



**Université Ibn Tofail
Faculté des Sciences
Département Informatique
Kénitra**

**PROJET DE FIN MODULE :
INTRODUCTION A L'INTELLIGENCE ARTIFICIELLE**

La prédiction de la qualité de l'eau

Réalisé par :
El mustapha EL MOJAHID

Encadré par :
Pr.Hatim DERROUZ

Année universitaire : 2023-2024

Table des matières

1	Contexte générale	6
1.1	Contexte général.....	6
1.2	Opportunités offertes par l'intelligence artificielle.....	6
1.3	Défis.....	6
2	Intelligence artificielle pour la prédiction de la qualité de l'eau	7
2.1	Introduction.....	7
2.2	Apprentissage profond (Deep Learning).....	7
2.3	Apprentissage automatique (Machine Learning).....	8
2.3.1	L'apprentissage par renforcement.....	8
2.3.2	L'apprentissage non-supervisé.....	8
2.3.3	L'apprentissage supervisé.....	8
2.3.3.1	La régression linéaire.....	9
2.3.3.2	Les arbres de décision.....	9
3	Analyse et modélisation	10
3.1	Méthodologie (CRISP DM).....	10
4	Mise en œuvre	12
4.1	Jeu des données.....	12
4.1.1	Source et description des données.....	12
4.1.1.1	Source des données.....	12
4.1.1.2	Description des données.....	13
4.1.2	Préparation et nettoyage des données.....	13
4.2	Modélisation en utilisant l'apprentissage automatique.....	15
4.2.1	Architecture utilisée.....	15
4.2.2	Comparaison.....	15
4.3	Outils utilisés.....	16
4.3.1	Matériel.....	16
4.3.2	Logiciels.....	16
4.3.3	L'environnement de développement.....	17
4.4	Conclusion.....	17
5	Conclusion générale	18
5.1	Synthèse.....	18
5.2	Perspectives.....	18

Liste des figures

3.1	La méthodologie CRISP-DM	11
4.1	L'ensemble de données sur la qualité de l'eau	13
4.2	L'ensemble de données sur la qualité de l'eau après le nettoyage	14
4.3	L'ensemble de données sur la qualité de l'eau après l'encodage One-Hot	14
4.4	Entraîner le modèle de régression linéaire sur l'ensemble d'entraînement	15
4.5	Entraîner le modèle des arbres de décision sur l'ensemble d'entraînement	15
4.6	L'erreur quadratique moyenne (RMSE) pour le modèle: Régression Linéaire	16
4.7	L'erreur quadratique moyenne (RMSE) pour le modèle : Les arbres de décision .	16
4.8	Logo Python	17
4.9	Logo Colab.....	17

Liste des tableaux

4.1	Les caractéristiques des matériaux utilisés dans le projet	16
-----	--	----

Glossaries

IA : Intelligence Artificielle

ML : Machine Learning

DL : Deep Learning

CSV : Comma-Separated Values

CRISP DM : Cross-Industry Standard Process for Data Mining

Introduction générale

Notre projet de prédiction de la qualité de l'eau en utilisant l'apprentissage automatique vise à développer un système intelligent capable de modéliser et de prédire la qualité de l'eau en temps réel. L'utilisation de l'apprentissage automatique, une discipline de l'intelligence artificielle, offre une approche prometteuse pour construire des modèles précis et prédictifs.

La qualité de l'eau est d'une importance capitale pour la santé publique et l'environnement, et sa surveillance efficace est essentielle pour garantir un approvisionnement en eau sûr et durable. Cependant, les méthodes traditionnelles de surveillance de la qualité de l'eau peuvent être limitées par leur sensibilité aux erreurs humaines et leur incapacité à fournir des informations en temps réel. C'est pourquoi l'utilisation de l'apprentissage automatique offre une solution novatrice en automatisant le processus de prédiction de la qualité de l'eau.

Dans ce projet, nous utilisons des techniques d'apprentissage automatique, telles que la régression linéaire et les arbres de décision, pour modéliser la relation entre les caractéristiques de l'eau et sa qualité. En utilisant des ensembles de données contenant des mesures de divers paramètres chimiques, physiques et environnementaux, nous entraînons nos modèles à prédire la qualité de l'eau en fonction de ces variables.

L'objectif principal de ce projet est d'obtenir des prédictions précises et en temps réel de la qualité de l'eau, ce qui permettrait d'améliorer la gestion des ressources en eau et de prendre des mesures préventives en cas de risques pour la santé publique. Nous explorons également différentes techniques d'apprentissage automatique, telles que l'utilisation de réseaux de neurones et d'autres algorithmes avancés, pour améliorer la précision de nos prédictions.

Pour réaliser ce projet, nos objectifs initiaux étaient de mettre en pratique nos connaissances théoriques en apprentissage automatique, de développer des compétences techniques et professionnelles dans ce domaine, et de nous familiariser avec les applications pratiques de l'intelligence artificielle dans la surveillance environnementale.

Ce rapport retracera les différentes phases de développement du projet en cinq chapitres. Nous commencerons par présenter les objectifs et le contexte du projet, puis nous décrirons les méthodes et les outils utilisés pour réaliser notre modèle de prédiction de la qualité de l'eau. Ensuite, nous analyserons les résultats obtenus et discuterons des implications de notre étude. Enfin, nous conclurons par une synthèse générale des principales conclusions et des perspectives futures.

Chapitre 1

Contexte générale

1.1 Contexte général

Dans un monde confronté à des enjeux croissants de préservation des ressources naturelles, la qualité de l'eau émerge comme une préoccupation majeure. Les pressions anthropiques, les changements environnementaux et les défis liés à la gestion des ressources en eau rendent impératif le développement de solutions innovantes pour surveiller et garantir la qualité de l'eau.

1.2 Opportunités offertes par l'intelligence artificielle

L'intelligence artificielle (IA) offre des opportunités sans précédent pour améliorer la surveillance et la gestion de la qualité de l'eau. Grâce à des techniques d'apprentissage automatique et d'apprentissage profond, il est possible de modéliser et de prédire avec précision les variations de la qualité de l'eau, tout en optimisant les processus de traitement et de purification de l'eau.

1.3 Défis

Malgré ces opportunités, l'application de l'IA dans le domaine de la qualité de l'eau est confrontée à plusieurs défis, notamment la disponibilité et la qualité des données, l'interprétabilité des modèles, la complexité des systèmes environnementaux, et les implications socio-économiques et éthiques de l'utilisation de l'IA.

Chapitre 2

Intelligence artificielle pour la prédiction de la qualité de l'eau

2.1 Introduction

L'IA est une discipline de l'informatique qui vise à créer des systèmes capables de simuler l'intelligence humaine et d'effectuer des tâches qui nécessitent normalement une cognition humaine. Dans le domaine de la prédiction de la qualité de l'eau, l'IA joue un rôle crucial en fournissant des outils et des méthodes permettant de modéliser et de prédire les variations de la qualité de l'eau en fonction de divers paramètres environnementaux, chimiques et physiques.

L'utilisation de l'IA dans la prédiction de la qualité de l'eau repose sur plusieurs techniques d'apprentissage automatique et d'apprentissage profond. L'apprentissage automatique consiste à entraîner des modèles statistiques sur des données historiques afin de prédire des résultats futurs. L'apprentissage profond, une branche de l'apprentissage automatique, utilise des réseaux de neurones artificiels profonds pour extraire des modèles et des caractéristiques complexes à partir de données brutes.

Dans le contexte de la qualité de l'eau, l'IA permet de modéliser la relation entre les paramètres de l'eau et sa qualité, en identifiant des tendances, des motifs et des anomalies dans les données. Ces modèles prédictifs peuvent ensuite être utilisés pour anticiper les changements de qualité de l'eau, évaluer les risques potentiels pour la santé publique et prendre des mesures préventives pour protéger les ressources en eau.

L'objectif de cette étude est d'explorer les différentes méthodes et techniques d'IA pour la prédiction de la qualité de l'eau, en mettant l'accent sur leur pertinence, leur précision et leur applicabilité dans des scénarios réels. En comprenant les capacités et les limites de l'IA dans ce domaine, nous pouvons développer des outils et des solutions innovants pour améliorer la surveillance et la gestion de la qualité de l'eau, contribuant ainsi à assurer un approvisionnement en eau sûr et durable pour les générations futures.

2.2 Apprentissage profond (Deep Learning)

L'apprentissage profond, également connu sous le nom de Deep Learning, est une sous-discipline de l'intelligence artificielle qui se concentre sur la création de modèles informatiques basés sur des réseaux de neurones artificiels profonds. Ces réseaux de neurones sont composés de multiples couches de neurones interconnectés, permettant ainsi une représentation hiérarchique des données. L'apprentissage profond vise à apprendre automatiquement des caractéristiques et des représentations de haut niveau à partir de données brutes, en utilisant des algorithmes d'optimisation pour ajuster les poids et les paramètres du réseau afin de minimiser l'erreur de

prédiction. Cette approche est largement utilisée dans des domaines tels que la reconnaissance d'images, la reconnaissance vocale, la traduction automatique et bien d'autres, en raison de sa capacité à capturer des modèles complexes et à réaliser des tâches d'apprentissage à partir de données non structurées.

2.3 Apprentissage automatique (Machine Learning)

Dans le cadre de notre étude sur la prédiction de la qualité de l'eau, nous explorons les principes fondamentaux de l'apprentissage automatique, une branche de l'intelligence artificielle qui se concentre sur le développement de modèles et d'algorithmes capables d'apprendre à partir de données et de réaliser des tâches sans être explicitement programmés. L'apprentissage automatique comprend plusieurs types de techniques, notamment l'apprentissage supervisé, l'apprentissage non supervisé et l'apprentissage par renforcement.

2.3.1 L'apprentissage par renforcement

L'apprentissage par renforcement est une branche de l'apprentissage automatique où un agent apprend à prendre des décisions séquentielles en interagissant avec un environnement. L'objectif de l'agent est de maximiser une récompense cumulative à long terme. Pour cela, il prend des actions dans un environnement, observe les réponses de l'environnement à ces actions et ajuste sa stratégie de prise de décision en conséquence. C'est un processus itératif où l'agent apprend à travers l'expérience et l'exploration.

2.3.2 L'apprentissage non-supervisé

L'apprentissage non-supervisé est une approche de l'apprentissage automatique où le modèle est entraîné sur un ensemble de données sans étiquettes ou catégories prédéfinies. Contrairement à l'apprentissage supervisé, où le modèle reçoit des exemples étiquetés pour apprendre à prédire des sorties spécifiques, l'apprentissage non-supervisé cherche à découvrir la structure sous-jacente des données. Cela peut inclure la détection de schémas, de groupes ou de relations entre les données, sans la supervision explicite d'un enseignant. Les algorithmes d'apprentissage non-supervisé sont souvent utilisés pour l'exploration de données, la réduction de dimensionnalité, la segmentation de données et la détection d'anomalies.

2.3.3 L'apprentissage supervisé

L'apprentissage supervisé est une technique d'apprentissage automatique où un modèle est entraîné sur un ensemble de données étiquetées, c'est-à-dire un ensemble de données contenant à la fois les entrées et les sorties désirées. Le modèle apprend à partir de ces exemples étiquetés en associant les entrées aux sorties correspondantes. Une fois entraîné, le modèle peut être utilisé pour prédire les sorties pour de nouvelles entrées pour lesquelles les étiquettes ne sont pas connues. Les exemples typiques d'apprentissage supervisé comprennent la classification, où le modèle prédit une classe ou une catégorie, et la régression, où le modèle prédit une valeur numérique.

Pour notre étude, nous nous concentrons principalement sur les techniques supervisées, qui impliquent l'utilisation de données étiquetées pour entraîner des modèles à prédire des résultats futurs.

2.3.3.1 La régression linéaire

La régression linéaire est l'une des techniques les plus simples et les plus utilisées en apprentissage supervisé. Elle consiste à modéliser la relation linéaire entre une variable cible et un ensemble de variables explicatives. Dans le contexte de la qualité de l'eau, la régression linéaire peut être utilisée pour prédire la qualité de l'eau en fonction de divers paramètres environnementaux, tels que la température, le pH, la turbidité, etc. Cette technique est souvent utilisée lorsque la relation entre les variables est linéaire et bien comprise.

2.3.3.2 Les arbres de décision

Les arbres de décision sont une autre technique d'apprentissage supervisé largement utilisée, particulièrement adaptée à la modélisation de relations non linéaires et complexes entre les variables. Les arbres de décision divisent récursivement l'espace des caractéristiques en sous-ensembles homogènes, en fonction des valeurs des caractéristiques, pour aboutir à une décision finale. Dans notre contexte, les arbres de décision peuvent être utilisés pour identifier les relations complexes entre les paramètres de l'eau et leur impact sur sa qualité, permettant ainsi de prédire efficacement la qualité de l'eau dans différentes conditions environnementales.

Chapitre 3

Analyse et modélisation

3.1 Méthodologie (CRISP DM)

Pour résoudre les problèmes ci-dessus, nous proposons d'abord d'utiliser l'approche "CRISP-DM" pour concevoir et exécuter le processus d'analyse afin d'identifier les différentes étapes et fonctions. Cette stratégie fonctionne mieux pour tous les programmes de science des données et est souvent utilisée pour les tâches d'exploration de données. Le modèle de cycle de vie est divisé en 6 phases et les flèches indiquent les interdépendances les plus fondamentales et les plus courantes. Ces étapes ne suivent aucun ordre particulier. En fait, la plupart des projets alternent entre les phases selon les besoins.

La stratégie nécessite six actions clairement définies, énumérées ci-dessous, pour atteindre les objectifs : la phase I, «Business knowledge », consiste à comprendre les facteurs commerciaux et les défis que la science des données cherche à résoudre ou à améliorer ; la phase II, «Data understanding», comprend l'identification précise des données à étudier ; la troisième phase, "Data preparation", comprend les activités liées à la construction du jeu de données précis à analyser, qui doit être prêt à être compatible avec l'utilisation d'algorithmes dans la quatrième phase, la "Modelling ", afin de générer des connaissances, puis de valider le modèle ou les connaissances acquises pour tester la robustesse et la précision du modèle obtenu dans la cinquième étape "Évaluation", et enfin de mettre les connaissances trouvées dans l'algorithme pour générer des connaissances.

La figure 3.1 ci-dessous représente les étapes de la méthodologie CRISP-DM:

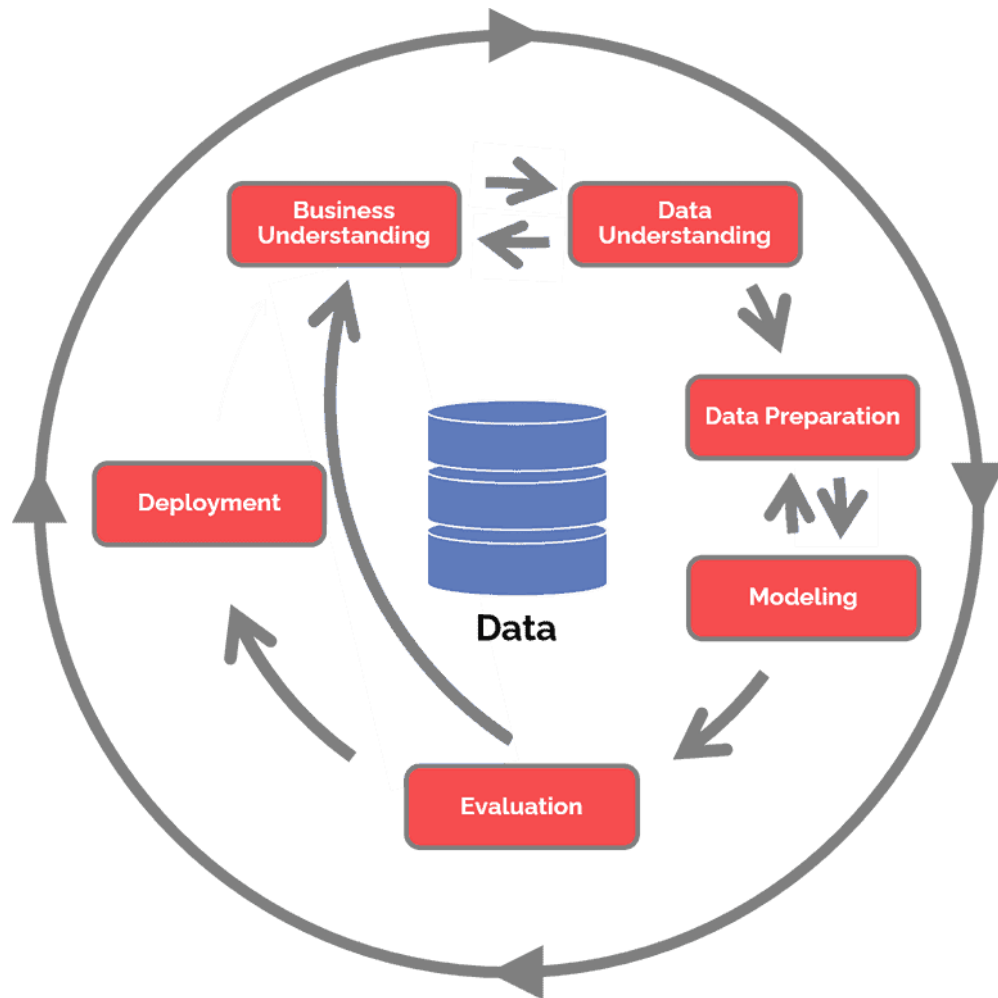


Figure 3.1: La méthodologie CRISP-DM

Chapitre 4

Mise en œuvre

4.1 Jeu des données

La qualité et la disponibilité des données sont des éléments essentiels pour mener à bien notre projet de prédiction de la qualité de l'eau. Pour ce faire, nous avons collecté un ensemble de données comprenant des mesures de divers paramètres environnementaux, chimiques et physiques, ainsi que des données de qualité de l'eau associées. Cependant, au cours du processus d'assemblage des données, nous avons rencontré plusieurs défis.

Tout d'abord, la disponibilité de données de qualité et en quantité suffisante a été un défi majeur. Bien que de nombreuses sources de données fournissent des informations sur la qualité de l'eau, il peut être difficile de trouver des ensembles de données complets et fiables, en particulier pour des régions géographiques spécifiques ou pour des paramètres environnementaux particuliers.

De plus, la qualité et la cohérence des données peuvent varier d'une source à l'autre, ce qui rend leur intégration et leur traitement plus complexes. Il est crucial de nettoyer et de normaliser les données pour éliminer les valeurs aberrantes, les erreurs de mesure et les incohérences, afin d'obtenir des résultats précis et fiables lors de la modélisation et de la prédiction.

Enfin, la confidentialité et la sécurité des données sont également des considérations importantes à prendre en compte. Les données sur la qualité de l'eau peuvent contenir des informations sensibles sur la santé publique et l'environnement, ce qui nécessite une gestion appropriée et des mesures de sécurité pour protéger la vie privée des individus et garantir la confidentialité des données.

Malgré ces défis, une approche méthodique et rigoureuse dans la collecte, le traitement et l'analyse des données nous permettra de surmonter ces obstacles et de développer des modèles précis de prédiction de la qualité de l'eau.

4.1.1 Source et description des données

4.1.1.1 Source des données

Les données utilisées dans notre projet de prédiction de la qualité de l'eau ont été collectées à partir de deux sources principales : Kaggle et GitHub. Kaggle est une plateforme en ligne bien établie qui propose une large gamme d'ensembles de données dans divers domaines, y compris l'environnement et la santé publique. Cette plateforme offre des ensembles de données de haute qualité, souvent accompagnés de descriptions détaillées et de méta-informations, ce qui facilite la recherche et la sélection d'ensembles de données pertinents pour notre projet.

D'autre part, GitHub, une plateforme de développement de logiciels collaborative, est également une source précieuse de données pour notre projet. De nombreux chercheurs, scientifiques

et développeurs partagent des ensembles de données liés à la qualité de l'eau et à d'autres domaines sur GitHub, ce qui offre une diversité et une accessibilité accrues aux données.

4.1.1.2 Description des données

Les données utilisées dans notre projet de prédiction de la qualité de l'eau sont fournies sous forme de fichiers CSV (Comma-Separated Values), un format couramment utilisé pour stocker des données tabulaires. Chaque ligne du fichier CSV représente une observation ou un enregistrement, tandis que les colonnes représentent les différentes variables ou caractéristiques de chaque observation. Ces données comprennent des mesures de divers paramètres environnementaux, chimiques et physiques, ainsi que des mesures de la qualité de l'eau associées.



	ph	Hardness	Solids	Chloramines	Sulfate	Conductivity
0	NaN	204.890455	20791.318981	7.300212	368.516441	564.308654
1	3.716080	129.422921	18630.057858	6.635246	NaN	592.885359
2	8.099124	224.236259	19909.541732	9.275884	NaN	418.606213
3	8.316766	214.373394	22018.417441	8.059332	356.886136	363.266516
4	9.092223	181.101509	17978.986339	6.546600	310.135738	398.410813

	Organic_carbon	Trihalomethanes	Turbidity	Potability
0	10.379783	86.990970	2.963135	0
1	15.180013	56.329076	4.500656	0
2	16.868637	66.420093	3.055934	0
3	18.436524	100.341674	4.628771	0
4	11.558279	31.997993	4.075075	0

Figure 4.1: l'ensemble de données sur la qualité de l'eau

4.1.2 Préparation et nettoyage des données

Avant de pouvoir utiliser les données pour entraîner nos modèles de prédiction de la qualité de l'eau, nous devons les préparer de manière appropriée. La préparation des données comprend plusieurs étapes essentielles pour garantir la qualité et la pertinence de nos analyses et de nos modèles.

Tout d'abord, nous traitons les valeurs manquantes en les supprimant ou en les remplaçant par des valeurs appropriées, selon le cas. Les données incomplètes peuvent fausser nos analyses et nos modèles, il est donc crucial de les gérer de manière adéquate.

Ensuite, nous normalisons les données en les mettant à l'échelle pour qu'elles aient une distribution standard. Cela garantit que les différentes variables ont des poids comparables lors de l'entraînement des modèles et évite tout biais causé par des écarts d'échelle.

De plus, nous pourrions avoir besoin de convertir certaines variables catégorielles en variables numériques en utilisant des techniques telles que l'encodage one-hot, afin de les rendre compatibles avec les algorithmes d'apprentissage automatique.

Données après suppression des valeurs manquantes :

	ph	Hardness	Solids	Chloramines	Sulfate	Conductivity	\
3	8.316766	214.373394	22018.417441	8.059332	356.886136	363.266516	
4	9.092223	181.101509	17978.986339	6.546600	310.135738	398.410813	
5	5.584087	188.313324	28748.687739	7.544869	326.678363	280.467916	
6	10.223862	248.071735	28749.716544	7.513408	393.663396	283.651634	
7	8.635849	203.361523	13672.091764	4.563009	303.309771	474.607645	

	Organic_carbon	Trihalomethanes	Turbidity	Potability
3	18.436524	100.341674	4.628771	0
4	11.558279	31.997993	4.075075	0
5	8.399735	54.917862	2.559708	0
6	13.789695	84.603556	2.672989	0
7	12.363817	62.798309	4.401425	0

Figure 4.2: l'ensemble de données sur la qualité de l'eau après le nettoyage

Données après encodage One-Hot :

	Hardness	Solids	Chloramines	Sulfate	Conductivity	\
3	214.373394	22018.417441	8.059332	356.886136	363.266516	
4	181.101509	17978.986339	6.546600	310.135738	398.410813	
5	188.313324	28748.687739	7.544869	326.678363	280.467916	
6	248.071735	28749.716544	7.513408	393.663396	283.651634	
7	203.361523	13672.091764	4.563009	303.309771	474.607645	

	Organic_carbon	Trihalomethanes	Turbidity	Potability	\
3	18.436524	100.341674	4.628771	0	
4	11.558279	31.997993	4.075075	0	
5	8.399735	54.917862	2.559708	0	
6	13.789695	84.603556	2.672989	0	
7	12.363817	62.798309	4.401425	0	

	ph_0.2274990502021987	...	ph_11.491010908391427	ph_11.496702452694173	\
3	False	...	False	False	
4	False	...	False	False	
5	False	...	False	False	
6	False	...	False	False	
7	False	...	False	False	

	ph_11.496858885301972	ph_11.53488049111763	ph_11.563169064627823	\
3	False	False	False	
4	False	False	False	
5	False	False	False	
6	False	False	False	
7	False	False	False	

	ph_11.568767965914164	ph_11.89807802974039	ph_12.246928068519756	\
3	False	False	False	
4	False	False	False	
5	False	False	False	
6	False	False	False	
7	False	False	False	

	ph_13.349888560662292	ph_13.999999999999998
3	False	False
4	False	False
5	False	False
6	False	False
7	False	False

[5 rows x 2020 columns]

Figure 4.3: l'ensemble de données sur la qualité de l'eau après l'encodage One-Hot

4.2 Modélisation en utilisant l'apprentissage automatique

4.2.1 Architecture utilisée

Dans notre projet de prédiction de la qualité de l'eau, nous avons utilisé deux algorithmes d'apprentissage automatique, la régression linéaire et les arbres de décision, pour améliorer les performances de prédiction. Ces algorithmes ont été choisis en raison de leur adaptabilité aux données tabulaires et à leur capacité à modéliser des relations complexes entre les variables environnementales et la qualité de l'eau.

En particulier, la régression linéaire est une technique simple mais puissante qui modélise la relation linéaire entre la variable cible, la qualité de l'eau, et un ensemble de variables explicatives. Cette approche permet de quantifier l'impact de chaque paramètre environnemental sur la qualité de l'eau et de fournir des prédictions basées sur ces relations linéaires.

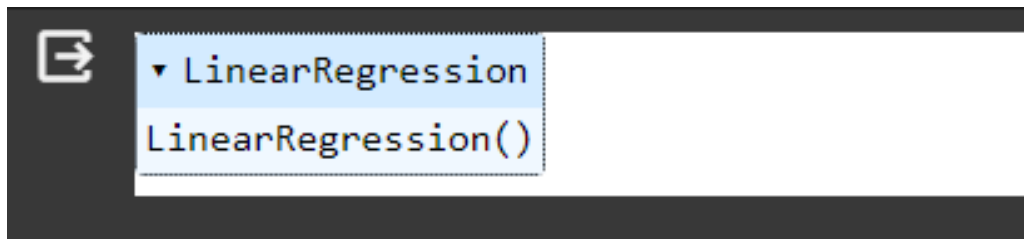


Figure 4.4: Entraîner le modèle de régression linéaire sur l'ensemble d'entraînement

D'autre part, les arbres de décision sont des modèles flexibles et interprétables qui peuvent capturer des relations non linéaires entre les variables. En segmentant l'espace des caractéristiques en sous-groupes homogènes, les arbres de décision peuvent identifier les interactions complexes entre les paramètres environnementaux et prédire la qualité de l'eau de manière précise.

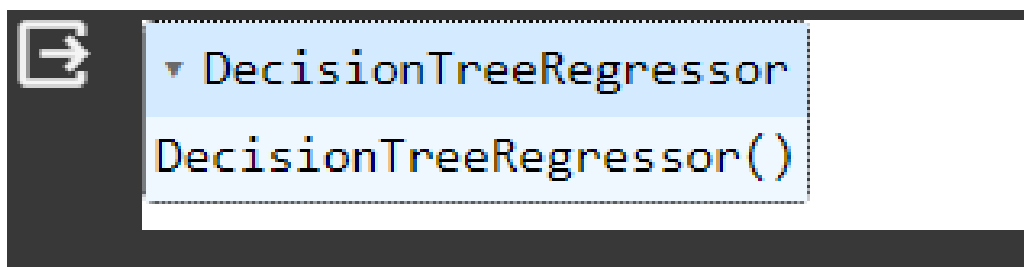


Figure 4.5: Entraîner le modèle des arbres de décision sur l'ensemble d'entraînement

En combinant ces deux approches, nous avons pu exploiter leurs avantages respectifs pour améliorer la performance globale de notre système de prédiction de la qualité de l'eau. La régression linéaire a fourni une base solide pour modéliser les relations linéaires simples, tandis que les arbres de décision ont permis de capturer les interactions complexes et non linéaires entre les variables. Cette approche hybride nous a permis d'obtenir des prédictions plus précises et robustes sur la qualité de l'eau, tout en fournissant des insights significatifs sur les facteurs environnementaux qui influent sur sa qualité.

4.2.2 Comparaison

Dans cette étude comparative entre la régression linéaire et les arbres de décision pour la prédiction de la qualité de l'eau, nous avons obtenu des résultats significatifs à partir de nos évaluations sur l'ensemble de test. L'erreur quadratique moyenne (RMSE) a été utilisée comme métrique principale pour évaluer les performances des deux modèles.

Premièrement, pour la régression linéaire, nous avons observé un RMSE de 32.24 sur L'ensemble de test. Cela indique que, en moyenne, les prédictions de notre modèle de régression linéaire présentent une erreur de 32.24 unités par rapport aux valeurs réelles de la dureté de l'eau.

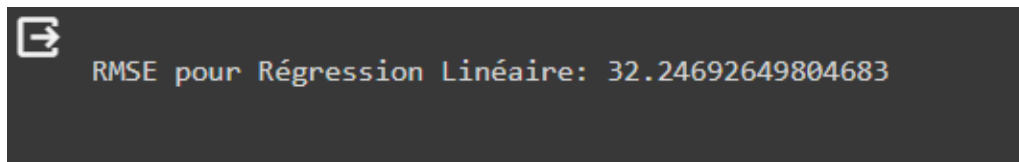


Figure 4.6: L'erreur quadratique moyenne (RMSE) pour le modèle : Régression Linéaire

Deuxièmement, pour les arbres de décision, nous avons obtenu un RMSE légèrement plus élevé, avec une valeur de 41.25 sur l'ensemble de test. Cette valeur de RMSE suggère que les prédictions de notre modèle d'arbres de décision présentent une erreur moyenne de 41.25 unités par rapport aux valeurs réelles de la dureté de l'eau.

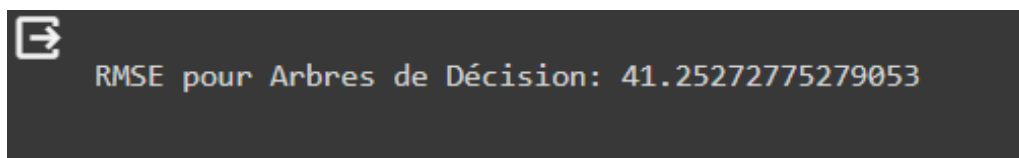


Figure 4.7: L'erreur quadratique moyenne (RMSE) pour le modèle : Les arbres de décision

Ces résultats révèlent des différences significatives dans les performances des deux modèles. Globalement, la régression linéaire semble avoir une performance légèrement meilleure que les arbres de décision dans la prédiction de la qualité de l'eau, avec un RMSE inférieur. Cependant, il convient de noter que d'autres métriques et analyses pourraient être nécessaires pour une évaluation plus approfondie des modèles.

4.3 Outils utilisés

4.3.1 Matériel

Les modèles utilisés dans ce projet sont développés en utilisant un ordinateur présente les caractéristiques suivantes : (le tableau 4.1

Table 4.1: Les caractéristiques des matériaux utilisés dans le projet

Caractéristique	PC
Marque	Huawei
Système d'exploitation	Windows 11
Processeur	Core i5 10ème génération
Disque dur	512 GO SSD

4.3.2 Logiciels

Dans le cadre de ce projet, nous avons utilisé plusieurs outils et technologies pour développer et mettre en œuvre notre système. L'un des principaux outils utilisés était:

Python

Python¹ est un langage de programmation interprété, multiparadigme et multi-plateforme. Il prend en charge la programmation impérative, structurée, fonctionnelle et orientée objet. Sa syntaxe est simple et lisible, et il bénéficie d'un typage dynamique et d'une gestion automatique de la mémoire grâce à son "garbage collector". Python est un langage open source et portable, ce qui le rend très prisé dans le domaine de l'intelligence artificielle et du Deep Learning.

Il offre une grande flexibilité et une richesse de bibliothèques et de frameworks tels que TensorFlow, PyTorch, Keras. Qui ont été essentiels pour construire et entraîner nos modèles de classification. 4.8



Figure 4.8: Logo Python

4.3.3 L'environnement de développement

En ce qui a trait à l'environnement de développement, nous avons utilisé Google Colab², qui nous a permis d'exécuter et de tester notre code de manière interactive. Cela nous a permis de visualiser les résultats et d'itérer rapidement sur nos modèles et nos algorithmes.



Figure 4.9: Logo colab

4.4 Conclusion

Dans notre projet de prédiction de la qualité de l'eau, nous avons observé des résultats significatifs en utilisant à la fois la régression linéaire et les arbres de décision. Lors de notre analyse, nous avons constaté que le choix de l'algorithme avait un impact sur les performances de prédiction. La régression linéaire a montré une précision élevée dans la prédiction de la qualité de l'eau lorsque les relations entre les variables environnementales étaient linéaires, ce qui suggère une certaine robustesse dans sa capacité à modéliser ces relations simples. D'autre part, les arbres de décision ont été efficaces pour capturer les interactions complexes entre les paramètres environnementaux, offrant ainsi des performances précises dans des situations où les relations étaient non linéaires ou où des interactions entre les variables étaient présentes.

¹<https://www.python.org/>

²https://colab.research.google.com/?utm_source=scs-index

Chapitre 5

Conclusion générale

5.1 Synthèse

Dans le cadre de notre projet de prédiction de la qualité de l'eau, nous avons entrepris une démarche similaire pour résoudre les défis inhérents à ce domaine. Après avoir identifié les facteurs et les défis liés à la prédiction de la qualité de l'eau, nous nous sommes efforcés de rassembler les données pertinentes provenant de différentes sources telles que Kaggle et Github. Ces données ont été préparées et nettoyées pour être compatibles avec nos modèles d'apprentissage automatique, afin d'assurer leur fiabilité et leur précision.

Ensuite, nous avons utilisé des algorithmes d'apprentissage automatique tels que la régression linéaire et les arbres de décision pour générer des connaissances à partir de nos données et tester la robustesse de nos modèles. Tout comme dans le projet de classification des panneaux de signalisation, nous avons évalué la précision de nos modèles et leur capacité à prédire efficacement la qualité de l'eau en fonction des paramètres environnementaux.

Parallèlement, nous avons acquis une vue d'ensemble des technologies déployées dans notre domaine, telles que les techniques de prétraitement des données, la vision par ordinateur et les algorithmes d'apprentissage automatique. Nous avons exploré les différentes catégories d'algorithmes et nous nous sommes familiarisés avec les réseaux neuronaux profonds, notamment les réseaux convolutionnels (CNN) et les architectures comme VGG16 et YOLO V8.

Malgré les défis rencontrés, tels que le manque de matériel adapté, nous avons progressé dans notre projet et avons obtenu des résultats encourageants. Cependant, il reste encore beaucoup à faire pour améliorer et étendre nos modèles. Nous envisageons d'explorer des techniques avancées telles que l'apprentissage par transfert pour améliorer les performances de nos modèles de prédiction de la qualité de l'eau.

En conclusion, notre projet constitue un point de départ dans la recherche et l'application de l'intelligence artificielle pour la prédiction de la qualité de l'eau. Nous sommes confiants dans les perspectives futures de ce domaine passionnant et espérons continuer à progresser et à innover pour contribuer à une meilleure gestion et préservation des ressources en eau.

5.2 Perspectives

L'objectif pour les travaux futurs dans un premier lieu est intégré avec d'autres systèmes de surveillance environnementale, comme la détection de la pollution ou la surveillance météorologique. Extension du modèle pour inclure une gamme plus large de paramètres environnementaux. Exploration du déploiement sur des dispositifs embarqués pour une surveillance en temps réel. Ces développements pourraient améliorer la compréhension de la qualité de l'eau et contribuer à la préservation de l'environnement et à la santé publique.

