IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 3, May 2022
Impact Factor: 6.252

Machine Learning Model for Water Quality Prediction using Python and AI framework

Dr. Kalaivazhi Vijayaragavan[1], N. Praveen[2], M. V. Sudharsan[3] and P. S. Vijayan[4]
Associate Professor, Department of Information Technology[1]
Students, B.Tech., Final Year, Department of Information Technology[2,3,4]
Anjalai Ammal Mahalingam Engineering College, Thiruvarur, India

Abstract: *During the last years, water quality has been threatened due to unprocessed effluents, municipal refuse, factory wastes, junking of compostable and non-compostable effluents has hugely contaminated nature-provided water bodies like rivers, lakes and ponds are pollutants. Therefore, it is necessity to look into the water standards before the usage. Hence modeling and predicting water quality have become very important in controlling water pollution. Safe drinking-water access is essential to health, a basic human right and a component of effective policy for health protection. It is important as a health and development issue at a national, regional and local level. Thus it is a problem that can greatly benefit from Artificial Intelligence (AI). Traditional methods require human inspection and is time consuming. Automatic Machine Learning (AutoML) facilities provide machine learning with push of a button, or, on a minimum level, ensure to retain algorithm execution, data pipelines, and code, generally, are kept from sight and are anticipated to be the stepping stone for normalizing AI. However, it is a field under research still. This project work aims to recognize the areas where an AutoML system falls short or outperforms a traditional expert system built by data scientists. Keeping this as the motive, this project work dives into the Machine Learning (ML) algorithms for comparing AutoML and an expert architecture built by this project for Water Quality Assessment to evaluate the Water Quality Index, which gives the general water quality, and the Water Quality Class, a term classified on the basis of the Water Quality Index using python. In this Project, we are going to implement a water quality prediction using machine learning techniques. In this project, our model predicts, that the water is safe to drink or not, using some parameters like PH value, conductivity, hardness, etc. Finally the results of accuracy level of AutoML and Python compared with conventional ML techniques.*

Keywords: Machine Learning, Classification Algorithm, Prediction, PyThon and AI framework

I. INTRODUCTION

Machine learning is an application of AI that enables systems to learn and improve from experience without being explicitly programmed. Machine learning focuses on developing computer programs that can access data and use it to learn for themselves. Similar to how the human brain gains knowledge and understanding, machine learning relies on input, such as training data or knowledge graphs, to understand entities, domains and the connections between them. With entities defined, deep learning can begin. The machine learning process begins with observations or data, such as examples, direct experience or instruction. It looks for patterns in data so it can later make inferences based on the examples provided. The primary aim of ML is to allow computers to learn autonomously without human intervention or assistance and adjust actions accordingly. Machine learning as a concept has been around for quite some time. The term “machine learning” was coined by Arthur Samuel, a computer scientist at IBM and a pioneer in AI and computer gaming. Samuel designed a computer program for playing checkers. The more the program played, the more it learned from experience, using algorithms to make predictions.

II. PYTHON AND AI FRAMEWORK

- **Python:** Python is an computer programming language often used to build websites and software, automate tasks, and conduct data analysis. Python is an general-purpose language, meaning it can be used to create a variety of different programs and isn’t specialized for any specific problems. This versatility, along with its beginner-

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 3, May 2022
Impact Factor: 6.252

friendlyness, has made it one of the most-used programming languages. A survey conducted by industry analyst found that it was the second-most popular programming language among developers in 2021.

- **AutoML:** Automated machine learning is the process of applying machine learning models to use real-world problems using automation. More specifically, it automates the selection, composition and parameterization of Machine Learning models. Automating the ML process makes it more user-friendly and often provides faster, more accurate outputs than hand-coded algorithms. AutoML is a typically platform or open source library that simplifies each step in the ML process, from handling a raw dataset to deploying a practical ML model. In traditional ML, models are developed by hand, and each step in the process must be handled separately.

III. Design Issues

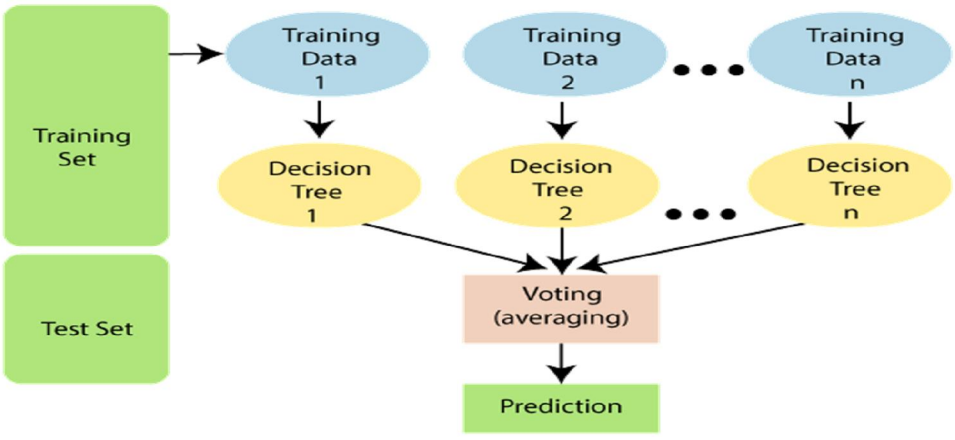
The challenge is aimed to make use of machine learning algorithm in Water Quality Assessment to evaluate the Water Quality Index of the dataset. In this project, we aim to impart the ability to get rid of biases in a machine algorithm and to predict the accuracy of the datasets.

- To evaluate the training speed of AutoML and Python based on Classification Algorithm.
- Design of a machine learning model, which can classify the different datasets.
- Datasets using Supervised and Unsupervised Learning techniques analyses the accuracy of the water quality based on parameters like PH value, conductivity and hardness.
- Machine learning algorithms use different methods to analyse training data and apply what they learn to new examples.
- When choosing a machine learning framework, it is important to consider whether this adjustment should be automatic or manual.
- AutoML library and Python platform to work with deep neural networks, testing array operations in order to get better accuracy.

3.1 Algorithm Used

A. Random Forest Classification:

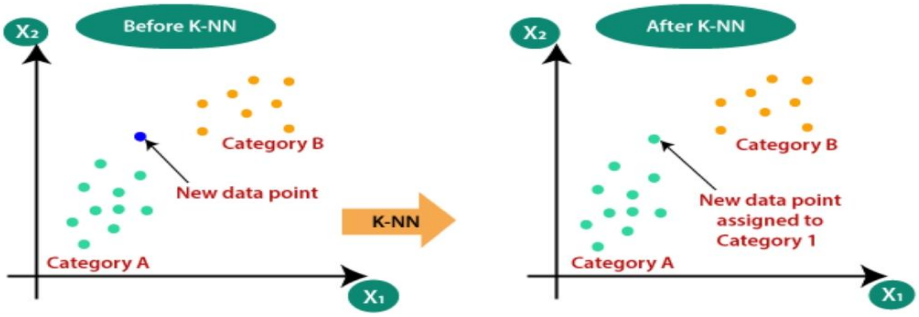
- A random forest is a machine learning technique, that is used to solve regression and classification problems. It utilizes ensemble learning, which is technique that combines many classifiers to provide solutions to complex problems.
- A random forest Classification algorithm consists of many decision trees. The ‘forest’ generated by the random forest classification algorithm is trained through bagging or bootstrap aggregating. Bagging is an ensemble meta-algorithm that improves the accuracy of ML algorithms.
- The (random forest classification) algorithm establishes the outcome based on the predictions of the decision trees. It predicts by taking average or mean of the output from various trees. Increasing the number of trees and increases the precision of the outcome.



IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 3, May 2022
Impact Factor: 6.252

B. K-Nearest Neighbour:

K Nearest Neighbor algorithm(KNN) falls under the Supervised Learning category and is used for classification and regression. It is a versatile algorithm and used for imputing missing values also resampling datasets. As the name (KNN) suggests it considers K Nearest Neighbors to predict the class or continuous value for the new Datapoint.



C. In AutoML using Tpot

TPOT (Tree-based Pipeline Optimization Tool) is a AutoML tool specifically designed for the efficient construction of optimal pipelines through genetic programming. TPOT is a open source library and makes use of scikit-learn components for data transformation, feature decomposition, feature selection and model selection .Although TPOT is classified as AutoML tool, as such it does not offer the "end-to-end" of an Machine Learning pipeline. TPOT is merely focused on the optimized automation of specific components of an Machine Learning pipeline. we can see the phases automated by TPOT and the ones specifically addressed by the Data Scientist or Machine Learning Engineer.

3.2 Development Model

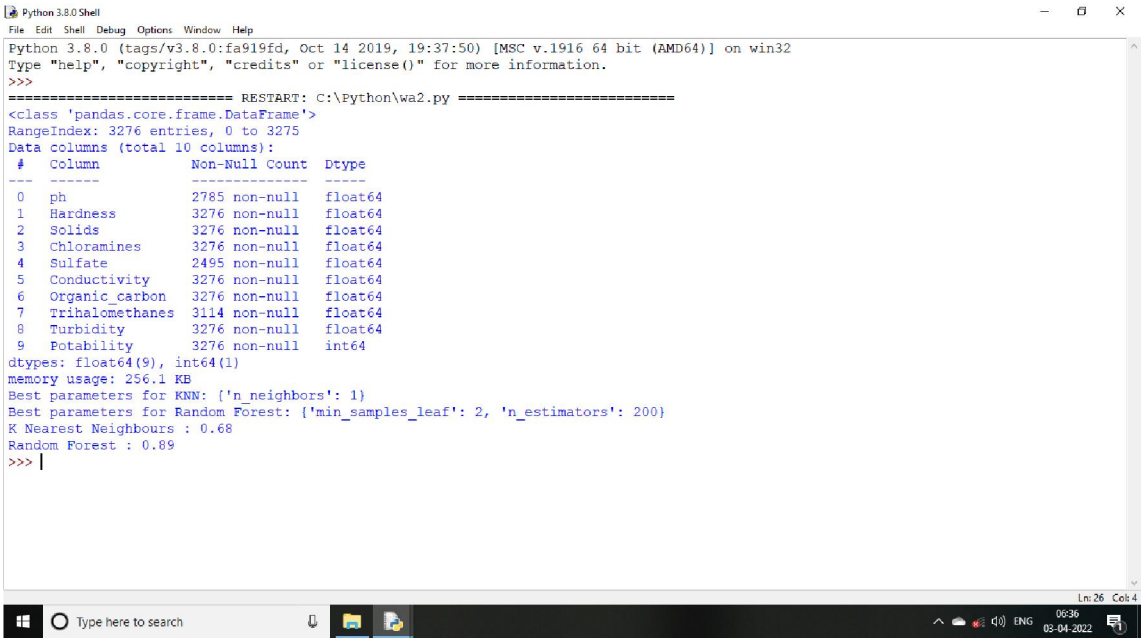
The first stage of development of Artificial Intelligence models is the preparation of the dataset. In this stage, the collected dataset shall be divided into two groups, training and testing. The training and testing dataset are used to the calibration and validation of applied models, respectively. Depending on the simulation conditions regarding time series modeling or function fitting, the approach of assigning a dataset for each group are different. In time series modeling, the history of collecting data shall be considered and shuffling the dataset is not correct, whereas for function fitting using data shuffling idea is allowed. Usually for both scenarios, about 70%–80% of the dataset is assigned for calibration and the remaining 20%–30% for validation. The next step for developing the AI models, such as Random forest classification, K-nearest neighbor and Tpot in AutoML is designing the architecture of the network.

3.3 Testing Analysis

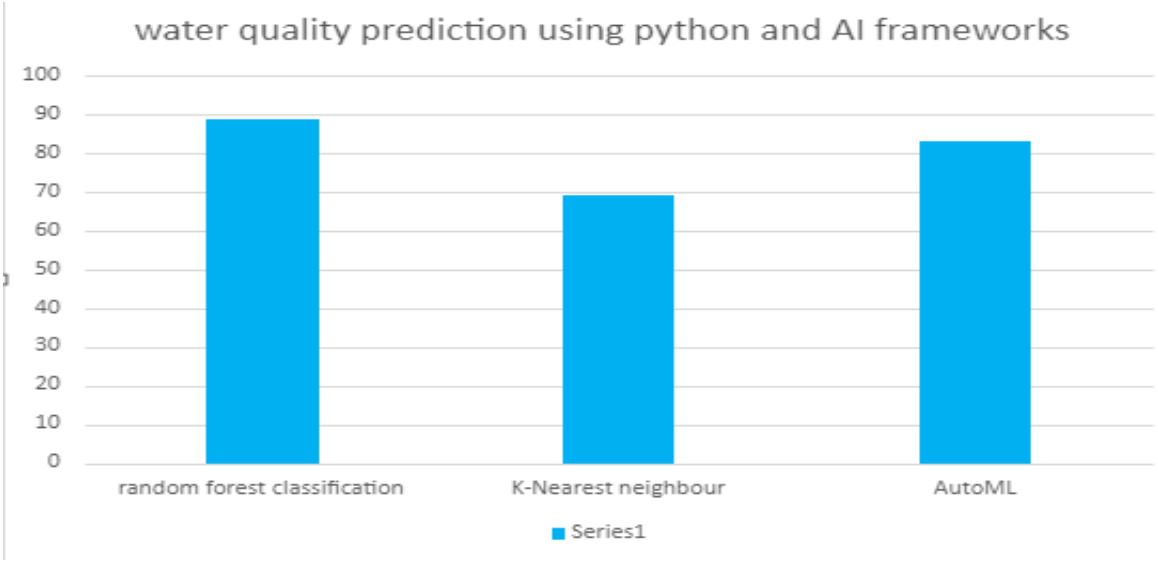
We are going to implement a water quality prediction using machine learning techniques. We will implement in this project in Random forest classification and K-nearest neighbor algorithm in supervised learning and Tpot in AutoML. Then we compare python and AI framework, Finally we find which one is accurate the Highest level.

ALGORITHM	ACCURACY LEVEL
Random forest Classification	0.89%
K-nearest neighbor	0.68%
TPOT in AutoML	0.83%

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 3, May 2022
Impact Factor: 6.252



IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 3, May 2022
Impact Factor: 6.252



IV. Result And Discussion

For validating the developed model, the dataset has been divided into 70% training and 30% testing subsets. While the Random forest classification, K-nearest neighbor and Tpot models were used to predict the Water Quality Prediction, the RFC,KNN and Tpot were utilized for the water quality classification prediction.

V. Conclusion

In this paper, the performance of Random forest classification, K-nearest neighbor in supervised learning and Tpot in AutoML were evaluated to predict the water quality. To this end most dataset related well-known components, such as pH, Conductivity, hardness etc., In the Comparison of Python and AIFramework, Python is predict more accuracy level than the AI framework.

References

[1]. Efficient Prediction of Water Quality Index (WQI) Using Machine Learning Algorithms.Author:Md. Mehedi Hassan,Laboni Akter3, Md. Mushfiqur Rahman4. Year:10 December 2021

[2]. Efficient Water Quality PredictionUsing Supervised Machine Learning Author:Umair Ahmed , Rafia Mumtaz , HirraAnwar , Asad A. Shah Year:24 October 2019

[3]. Predictive Analysis of Water Quality Parameters using Deep Learning Author:AcharanaSolanki,Himanshuagrawal,kanchan kare. Year:9 September 2015

[4]. Analysis and prediction of water quality using deep learning and auto deep learning techniques.Authors:D. Venkata VaraPrasad,Lokes wari P. SenthilKumar Year:20 January 2021

[5]. Automating water quality analysis using ML and auto ML technique Authors:G.Prasannamedha,S.harshana,S.Jahnavi Srividya Year:20 July 2021

[6]. Water Quality Prediction Using Artificial Intelligence Algorithms Authors:Fahd Abd Algalil Year: 30 Dec 2020

[7]. Predicting and Analyzing Water Quality using Machine Learning: A Comprehensive Model Author:Yafra Khan,Chai Soo See Year: 21 Nov 2019

[8]. Water Quality Prediction Method Based on LSTM Neural Network Author:Yuanyuan Wang, Jian Zhou, Kejia Chen Year:12 Dec 2017

[9]. Water quality prediction and classification based on principal component regression and gradient boosting classifier approach Author:Md.SaikatIslam Khan,Nazrul Islam,Sifatul Islam Year:3 June 2021

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 3, May 2022
Impact Factor: 6.252

[10]. River water quality index prediction and uncertainty analysis: A comparative study of machine learning models Author:Seyed Babak Haji SeyedAsadollah,AhmadSharafati,DavideMotta Year:24 January 2021

[11]. Ahmed et al., 2019 Umair Ahmed, et al.Efficient water quality prediction using supervised Machine Learning Water, 11 (11) (2019), p. 2210

[12]. Seyed Babak Asadollah, Haji Seyed, et al.River water quality index prediction and uncertainty analysis: A comparative study of machine learning modelsJournal of Environmental Chemical Engineering, 9 (1) (2021), Article

[13]. Ali Najah Ahmed, et al. Machine learning methods for better water quality prediction

[14]. Shweta Tyagi, et al.Water quality assessment in terms of water quality index American Journal of Water Resources, 1 (3) (2013), pp. 34-38

[15]. Xiaoping Wang, Fei Zhang, Jianli Ding Evaluation of water quality based on a machine learning algorithm and water quality index for the Ebinur Lake Watershed, China Scientific Reports, 7 (1) (2017), pp. 1-18

BIBLIOGRAPHY

Kalaivazhi Vijayaragavan obtained her Bachelor's degree in Electronics and Communication Engineering from REC – Trichy (Presently NIT – Trichy) in 1996 and her Master's Degree in Computer Science and Engineering from SASTRA University, Thanjavur in 2003.She is currently working as an Associate Professor in the Department of Information Technology at Anjalai Ammal Mahalingam Engineering College, Kovilvenni. She obtained her Doctoral Degree from Anna University, Chennai. Her areas of interest include Machine Learning, Data Analytics, IoTMobile Communication and Computer Networks.

N.Praveen, Pursuing B.Tech – Information Technology (IT) Final year in Anjalai Ammal Mahalingam Engineering College,T HIRUVAR

M.V.Sudharsan, Pursuing B.Tech – Information Technology (IT) Final year in Anjalai Ammal Mahalingam Engineering College, Thiruvarur

P.S.Vijayan, Pursuing B.Tech – Information Technology (IT) Final year in Anjalai Ammal Mahalingam Engineering College, Thiruvarur