

Higher National School of Hydraulic

The Library

Digital Repository of ENSH



المدرسة الوطنية العليا للري

المكتبة

المستودع الرقمي للمدرسة العليا للري



The title (العنوان):

Prédiction de la qualité de l'eau en utilisant des méthodes machine learning.

The paper document Shelf mark (الشفرة) : 6-0046-21

APA Citation (توثيق APA):

Babesse, Ahmed Anis (2021). Prédiction de la qualité de l'eau en utilisant des méthodes machine learning.[Thèse de master, ENSH].

The digital repository of the Higher National School for Hydraulics "Digital Repository of ENSH" is a platform for valuing the scientific production of the school's teachers and researchers.

Digital Repository of ENSH aims to limit scientific production, whether published or unpublished (theses, pedagogical publications, periodical articles, books...) and broadcasting it online.

Digital Repository of ENSH is built on the open DSpace software platform and is managed by the Library of the National Higher School for Hydraulics. <http://dspace.ensh.dz/jspui/>

المستودع الرقمي للمدرسة الوطنية العليا للري هو منصة خاصة بثمين الإنتاج العلمي لأساتذة و باحثي المدرسة.

يهدف المستودع الرقمي للمدرسة إلى حصر الإنتاج العلمي سواء كان منشورا أو غير منشور (أطروحات، مطبوعات، مبداعات، مقالات، دوريات، كتب....) و بثه على الخط.

المستودع الرقمي للمدرسة مبني على المنصة المفتوحة DSpace و يتم إدارته من طرف مديرية المكتبة للمدرسة العليا للري.

كل الحقوق محفوظة للمدرسة الوطنية العليا للري.



## MEMOIRE DE MASTER

*Pour l'obtention du diplôme de Master en Hydraulique*

**Option:ALIMENTATION EN EAU POTABLE**

**THEME :**

**PREDICTION DE LA QUALITE DE L'EAU EN UTILISANT  
DES METHODES MACHINE LEARNING**

**Présenté par :**

**BABESSE Ahmed Anis**

**Devant les membres du jury**

<b>Nom et Prénoms</b>	<b>Grade</b>	<b>Qualité</b>
MOKRANE Wahiba	M.C.B	Président
DERNOUNI YOUCEF	M.A.A	Examineur
AMIRI Djamel	M.A.A	Examineur
BENLATRECH Tarek	M.A.A	Examineur
KAHLERRAS Malika	M.C.B	Promoteur

**Session 2020/2021**

## ملخص :

في السنوات الأخيرة، شكلت الملوثات المختلفة تهديدا لنوعية المياه ولذلك أصبح وضع نماذج نوعية المياه و التنبؤ بها أمرا أساسيا في مكافحة التلوث ؛ في هذه الدراسة المقارنة يتم شرح خوارزميات الذكاء الاصطناعي المتقدمة، بما في ذلك كيفية التنبؤ بنوعية المياه باستخدام مؤشر جودة المياه ، و تصنيف جودة المياه ، تحدثنا على وجه التحديد عن الشبكات العصبية الاصطناعية و أخذنا مثالين لنماذج بالنسبة لمؤشر الجودة ، حيث تم السخدا م الخوارزميا الخاصة بالتعلم العميق ، التنظيم الذاتي للشبكة العصبية الغير خطية و الذاكرة طويلة-قصيرة الأمد ، و تضم مجموعة البيانات سبعة مؤشرات هامة.جرى تقييم النتائج باستخدام تحليلات إحصائية و كشفت النتائج ان النماذج المقترحة قادرة على التنبؤ بدقة و أن بإمكان هذا النوع من البحوث الواعدة أن يساهم كثيرا في إدارة الموارد المائية.

## الكلمات المفتاحية:

المدى قصير, خوارزمية الذاكرة طويلة, خوارزمية غير خطية, تصنيف جودة الماء, الشبكات العصبية, الجودة, مؤشر, تنبؤ

## Résumé

Ces dernières années, divers polluants ont représenté une menace pour la qualité de l'eau. Par conséquent, la modélisation et la prévision de la qualité de l'eau sont devenues essentielles dans la lutte contre la pollution. Les algorithmes avancés d'intelligence artificielle (IA) sont expliqués dans cette étude comparative, y compris la façon dont ils prédisent la qualité de l'eau à l'aide de l'indice de qualité de l'eau (IQE) et de la classification de la qualité de l'eau (CQE). Nous avons spécifiquement parlé des réseaux neuronaux artificiels et nous avons pris deux exemples de modèles. Pour l'IQE, les algorithmes d'apprentissage profond réseau neuronal non linéaire autorégressif (NARNET) et mémoire à long terme (LSTM) ont été utilisés. L'ensemble de données utilisé comporte sept paramètres significatifs, et les modèles développés ont été évalués à l'aide de paramètres statistiques. Les résultats ont révélé que les modèles proposés sont capables de prédire avec précision l'IQE et de classer la qualité de l'eau en fonction de la robustesse. D'après les résultats, le modèle NARNET est légèrement plus performant que le modèle LSTM pour prédire les valeurs de l'IQE. En outre, les modèles NARNET et LSTM ont atteint une précision similaire pour la phase de test avec une légère différence dans le coefficient de régression ( $R_{\text{NARNET}}=96:17\%$  et  $R_{\text{LSTM}}=94:21\%$ ). Ce type de recherche prometteur peut contribuer de manière significative à la gestion de l'eau.

**Mots clés:** Prédiction, IQE, CQE, ANN, Qualité de l'eau, NARNET, LSTM

## **Abstract**

Various pollutants have presented a threat to water quality in recent years. As a result, water quality modeling and prediction have become critical in the fight against pollution. Advanced artificial intelligence (AI) algorithms are explained in this comparative study, including how they predict water quality using the water quality index (WQI) and water quality classification (WQC). Specifically, we talked about artificial neural networks and we took two examples of the models. For the WQI, the deep learning algorithms nonlinear autoregressive neural network (NARNET) and long short-term memory (LSTM) were used. The dataset used has seven significant parameters, and the developed models were assessed using statistical parameters. The findings revealed that the proposed models are capable of accurately predicting WQI and classifying water quality based on robustness. The NARNET model performed slightly better than the LSTM model in predicting WQI values, according to the results. Furthermore, the NARNET and LSTM models have achieved similar accuracy for the testing phase with a slight difference in the regression coefficient ( $R_{\text{NARNET}}=96.17\%$  and  $R_{\text{LSTM}}=94.21\%$ ). This kind of promising research can contribute significantly to water management.

**Keys words:** Prediction, WQI, WQC, ANN, Water quality, NARNET, LSTM

# Sommaire

## Liste des tableaux

## Liste des figures

## INTRODUCTION GENERALE

### Partie I : Etat de l'art et recherche bibliographique

I. Etat de l'art et recherche bibliographique .....	1
-----------------------------------------------------	---

### Partie II : Méthodes et Procédures

II. Les ANN.....	5
II.1. Trois structures de modèles de base pour la prévision de la qualité de l'eau .....	6
II.1.1 Architectures avec une approche "feed-forward" .....	6
II.1.2 Architectures récurrentes.....	8
II.1.3 Architectures hybrides.....	9
II.1.4 Méthodes émergentes .....	11
II.1.5 Le réseau de croyance profond (DBN).....	12
II.2. Les étapes à suivre dans les modèles ANN .....	13
II.3. Collection de données .....	14
II.4 Sélection des Inputs .....	17
II.5 Division des données.....	18
II.6 Prétraitement des données .....	18
II.7 Détermination de la structure du modèle .....	19
II.8 L'entraînement des modèles.....	20

### Partie III : Exemples et résultats

III.1. Normalisation Z-score .....	23
III.2. Prédiction de l'indice de qualité de l'eau .....	23
III.3. Modèle d'un réseau neuronal artificiel (ANN) .....	23
III.4. Modèle d'un réseau neuronal profond (DNN) .....	25

**Partie IV : Résultats et discussion**

IV. Résultats et discussion .....32

**CONCLUSION GENERALE**

**Référence bibliographique**

## INTRODUCTION GENERALE

L'eau est un élément vital de l'écosystème de la Terre, elle est nécessaire à la survie de toutes les espèces vivantes, y compris les humains, les animaux, les végétaux et les plantes. Il n'y aurait pas de vie sur Terre sans eau (Grifths et al. 2010), effectivement L'eau est, sans aucun doute, une ressource importante qui répond aux besoins de nombreuses activités quotidiennes.

La qualité de la plupart des masses d'eau ambiante, comme les rivières, les lacs et les ruisseaux, est déterminée par des normes de qualité précises. Les spécifications de l'eau pour diverses applications/utilisations ont également leur propre série de normes. L'eau d'irrigation, par exemple, ne doit pas être trop salée et ne doit pas contenir d'éléments dangereux susceptibles d'être transmis aux plantes ou au sol pour ne pas nuire les écosystèmes. La qualité de l'eau pour les utilisations industrielles requiert également des propriétés différentes en fonction des processus industriels spécifiques, donc une eau de qualité suffisante est nécessaire à la survie des êtres vivants. Les espèces aquatiques ne peuvent supporter qu'une certaine quantité de pollution. Le dépassement de ces limites a un impact sur l'existence de ces organismes et met leur vie en péril (Theyazn H. H aldhvani et al. 2021).

Les ressources naturelles en eau, telles que les eaux souterraines et de surface, sont parmi les sources d'eau douce les moins chères. L'activité humaine/industrielle, ainsi que d'autres processus naturels, peuvent polluer ces ressources.

En raison de l'expansion industrielle croissante, la qualité de l'eau s'est détériorée à un rythme alarmant. En outre, la rapidité de l'urbanisation et l'industrialisation ont un impact considérable sur la qualité de l'eau potable en raison d'un manque de sensibilisation du public et d'une diminution des attributs hygiéniques (P. Zeilhofer et al.2007). En effet, les effets d'une eau polluée sont très néfastes et constituent une menace sérieuse pour la santé humaine, et l'environnement.

Selon un rapport des Nations unies (ONU .2019), 1,5 million de personnes meurent chaque année suite aux maladies causées par de l'eau contaminée.

La qualité de l'eau a traditionnellement été déterminée par des analyses statistiques et de laboratoires longs et coûteux, ce qui vide de son sens le concept moderne de surveillance en temps réel, et qui nécessitent des prélèvements des échantillons, leurs transport jusqu'aux laboratoires, ainsi qu'un temps énorme et de nombreux efforts et calculs, Cette solution est inefficace car l'eau est un milieu hautement transmissible et le temps est un facteur essentiel si l'eau est contaminée par des substances pathogènes (Gazzaz, N.M et al.2012) Les

conséquences dévastatrices de la pollution de l'eau nécessitent une solution plus rapide et moins coûteuse. Par conséquent, il est essentiel de proposer des méthodes innovantes pour analyser et, si possible, prévoir la qualité de l'eau (QE). Pour assurer le suivi des changements saisonniers de la qualité de l'eau, il est conseillé de tenir compte de la dimension temporelle lors de la prévision des schémas de qualité de l'eau (K. Farrell-Poe.2000), Dans cette optique, cette recherche bibliographique examine un ensemble d'algorithmes d'apprentissage automatique basant sur les réseaux de neurones artificielles (ANN) pour estimer l'indice de qualité des eaux (IQE), un indice unique qui décrit la qualité générale de l'eau, et la classe de qualité des eaux (CQE), une classe unique définie à l'aide de l'IQE (Umair, A et al.2019).