**CHAITANYA BHARATHI INSTITUTE OF TECHNOLOGY**

**Department of Artificial Intelligence and Machine Learning**

**WATER QUALITY ANALYSIS USING ML**

**Team members**

1601-22-748-026 K. Rajesh

1601-22-748-027 K. Tharun Kumar Reddy

1601-22-748-031 K. Karthik

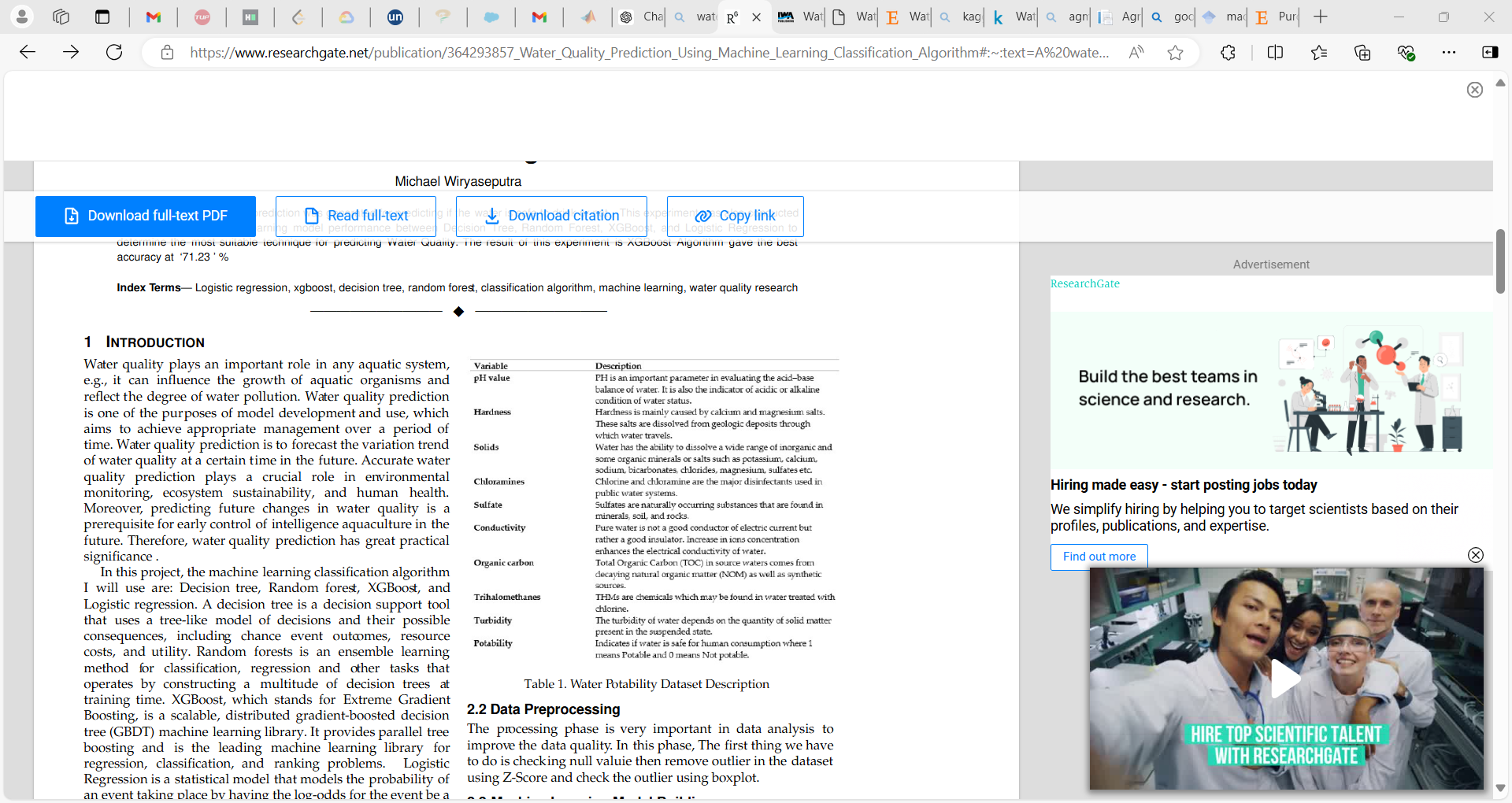
**Project Supervisor**

**Sri. M. Vishnu Chaitanya**

Assistant Professor

**Abstract**

A water quality prediction was generated for predicting if the water is safe to drink or not. This experiment was also conducted to compare the machine learning model performance between Decision Tree, Random Forest, and Logistic Regression to determine the most suitable technique for predicting Water Quality.

Water is the most crucial resource of life and it is necessary for the survival of all living creatures including human beings. The survival of business and agriculture depends on freshwater. An essential step in managing freshwater assets is the evaluation of the quality of the water. Before using water for anything, including drinking, chemical spraying (pesticides, etc.), or animal hydration, it is crucial to assess its purity. The ecosystem and the general public’s health are directly impacted by water quality. Therefore, analysing and predicting water quality is necessary for both environmental and human protection. Machine learning can be used to analyse and predict the water quality based on the parameters like PH value, turbidity, hardness, conductivity, dissolved solids in water and other parameters. In this work, the water quality is predicted by giving the concentration of various parameters as input to machine learning algorithms and the water is classified as safe or unsafe for the usage of domestic purposes.

The dataset used in this research are collected from some water condition checking. It contained 3276 samples and the dataset has 10 parameters, they are: ph, Hardness, Solids, Chloramines, Sulphate, Conductivity,Organic Carbon, Trihalomethanes, Turbidity, Potability. This dataset was obtained from kaggle:

[Water Quality (kaggle.com)](https://www.kaggle.com/datasets/adityakadiwal/water-potability)

**Keywords:**

*Logistic regression, decision tree, random forest, classiﬁcation algorithm, machine learning, water quality research, K nearest neighbour, SVM etc****.***

**Problem Statement:**

Develop a machine learning model to check the water quality of a sample and to classify water samples as potable or non-potable based on physicochemical properties, including pH, hardness, and conductivity etc. to ensure safe drinking water availability.

The Water Quality Analysis project aims to develop a machine learning model to classify water samples as either potable (safe for drinking) or non-potable (unsafe) based on various physicochemical attributes. The dataset includes features such as pH, hardness, total dissolved solids, chloramines, sulphate, conductivity, and nitrates. By analysing these attributes, the objective is to build a predictive model that accurately determines water safety. This model can assist in public health efforts by identifying potentially unsafe water sources, ensuring timely interventions, and improving water quality management practices to safeguard human health.

**Project Outline & Methodologies:**

In our project, we will employ a systematic and data-driven approach to develop the machine learning model for water quality analysis. First, we will collect a comprehensive dataset of water samples, which will include various physicochemical properties such as pH, hardness, conductivity, turbidity, and total dissolved solids (TDS). Once the data is collected, we will preprocess it by handling missing values, normalizing the data, and possibly applying feature selection techniques to identify the most significant indicators of water quality.

Next, we will split the dataset into training and testing subsets to ensure the model's ability to generalize to unseen data. we will experiment with different machine learning algorithms, such as decision trees, random forests, support vector machines (SVM), and neural networks, to identify the best-performing model for this classification task. The performance of these models will be evaluated using metrics such as accuracy, precision, recall, and F1-score.

After selecting the most effective model, we will fine-tune its hyperparameters to optimize its performance further. Finally, we will validate the model using the testing dataset and cross-validation techniques to ensure its robustness and reliability. Throughout this process, we will also consider the interpretability of the model to ensure that the results can be easily understood and applied by water management authorities for real-world decision-making.

**Objective of the project**

The objective of this project is to develop a robust machine learning model capable of assessing water quality by classifying water samples as potable or non-potable based on their physicochemical properties. These properties include key indicators of water quality such as pH, hardness, conductivity, total dissolved solids (TDS), turbidity, and other relevant factors. By leveraging historical data and advanced machine learning algorithms, the model aims to provide an accurate and efficient method for determining water safety, which is crucial for public health. The ultimate goal is to ensure the availability of safe drinking water by enabling timely and reliable water quality assessments, thereby supporting water management authorities and ensuring the protection of communities from waterborne diseases.

**References**

1. Khan, Y., & See, C. S. (2016, April). Predicting and analysing water quality using machine learning: a comprehensive model. In *2016 IEEE Long Island Systems, Applications and Technology Conference (LISAT)* (pp. 1-6). IEEE.
2. Ahmed, A. N., Othman, F. B., Afan, H. A., Ibrahim, R. K., Fai, C. M., Hossain, M. S., ... & Elshafie, A. (2019). Machine learning methods for better water quality prediction. *Journal of Hydrology*, *578*, 124084.
3. Vergina, S. A., Kayalvizhi, S., Bhavadharini, R., & Kalpana Devi, S. (2020). A real time water quality monitoring using machine learning algorithm. *Eur. J. Mol. Clin. Med*, *7*(8), 2035-2041.
4. Melesse, A. M., Khosravi, K., Tiefenbacher, J. P., Heddam, S., Kim, S., Mosavi, A., & Pham, B. T. (2020). River water salinity prediction using hybrid machine learning models. *Water*, *12*(10), 2951.
5. Kuthe, A., Bhake, C., Bhoyar, V., Yenurkar, A., Khandekar, V., & Gawale, K. (2022). Water quality analysis using machine learning. *International Journal for Research in Applied Science and Engineering Technology*, *10*(12), 581-585.
6. Akshay, R., Tarun, G., Kiran, P. U., Devi, K. D., & Vidhyalakshmi, M. (2022, December). Water-Quality-Analysis using Machine Learning. In *2022 11th International Conference on System Modeling & Advancement in Research Trends (SMART)* (pp. 13-18). IEEE.
7. Kaddoura, S. (2022). Evaluation of machine learning algorithm on drinking water quality for better sustainability. *Sustainability*, *14*(18), 11478.
8. Poudel, D., Shrestha, D., Bhattarai, S., & Ghimire, A. (2022). Comparison of machine learning algorithms in statistically imputed water potability dataset. *Journal of Innovations in Engineering Education*, *5*(1), 38-46.
9. Wang, X., Li, Y., Qiao, Q., Tavares, A., & Liang, Y. (2023). Water quality prediction based on machine learning and comprehensive weighting methods. *Entropy*, *25*(8), 1186.
10. Patel, S., Shah, K., Vaghela, S., Aglodiya, M., & Bhattad, R. (2023). Water Potability Prediction Using Machine Learning.
11. Brindha, D., Puli, V., NVSS, B. K. S., Mittakandala, V. S., & Nanneboina, G. D. (2023, February). Water quality analysis and prediction using machine learning. In *2023 7th International Conference on Computing Methodologies and Communication (ICCMC)* (pp. 175-180). IEEE.