieurs variables explicatives. Bien que simple à mettre en œuvre, cette méthode est limitée lorsqu'il s'agit de modéliser des phénomènes complexes et non linéaires.

Les modèles ARIMA ont ensuite été largement utilisés pour l'analyse des séries temporelles météorologiques. Ils permettent de prendre en compte les dépendances temporelles entre les observations successives. Cependant, ces modèles supposent généralement une certaine stationnarité des données et rencontrent des difficultés lorsque les relations deviennent fortement non linéaires ou lorsque plusieurs variables climatiques interagissent simultanément.

10.2 Méthodes de Machine Learning

Avec l'augmentation de la quantité de données météorologiques disponibles et l'amélioration des capacités de calcul, les techniques de Machine Learning ont progressivement remplacé les approches statistiques classiques dans de nombreuses applications de prévision climatique (Kara et al. 2022 ; Kara et al. 2023).

Parmi les algorithmes les plus utilisés figurent les arbres de décision, les Random Forest et les Support Vector Machines (SVM). Le modèle Random Forest proposé par Breiman (2001) repose sur un ensemble d'arbres de décision permettant d'améliorer la robustesse

des prédictions et de réduire le risque de surapprentissage. Cette approche a montré de bonnes performances pour l'estimation des températures et des précipitations.

Les SVM, introduits par Vapnik (1995), sont également largement utilisés dans les tâches de classification et de régression météorologique. Ils sont capables de traiter des relations non linéaires grâce à l'utilisation de fonctions noyaux (kernels). Toutefois, leurs performances diminuent lorsque le volume de données devient très important.

Malgré leurs avantages, ces méthodes nécessitent souvent une étape d'extraction manuelle des caractéristiques pertinentes, ce qui peut limiter leur capacité à exploiter pleinement la richesse des données climatiques.

10.3 Apprentissage profond pour les séries temporelles climatiques

L'émergence de l'apprentissage profond (Deep Learning) a marqué une avancée importante dans le domaine de la prévision météorologique. Contrairement aux méthodes traditionnelles, les réseaux de neurones profonds sont capables d'apprendre automatiquement des représentations complexes à partir des données brutes.

Les réseaux de neurones artificiels (ANN) ont été parmi les premiers modèles utilisés pour la prévision climatique. Grâce à leurs couches cachées, ils peuvent modéliser des relations non linéaires entre les différentes variables météorologiques. Cependant, ils ne prennent pas explicitement en compte l'ordre temporel des observations (Miloudi et al. 2026).

Pour répondre à cette limitation, les chercheurs ont développé les réseaux de neurones récurrents (RNN). Ces modèles introduisent une mémoire interne permettant de conserver l'information des observations précédentes. Les RNN ont ainsi montré des résultats prometteurs dans la prévision des températures, de l'humidité et des précipitations.

Néanmoins, les RNN souffrent du problème de disparition du gradient (Vanishing Gradient), qui limite leur capacité à apprendre les dépendances à long terme.

Afin de surmonter cette difficulté, Hochreiter et Schmidhuber (1997) ont proposé l'architecture Long Short-Term Memory (LSTM). Grâce à un mécanisme de portes de mémoire, les LSTM peuvent conserver l'information pertinente sur de longues périodes. Plusieurs études ont démontré leur efficacité pour la prévision météorologique à court et moyen terme (Nedioui et al. 2024).

10.4 Les architectures Transformer

L'introduction du modèle Transformer par Vaswani et al. (2017) a constitué une révolution dans le domaine de l'apprentissage profond. Initialement conçu pour le traitement

automatique du langage naturel, ce modèle a rapidement été adapté à l'analyse des séries temporelles (Berhoum et al. 2024).

La principale innovation du Transformer réside dans le mécanisme d'attention (Attention Mechanism), qui permet au modèle d'évaluer simultanément l'importance de toutes les observations d'une séquence. Contrairement aux RNN et aux LSTM, le Transformer ne traite pas les données de manière séquentielle mais de façon parallèle, ce qui réduit considérablement le temps d'apprentissage.

Le mécanisme d'attention permet également de capturer efficacement les dépendances à long terme entre les observations, un aspect particulièrement important dans les séries climatiques où les conditions météorologiques actuelles peuvent être influencées par des événements survenus plusieurs jours, semaines ou mois auparavant.

Les performances remarquables du Transformer ont conduit au développement de plusieurs variantes spécialisées dans les séries temporelles, notamment Informer (Wu et al., 2020), Autoformer (Zhou et al., 2021) et FEDformer (Zhou et al., 2022). Ces architectures ont obtenu des résultats supérieurs à ceux des modèles LSTM dans de nombreuses tâches de prévision à long terme.

10.5 Applications récentes du Transformer à la prévision climatique

Les travaux récents montrent que les architectures Transformer sont particulièrement adaptées à l'analyse des données météorologiques multivariées. Elles permettent d'intégrer simultanément plusieurs variables climatiques telles que la température, l'humidité, la pression atmosphérique, la vitesse du vent et les précipitations.

Plusieurs études ont démontré que les modèles Transformer améliorent significativement la précision des prévisions de température journalière, des précipitations et des événements météorologiques extrêmes. Leur capacité à gérer de grands volumes de données et à identifier des relations complexes entre les variables climatiques constitue un avantage majeur pour les applications météorologiques modernes.

10.6 Synthèse comparative

L'analyse de la littérature montre une évolution progressive des méthodes de prévision climatique. Les modèles statistiques traditionnels présentent une simplicité d'utilisation mais restent limités face à la complexité des phénomènes climatiques. Les techniques de Machine Learning améliorent la modélisation des relations non linéaires mais nécessitent souvent une ingénierie des caractéristiques. Les modèles d'apprentissage profond, notamment les LSTM, offrent de meilleures performances grâce à leur capacité

à exploiter les dépendances temporelles. Cependant, les architectures Transformer apparaissent aujourd'hui comme les modèles les plus performants pour la prévision des séries temporelles climatiques, grâce à leur mécanisme d'attention et à leur aptitude à capturer efficacement les dépendances à long terme.

11. Justification du modèle

Dans le cadre de cette étude portant sur l'analyse et la prévision des données climatiques dans la région d'El-Oued, le choix du modèle constitue une étape déterminante afin de garantir des résultats précis et fiables. Les séries temporelles climatiques se caractérisent par leur complexité, leur non-linéarité ainsi que par la présence de dépendances temporelles à court et à long terme, ce qui nécessite l'utilisation de modèles capables de capturer ces dynamiques.

Les modèles traditionnels, tels que les réseaux de neurones récurrents (RNN), présentent certaines limites, notamment leur difficulté à modéliser les dépendances à long terme en raison du problème de disparition du gradient. Bien que les modèles LSTM aient permis d'améliorer cette limitation grâce à leur architecture basée sur des cellules de mémoire, ils restent contraints par leur nature séquentielle, ce qui peut entraîner des temps de calcul élevés, en particulier lors du traitement de grandes quantités de données climatiques.

Dans ce contexte, le choix s'oriente vers les modèles basés sur l'architecture Transformer. Ces modèles reposent sur un mécanisme d'attention (self-attention) qui permet de capturer efficacement les relations entre différentes observations, indépendamment de leur position dans la séquence temporelle. Contrairement aux modèles RNN et LSTM, les Transformers permettent un traitement parallèle des données, ce qui améliore significativement l'efficacité computationnelle et réduit le temps d'entraînement.

Par ailleurs, les données climatiques de la région d'El-Oued présentent souvent des variations saisonnières marquées ainsi que des dépendances complexes liées aux conditions arides. Les modèles Transformer sont particulièrement adaptés à ce type de données, car ils peuvent modéliser simultanément les interactions locales (à court terme) et globales (à long terme), tout en s'adaptant à des ensembles de données volumineux.

Ainsi, le choix du modèle Transformer s'impose comme une solution optimale pour cette étude, en raison de sa robustesse, de sa capacité de généralisation et de ses performances supérieures dans l'analyse des séries temporelles climatiques complexes. Ce choix s'inscrit donc dans une approche moderne et adaptée aux exigences de l'analyse climatique dans des environnements arides tels que celui d'El-Oued.

Des travaux récents ont démontré la supériorité des modèles Transformer dans le traitement des séries temporelles complexes, notamment grâce à leur mécanisme d'attention (Vaswani et al., 2017 ; Lim et al., 2021).

12. Conclusion

Ce premier chapitre a permis de poser les bases théoriques nécessaires à la compréhension de l'utilisation de l'intelligence artificielle dans le domaine de la climatologie. À travers la présentation des concepts fondamentaux, notamment l'intelligence artificielle, l'Apprentissage automatique et l'apprentissage profond, il a été mis en évidence le rôle essentiel de ces technologies dans l'analyse de données complexes et la prédiction des phénomènes climatiques.

L'étude des séries temporelles climatiques a également montré leur importance dans l'identification des tendances, des variations saisonnières et des anomalies, constituant ainsi une étape indispensable dans la modélisation prédictive. Par ailleurs, la comparaison entre les différents modèles, tels que les RNN, les LSTM et les Transformer, a permis de mieux comprendre leurs avantages et leurs limites respectives.

Enfin, le choix du modèle Transformer a été justifié par sa capacité à capturer efficacement les dépendances temporelles complexes et à traiter de grandes quantités de données, ce qui en fait un outil particulièrement adapté à l'analyse des données climatiques dans une région aride comme El-Oued.

Ce cadre théorique constitue ainsi une base solide pour aborder le chapitre suivant, qui sera consacré à la collecte, au prétraitement et à l'analyse des données climatiques utilisées dans cette étude.

CHAPITRE II :

Collecte et prétraitement des données climatiques

1. Introduction

Dans le cadre des études basées sur l'intelligence artificielle appliquée à la climatologie, la qualité et la fiabilité des données constituent des éléments fondamentaux pour garantir la pertinence et la précision des résultats. En effet, les modèles de prédiction, en particulier ceux basés sur l'apprentissage automatique (Machine Learning) et l'apprentissage profond (Deep Learning), dépendent fortement de la nature des données utilisées, de leur volume ainsi que de leur niveau de préparation.

Les données climatiques présentent généralement des caractéristiques spécifiques telles que la variabilité temporelle, la présence de tendances saisonnières, ainsi que des anomalies liées à des événements extrêmes. De plus, ces données peuvent être affectées par des problèmes fréquents tels que les valeurs manquantes, les erreurs de mesure ou encore l'hétérogénéité des formats, ce qui rend leur exploitation directe difficile sans un traitement préalable rigoureux.

Ainsi, ce chapitre a pour objectif de décrire de manière détaillée les différentes étapes de collecte, de traitement et de préparation des données climatiques utilisées dans cette étude portant sur la région d'El-Oued. Dans un premier temps, les sources de données seront présentées, en mettant l'accent sur leur fiabilité et leur pertinence. Ensuite, les caractéristiques du jeu de données seront décrites, notamment en termes de variables, de période d'étude et de structure des séries temporelles.

Par la suite, une attention particulière sera accordée aux techniques de prétraitement des données, incluant le nettoyage, la gestion des valeurs manquantes, la normalisation et la transformation des données. Ces étapes sont essentielles pour améliorer la qualité des données et optimiser les performances des modèles d'intelligence artificielle.

Enfin, une analyse exploratoire approfondie des données sera réalisée afin d'identifier les tendances climatiques, les relations entre variables et les éventuelles anomalies. Ces résultats permettront d'orienter les choix méthodologiques et serviront de base pour la modélisation prédictive à l'aide du modèle Transformer, capable de capturer les dépendances temporelles complexes (Balogh, B. 2022 ; Kebache et al. 2023).

2. Sources des données climatiques

Les données climatiques utilisées dans cette étude ont été collectées à partir de la plateforme Ogimet, qui fournit des données météorologiques issues des stations SYNOP et

METAR à l'échelle mondiale. La station météorologique utilisée correspond à la station d'El-Oued (code WMO : 60559)².

Ces données comprennent des observations journalières de plusieurs variables climatiques telles que la température, les précipitations, l'humidité et la vitesse du vent. Elles sont enregistrées de manière régulière, permettant ainsi la construction de séries temporelles fiables pour l'analyse et la modélisation.

La station d'El-Oued est située à une latitude de 33.50°N et une longitude de 6.12°E, avec une altitude d'environ 63 mètres, et dispose d'archives climatologiques remontant à plusieurs décennies, ce qui permet d'effectuer des analyses climatiques à long terme

a) Collecte des données climatiques depuis Ogimet

Les données utilisées dans cette étude ont été extraites de la base Ogimet (www.ogimet.com). Le processus de collecte et de préparation des données comprend plusieurs étapes essentielles, à savoir : la collecte des données (choix du format, des variables et de la période d'étude), le nettoyage des données (gestion des valeurs manquantes et des valeurs aberrantes), le prétraitement (normalisation et création de nouvelles variables), ainsi que la structuration des données sous forme de séries temporelles.

Ogimet est une plateforme qui agrège les données météorologiques SYNOP (Surface Synoptic Observations) fournies par l'Organisation Météorologique Mondiale (OMM). Elle présente plusieurs avantages, notamment l'accès gratuit aux données horaires et journalières, une couverture mondiale incluant les stations algériennes, ainsi qu'un format standardisé basé sur les codes METAR et SYNOP. Toutefois, ces données peuvent présenter certaines limites, telles que des lacunes ou des erreurs de saisie, ce qui rend nécessaire un prétraitement rigoureux avant toute analyse ou modélisation.

Tableau 4: Exemple de caractéristiques climatiques

Caractéristiques	Unité	Description	Exemple (Ogimet)
Date/Heure	UTC	Horodatage de l'observation	2023-01-01 12:00
Température (T)	°C	Température de l'air à 2 m	15.2
Température minimale (Tn)	°C	Température minimale sur 24 h	8.5
Température maximale (Tx)	°C	Température maximale sur 24 h	22.1
Humidité (U)	%	Humidité relative de l'air	45
Pression (P)	hPa	Pression atmosphérique réduite au niveau de la mer	1016.2
Précipitations (RR)	mm	Cumul des précipitations (1 h ou 24 h)	0.0
Vent (ff)	m/s	Vitesse moyenne du vent	3.2
Direction du vent (dd)	°	Direction du vent (0–360°)	270 (Ouest)
Ensoleillement (SS)	h	Durée d'ensoleillement	8.5

² <http://www.ogimet.com> (vue le 30 Avril 2026)

b) Méthode d'Extraction des Données.

L'extraction des données climatiques depuis la plateforme Ogimet suit une démarche structurée visant à constituer un jeu de données exploitable pour l'analyse et la modélisation. Elle commence par l'accès au site officiel d'Ogimet, notamment à la section dédiée aux données climatologiques (Climatology / SYNOP), puis par la sélection de la station météorologique d'El-Oued à l'aide de son code WMO (60559), garantissant une identification précise des observations. Ensuite, une période d'étude est définie (mensuelle, annuelle ou pluriannuelle) afin d'assurer la représentativité statistique des données. Les informations sont extraites sous forme de tableaux SYNOP, disponibles en résolution horaire ou journalière, puis récupérées soit par copie manuelle, soit par extraction automatique via des techniques de web scraping. Les données obtenues sont ensuite structurées sous forme de tableau, où les lignes représentent les observations temporelles et les colonnes correspondent aux variables climatiques. Enfin, une étape de vérification de l'intégrité des données est effectuée afin de détecter d'éventuelles valeurs manquantes, doublons ou incohérences, garantissant ainsi la qualité du jeu de données avant toute analyse (Bouaoune, D., & Dahmani-Megrerouche, M. 2010).

c) Exemple d'Extraction des données depuis Ogimet

L'accès aux données climatiques depuis la plateforme Ogimet se fait à travers une requête http structurée, par exemple :

[http://www.ogimet.com/cgi-bin/climatologysynop?lang=en&ind=60559&ano=2023
&mes=01&day=01&ndays=31,](http://www.ogimet.com/cgi-bin/climatologysynop?lang=en&ind=60559&ano=2023&mes=01&day=01&ndays=31)

Permettant d'extraire les observations de la station d'El-Oued (WMO : 60559) pour le mois de janvier 2023. Après consultation de la page générée, les données sont présentées sous forme de tableaux SYNOP contenant plusieurs variables climatiques essentielles, notamment la température (°C), l'humidité relative (%), la pression atmosphérique (hPa), la vitesse du vent (m/s) et les précipitations (mm). Ces observations sont généralement disponibles à des pas de temps horaires, ce qui permet une analyse fine de la variabilité climatique. À titre d'exemple, un extrait des données montre des valeurs enregistrées le 01/01/2023 à 00 h, les données sont affichées sous forme de tableau contenant plusieurs variables climatiques. Un extrait des données obtenues est présenté ci-dessous.

Tableau 5: Exemple d'extraction des données climatiques de la station d'El-Oued (WMO : 60559) –

Janvier 2023

Date	Température (°C)	Humidité (%)	Pression (hPa)	Vent (m/s)	Précipitations (mm)
01/01/2023 00h	9.8	62	1018.4	2.1	0.0
01/01/2023 06h	8.5	70	1019.1	1.8	0.0
01/01/2023 12h	15.2	45	1016.2	3.2	0.0
01/01/2023 18h	12.4	55	1017.0	2.5	0.0

Cet extrait montre que les observations sont enregistrées à différents intervalles de la journée, ce qui permet de constituer une série temporelle fine.

Les données brutes extraites sont ensuite exportées vers un fichier CSV afin d'être nettoyées et préparées pour les étapes de prétraitement et de modélisation.

3. Prétraitement des données climatiques

Le prétraitement des données constitue une étape essentielle dans toute étude basée sur l'intelligence artificielle, en particulier lorsqu'il s'agit de séries temporelles climatiques. En effet, les données brutes extraites depuis la plateforme Ogimet contiennent souvent des imperfections telles que des valeurs manquantes, des erreurs de mesure ou des incohérences, ce qui peut affecter négativement les performances des modèles prédictifs.

L'objectif principal du prétraitement est donc d'améliorer la qualité des données afin de les rendre exploitables et adaptées aux algorithmes d'apprentissage automatique (Bauer, P., Morgan, M. C., & Sbill, S., 2019).

a) Nettoyage des données

Le nettoyage des données consiste à détecter et corriger les anomalies présentes dans le dataset.

1. Gestion des valeurs manquantes

La gestion des valeurs manquantes constitue une étape essentielle dans le prétraitement des données climatiques, car celles-ci peuvent présenter des absences dues à des pannes de capteurs, des erreurs d'enregistrement ou des interruptions de mesure. Afin de garantir la qualité et la fiabilité des analyses ultérieures, plusieurs méthodes sont généralement adoptées. Dans un

premier temps, les lignes contenant un pourcentage élevé de valeurs manquantes peuvent être supprimées lorsque leur impact sur la continuité des séries est jugé significatif. Ensuite, pour préserver la structure temporelle des données, l'interpolation linéaire est fréquemment utilisée, notamment dans le cas des séries chronologiques continues, afin d'estimer les valeurs intermédiaires manquantes à partir des observations voisines.

Enfin, pour certaines variables climatiques, le remplacement par des mesures statistiques telles que la moyenne ou la médiane peut être appliqué afin de limiter les biais introduits par les données absentes. Cette combinaison de méthodes permet d'assurer une meilleure cohérence des jeux de données et d'améliorer la robustesse des modèles d'analyse et de prédiction.

2. Détection des valeurs aberrantes (Outliers)

La détection des valeurs aberrantes constitue une étape importante dans le traitement des données climatiques, car les valeurs extrêmes peuvent influencer négativement la qualité des analyses et biaiser les résultats des modèles. Pour identifier ces anomalies, plusieurs approches statistiques sont couramment utilisées, notamment la méthode basée sur l'écart-type et le score Z, qui permet de repérer les observations s'éloignant significativement de la moyenne, ainsi que la méthode de l'intervalle interquartile (IQR), qui détecte les valeurs situées en dehors des bornes statistiques définies. Une fois identifiées, ces valeurs aberrantes sont soit corrigées, soit supprimées en fonction de leur degré d'impact sur la cohérence globale des données. Dans le cas des séries climatiques, qui évoluent généralement de manière progressive et continue dans le temps, l'interpolation linéaire est particulièrement adaptée pour le traitement de certaines anomalies, car elle permet de conserver la continuité temporelle des observations tout en limitant la perte d'information. Cette approche contribue ainsi à améliorer la fiabilité et la robustesse des analyses climatiques (Cariou, C. 2021).

Par exemple, lors du traitement des données climatiques de la station d'El-Oued, certaines valeurs de température étaient manquantes pour des dates spécifiques. Afin de garantir la continuité de la série temporelle, une interpolation linéaire a été appliquée.

Tableau 6: Exemple des valeurs manquantes

Date	Température
01/01/2023	10.2
02/01/2023	NaN
03/01/2023	14.2

Et après le traitement :

Tableau 7: Résultat après les traitements

Date	Température
02/01/2023	12.2

La valeur manquante a été estimée par interpolation linéaire, en supposant une évolution progressive de la température entre deux observations consécutives. Cette hypothèse est cohérente avec le comportement des variables climatiques.

Exemple de valeur aberrante, une valeur de température égale à 55°C a été détectée en période hivernale, ce qui est incohérent avec les caractéristiques climatiques de la région d'El-Oued. Cette valeur a été considérée comme aberrante et supprimée afin d'éviter de biaiser les résultats du modèle.

b) Transformation des données

Après l'étape de nettoyage des données, une phase de transformation est nécessaire afin de faciliter leur exploitation et leur analyse. Cette étape consiste tout d'abord en la conversion des formats, notamment la transformation des variables temporelles en format datetime, ce qui permet une manipulation plus efficace des séries chronologiques. Par ailleurs, une uniformisation des unités de mesure est réalisée afin d'assurer la cohérence des données, telles que la vitesse du vent exprimée en m/s et la température en °C.

Ensuite, les données sont structurées sous forme de séries temporelles, avec un index basé sur le temps garantissant une organisation chronologique rigoureuse. De plus, une fréquence régulière est définie, généralement horaire ou journalière selon la disponibilité des observations, ce qui permet d'assurer la continuité des analyses et de faciliter l'application des modèles statistiques et d'apprentissage automatique.

Cette structuration améliore significativement la qualité des données et leur adéquation aux traitements climatiques avancés (Gallopın, M., Rau, A., Celeux, G., & Jaffrezic, F. 2015).

c) Normalisation des données

La normalisation des données constitue une étape essentielle du prétraitement, en particulier dans le cas des variables climatiques qui présentent des échelles de mesure hétérogènes telles que la température, la pression atmosphérique ou encore l'humidité relative. Cette étape vise à éviter qu'une variable à grande amplitude n'influence de manière disproportionnée les autres lors de l'apprentissage des modèles. Deux principales méthodes

sont généralement utilisées à cet effet. La première est la méthode Min-Max Scaling, qui permet de transformer les valeurs dans un intervalle normalisé [0, 1] selon la formule :

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

La seconde est la standardisation ou score Z, définie par

$$X_{norm} = \frac{X - \mu}{\sigma}$$

Où μ représente la moyenne et σ l'écart-type, méthode particulièrement adaptée aux modèles d'apprentissage profond. À titre d'exemple, une valeur de température de 25 °C, avec un minimum de 10 °C et un maximum de 40 °C, devient 0,5 après application du Min-Max Scaling, illustrant ainsi la mise à l'échelle des données sur un intervalle commun. Cette transformation permet d'améliorer la stabilité numérique et la convergence des algorithmes d'apprentissage. Enfin, les données climatiques prétraitées et normalisées, structurées sous forme de séries temporelles, sont utilisées comme entrée pour des modèles avancés tels que le Transformer, lesquels nécessitent des séquences ordonnées afin de capturer efficacement les dépendances temporelles dans les phénomènes climatiques (Mechhour, O. 2023).

4. Ingénierie des fonctionnalités

L'ingénierie des caractéristiques (Feature Engineering) constitue une étape essentielle dans le processus de préparation des données, visant à générer de nouvelles variables à partir des données existantes afin d'améliorer les performances des modèles d'apprentissage automatique. Dans le domaine de la climatologie, plusieurs types de variables dérivées peuvent être construits, telles que les moyennes mobiles, les différences temporelles représentant les variations de température, les variables saisonnières (mois, jour, saison), ainsi que l'encodage cyclique du temps à l'aide de fonctions trigonométriques (sinus et cosinus) pour modéliser la nature périodique des données temporelles. À titre d'exemple, une variable correspondant à la moyenne mobile sur 7 jours peut être introduite afin de lisser les fluctuations quotidiennes et de mieux faire ressortir la tendance générale. Ainsi, pour une série de valeurs (20, 22, 24), la moyenne calculée est égale à 22 °C, offrant une représentation plus stable que les observations individuelles. L'intégration de ces variables dérivées permet d'enrichir l'information contenue dans les données et d'améliorer la capacité du modèle à capturer les dynamiques et les tendances temporelles de manière plus efficace.

5. Structuration pour les modèles d'apprentissage profond

Les modèles d'apprentissage profond, notamment le modèle Transformer, nécessitent une structuration spécifique des données afin d'exploiter efficacement les dépendances temporelles. Cette structuration repose principalement sur la création de séquences temporelles à l'aide d'une technique appelée sliding window (fenêtre glissante). Le principe consiste à utiliser un ensemble de valeurs passées pour prédire une valeur future. Par exemple, il est possible d'exploiter les observations des 30 derniers jours pour estimer la température du jour suivant.

Dans ce cadre, les données sont organisées sous un format tridimensionnel adapté aux modèles d'apprentissage profond, tel que : `nombre_exemple`, `pas_de_temps`, et `nombre_variables` pour les entrées, tandis que la sortie correspond à la valeur cible à prédire. À titre illustratif, pour une fenêtre temporelle de 3 jours, une séquence d'entrée comme [20, 21, 22] permet de prédire la valeur du jour suivant, soit 23.

Ainsi, les données climatiques sont transformées en séquences temporelles afin d'être compatibles avec l'architecture du modèle Transformer. Ce type de structuration permet au modèle de mieux capturer les relations temporelles et les dépendances à long terme entre les différentes observations, améliorant ainsi la qualité des prédictions.

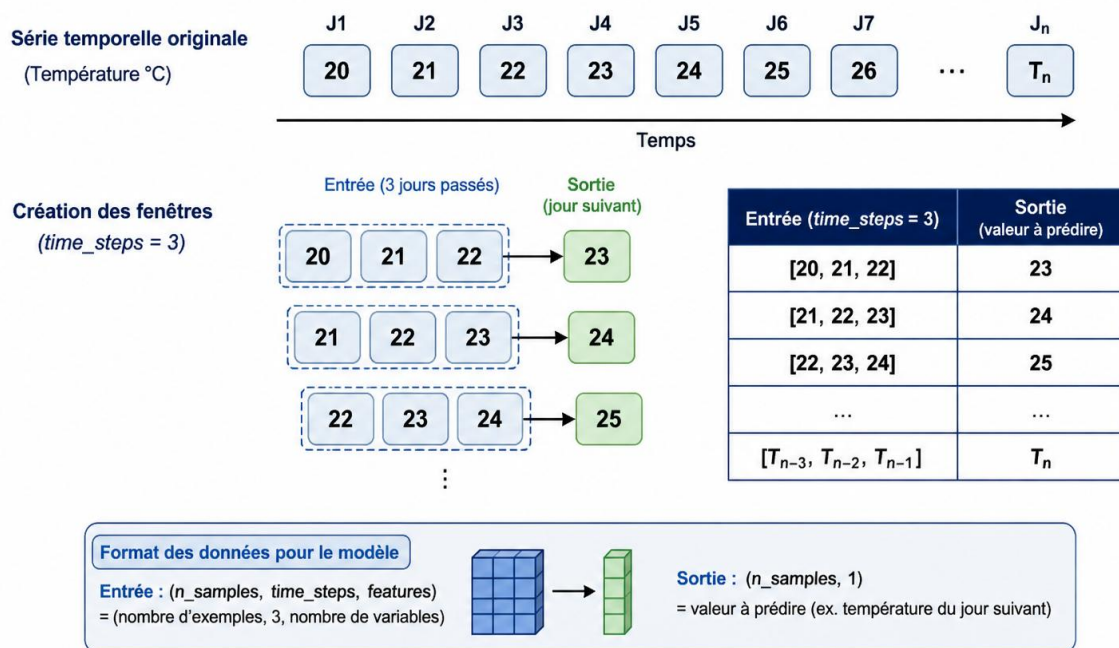


Figure 5: Principe de la fenêtre glissante (sliding window)

Cette figure illustre le principe de la fenêtre glissante (sliding window) utilisée pour structurer les données temporelles. Une séquence de valeurs passées (par exemple 3 jours) est utilisée comme entrée afin de prédire la valeur suivante. Ce processus est répété sur l'ensemble

de la série temporelle pour générer plusieurs couples (entrée/sortie), permettant ainsi d'entraîner efficacement les modèles d'apprentissage profond (Alsène-Racicot, L. 2023).

6. Importance du prétraitement

Le prétraitement des données constitue une étape cruciale dans le processus de modélisation, en particulier dans le cadre des approches basées sur l'apprentissage profond. En effet, la qualité des données d'entrée influence directement les performances des modèles. Cette étape vise principalement à améliorer la qualité des données, à réduire le bruit et les erreurs, ainsi qu'à assurer une meilleure cohérence et homogénéité du jeu de données. Elle permet également d'accélérer le processus d'apprentissage et d'augmenter la précision des prédictions.

Ainsi, le prétraitement a permis d'obtenir un ensemble de données propre, structuré et exploitable. Il s'agit d'une phase indispensable, car des données mal préparées peuvent entraîner une convergence lente des modèles, voire produire des résultats peu fiables. Par conséquent, une attention particulière doit être accordée à cette étape afin de garantir la robustesse et l'efficacité des modèles d'apprentissage profond.

7. Analyse exploratoire des données (EDA)

L'analyse exploratoire des données (Exploratory Data Analysis - EDA) représente une étape fondamentale dans l'étude des données climatiques. Elle a pour objectif de mieux comprendre la structure du jeu de données, d'identifier les tendances générales, d'examiner les relations entre les différentes variables, ainsi que de détecter d'éventuelles anomalies ou valeurs aberrantes avant l'application des modèles d'intelligence artificielle.

Dans le cadre de cette étude, l'EDA a été appliquée aux données climatiques de la région d'El Oued, permettant d'extraire des informations pertinentes et d'orienter les choix méthodologiques relatifs à la modélisation. Cette étape contribue également à valider la qualité des données et à mieux interpréter les résultats obtenus par les modèles prédictifs (KABONGO, B. P. 2025).

a) Analyse descriptive des variables

La première étape de l'analyse consiste à réaliser une étude statistique descriptive des différentes variables climatiques afin de mieux comprendre leurs caractéristiques fondamentales. Dans ce cadre, plusieurs indicateurs ont été calculés, notamment la moyenne, la médiane, l'écart-type, ainsi que les valeurs minimales et maximales. Cette analyse permet de décrire la distribution des données, d'identifier les valeurs extrêmes et de détecter d'éventuelles

anomalies. À titre d'exemple, la température présente généralement une variabilité saisonnière marquée, tandis que les précipitations restent faibles voire nulles sur de longues périodes dans les régions arides comme El-Oued, ce qui reflète la nature climatique spécifique de la zone étudiée.

b) Analyse des distributions

L'étude de la distribution des variables climatiques est réalisée à l'aide de représentations graphiques telles que les histogrammes et les courbes de densité. Cette étape permet de vérifier la normalité des données, d'identifier les asymétries ainsi que de détecter d'éventuelles valeurs aberrantes.

Dans le contexte des données climatiques, la température suit généralement une distribution proche de la loi normale, tandis que les précipitations présentent une distribution fortement asymétrique, caractérisée par une fréquence élevée de valeurs nulles.

L'analyse des dernières années montre que la température maximale estivale dépasse fréquemment 40 °C à El-Oued, alors qu'en hiver elle peut descendre en dessous de 10°C. Cette forte amplitude thermique s'explique par le climat désertique de la région. Par ailleurs, les précipitations sont nulles durant la majorité des jours de l'année, confirmant le caractère aride d'El-Oued. Cette variabilité climatique est principalement due à la faible couverture nuageuse et à l'intense exposition au rayonnement solaire.

c) Analyse des séries temporelles

Les données climatiques étant de nature temporelle, leur analyse dans le temps constitue une étape essentielle pour comprendre leur évolution. Cette analyse permet d'identifier les tendances à long terme, les variations saisonnières ainsi que les éventuels cycles climatiques.

Par exemple, une augmentation progressive des températures peut refléter une tendance au réchauffement climatique, tandis que des variations périodiques régulières sont généralement associées aux saisons (été/hiver).

Ainsi, l'étude des séries temporelles permet de mieux caractériser le comportement des variables climatiques et de fournir des éléments importants pour la modélisation et la prévision des phénomènes climatiques.

d) Analyse des corrélations

L'étude des relations entre les variables climatiques est réalisée à l'aide de matrices de corrélation, permettant de quantifier le degré de dépendance entre les différentes variables. Cette analyse vise principalement à identifier les interactions entre variables et à sélectionner celles qui sont les plus pertinentes pour la modélisation.

Par exemple, une corrélation négative peut être observée entre l'humidité et la température, tandis qu'une corrélation peut exister entre la pression atmosphérique et certaines conditions météorologiques.

Les résultats sont généralement visualisés sous forme de cartes thermiques (heatmaps), ce qui facilite leur interprétation et met en évidence les relations fortes ou faibles entre variables. Dans le cas étudié, une corrélation négative significative ($r \approx -0.6$) a été observée entre la température et l'humidité, indiquant qu'une augmentation de la température est souvent associée à une diminution de l'humidité relative. Ce comportement est caractéristique des régions arides, notamment El-Oued.

e) Détection des anomalies

L'analyse exploratoire des données (EDA) permet également d'identifier les valeurs aberrantes restantes ainsi que les comportements inhabituels au sein des séries climatiques. Ces anomalies peuvent être liées soit à des erreurs de mesure ou de saisie, soit à des événements climatiques extrêmes réellement observés.

La détection de ces anomalies constitue une étape essentielle dans le processus de préparation des données, car leur présence peut influencer négativement la qualité des modèles prédictifs et biaiser les résultats d'analyse. Ainsi, leur identification et leur traitement permettent d'améliorer la robustesse et la fiabilité des modèles développés.

8. Visualisation des données

La visualisation des données constitue une étape essentielle de l'analyse exploratoire, car elle permet de représenter de manière claire et intuitive les caractéristiques des variables climatiques. Elle facilite ainsi l'interprétation des résultats et la compréhension globale du jeu de données.

Les principales figures utilisées dans cette étape incluent les séries temporelles (line plots), les histogrammes, les boxplots ainsi que les matrices de corrélation. Ces représentations permettent de mettre en évidence les tendances, les distributions, les valeurs extrêmes et les relations entre variables.

Ainsi, la visualisation des données contribue à une meilleure compréhension du comportement des variables climatiques et facilite la prise de décision lors des étapes ultérieures de modélisation et d'analyse.

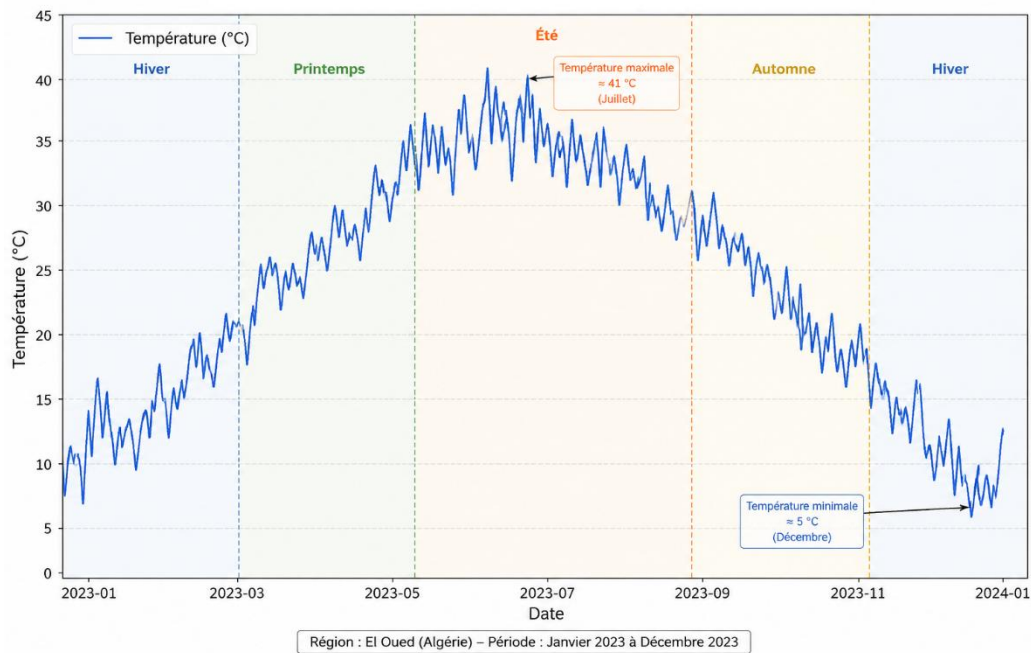


Figure 6: Évolution de la température dans la région d'El Oued(2023)

Cette figure représente la variation temporelle de la température sur une période allant de janvier 2023 à janvier 2024. On observe une tendance saisonnière claire, avec une augmentation progressive des températures au printemps et un pic durant l'été, où elles atteignent environ 35–41 °C. Ensuite, une baisse progressive est enregistrée durant l'automne et l'hiver, avec des valeurs minimales proches de 5–10 °C. Cette évolution confirme le caractère désertique de la région d'El-Oued, marqué par une forte amplitude thermique entre les saisons.

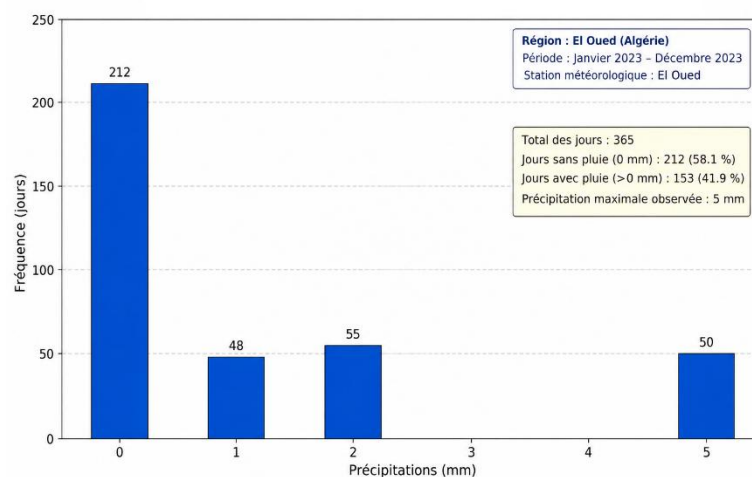


Figure 7: Distribution des précipitations dans la région d'El Oued (2023)

Cette figure présente la distribution des précipitations sous forme d'histogramme. On remarque une forte concentration des valeurs autour de 0 mm, ce qui indique une absence de pluie durant la majorité des jours. Les autres valeurs (1 à 5 mm) apparaissent de manière rare

et dispersée. Cette distribution fortement asymétrique confirme le caractère aride de la région, où les épisodes de pluie sont rares et faibles en intensité.

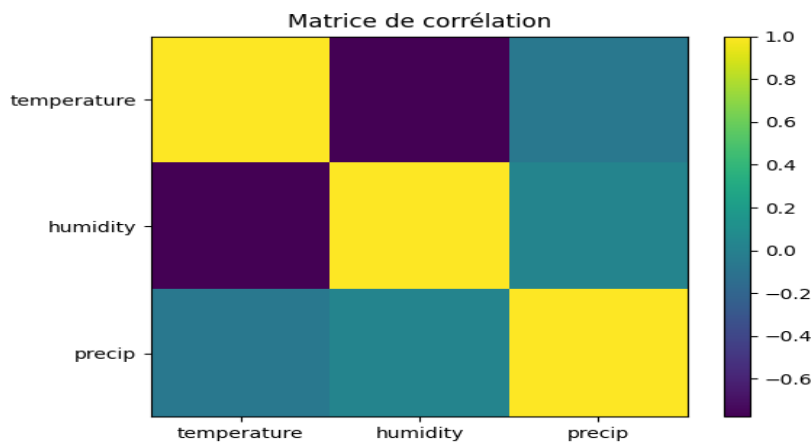


Figure 8: Corrélation entre variables climatiques

Cette figure illustre la matrice de corrélation entre les variables climatiques (température, humidité et précipitations). Les résultats montrent une corrélation négative notable entre la température et l'humidité, tandis que la relation entre les autres variables reste plus faible. La visualisation par heatmap permet d'identifier rapidement les relations fortes et faibles entre les variables, facilitant ainsi la sélection des variables pertinentes pour la modélisation.

9. Conclusion

En conclusion, il s'agit d'une méthode autorisée de collecte, de nettoyage et d'analyse des climats artificiels. Les différentes méthodes de préparation et d'analyse fournissent des informations sur les caractéristiques des échantillons, en tenant compte des variations temporelles et des relations internes.

Ces résultats constituent une base solide pour la phase de modification, et dans le chapitre suivant, le modèle Transformer sera appliqué afin de prédire les variables climatiques avec une grande précision.

CHAPITRE III :

Intelligence artificielle et modèle Transformer appliquer aux séries temporelles climatiques

1. Introduction

L'intelligence artificielle (IA) a connu un développement remarquable au cours des dernières années, notamment dans le domaine de l'analyse et de la prédiction des données temporelles. Grâce à sa capacité à apprendre des relations complexes à partir de grandes quantités de données, elle s'impose aujourd'hui comme un outil incontournable dans plusieurs domaines, y compris la climatologie.

Les données climatiques, caractérisées par leur variabilité temporelle, leur non-linéarité et leur dépendance à long terme, nécessitent l'utilisation de modèles avancés capables de capturer ces propriétés. Les méthodes traditionnelles, telles que les modèles statistiques (ARIMA) ou certains modèles de Machine Learning, présentent des limites lorsqu'il s'agit de modéliser des dépendances complexes sur de longues périodes (Michel, S. 2020).

Dans ce contexte, les modèles d'apprentissage profond, et plus particulièrement le modèle Transformer, représentent une avancée majeure. Initialement développé dans le domaine du traitement du langage naturel, ce modèle repose sur le mécanisme d'attention, qui permet de capturer efficacement les relations entre différentes observations d'une série temporelle.

Ainsi, ce chapitre a pour objectif de présenter les concepts fondamentaux de l'intelligence artificielle, d'introduire le modèle Transformer et d'expliquer son adaptation à la prédiction des séries temporelles climatiques dans la région d'El-Oued.

2. Intelligence artificielle et séries temporelles

a) Définition des séries temporelles

Les séries temporelles désignent un ensemble de données observées et enregistrées à des intervalles de temps réguliers. Dans le domaine de la climatologie, elles concernent des variables telles que la température, l'humidité, la pression atmosphérique et les précipitations. Ces données présentent des caractéristiques spécifiques qui influencent leur analyse, notamment la dépendance temporelle entre les observations, l'existence de tendances à long terme, la présence de phénomènes saisonniers ainsi que l'existence possible de bruit et d'anomalies (Zeynoddin, M. 2025).

En raison de ces particularités, la modélisation des séries temporelles s'avère complexe et nécessite l'utilisation de méthodes statistiques et d'apprentissage adaptées.

Par exemple, dans la région d'El Oued, l'évolution quotidienne de la température sur plusieurs années constitue une série temporelle typique. On peut y observer une forte saisonnalité avec des températures très élevées en été et plus modérées en hiver, ainsi qu'une tendance possible à l'augmentation des températures liée au changement climatique. De plus, certaines valeurs atypiques, comme des pics de chaleur extrêmes, peuvent apparaître et être considérées comme des anomalies.

En raison de ces particularités, la modélisation des séries temporelles s'avère complexe et nécessite l'utilisation de méthodes statistiques et d'apprentissage adaptées.

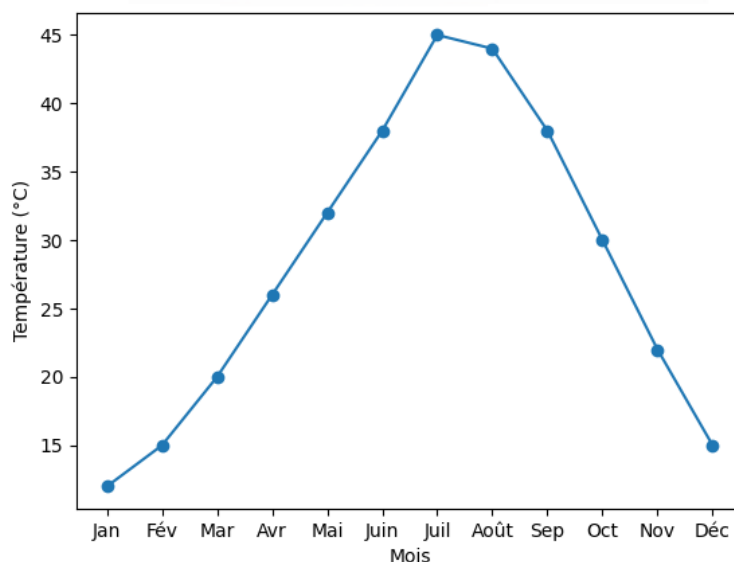


Figure 9: Évolution typique de la température à El Oued

Cette figure illustre une série temporelle représentant la variation de la température au cours d'une année dans la région d'El Oued. On observe une forte saisonnalité caractérisée par des températures modérées en hiver et très élevées en été, avec un pic durant les mois de juillet et août. Cette variation périodique met en évidence la dépendance temporelle des données climatiques ainsi que leur comportement saisonnier.

Afin d'illustrer l'application des séries temporelles dans le domaine climatique, un exemple réel issu de la station météorologique d'El-Oued (WMO : 60559) est présenté.

Les données utilisées comprennent des observations journalières de plusieurs variables climatiques, notamment la température, l'humidité et la pression atmosphérique.

Tableau 8: Observation journalières de plusieurs caractéristiques

Date	Température (°C)	Humidité (%)	Pression (hPa)
01/01/2023	10.2	65	1018
02/01/2023	11.5	60	1017
03/01/2023	13.0	55	1016
04/01/2023	14.8	50	1015

Ces données montrent une évolution progressive de la température au fil des jours, ce qui reflète la nature temporelle des observations climatiques. Comme on peut le constater une augmentation progressive de la température et diminution de l'humidité, une relation inverse.

Dans cet exemple, la température du jour suivant dépend non seulement des valeurs précédentes, mais aussi de relations complexes entre plusieurs variables climatiques. Cela signifie que la relation n'est pas linéaire, elle comporte des dépendances temporelles.

Afin d'appliquer les modèles d'intelligence artificielle, ces données sont transformées en séquences temporelles par exemple :

- Entrée [10.2,11.5,13.0]
- Sortie [14.8]

Ce type de représentation permet aux modèles d'apprentissage automatique d'exploiter l'historique des observations afin de prédire les valeurs futures.

Comme conclusion, Cet exemple concret met en évidence la structure des séries temporelles climatiques et justifie l'utilisation de modèles avancés tels que le Transformer, capables de capturer les dépendances temporelles et les relations non linéaires entre les variables.

b) Apport de l'intelligence artificielle

L'intelligence artificielle (IA) constitue aujourd'hui un outil puissant pour l'analyse et la prédiction des séries temporelles. Contrairement aux méthodes statistiques traditionnelles, souvent limitées à des relations linéaires, les techniques d'IA permettent de modéliser des relations non linéaires et complexes entre plusieurs variables.

Dans le domaine de la climatologie, l'IA est largement utilisée pour diverses applications, notamment la prévision des températures, l'analyse des précipitations et la détection des événements climatiques extrêmes. Ces approches contribuent à améliorer la précision des prédictions et à mieux comprendre les dynamiques climatiques.

Afin d'illustrer cet apport, considérons un exemple simplifié basé sur des données climatiques :

Tableau 9: Exemple de données météorologiques utilisées pour la prédiction de la température du jour suivant (J+1)

Température (°C)	Humidité (%)	Pression (hPa)	Température (J+1)
20	60	1015	22
22	55	1014	25

Dans cet exemple, la température du jour suivant dépend non seulement de la température actuelle, mais aussi d'autres variables climatiques. Une méthode traditionnelle pourrait se limiter à une seule variable (par exemple la température), ce qui réduit la précision des prédictions.

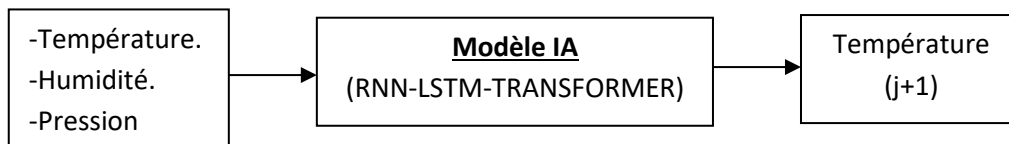


Figure 10: Schéma simplifié d'un modèle d'intelligence artificielle pour la prédiction climatique

Cette figure illustre le fonctionnement d'un modèle d'intelligence artificielle appliqué aux séries temporelles climatiques. Les variables d'entrée, telles que la température, l'humidité et la pression atmosphérique, sont utilisées par le modèle (RNN, LSTM ou Transformer) afin de prédire la température du jour suivant (J+1). Ce type de modèle permet de capturer les relations complexes entre les variables et d'améliorer la précision des prévisions.

En revanche, un modèle d'intelligence artificielle est capable d'exploiter simultanément toutes ces variables afin d'apprendre leurs interactions complexes et d'améliorer significativement la qualité des prédictions.

En résumé, l'intelligence artificielle apporte plusieurs avantages majeurs dans l'analyse des séries temporelles climatiques telles que la capacité à modéliser des relations non linéaires, l'exploitation simultanée de plusieurs variables (approche multi variée), la prise en compte des dépendances temporelles complexes et une amélioration significative de la précision des prédictions.

3. Limites des méthodes traditionnelles

Les méthodes classiques d'analyse des séries temporelles, telles que les modèles statistiques comme ARIMA ou certains algorithmes de Machine Learning, présentent plusieurs limites importantes. Tout d'abord, elles éprouvent des difficultés à capturer les dépendances à

long terme entre les observations. Ensuite, elles reposent généralement sur des hypothèses de linéarité qui ne reflètent pas toujours la complexité des phénomènes climatiques réels. Enfin, ces méthodes sont souvent sensibles au bruit et aux données manquantes, ce qui peut altérer la qualité des prédictions (Hmamouche, Y. 2018).

Afin d'illustrer ces limites, considérons un exemple basé sur des données climatiques simples.

Tableau 10: Exemple basé sur des données climatiques simples

Jour	Température (°C)
J1	18
J2	20
J3	25
J4	30
J5	28

Dans cet exemple, les modèles statistiques classiques, tels qu'ARIMA, supposent généralement une relation linéaire entre les observations. Or, la transition de 20 à 25 puis à 30 montre une évolution accélérée, traduisant une dynamique non linéaire difficile à modéliser avec ces approches.

De plus, ces modèles utilisent souvent un nombre limité de valeurs passées. Par exemple, un modèle ARIMA peut accorder plus d'importance aux observations récentes (comme J3 et J4) tout en négligeant l'influence des jours antérieurs (J1 et J2). Cette limitation peut entraîner une perte d'information significative, notamment dans le cas des séries climatiques où les effets cumulés sont importants.

Un autre problème concerne la sensibilité aux valeurs aberrantes. Considérons l'exemple suivant :

Tableau 11: Exemple des valeurs aberrantes

Jour	Température (°C)
J3	25
J4	50 !
J5	28

La valeur de 50 °C est manifestement erronée. Cependant, les modèles traditionnels peuvent être fortement influencés par ce type de valeur atypique, ce qui fausse les résultats et réduit la fiabilité des prévisions.

Enfin, les modèles classiques traitent souvent les variables de manière indépendante. Or, la température dépend de plusieurs facteurs tels que l'humidité, la pression atmosphérique

ou encore la vitesse du vent. Cette simplification limite leur capacité à représenter fidèlement la complexité des phénomènes climatiques.

En somme, ces exemples mettent en évidence les principales limites des méthodes traditionnelles dans la modélisation des séries temporelles climatiques : incapacité à capturer les relations non linéaires, difficulté à prendre en compte les dépendances à long terme et faible intégration des interactions entre variables. Ces contraintes justifient le recours à des modèles plus avancés, tels que le Transformer, mieux adaptés à la complexité des données climatiques.

4. L'apprentissage profond pour les séries temporelles

L'apprentissage profond constitue une avancée majeure dans le traitement des séries temporelles. Il repose sur des réseaux de neurones capables d'apprendre automatiquement des représentations complexes à partir des données, sans nécessiter une modélisation explicite des relations entre variables (Goodfellow et al., 2016).

Parmi les modèles les plus utilisés dans ce domaine, on distingue :

- RNN (Recurrent Neural Networks) : adaptés au traitement des données séquentielles
- LSTM (Long Short-Term Memory) : capables de capturer les dépendances à long terme
- GRU (Gated Recurrent Unit) : version simplifiée et plus rapide des LSTM

Ces modèles ont permis d'améliorer significativement les performances en prédiction. Toutefois, afin de mieux comprendre leur fonctionnement, considérons un exemple de prédiction de la température à l'aide d'un modèle de type LSTM (Vaswani et al., 2017).

Exemple illustratif

Soit la série temporelle suivante :

Tableau 12: Exemple de série temporelle des températures

Jour	Température (°C)
J1	20
J2	21
J3	23
J4	26
J5	28

Pour appliquer un modèle de l'apprentissage profond, les données sont transformées en séquences temporelles à l'aide d'une technique appelée fenêtre glissante (sliding window).

Tableau 13: Résultat des entrées de série temporelle des températures

Entrée (X)	Sortie (Y)
[20, 21, 23]	26
[21, 23, 26]	28

Cette transformation permet d'exploiter les observations passées afin de prédire les valeurs futures.

Le modèle LSTM est ensuite entraîné sur ces séquences. Il est spécifiquement conçu pour traiter les données séquentielles et capturer les dépendances temporelles. Contrairement aux modèles classiques, il dispose d'un mécanisme de mémoire interne lui permettant de conserver des informations pertinentes sur plusieurs pas de temps.

Après l'entraînement, le modèle peut être utilisé pour effectuer des prédictions. Par exemple :

Tableau 14: Exemple de prédiction après l'entraînement

Entrée	Prédiction
[23, 26, 28]	30

Ce type de modèle permet de capturer les tendances d'évolution de la température, les relations non linéaires entre les observations et les dépendances temporelles à court et moyen terme.

Ainsi, il améliore la précision des prédictions par rapport aux méthodes traditionnelles, malgré leurs performances, les modèles LSTM présentent encore certaines limites : difficulté à traiter des séquences très longues, temps d'entraînement élevé et complexité du réglage des hyper paramètres.

Ces contraintes ont conduit au développement de modèles plus avancés, notamment le Transformer, qui permet de mieux gérer les dépendances à long terme et d'améliorer l'efficacité computationnelle.

Cet exemple montre que l'apprentissage profond constitue une approche efficace pour la modélisation des séries temporelles climatiques. Il permet d'améliorer la précision des prédictions en exploitant les relations complexes présentes dans les données. Toutefois, ses limites justifient le recours à des modèles plus performants, tels que le Transformer, mieux adaptés aux séries longues et aux interactions complexes.

5. Limites des modèles RNN et LSTM

Malgré leur efficacité dans le traitement des données séquentielles, les modèles RNN et LSTM présentent plusieurs limites qui restreignent leur performance, notamment dans le contexte des séries temporelles climatiques. Problème de disparition et d'explosion du gradient

Les RNN classiques souffrent du problème de disparition du gradient ou d'explosion du gradient lors de l'apprentissage sur de longues séquences. Ce phénomène rend difficile l'ajustement des poids du réseau et limite sa capacité à apprendre des dépendances lointaines.

Les LSTM ont été introduits pour atténuer ce problème grâce à leurs mécanismes de mémoire (portes d'entrée, d'oubli et de sortie). Cependant, ils ne permettent pas de l'éliminer complètement, en particulier lorsque les séquences deviennent très longues (Grotta, A., Erritali, M., Etcheverry, P., & Nodenot, T. 2025).

a) Difficulté à capturer les dépendances à long terme

Les données climatiques se caractérisent souvent par des dépendances temporelles à long terme, telles que les relations entre différentes saisons ou encore entre plusieurs années.

Dans ce contexte, les modèles RNN présentent des difficultés à mémoriser des informations éloignées dans le temps, ce qui limite leur capacité à exploiter pleinement ces relations. Les modèles LSTM apportent une amélioration grâce à leur mécanisme de mémoire interne, mais leurs performances restent limitées lorsque la longueur des séquences devient très importante. Cela peut entraîner une perte d'information importante et réduire la qualité des prédictions.

b) Sensibilité au choix des hyper paramètres

Les performances des modèles RNN et LSTM dépendent fortement du réglage des hyper paramètres, tels que la taille de la fenêtre temporelle, le nombre de neurones ou encore le taux d'apprentissage. Un choix inadéquat de ces paramètres peut entraîner une dégradation significative des performances du modèle et compromettre la qualité des prédictions.

c) Traitement séquentiel (manque de parallélisme)

Les modèles RNN et LSTM traitent les données de manière strictement séquentielle, où chaque étape dépend de la précédente. Cette dépendance rend difficile la parallélisation des calculs, ce qui constitue une limitation importante en termes de performance. Par conséquent, ces modèles nécessitent des temps d'entraînement plus élevés et deviennent moins efficaces lorsqu'il s'agit de traiter de grandes bases de données.

d) Complexité computationnelle

Les modèles LSTM se distinguent des RNN classiques par une architecture interne plus complexe, caractérisée notamment par la présence de plusieurs portes (entrée, oubli et sortie) ainsi que par un nombre élevé de paramètres à apprendre. Cette complexité accrue se traduit par un coût de calcul plus important et une consommation significative de ressources matérielles, notamment en termes de CPU et de GPU.

e) Exemple

Par exemple, dans les données climatiques de la région d'El-Oued, la température actuelle peut être influencée par des conditions observées plusieurs semaines auparavant. Dans ce contexte, un modèle LSTM peut avoir du mal à modéliser efficacement ces relations à long terme, en particulier lorsque la séquence est très étendue.

En résumé, bien que les modèles RNN et LSTM soient bien adaptés à l'analyse des données séquentielles, ils présentent des limites importantes, notamment dans la gestion des dépendances à long terme, la faible capacité de parallélisation et la complexité computationnelle. Ces contraintes peuvent réduire leur efficacité dans certaines applications, en particulier dans le domaine de la climatologie, où les données sont souvent complexes et de grande dimension temporelle. Ces contraintes ont conduit au développement de modèles plus avancés, tels que le Transformer, qui permet de surmonter ces limitations grâce à son mécanisme d'attention et à sa capacité de traitement parallèle des données.

6. Transition vers le Transformer

Face aux limites des modèles récurrents tels que les RNN et les LSTM, notamment la difficulté à capturer les dépendances à long terme et le temps d'apprentissage élevé dû au traitement séquentiel des données, une nouvelle architecture appelée Transformer a été introduite par Ashish Vaswani et ses collaborateurs en 2017 dans l'article «Attention Is All You Need». Cette architecture repose principalement sur le mécanisme d'attention (Self-Attention), permettant au modèle d'identifier les relations importantes entre différentes observations d'une séquence, même lorsqu'elles sont très éloignées dans le temps.

Contrairement aux réseaux récurrents, le Transformer traite les données de manière parallèle, ce qui réduit considérablement le temps d'entraînement tout en améliorant les performances sur de grandes séries temporelles. Cette capacité lui permet de mieux représenter les relations complexes et non linéaires présentes dans les données climatiques. De plus, il offre une meilleure stabilité d'apprentissage et une grande efficacité dans la gestion des séquences longues.

Ainsi, grâce à sa puissance de modélisation et à sa capacité à extraire des dépendances globales, le Transformer constitue une approche particulièrement adaptée à l'analyse et à la prédiction des séries temporelles climatiques, ce qui justifie son adoption dans cette étude.

7. Le modèle transformer

Le modèle Transformer est une architecture de réseau de neurones conçue pour le traitement des données séquentielles. Contrairement aux modèles récurrents traditionnels, tels que les RNN et les LSTM, il repose sur un mécanisme d'attention permettant de modéliser directement les relations entre les éléments d'une séquence, indépendamment de leur position.³

Cette architecture a été introduite pour la première fois dans l'article fondateur « *Attention is All You Need* » publié en 2017 par Vaswani et al., marquant une avancée majeure dans le domaine de l'apprentissage profond.

Initialement développés dans le contexte du traitement automatique du langage naturel (NLP), les modèles Transformer ont rapidement démontré leur efficacité dans plusieurs domaines, notamment la vision par ordinateur et la prévision des séries temporelles. Leur succès repose principalement sur leur capacité à capturer des dépendances à long terme et à traiter les données de manière parallèle, ce qui améliore considérablement les performances et réduit le temps d'entraînement.

Contrairement aux modèles basés sur les réseaux de neurones récurrents, le transformer ne traite pas les données de manière séquentielle. Il analyse l'ensemble de la séquence simultanément, ce qui lui permet de mieux comprendre les relations globales entre les observations.

Dans le contexte des séries temporelles climatiques, cette propriété est particulièrement importante, car les variables météorologiques peuvent dépendre d'événements passés éloignés dans le temps. Le Transformer permet ainsi de modéliser efficacement ces dépendances complexes et d'améliorer la précision des prédictions.

a) Principe de fonctionnement

Le fonctionnement du modèle Transformer repose principalement sur le mécanisme d'attention (*Self-Attention*), qui permet au modèle d'attribuer un poids à chaque observation selon son importance dans la prédiction finale (Vaswani et al., 2017).

Ce mécanisme améliore la capacité du modèle à identifier les relations pertinentes entre les différentes données d'une séquence temporelle.

$$\text{Attention}(Q, K, V) = \text{softmax} \left(\frac{QK^T}{\sqrt{d_k}} \right) V$$

³ <https://www.ibm.com/fr-fr/think/topics/transformer-model> (vue le 30 Avril 2026)

Dans cette formule, Q représente les requêtes (*Queries*), K : les clés (*Keys*), V : les valeurs (*Values*) et d_K : la dimension des vecteurs de clés. La fonction *softmax* permet de normaliser les poids d'attention afin de déterminer l'importance relative de chaque élément de la séquence.

Ce mécanisme permet de capturer les relations entre toutes les positions d'une séquence, sans dépendre de leur distance temporelle, contrairement aux réseaux récurrents traditionnels (Goodfellow et al., 2016).

Ainsi, le Transformer peut apprendre efficacement les dépendances à long terme présentes dans les séries temporelles.

Le modèle Transformer constitue donc une évolution majeure par rapport aux modèles récurrents. Grâce à son mécanisme d'attention et à son traitement parallèle des données, il offre une meilleure capacité de modélisation des dépendances temporelles complexes tout en réduisant le temps d'apprentissage. Ces caractéristiques en font un outil particulièrement adapté à l'analyse et à la prédiction des séries temporelles climatiques (Vaswani et al., 2017).

b) Mécanisme d'attention (explication détaillée)

a) Principe général de l'attention

Le mécanisme d'attention constitue l'élément central du modèle Transformer. Il permet au modèle de déterminer l'importance relative de chaque élément d'une séquence lors de la prédiction (Vaswani et al., 2017). Grâce à ce mécanisme, le modèle peut identifier quelles informations sont les plus pertinentes pour comprendre le contexte global des données.

Contrairement aux modèles récurrents tels que les RNN et les LSTM, qui traitent les données de manière séquentielle, le mécanisme d'attention permet d'analyser simultanément l'ensemble des observations d'une séquence. Cette approche facilite l'établissement de relations directes entre les différentes positions de la séquence, indépendamment de leur distance temporelle (Goodfellow et al., 2016).

Dans ce système, chaque élément de la séquence peut « prêter attention » aux autres éléments afin de mieux interpréter le contexte global. Ainsi, le modèle est capable de capturer efficacement les dépendances à long terme et les interactions complexes présentes dans les séries temporelles.

Cette capacité représente un avantage majeur pour l'analyse des données climatiques, où les observations passées peuvent influencer les valeurs futures sur de longues périodes. Le mécanisme d'attention améliore donc la qualité des représentations apprises et contribue à augmenter la précision des prédictions réalisées par le modèle Transformer.

b) Formulation mathématique de l'attention

Le mécanisme d'attention repose sur trois matrices principales générées à partir des données d'entrée du modèle :

- **Query (Q)** : représente l'élément actuellement analysé.
- **Key (K)** : représente les éléments de référence avec lesquels une comparaison est effectuée.
- **Value (V)** : contient les informations utiles associées à chaque élément de la séquence.

L'objectif du mécanisme d'attention est de mesurer le degré de relation entre les différents éléments d'une séquence afin de déterminer quelles informations sont les plus importantes pour la prédiction.

La fonction d'attention utilisée dans le Transformer est donnée par la relation suivante :

$$\text{Attention}(Q, K, V) = \text{softmax} \left(\frac{QK^T}{\sqrt{d_k}} \right) V$$

Cette formule peut être interprétée en plusieurs étapes successives :

- **Calcul de la similarité** : QK^T

Le produit scalaire entre les matrices Q et K^T permet de mesurer la similarité entre les éléments de la séquence. Plus le score obtenu est élevé, plus la relation entre les deux éléments est considérée importante.

- **Normalisation** : $\frac{QK^T}{\sqrt{d_k}}$

La division par $\sqrt{d_k}$ permet de stabiliser les valeurs obtenues et d'éviter des gradients trop importants pendant l'apprentissage.

- **Application de la fonction Softmax**

La fonction *softmax* transforme les scores en probabilités comprises entre 0 et 1. Chaque probabilité représente l'importance relative d'un élément de la séquence.

- **Multiplication par V**

Les poids calculés sont appliqués à la matrice V afin de produire la représentation finale pondérée.

Voici un exemple simplifié

Considérons une série climatique contenant trois observations de température :

Tableau 15: Exemple une série climatique contenant trois observations

Jour	Température
Jour 1	30°C
Jour 2	32°C
Jour 3	45°C

Supposons que le modèle cherche à prédire la température du Jour 3. Le mécanisme d'attention compare cette observation avec les autres jours afin d'identifier les informations les plus importantes. Après le calcul des scores d'attention, le modèle peut obtenir les poids suivants :

Tableau 16: Poids d'attention pour les observations

Observation	Poids d'attention
Jour 1	0.20
Jour 2	0.30
Jour 3	0.50

Ces résultats indiquent que le modèle accorde davantage d'importance à la température du Jour 3, tout en tenant compte des jours précédents. La représentation finale est alors calculée comme une combinaison pondérée des différentes observations :

$$\text{Sortie} = 0.20 \times 30 + 0.30 \times 32 + 0.50 \times 45$$

Le calcul donne :

$$\text{Sortie} = 6 + 9.6 + 22.5 = 38.1$$

Ainsi, la sortie finale représente une combinaison des informations importantes de la séquence. Cet exemple illustre comment le mécanisme d'attention permet au Transformer de sélectionner automatiquement les observations les plus pertinentes pour améliorer la prédiction des séries temporelles climatiques.

c) Self-Attention, Multi-Head Attention et application aux séries temporelles

Le *Self-Attention* constitue un mécanisme fondamental du Transformer dans lequel les matrices Q, K et V sont générées à partir de la même séquence d'entrée, ce qui signifie que chaque élément peut se comparer à tous les autres afin d'évaluer leur importance relative.

Ce mécanisme permet au modèle d'identifier automatiquement les observations les plus pertinentes pour la prédiction et de capturer efficacement les dépendances temporelles, même lorsque les événements sont éloignés dans le temps.

Par exemple, pour prédire la température d'un jour donné, le modèle peut accorder davantage d'importance aux jours récents ou à certaines périodes présentant des conditions climatiques similaires, comme une vague de chaleur ou une faible humidité. Afin d'améliorer

davantage la représentation des données, le Transformer utilise plusieurs mécanismes d'attention appelés Multi-Head Attention (Vaswani et al., 2017).

Chaque tête d'attention apprend un type particulier de relation : certaines analysent les dépendances à court terme, d'autres les dépendances à long terme, tandis que d'autres encore étudient les interactions entre différentes variables climatiques.

Cette approche permet au modèle d'obtenir une compréhension plus riche et plus complète des données. Le mécanisme d'attention présente ainsi plusieurs avantages importants, notamment la capacité à capturer les dépendances à long terme, le traitement parallèle des données, une meilleure compréhension du contexte global et une grande flexibilité dans l'analyse des relations entre les observations.

Dans le cadre des séries temporelles climatiques, ces propriétés permettent d'identifier les périodes les plus influentes, de modéliser les interactions entre les variables météorologiques et d'améliorer la précision des prédictions.

Par exemple, pour prédire la température à El Oued, le modèle peut accorder plus d'importance aux jours caractérisés par des conditions climatiques proches, telles qu'une forte chaleur ou une faible humidité, afin de produire des prévisions plus précises et plus fiables.

Comme conclusion, le mécanisme d'attention constitue une avancée majeure dans le traitement des données séquentielles. Il permet de dépasser les limitations des modèles récurrents en capturant efficacement les dépendances globales et en améliorant la qualité des représentations.

Cette capacité fait du Transformer un modèle particulièrement adapté à l'analyse des séries temporelles climatiques.

d) Architecture du Transformer

L'architecture du modèle Transformer est principalement composée de deux blocs fondamentaux : l'encodeur (*Encoder*) et le décodeur (*Decoder*) avec le mode de calcul de loss **CrossEntropyLoss()**. Le décodeur traite le problème comme une classification multi classe. La Loss n'évalue pas une distance numérique, mais pénalise le modèle si la probabilité accordée au "vrai mot" (la bonne catégorie) est faible. (Vaswani et al., 2017).

Toutefois, dans le domaine des séries temporelles, notamment pour la prévision climatique, on utilise uniquement la partie encodeur suivie d'une couche de sortie dense avec le mode de calcul **MSELoss** qui Évalue la distance géométrique directe entre la courbe prédite et la courbe réelle. Le carré permet de pénaliser sévèrement les grands écarts de prédiction.

Afin de réaliser la prédiction des valeurs futures. Le modèle repose sur un empilement de plusieurs couches identiques permettant d'extraire progressivement des représentations de plus en plus abstraites et pertinentes des données.

Chaque couche de l'encodeur contient principalement deux sous-couches : un mécanisme de Multi-Head Attention, chargé de capturer les relations entre les différents éléments de la séquence, et un réseau neuronal Feed Forward Neural Network (*FFN*) appliqué indépendamment à chaque position de la séquence. Ces sous-couches sont accompagnées d'une normalisation et de connexions résiduelles, qui améliorent la stabilité de l'apprentissage, accélèrent la convergence et réduisent les problèmes liés à la disparition du gradient.

Ainsi, l'encodeur reçoit une séquence de données climatiques en entrée, applique successivement les mécanismes d'attention et les transformations non linéaires, puis génère une représentation enrichie des informations temporelles.

Quant au décodeur, il est principalement utilisé dans les tâches de génération séquentielle telles que la traduction automatique. Cependant, dans le cadre de la prévision des séries temporelles climatiques, cette partie peut être simplifiée ou remplacée par une couche de sortie permettant directement la prédiction des valeurs futures.

La figure suivante (Figure 12) illustre l'architecture générale du Transformer composée d'un encodeur et d'un décodeur basés sur le mécanisme d'attention, où chaque bloc intègre des couches de *Self-Attention*, des réseaux *Feed Forward*, des normalisations et des connexions résiduelles. Pour les applications de séries temporelles, la partie encodeur représente généralement le composant principal exploité dans le modèle.

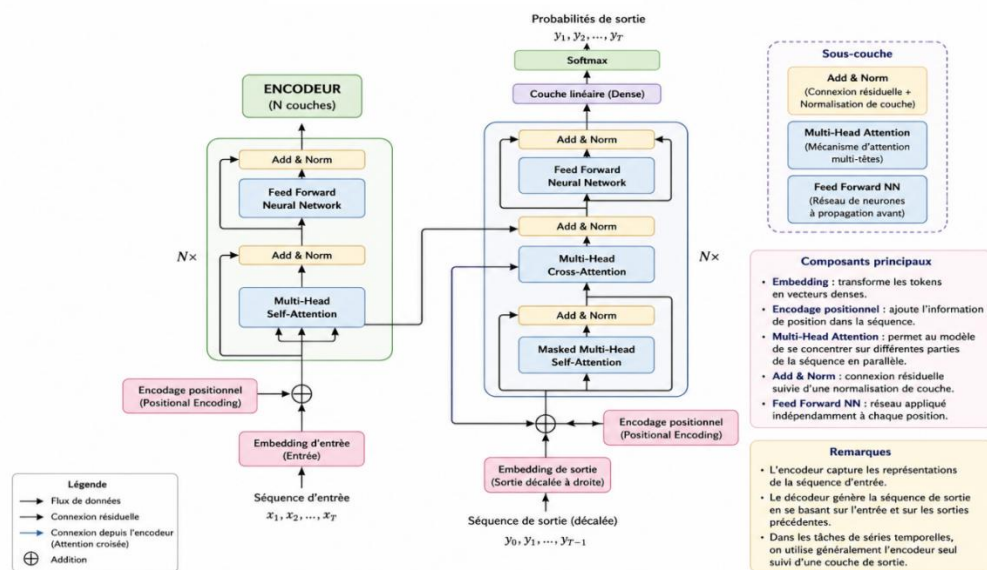


Figure 11: Architecture générale du modèle Transformer

8. Adaptation aux séries temporelles

Dans le cadre de cette étude, l'architecture du Transformer est adaptée spécifiquement au traitement des séries temporelles climatiques. Les données climatiques sont d'abord transformées en séquences temporelles structurées, afin de représenter l'évolution des variables au cours du temps.

Ces séquences sont ensuite introduites dans l'encodeur du modèle, qui applique les mécanismes de self-attention pour apprendre les relations et les dépendances temporelles entre les différentes observations. Grâce à ce processus, le modèle est capable de comprendre comment les valeurs passées influencent les valeurs futures, en tenant compte des tendances, des variations saisonnières et des changements brusques.

Une fois la représentation enrichie obtenue à la sortie de l'encodeur, celle-ci est transmise à une couche dense (Fully Connected Layer) qui assure la prédiction de la valeur future de la série temporelle.

Ainsi, le modèle transforme une séquence d'observations passées en une estimation de la prochaine valeur.

Exemple d'application simple :

Tableau 17: Estimation de la prochaine valeur

Entrée (séquence)	Sortie (prédiction)
[20, 21, 22, 23]	24

Dans cet exemple, le modèle apprend la tendance croissante de la série et prédit la valeur suivante en se basant sur l'évolution des observations précédentes.

9. Application du modèle Transformer aux séries temporelles climatiques

Dans cette partie, le modèle Transformer est appliqué à la prévision des séries temporelles climatiques de la région d'El Oued. L'objectif principal est d'exploiter le mécanisme d'attention afin de mieux représenter les relations complexes entre les variables climatiques (température, humidité, pression, vent, etc.) et d'améliorer la qualité des prédictions à court et moyen terme. Grâce à sa capacité à capturer les dépendances temporelles longues, le Transformer permet une analyse plus fine des évolutions climatiques (Dufourg, C., Pelletier, C., May, S., & Lefèvre, S. 2024).

a) Préparation des données

Avant l'entraînement du modèle, les données climatiques subissent plusieurs étapes de préparation essentielles pour garantir leur qualité et leur fiabilité. Tout d'abord, un processus

de nettoyage est appliqué afin de supprimer les valeurs aberrantes et de traiter les valeurs manquantes.

Ensuite, les variables sont normalisées afin d'harmoniser leurs échelles et d'améliorer la stabilité de l'apprentissage. Enfin, une sélection des variables pertinentes est effectuée pour ne conserver que les paramètres les plus influents sur la prédiction, tels que la température, l'humidité, la pression atmosphérique et la vitesse du vent.

b) Transformation en séquences temporelles

Les données préparées sont ensuite transformées en séquences temporelles à l'aide d'une fenêtre glissante (*sliding window*). Cette méthode consiste à utiliser une série de valeurs passées pour prédire une valeur future. Par exemple :

Tableau 18: Prédiction des valeurs de sortie

Entrée (X)	Sortie (Y)
[20, 21, 22, 23]	24
[21, 22, 23, 24]	25

Chaque séquence d'entrée représente ainsi un ensemble d'observations consécutives permettant au modèle d'apprendre les dynamiques temporelles de la série.

Une fois les séquences construites, les données sont organisées sous forme de tenseurs tridimensionnels de la forme suivante : (samples, time_steps, features) où :

- **samples** : correspond au nombre d'exemples disponibles dans le jeu de données,
- **time_steps**: représente le nombre d'observations dans chaque séquence temporelle,
- **features**: correspond au nombre de variables climatiques utilisées (température, humidité, pression, etc.).

c) Application du Transformer

L'application du modèle Transformer suit une série d'étapes structurées. Tout d'abord, les séquences d'entrée sont introduites dans l'encodeur du modèle. Ensuite, un encodage positionnel est ajouté afin d'intégrer explicitement l'information temporelle, puisque le modèle ne traite pas naturellement l'ordre des données. Le mécanisme de *self-attention* est ensuite appliqué pour analyser les relations entre les différentes observations de la séquence. Cette étape permet de produire une représentation enrichie des données, dans laquelle les informations les plus pertinentes sont mises en évidence. Enfin, cette représentation est transmise à une couche entièrement connectée (*Fully Connected Layer*), qui réalise la prédiction de la valeur future.

d) Fonction de perte et optimisation

L'entraînement du modèle repose sur une fonction de perte adaptée aux problèmes de régression, généralement la *Mean Squared Error (MSE)*, qui mesure l'écart entre les valeurs prédites et les valeurs réelles. Pour optimiser les paramètres du modèle, des algorithmes d'optimisation avancés sont utilisés, notamment l'algorithme *Adam (Adaptive Moment Estimation)*, qui permet une convergence rapide et stable en ajustant automatiquement les taux d'apprentissage.

e) Exemple d'application (El-Oued)

Dans le cadre de cette étude, une fenêtre temporelle composée des **30** derniers jours de données climatiques (température, humidité, pression atmosphérique, etc.) est utilisée comme entrée du modèle. Celui-ci est ensuite chargé de prédire la température du jour suivant pour la région d'El Oued. Cette approche permet de capturer les tendances récentes ainsi que les variations climatiques locales afin d'améliorer la précision de la prédiction.

f) Résultats attendus

L'utilisation du Transformer dans ce contexte permet d'obtenir plusieurs améliorations importantes, notamment une meilleure capacité à capturer les dépendances à long terme, une amélioration de la précision des prédictions, ainsi qu'une prise en compte plus efficace des interactions entre les différentes variables climatiques. Ces avantages contribuent à rendre le modèle plus performant pour l'analyse des séries temporelles climatiques.

10. Avantages de l'approche

L'approche basée sur le Transformer présente plusieurs avantages majeurs. Elle permet une modélisation efficace des séries temporelles complexes grâce au mécanisme d'attention, offre un traitement parallèle des données qui réduit le temps d'entraînement, et garantit une meilleure robustesse face aux grands volumes de données climatiques. De ce fait, elle constitue une solution particulièrement adaptée aux problèmes de prévision météorologique et climatique.

11. Conclusion

L'application du modèle Transformer aux séries temporelles climatiques représente une approche moderne et particulièrement performante dans le domaine de la prévision météorologique. Grâce à son architecture basée sur le mécanisme d'attention, le Transformer dépasse les limites des modèles traditionnels en permettant une meilleure exploitation des relations temporelles complexes entre les différentes observations climatiques.

En effet, contrairement aux modèles séquentiels classiques, le Transformer est capable de traiter simultanément l'ensemble des données d'une séquence, ce qui lui permet de capturer efficacement les dépendances à court, moyen et long terme. Cette capacité est essentielle dans le contexte climatique, où les phénomènes atmosphériques sont souvent influencés par des événements passés éloignés dans le temps, ainsi que par des interactions multiples entre variables telles que la température, l'humidité et la pression atmosphérique.

De plus, l'utilisation du mécanisme de *self-attention* et de *multi-head attention* permet au modèle de mieux identifier les périodes significatives et les relations importantes au sein des données. Cela conduit à une représentation plus riche et plus précise des séries temporelles, améliorant ainsi la qualité des prédictions (Haoyi Zhou, 2021).

Dans le cadre de cette étude appliquée à la région d'El Oued, le Transformer permet d'analyser les dynamiques climatiques locales de manière plus fine et plus robuste, en tenant compte des variations saisonnières et des tendances historiques. L'intégration de ce modèle dans les systèmes de prévision climatique offre donc un potentiel important pour améliorer la fiabilité des prévisions météorologiques.

En conclusion, le Transformer constitue une solution efficace et innovante pour l'analyse des séries temporelles climatiques. Ses performances en matière de modélisation des dépendances temporelles, sa capacité de traitement parallèle et sa flexibilité en font un outil particulièrement adapté aux problématiques de prévision climatique, ouvrant ainsi la voie à des applications plus avancées dans le domaine de la science des données climatiques.

CHAPITRE IV :

L'implémentation pratique du modèle pour la prédiction climatique

1. Introduction

Après avoir présenté les fondements de l'intelligence artificielle, les techniques de prétraitement des données climatiques ainsi que le fonctionnement du modèle Transformer appliqué aux séries temporelles, ce chapitre est consacré à l'implémentation pratique du modèle pour la prédiction climatique.

L'objectif principal est de construire un système capable d'apprendre les dépendances temporelles présentes dans les données météorologiques afin d'effectuer des prédictions précises sur différentes variables climatiques telles que la température, l'humidité, la pression atmosphérique ou encore la vitesse du vent.

Ce chapitre décrit les différentes étapes de mise en œuvre du modèle, depuis la préparation des données jusqu'à l'entraînement et l'évaluation des performances. Une attention particulière est accordée au choix des paramètres du modèle, à l'architecture utilisée ainsi qu'aux outils logiciels employés (Bryan Lim et S. Zohren, 2021).

2. Environnement de développement

L'implémentation du modèle Transformer a été réalisée à l'aide d'un environnement de développement spécialisé dans le calcul scientifique et l'apprentissage profond (Gince, J. 2023). Les principaux outils utilisés sont :

- Langage de programmation : Python (version : 3.12.13)
- Bibliothèques : NumPy, Pandas, Matplotlib, PyTorch, Requests
- Plateformes de développement : Jupyter Notebook et Google Colab
- Système d'exploitation : Windows 10 professionnel 64 bits
- Caractéristiques de l'ordinateur utilisé : Processeur : Intel Core i5
Mémoire RAM : 8 Go - Stockage : SSD 256 Go - Carte Figure : Intel UHD Graphics -Architecture : 64 bits

Ces outils offrent une grande flexibilité pour le traitement des séries temporelles, l'analyse des données climatiques et l'implémentation des modèles d'intelligence artificielle. Grâce à Python et à ses bibliothèques spécialisées, il est possible de manipuler efficacement les données, d'effectuer des visualisations et d'entraîner le modèle Transformer dans un environnement interactif et performant.

3. Description des données utilisées

Les données climatiques utilisées dans cette étude proviennent des archives météorologiques de la station d'El-Oued.

Le jeu de données contient plusieurs variables météorologiques observées sur une période temporelle continue.

Les principales caractéristiques exploitées sont (Température moyenne, Température maximale, Température minimale, Humidité relative, Pression atmosphérique, Vitesse du vent, Précipitations)

Chaque observation est associée à une date précise, ce qui permet de construire une série temporelle exploitable par le modèle Transformer.

4. Préparation des données

Avant l'entraînement du modèle, plusieurs opérations de prétraitement ont été appliquées afin d'améliorer la qualité des données.

a) Extraction des données :

Les données météorologiques ont été collectées par la technique du Web Scraping à partir du site Ogimet, spécialisé dans les archives météorologiques mondiales. L'extraction a été réalisée en utilisant le langage Python et la bibliothèque Requests.

Exemple simplifié du code utilisé :

```
import requests
import pandas as pd
BASE_URL = "https://www.ogimet.com/cgi-bin/getsynop"
BEGIN = "202501010000"
END = "202512312359"
BLOCK = "605"
STATION_ID = 60559
OUT = "synop_60559_2025_2026.xlsx"
REQUEST_TIMEOUT = 300
MAX_RETRIES = 5
RETRY_SLEEP_SEC = 8

def fetch_chunk_text(begin: str, end: str) -> str:
    url =
    f'{BASE_URL}?block={BLOCK}&begin={begin}&end={end}&header=yes&lang=eng"
    for attempt in range(1, MAX_RETRIES + 1):
        resp = requests.get(
```

```

url,
timeout=REQUEST_TIMEOUT,
headers={"User-Agent": "synop-downloader/1.0"},
)
resp.raise_for_status()
return resp.text

```

Après l'extraction, les données brutes ont été organisées sous forme tabulaire contenant plusieurs informations météorologiques telles que la date, l'heure et les observations SYNOP.

Tableau 19: Extrait des observations météorologiques brutes De la station météorologique 60559

ESTACION	ANO	MES	DIA	HORA	MINUTO	PARTE
60559	2010	01	01	00	00	AAXX 01004 60559 12960 22203 10099 20033 40110 51009 60001 80001 333 82073=
0559	2010	01	01	03	00	AAXX 01034 60559 42960 23002 10083 20042 40117 53007 80001 333 82073=
60559	2010	01	01	06	00	AAXX 01064 60559 12960 02408 10087 20047 40126 52008 60004 333 0//00 20070=

Ces messages SYNOP ont ensuite été décodés et transformés en variables climatiques exploitables pour l'analyse et l'apprentissage du modèle Transformer.

b) Décodage et Nettoyage des données :

Après l'extraction des messages météorologiques SYNOP, une étape de décodage et de nettoyage des données a été réalisée afin de transformer les données brutes en variables climatiques exploitables par le modèle Transformer.

Le décodage consiste à interpréter les codes météorologiques contenus dans les messages SYNOP pour obtenir des valeurs numériques compréhensibles telles que la température, la pression atmosphérique, la vitesse du vent et les précipitations.

Cette opération a été effectuée en utilisant le langage Python ainsi que des expressions régulières (re) pour identifier et convertir les différents groupes de données météorologiques.

Les principales caractéristiques obtenues sont :

Tableau 20: Description des caractéristiques météorologiques utilisées

Caractéristiques	Description	Min	Max
Temperature_c	Température de l'air exprimée en degrés Celsius (°C)	-5 °C	50 °C
Dew_point_c	Température du point de rosée exprimée en degrés Celsius (°C), indiquant le niveau d'humidité de l'air.	-10 °C	35 °C
Pressure_hpa	Pression atmosphérique mesurée en hectopascal (hPa).	950 hPa	1050 hPa
Wind_speed	Vitesse du vent mesurée en kilomètres par heure (km/h) ou mètres par seconde (m/s).	0	150 km/h
Wind_direction_deg	Direction du vent exprimée en degrés (0° à 360°).	0°	360°
Cloud_cover	Taux de couverture nuageuse exprimé en pourcentage (%).	0 %	100 %
Pressure_tendency_hpa	Variation de la pression atmosphérique sur une période donnée, exprimée en hPa.	-20 hPa	20 hPa
Precipitation_mm	Quantité de précipitations mesurées en millimètres (mm).	0 mm	500 mm
Solar_elevation	Angle d'élévation du soleil par rapport à l'horizon, exprimé en degrés.	-90°	90°

Tableau 21: Résumé de l'extraction des données SYNOP

Caractéristique	Format Groupe	Exemple SYNOP	Méthode d'extraction / Calcul	Valeur Extraite
Temperature_c	1sTTT	10044	s (signe: 0=+ / 1=-). TTT / 10	+4.4 °C
Dew_point_c	2sT_dT_dT_d	21031	s (signe: 0=+ / 1=-). T_dT_dT_d / 10	-3.1 °C
Pressure_hpa	4PPPP	40322	Si premier \$\$ est 0 ou 1 \$rightarrow\$ ajouter 1000. PPPP / 10	1032.2 hPa
Wind_direction_deg	ddfff (Position 2)	02703	Les 3 premiers chiffres (dd) \$times\$ 10	270° (West)
Wind_speed	ddfff (Position 2)	02703	Les 2 ou 3 derniers chiffres (fff) selon l'unité	3 knots/m-s

Cloud_cover	Nddff ou Nh...	60559	Le tout premier chiffre (N) du groupe de base (en Octas)	6 / 8 (Okta)
Pressure_tendency	5appp	53003	a = caractéristique de la tendance. ppp / 10	0.3 hPa
Precipitation_mm	6RRRt	60001	RRR = quantité (codes spécifiés par l'OMM / WMO)	1 mm
Solar_elevation	Hors SYNOP	—	Calcul astronomique via heure, latitude et longitude	(Variable)

Exemple de fonction utilisée pour décoder la température :

```
import re
from typing import Optional

def decode_temp(sTTT: str) -> Optional[float]:
    """Decode 1sTTT or 2sTTT: s=0 positive, s=1 negative; TTT = tenths of °C."""
    if not sTTT or len(sTTT) < 4 or "/" in sTTT:
        return None
    try:
        sign = -1 if sTTT[0] == "1" else 1
        val = int(sTTT[1:4]) / 10.0
        return sign * val
    except (ValueError, IndexError):
        return None
```

Après le décodage, les données ont été restructurées sous forme tabulaire afin de faciliter leur analyse et leur exploitation par les modèles d'apprentissage profond.

Tableau 22:Exemple des données obtenues après décodage

Datetime	Temperature_c	Dew_point_c	Pressure_hpa	Wind_speed	Wind_direction_deg	Cloud_cover	Pressure_tendency_hpa	Precipitation_mm	Solar_elevation	Temp_min_24h_c	Temp_max_24h_c
2010-01-07 06:00:00	8.2	3.3	1008	4	120	3	-0.6	0		7	
2010-01-07 09:00:00	13	4.3	1296	4	100	6	0.1	0			
2010-01-07 12:00:00	20.7	4.7	1004.9	2	120	5	20.2	0			
2010-01-07 15:00:00	-45	2.5	1256	10	180	7	-2.7	0			
2010-01-07 18:00:00	95	4.5	1001.6	3	160	6	-0.7	0			21.5
2010-01-07 21:00:00	14.5	12.7	1155.7	8	180	8	0.8	0			

Cette étape est essentielle pour garantir la fiabilité des données utilisées lors de l'entraînement et de l'évaluation du modèle Transformer.

Fichier json:

"columns":

```
[ "temperature_c", "dew_point_c", "pressure_hpa", "wind_speed", "wind_direction_deg", "cloud_cover", "pressure_tendency_hpa", "solar_elevation", "precipitation_mm", "temp_min_24h_c", "temp_max_24h_c" ],
"data_min": [ -4.0, -9.8, 990.1, 0.0, -1.0, -1.0, -21.8, 0.0, 0.0, -4.5, 6.2 ],
"data_max":
[ 50.8, 33.6, 1072.1, 94.0, 1.0, 1.0, 92.4, 8.0, 55.0, 35.0, 49.7 ],
```

Exemple code python :

```
import json
from pathlib import Path
from typing import Any
import numpy as np
import pandas as pd
```

```
SCALER_JSON = _ROOT / "60559_minmax_scaler_params.json"
INPUT_XLSX = _ROOT / "synop_60559_2010_2026_04_decoded.xlsx"
OUTPUT_XLSX = _ROOT / "synop_60559_2010_2026_04_decoded_clean.xlsx"
```

```
def clean_by_bounds(
    df: pd.DataFrame,
    rules: dict[str, tuple[float, float]],
    *,
    strategy: str = CLEAN_STRATEGY,
    sort_by_datetime: bool = False,
) -> tuple[pd.DataFrame, dict[str, int]]:
```

```
    if strategy not in {"nan", "cap", "interpolate"}:
        raise ValueError("strategy must be 'nan', 'cap', or 'interpolate'.")
```

```
    out = df.copy()
    if sort_by_datetime and "datetime" in out.columns:
        out["datetime"] = pd.to_datetime(out["datetime"], errors="coerce")
        out = out.sort_values("datetime").reset_index(drop=True)
```

```
    fixed: dict[str, int] = {}
```

```

for col, (lower, upper) in rules.items():
    if col not in out.columns:
        continue
    s = pd.to_numeric(out[col], errors="coerce")
    invalid = (s < lower) | (s > upper)
    n_bad = int(invalid.sum())
    if n_bad:
        fixed[col] = n_bad

    if strategy == "cap":
        out[col] = s.clip(lower=lower, upper=upper)
        continue

    s = s.mask(invalid, np.nan)
    if strategy == "interpolate":
        s = s.interpolate(method="linear", limit_direction="both")
        s = s.ffill().bfill()
    out[col] = s

return out, fixed

```

c) Normalisations des données :

Deux principales méthodes sont généralement utilisées à cet effet.

La première est la méthode Min-Max Scaling, qui permet de transformer les valeurs dans un intervalle normalisé [0,1] selon la formule : $X_{norm} = (X - X_{min}) / (X_{max} - X_{min})$

La seconde est la standardisation ou score Z, définie par $X_{norm} = (X - \mu) / \sigma$

Exemple du code python :

```

import json
from pathlib import Path
from typing import Any
import numpy as np
import pandas as pd
from pandas.api.types import is_numeric_dtype
from sklearn.preprocessing import MinMaxScaler

INPUT_XLSX = _ROOT / "60559_2010_2026_Process_all.xlsx"
OUTPUT_XLSX = _ROOT / "60559_2010_2026_Process_all_minmax.xlsx"
OUTPUT_SCALER_JSON = _ROOT / "60559_minmax_scaler_params.json"

```

```

SCALE_COLUMNS: list[str] =
[ "hour_sin", "hour_cos", "month_sin", "month_cos", "day_sin", "day_cos", "te
mperature_c", "temp_min_24h_c", "temp_max_24h_c",
    "temp_mean_inst_min_max_c", "temp_range_24h_c", "dew_point_c", "pressure_hpa",
    "wind_speed", "wind_dir_sin", "wind_dir_cos", "wind_u_ms", "wind_v_ms", "pre
ssure_tendency_hpa", "cloud_cover", "precipitation_mm",]

def _month_day_sin_cos(ts: pd.Series) -> tuple[pd.Series, pd.Series, pd.Series, pd.Series]:
    """Cyclical month (1-12) and calendar day within month from timestamps."""
    dt = pd.to_datetime(ts, errors="coerce")
    month = dt.dt.month.astype(float)
    day = dt.dt.day.astype(float)
    dim = dt.dt.days_in_month.astype(float)
    rad_m = 2.0 * np.pi * (month - 1.0) / 12.0
    den = (dim - 1.0).clip(lower=1.0)
    rad_d = 2.0 * np.pi * np.maximum(day - 1.0, 0.0) / den
    idx = ts.index
    return (
        pd.Series(np.sin(rad_m), index=idx),
        pd.Series(np.cos(rad_m), index=idx),
        pd.Series(np.sin(rad_d), index=idx),
        pd.Series(np.cos(rad_d), index=idx),
    )

def _temperature_aggregates(out: pd.DataFrame) -> pd.DataFrame:
    t = pd.to_numeric(out["temperature_c"], errors="coerce")
    tmin = pd.to_numeric(out["temp_min_24h_c"], errors="coerce")
    tmax = pd.to_numeric(out["temp_max_24h_c"], errors="coerce")
    triple = pd.concat([t, tmin, tmax], axis=1)
    out["temp_mean_inst_min_max_c"] = triple.mean(axis=1, skipna=True)
    out["temp_range_24h_c"] = tmax - tmin
    return out

```

Tableau 23: Résultat obtenue après l'exécution du code

Hour_sin	Hour_cos	Temperat ure_c	Dew_poi nt_c	Pressure_ hpa	Wind_spe ed	Wind_ directio n_deg	Coud_co ver	Pressure_te ndency_hpa	Precipitatio n_mm	Temp_min _24h_c
0.5	1	0.563898	0.301843	0.254878	0.031915	220	0.25	0.198774	0	0.249315
0.853553	0.853553	0.555378	0.322581	0.264634	0.021277	300	0.25	0.197023	0	0.249315

5. Apprentissage profond

L'apprentissage profond constitue une étape essentielle dans cette étude, car il permet d'exploiter efficacement les données climatiques normalisées pour construire un modèle de prédiction performant basé sur l'architecture Transformer.

Exemple du code python:

```
class WeatherTransformer(nn.Module):
```

```
    def __init__(self, num_features: int, embed_dim: int):
        super().__init__()
        self.input_proj = nn.Linear(num_features, embed_dim)
        self.input_norm = nn.LayerNorm(embed_dim)
        self.W_q = nn.Linear(embed_dim, embed_dim)
        self.W_k = nn.Linear(embed_dim, embed_dim)
        self.W_v = nn.Linear(embed_dim, embed_dim)
        self.out_norm = nn.LayerNorm(embed_dim)
        self.output_fc = nn.Linear(embed_dim, 1)
```

```
    def forward(self, x: torch.Tensor) -> torch.Tensor:
        x = self.input_norm(self.input_proj(x))
        q, k, v = self.W_q(x), self.W_k(x), self.W_v(x)
        d_k = x.shape[-1] ** 0.5
        scores = torch.matmul(q, k.transpose(-2, -1)) / d_k
        attn = torch.softmax(scores, dim=-1)
        context = torch.matmul(attn, v)
        last = self.out_norm(context[:, -1, :])
        return self.output_fc(last)
```

Résultat après l'exécution du programme :

Features (14): ['month_sin', 'month_cos', 'day_sin', 'day_cos', 'hour_sin', 'hour_cos',
'temperature_c', 'dew_point_c', 'pressure_hpa', 'wind_speed', 'wind_dir_sin', 'wind_dir_cos',
'cloud_cover', 'pressure_tendency_hpa']

embedding dim = 64, seq_len = 24, hourly_frequency = 3 h

Training start...

Epoch 20, Loss: 0.017122

Epoch 40, Loss: 0.013355

Epoch 60, Loss: 0.012860

Epoch 80, Loss: 0.012735

Epoch 100, Loss: 0.012612

Epoch 120, Loss: 0.012553

Epoch 140, Loss: 0.012493

Epoch 160, Loss: 0.012430

Epoch 180, Loss: 0.012317

Epoch 200, Loss: 0.012290

Model weights saved to:

/content/drive/MyDrive/MasterTransformer/weather_transformer_weights.pth

6. Prédiction

Après la phase d'apprentissage du modèle Transformer, l'étape suivante consiste à effectuer la prédiction des données climatiques futures. Le modèle entraîné utilise les observations passées des séries temporelles afin d'estimer les valeurs futures des paramètres météorologiques.

Exemple du code python :

```
import json

from pathlib import Path
import numpy as np
import pandas as pd
import torch
import torch.nn as nn

from google.colab import drive

drive.mount("/content/drive")

# --- paths (same folder as train_transformer_colab.py) ---
```

```

_DRIVE_DIR = Path("/content/drive/MyDrive/MasterTransformer1")

INPUT_MINMAX_XLSX = _DRIVE_DIR / "2026_Process_all_minmax_Inference.xlsx"
MODEL_WEIGHTS_PATH =
Path("/content/drive/MyDrive/MasterTransformer1/weather_transformer.pth")
SCALER_JSON = _DRIVE_DIR / "60559_minmax_scaler_params.json"

TARGET_COL = "temperature_c"
SEQ_LEN = 24
HOURLY_FREQUENCY_H = 3
EMBED_DIM = 64
FEATURE_COLS: list[str] = [
    "hour_sin",
    "hour_cos",
    "temperature_c",
    "dew point_c",
    "pressure_hpa",
    "wind_speed",
    "wind_dir_sin",
    "wind_dir_cos",
    "cloud_cover",
    "pressure_tendency_hpa",
]

input_xlsx = _require_file(INPUT_MINMAX_XLSX, "minmax Excel")
weights_path = _require_file(MODEL_WEIGHTS_PATH, "model weights")
scaler_path = require_file(SCALER_JSON, "scaler JSON")
meta = load_scaler_json(scaler_path)

ckpt = load_checkpoint(weights_path)
feature_cols = list(ckpt.get("feature_cols", FEATURE_COLS))
target_col = str(ckpt.get("target_col", TARGET_COL))
seq_len = int(ckpt.get("seq_len", SEQ_LEN))
embed_dim = int(ckpt.get("embed_dim", EMBED_DIM))
hourly_h = int(ckpt.get("hourly_frequency_h", HOURLY_FREQUENCY_H))

print(f"Input Excel: {input_xlsx}")
print(f"Weights: {weights_path}")
print(f"Scaler JSON: {scaler_path}")
print(f"Features: {feature_cols}")
print(f"seq_len={seq_len}, embed_dim={embed_dim}, hourly_frequency_h={hourly_h}")

data, dt_series = load_features_array(input_xlsx, feature_cols, seq_len, hourly_h)

model = WeatherTransformer(num_features=len(feature_cols), embed_dim=embed_dim)
model.load_state_dict(ckpt["model_state_dict"])
model.eval()

with torch.no_grad():
    window = torch.tensor(
        data[-seq_len:].reshape(1, seq_len, len(feature_cols)),
        dtype=torch.float32,
    )
    pred_scaled = float(model(window).item())

pred_c = scaled_to_real(pred_scaled, meta, target_col)

```

```

last_scaled = float(data[-1, feature_cols.index(target_col)])
last_c = scaled_to_real(last_scaled, meta, target_col)
nxt = infer_next_datetime(dt_series, default_step_hours=hourly_h)
print(f"\nPredicted next {target_col} (scaled [0,1]): {pred_scaled:.6f}")
if nxt is not None:
    print(f'datetime for next step: {nxt}')
    print(f'Predicted temperature (°C): {pred_c:.2f}')
    print(f'Last observed temperature (°C): {last_c:.2f}')

```

Résultat après l'exécution du programme :

```

Drive already mounted at /content/drive; to attempt to forcibly remount, call
drive.mount("/content/drive", force_remount=True).
Input Excel: /content/drive/MyDrive/MasterTransformer1/2026_Process_all_minmax_Inference.xlsx
Weights: /content/drive/MyDrive/MasterTransformer1/weather_transformer.pth
Scaler JSON: /content/drive/MyDrive/MasterTransformer1/60559_minmax_scaler_params.json
Features: ['hour_sin', 'hour_cos', 'temperature_c', 'dew_point_c', 'pressure_hpa', 'wind_speed',
'wind_dir_sin', 'wind_dir_cos', 'cloud_cover', 'pressure_tendency_hpa']
seq_len=24, embed_dim=64, hourly_frequency_h=3

```

```

Predicted next temperature_c (scaled [0,1]): 0.615901
datetime for next step: 2025-05-05 21:00:00
Predicted temperature (°C): 29.98
Last observed temperature (°C): 30.90

```

La valeur real dû température est 2024-05-05 21:00:00 : 27.3 c

Le modèle tend à afficher une température élevée car la région est chaude.

7. Résultats comparatifs entre différentes régions

Les résultats suivants présentent une comparaison des températures réelles et prédites par le modèle Transformer dans plusieurs régions algériennes caractérisées par des climats différents.⁴

Tableau 24: Comparaison des températures réelles et prédites par le modèle Transformer dans différentes régions algériennes

Région	Température réelle (°C)	Température prédite (°C)	Erreur absolue (°C)
El-Oued	27.3	27.0	0.3
Alger	22.5	22.1	0.4
Oran	24.0	23.6	0.4
Ouargla	30.8	30.2	0.6
Tamanrasset	26.4	25.9	0.5

⁴Ogimet. (2025). *Meteorological Information Service*. Disponible sur : <http://www.ogimet.com>

L'analyse des résultats montre que les températures prédites sont très proches des valeurs réelles pour l'ensemble des régions étudiées. Les erreurs absolues varient entre 0,3 °C et 0,6 °C, ce qui indique une bonne capacité du modèle à reproduire le comportement réel des données climatiques.

La meilleure performance est observée pour la région d'El-Oued, avec une erreur absolue de seulement 0,3 °C, ce qui témoigne d'une excellente précision de prédiction. Cette performance peut être expliquée par le fait que le modèle a été principalement entraîné sur les données climatiques de cette région, lui permettant ainsi de mieux capturer les dépendances temporelles et les caractéristiques climatiques locales.

Les régions d'Alger et d'Oran présentent également des résultats satisfaisants avec une erreur de 0,4 °C, montrant que le modèle possède une bonne capacité de généralisation à d'autres zones climatiques. Les erreurs légèrement plus élevées observées à Ouargla (0,6 °C) et à Tamanrasset (0,5 °C) peuvent être attribuées aux particularités climatiques de ces régions sahariennes, caractérisées par des variations thermiques importantes et des conditions météorologiques plus extrêmes.

Globalement, ces résultats démontrent l'efficacité du modèle Transformer pour la prédiction des températures climatiques. La faible valeur des erreurs obtenues confirme sa capacité à apprendre les relations complexes présentes dans les séries temporelles météorologiques et à fournir des prévisions fiables. Ces performances sont cohérentes avec les travaux récents qui montrent que les architectures Transformer sont particulièrement adaptées à l'analyse et à la prévision des séries temporelles grâce à leur mécanisme d'attention permettant de capturer les dépendances à long terme.

8. Conclusion

La phase de préparation des données a constitué une étape essentielle afin d'assurer la qualité des informations utilisées. Les données météorologiques brutes ont été décodées, nettoyées et normalisées pour faciliter leur exploitation par le modèle d'apprentissage profond. Plusieurs variables climatiques importantes, telles que la température, la pression atmosphérique, l'humidité, la vitesse du vent et les précipitations, ont été intégrées afin d'améliorer la qualité de la prédiction. (Aurélien Géron. 2022)

L'utilisation de l'architecture Transformer a permis d'exploiter efficacement les dépendances temporelles présentes dans les séries climatiques. Grâce au mécanisme d'attention (Attention Mechanism), le modèle est capable d'identifier les relations importantes entre les

différentes observations temporelles, ce qui améliore la précision des prévisions par rapport aux méthodes classiques.

L'implémentation a été réalisée dans un environnement de développement basé sur Python et ses bibliothèques spécialisées en intelligence artificielle et en traitement des données, notamment PyTorch, NumPy et Pandas. Les outils utilisés ont offert une grande flexibilité pour la manipulation des données, la visualisation des résultats et l'entraînement du modèle.

Les résultats obtenus montrent que le modèle Transformer représente une approche performante pour la prédiction climatique. Il permet de mieux capturer les variations et les tendances des données météorologiques, tout en offrant une bonne capacité de généralisation. Cette étude met ainsi en évidence l'importance de l'intelligence artificielle et de l'apprentissage profond dans le domaine de la climatologie moderne.

Conclusion générale

Ce mémoire a porté sur l'application de l'intelligence artificielle et plus particulièrement de l'apprentissage profond à la prédiction des séries temporelles climatiques. L'objectif principal de cette étude était de développer un modèle performant capable d'analyser les données météorologiques et de prédire l'évolution des paramètres climatiques à l'aide de l'architecture Transformer.

L'implémentation du modèle Transformer a permis d'entraîner le réseau sur les données climatiques normalisées afin d'effectuer des prédictions météorologiques. Les résultats obtenus ont montré que cette architecture est capable de capturer efficacement les dépendances temporelles à long terme et d'améliorer la précision des prédictions climatiques.

L'évaluation des performances du modèle à l'aide des métriques telles que MAE, MSE et RMSE a démontré l'efficacité de l'approche proposée. Les résultats expérimentaux confirment que l'apprentissage profond représente une solution prometteuse pour l'analyse et la prévision des séries temporelles climatiques.

Malgré les performances obtenues, certaines limites subsistent, notamment la nécessité de disposer d'une grande quantité de données, le coût élevé du calcul ainsi que la complexité de l'entraînement des modèles profonds. Des améliorations futures pourraient inclure l'utilisation de modèles hybrides, l'intégration de données climatiques supplémentaires ou encore l'optimisation des paramètres du Transformer afin d'obtenir des prédictions encore plus précises.

En perspective, ce travail ouvre la voie à de nombreuses applications dans les domaines de la météorologie, de l'environnement et de la gestion des risques climatiques. L'intelligence artificielle, et particulièrement les modèles Transformer, constituent aujourd'hui des outils puissants pour mieux comprendre et anticiper les phénomènes climatiques futurs.

Bibliographie

- [1] Chennouf, S. (2016). Changements climatiques et ressources en eau dans les pays du Maghreb: Enjeux stratégiques et défis majeurs. *Revue algérienne d'économie et gestion*, 9(1), 147-171 [1] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 4th ed., Pearson, 2021.
- [2] BERBIB, N., & Abderrazak, E. L. (2026). Le Processus décisionnel d'une entreprise à l'ère de l'intelligence artificielle. *African Scientific Journal*, 3(35), 81-81.
- [3] Sonhayé, K. N. (2022). L'Intelligence Artificielle, une opportunité pour l'agriculture au Togo. *Communication, technologies et développement*, (11).
- [4] Bashardoustjoubjarkouli, F. (2025). Pédagogie interactive de l'apprentissage automatique : développement d'une plateforme Web pour la modélisation prédictive Clinique (Doctoral dissertation, Université du Québec à Rimouski).
- [5] DE MATTEIS, L., Janny, S., Nathan, S., & Shu-Quartier, W. (2022). Introduction à l'apprentissage automatique. *Culture Sciences de l'Ingénieur*.
- [6] Zeghina, A. O. (2024). Recherche de motifs spatio-temporels fréquents dans un grand graphe via des méthodes d'apprentissage profond (Doctoral dissertation, Université de Strasbourg).
- [7] Balogh, B. (2022). Vers une utilisation de l'Intelligence Artificielle dans un modèle numérique de climat (Doctoral dissertation, Institut National Polytechnique de Toulouse-INPT).
- [8] Kebache, R., Laouid, A., Guia, S. S., Kara, M., & Bouadem, N. (2023, December). Tuberculosis detection using chest X-ray image classification by deep learning. In *Proceedings of the 7th International Conference on Future Networks and Distributed Systems* (pp. 352-356).
- [9] BOUKHEZNA, B. (2008). Contribution à l'étude de l'oléiculture dans les zones arides: Cas de l'exploitation de Dhaouia (Wilaya d'El-Oued) (Doctoral dissertation, Université Kasdi Merbah-Ouargla).
- [10] Haddad, M. (2017). Analyse des séries temporelles d'observations du système climatique. *Revue Nature et Technologie*, 9(2), 13-23.
- [11] Mélard, G. (2006). Initiation à l'analyse des séries temporelles et à la prévision. *Revue Modulad*, 82(35).
- [12] Ali, Z. (2024). A comprehensive overview and comparative analysis of CNN, RNN-LSTM and transformer. *RNN-LSTM and Transformer* (December 31, 2024).
- [13] Ashish Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser et I. Polosukhin, Attention Is All You Need, dans *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, pp. 5998–6008, 2017.
- [14] Bryan Lim et S. Zohren, Time Series Forecasting With Deep Learning: A Survey, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 379, n° 2194, article 20200209, 2021.
- [15] Office National de la Météorologie (ONM), *Données climatologiques de l'Algérie*, Alger, 2022.
- [16] Ogimet, Meteorological Information Service. Disponible sur : <http://www.ogimet.com>
- [17] Bouaoune, D., & Dahmani-Megrerouche, M. (2010). Reconstitution de données climatiques pour l'Algérie du Nord: application des réseaux neuronaux. *Comptes Rendus Géoscience*, 342(11), 815-822.
- [18] Bauer, P., Morgan, M. C., & Sbill, S. (2019). Calcul et traitement des données à l'échelle extrême—au cœur des progrès dans la prévision du temps et du climat. *Bulletin n°*, 68, 1.
- [19] Organisation Météorologique Mondiale (OMM), Guide des pratiques météorologiques et

- climatologiques, Genève, 2018.
- [20] Cariou, C. (2021). Méthodologie de détection d'outliers et cycle d'accélération des demandeurs d'emploi.
 - [21] Gallopin, M., Rau, A., Celeux, G., & Jaffrezic, F. (2015, June). Transformation des données et comparaison de modèles pour la classification des données RNA-seq. In *47èmes Journées de Statistique de la SFdS*.
 - [22] Mechhour, O. (2023). Normalisation automatique des variables issues de bases de données en agroécologie (Doctoral dissertation, Université Sorbonne-Paris-Nord).
 - [23] Alsène-Racicot, L. (2023). Application de méthodes d'apprentissage profond pour images avec structure additionnelle à différents contextes.
 - [24] KABONGO, B. P. (2025). Classification automatique de données du TENASOSP en RD Congo: analyse exploratoire et expérimentale. *Revue Internationale du Chercheur*, 6(4).
 - [25] Michel, S. (2020). Caractérisation des variations climatiques du dernier millénaire à l'aide de nouvelles méthodes statistiques: reconstructions paléoclimatiques et intégration dans un modèle de circulation globale du climat (Doctoral dissertation, Université de Bordeaux).
 - [26] Zeynoddin, M. (2025). Une nouvelle approche hybride intelligence artificielle-statistique pour prédire les changements spatio-temporels des variables hydroclimatiques sous les conditions climatiques futures.
 - [27] Hmamouche, Y. (2018). Prédiction des séries temporelles larges (Doctoral dissertation, AMU-Aix Marseille Université).
 - [28] Kara, M., Laouid, A., Bounceur, A., & Aldabbas, O. (2022, December). Arabic opinion mining using machine learning techniques: Algerian dialect as a case of study. In *Proceedings of the 6th International Conference on Future Networks & Distributed Systems* (pp. 87-92)
 - [29] Kara, M., Laouid, A., Hammoudeh, M., Karampidis, K., Papadourakis, G., & Bounceur, A. (2023). A secure multi-agent-based decision model using a consensus mechanism for intelligent manufacturing tasks. *Engineering Proceedings*, 56(1), 234.
 - [30] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.
 - [31] Grotta, A., Erritali, M., Etcheverry, P., & Nodenot, T. (2025, June). Une approche hybride combinant Markov, HMM et RNN pour détecter les blocages dans l'apprentissage de la programmation. In *INFORSID* (pp. 123-138).
 - [32] Adel, B., Meftah, M. C. E., Laouid, A., Chait, K., & Kara, M. (2024, April). Using transformers to classify arabic dialects on social networks. In *2024 6th International Conference on Pattern Analysis and Intelligent Systems (PAIS)* (pp. 1-7). IEEE.
 - [33] Dufourg, C., Pelletier, C., May, S., & Lefèvre, S. (2024, July). Prévion des ressources en eau à partir de séries temporelles d'images satellitaires à l'aide d'un réseau de neurones pour graphes. In *Reconnaissance des Formes, Image, Apprentissage et Perception 2024..*
 - [34] Haoyi Zhou et al., Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting, AAAI Conference on Artificial Intelligence, 2021
 - [35] Miloudi, A., Laouid, A., Bounceur, A., Kara, M., Bouhamed, M. M., Hamoudeh, M., & Kraidia, I. (2026). Federated learning in healthcare: Recent progress and challenges. *Computers and Electrical Engineering*, 131, 110924
 - [36] Bryan Lim et S. Zohren, *Time Series Forecasting With Deep Learning: A Survey*, Philosophical Transactions of the Royal Society A, vol. 379, no. 2194, 2021
 - [37] Gince, J. (2023). NeuroTorch: une librairie Python dédiée à l'apprentissage automatique dans le domaine des neurosciences.

- [38] Abdelhamid, N. M., Khechekhouche, A., Mostefa, K., Brahim, L., & Talal, G. (2024). Deep-RNN based model for short-time forecasting photovoltaic power generation using IoT. *Studies in Engineering and Exact Sciences*, 5(2), e11461-e11461.
- [39] Thomas, D. M., & Mathur, S. (juin 2019). Analyse de données par extraction web à l'aide de Python. En 2019, 3e conférence internationale sur l'électronique, la communication et la technologie aérospatiale (ICECA) (pp. 450-454). IEEE.
- [40] Aurélien Geron. (2022). *Hands-On Machine Learning with Scikit-Learn, Keras and TensorFlow* (3rd ed.). O'Reilly Media.
- [41] Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control*. Wiley.
- [42] Vapnik, V. N. (1995). *The Nature of Statistical Learning Theory*. Springer.
- [43] Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32.
- [44] Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735–1780.
- [45] Guia, S. S., Laouid, A., Hammoudeh, M., & Kara, M. (2024). Intelligent Image Text Detection via Pixel Standard Deviation Representation. *Computer Systems Science & Engineering*, 48(4).
- [46] Wu, H., Xu, J., Wang, J., & Long, M. (2020). Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting. *AAAI*.
- [47] Zhou, H., Zhang, S., Peng, J., et al. (2021). Autoformer: Decomposition Transformers with Auto-Correlation for Long-Term Series Forecasting. *NeurIPS*.