**Predicting and Analysing Urban Water Quality using**

**Machine learning.**

**1. INTRODUCTION**

1.1 OVERVIEW

Water is the most significant resource of life, crucial for supporting the life of most existing creatures and human beings. Living organisms need water with enough quality to continue their lives. There are certain limits of pollutions that water species can tolerate. Exceeding these limits affects the existence of these creatures and threatens their lives.

Most ambient water bodies such as rivers, lakes, and streams have specific quality standards that indicate their quality. Moreover, water specifications for other applications/usages possess their standards. For example, irrigation water must be neither too saline nor contain toxic materials that can be transferred to plants or soil and thus destroying the ecosystems. Water quality for industrial uses also requires different properties based on the specific industrial processes. Some of the low-priced resources of fresh water, such as ground and surface water, are natural water resources. However, such resources can be polluted by human/industrial activities and other natural processes.

1.2 PURPOSE

Therefore, it is very important to suggest new approaches to analyze and, if possible, to predict the water quality (WQ). It is recommended to consider the temporal dimension for forecasting the WQ patterns to ensure the monitoring of the seasonal change of the WQ. However, using a special variation of models together to predict the WQ grants better results than using a single model. There are several methodologies proposed for the prediction and modeling of the WQ. These methodologies include statistical approaches, visual modeling, analyzing algorithms, and predictive algorithms. For the sake of the determination of the correlation and relationship among different water quality parameters, multivariate statistical techniques have been employed. The geostatistical approaches were used for transitional probability, multivariate interpolation, and regression analysis.

Massive increases in population, the industrial revolution, and the use of fertilizers and pesticides have led to serious effects on the WQ environments. Thus, having models for the prediction of the WQ is of great help for monitoring water contamination.

Several studies have been performed to model and predict the water quality using different ANN models. These studies have approved the feasibility and effectiveness of employing ANN applications to predict the quality of drinking water.

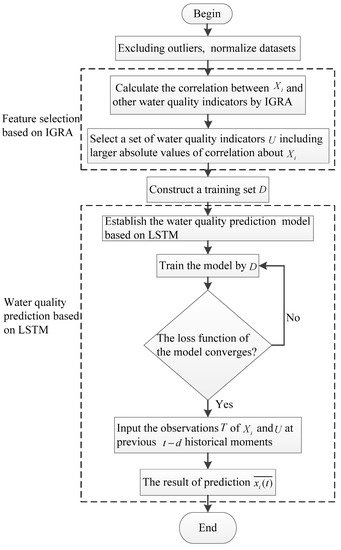
However, the contributions of the current study can be summarized as follows:

1. Developing highly efficient advanced artificial intelligence models to predict the water quality index (WQI) based on artificial neural networks and deep learning algorithms
2. Applying some machine learning models, namely, support vector machine (SVM), -nearest neighbour (K-NN), and Naive Bayes algorithms, for the prediction of water quality classification (WQC)

**2. LITERATURE SURVEY**

2.1 EXISTING PROBLEM

The procedure of water quality prediction method based on IGRA and LSTM is shown in the below figure



The specific steps for the prediction of the water quality indicator Xi are shown as follows:

Step 1.

Exclude outliers based on Pauta criterion and normalize datasets.

Step 2.

Calculate the correlation between Xi and other water quality indicators by IGRA.

Step 3.

Select a set of water quality indicators U including larger absolute values of correlation about Xi. After that, construct a training set D.

Step 4.

Establish the water quality prediction model based on LSTM, train the model by D until the loss function of the model converges.

Step 5.

Input the observations T of Xi and U at previous t−d historical moments to the model to acquire the prediction value xi(t)¯¯¯¯¯¯¯ of Xi at the *t*th moment.

##### PROPOSED SOLUTION

# Logistic regression

Logistic regression is a statistical analysis method used to predict a data value based on prior observations of a [data set](https://whatis.techtarget.com/definition/data-set). Logistic regression has become an important tool in the discipline of [machine learning](https://searchenterpriseai.techtarget.com/definition/machine-learning-ML). The approach allows an [algorithm](https://whatis.techtarget.com/definition/algorithm) being used in a machine learning application to classify incoming data based on historical data. As more relevant data comes in, the algorithm should get better at predicting classifications within data sets. Logistic regression can also play a role in [data preparation](https://searchbusinessanalytics.techtarget.com/definition/data-preparation) activities by allowing data sets to be put into specifically predefined buckets during the extract, transform, load ([ETL](https://searchdatamanagement.techtarget.com/definition/Extract-Load-Transform-ELT)) process in order to stage the information for analysis.

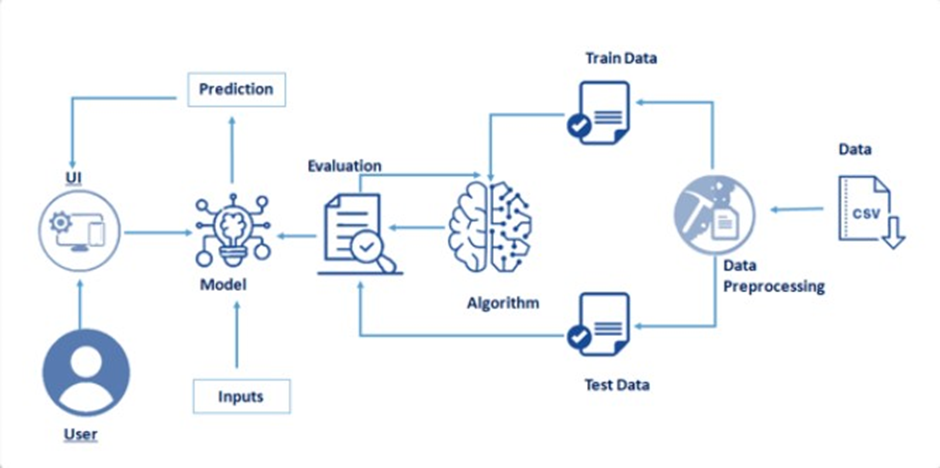
A logistic regression model predicts a [dependent data variable](https://whatis.techtarget.com/definition/dependent-variable) by analyzing the relationship between one or more existing independent variables. For example, a logistic regression could be used to predict whether a political candidate will win or lose an election or whether a high school student will be admitted to a particular college.

The resulting analytical model can take into consideration multiple input criteria. In the case of college acceptance, the model could consider factors such as the student’s grade point average, SAT score and number of extracurricular activities. Based on [historical data](https://whatis.techtarget.com/definition/historical-data) about earlier outcomes involving the same input criteria, it then scores new cases on their probability of falling into a particular outcome category.

**3. THEORITICAL ANALYSIS**

3.1 BLOCK DIAGRAM

ARCHITECTURE



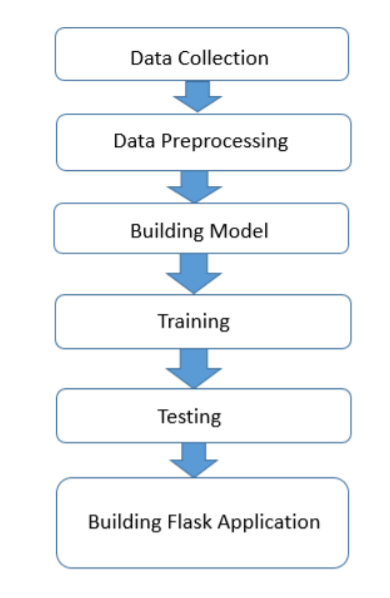
REQUIREMENTS:

The following are the requirements needed to finish off the project:

* Jupyter notebook for programming , which can be installed by Anaconda IDE.
* Python packages.

EXPERIMENTAL INVESTIGATIONS:

Water is considered as a vital resource that affects various aspects of human health and lives. The quality of water is a major concern for people living in urban areas. Quality of water serves as a powerful environmental determinant and a foundation for the prevention and control of waterborne diseases. However predicting the urban water quality is a challenging task since the water quality varies in urban spaces non-linearly and depends on multiple factors, such as meteorology, water usage patterns, and land uses, so this project aims at building a Machine Learning (ML) model by considering all water quality standard indicators.

CONTROL FLOW OF THE SOLUTION:

RESULT:

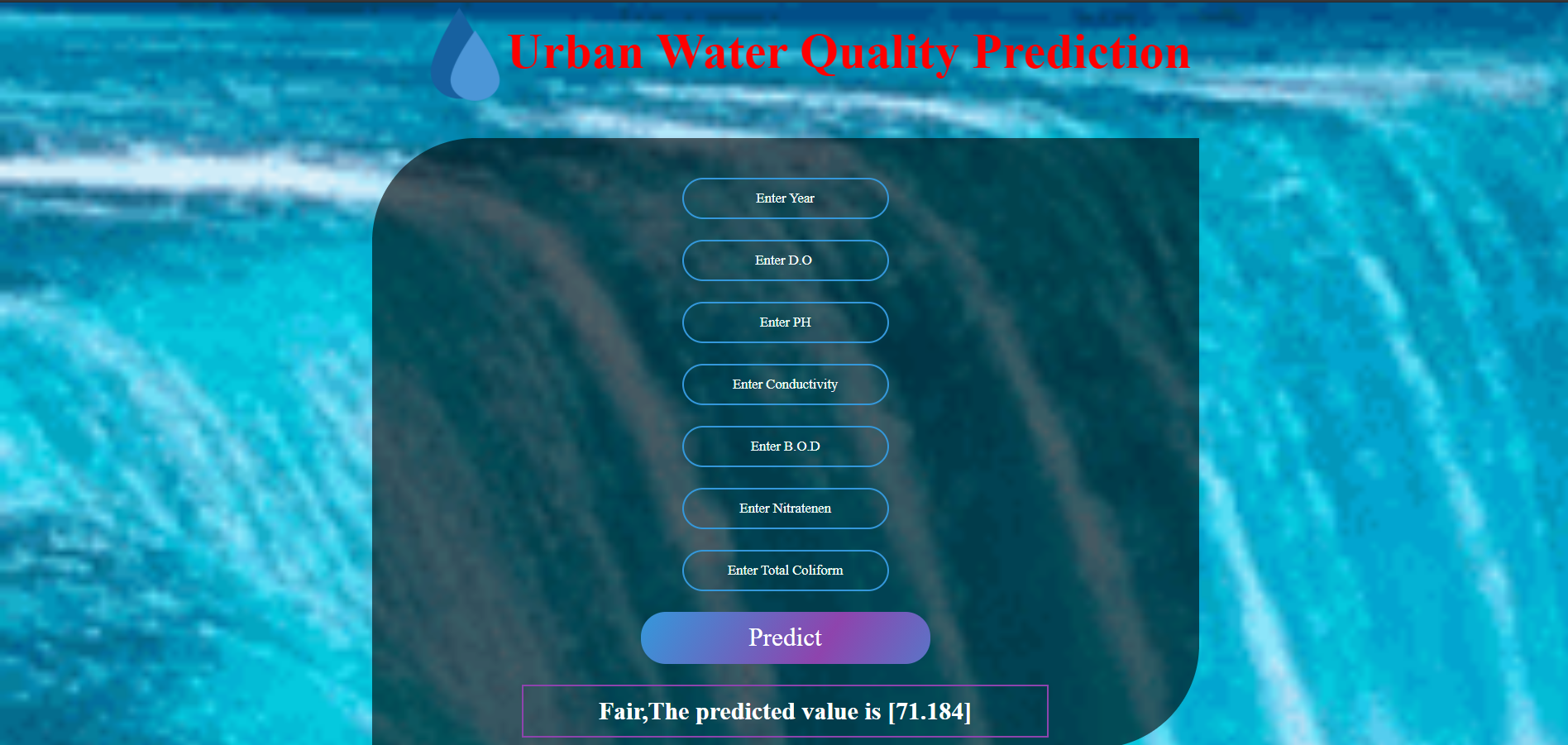
Here we used logistic regression to work out on a predicting and analysing urban water quality using machine learning to tell whether the water is sustainable or not.

This can be checked by giving the details of the contents of water and predict the water quality.

The output is as follows:



After giving the suitable inputs we get the output as shown below:



ADVANTAGES OF PROPOSED SOLUTION:

Water is an essential element of our lives, giving life to all living creatures on Earth. The importance of ensuring that this water is of sound quality is extremely important, particularly if that water is intended for consumption. Water quality testing can provide valuable data on the condition of a particular body of water, and whether it may need special treatment before use. Examing factors such as the pH level, nutrient levels, amount of dissolved oxygen, alkalinity and bacteria are all useful in understanding the health of a waterbody, allowing you to accurately create a water management plan with the data.

This article from ECO Environmental will be examining the advantages of water quality testing, before assessing some of the popular water quality meters available on the market today.

Whether it be for groundwater, surface water or open water, there are a number of reasons why it is important for you to undertake regular water quality testing. If you’re wanting to create a solid foundation on which to build a broader water management plan, then investing in water quality testing should be your first point of action.

This testing will also allow you to adhere to strict permit regulations and be in compliance with Australian laws.

Identifying the health of your water will help you to discover where it may need some help. Ultimately, finding a source of pollution, or remaining proactive with your monitoring will enable you to save money in the long term. The more information that you can obtain will assist you with your decision on what product you may need to improve the condition of your water. Simply guessing and buying products based on a hunch or a general trend is ill-advised, as each body of water has unique properties that can only be discovered through testing.

Measuring the amount of dissolved oxygen in your water is another important advantage of water quality testing, as typically the less oxygen, the higher the water temperature, resulting in a more harmful environment for aquatic life. These levels do fluctuate slightly across the seasons, but regular monitoring of your water quality will allow you to discover trends over time, and whether there are other factors that may be contributing to the results you discover.

DISADVANTAGES:

**STORET Data template:**

Training necessary Somewhat difficult to manage over time and with large data sets

**WQX Web:**

Requires manual operation to submit data, some configuration required

**Network Node:**

Costly, usually only feasible under Exchange Network grants Technical expertise and network server required

**Network Node Client:**

Requires manual operation to submit data Cannot respond to data queries from other nodes, and therefore cannot interact with the Exchange Network Technical expertise and network server required

**APPLICATIONS:**

1. Continuous online water quality monitoring in the water industry is a necessity to ensure efficient potable water production and that any out of threshold events are controlled and alarmed. Many specific parameters in the distribution system are monitored for regulatory requirements to ensure that water companies delivery safe drinking water to its communities . Our range of instrumentation can be used for:
   * Raw Water Intake Monitoring
   * Coagulation, Flocculation & Clarification Monitoring
   * Filter Monitoring
   * Disinfection Process Control
   * Clean Water Network Monitoring
2. Ongoing water quality monitoring gives you an **idea of your building's normal turbidity level**; the data will show immediately if conditions change. If that happens, you can either employ another sensor to look for specific contaminants, or send a water sample out for testing in a lab.



**CONCLUSION**

Modeling and prediction of water quality are very important for the protection of the environment. Developing a model by using advanced artificial intelligence algorithms can be used to measure the future water quality. In this proposed methodology, the advanced artificial intelligence algorithms, namely, NARNET and LSTM models were used to predict the WQI. Moreover, machine learning algorithms such as SVM, KNN, and Naive Bayes were used to classify the WQI data. The proposed models were evaluated and examined by some statistical parameters. For the WQI prediction, the result has revealed that the performance of the NARNET model is slightly better than the LSTM model based on the obtained  value. However, the SVM algorithm has achieved the highest accuracy of the prediction of the WQC as compared with KNN and Naive Bayes algorithms. After examining the robustness and efficiency of the proposed model for predicting the WQI, in future work, the developed models will be implemented to predict the water quality in India for different types of water.

**FUTURE SCOPE**

The Algorithms may be more advance and short in the future for the prediction and analysis of the water. And the water prediction may be more accurate and easy in the future.

BIBLOGRAPHY:

* P. Zeilhofer, L. V. A. C. Zeilhofer, E. L. Hardoim, Z. M. . Lima, and C. S. Oliveira, “GIS applications for mapping and spatial modeling of urban-use water quality: a case study in District of Cuiabá, Mato Grosso, Brazil,” *Cadernos de Saúde Pública*, vol. 23, no. 4, pp. 875–884, 2007.View at: [Publisher Site](https://doi.org/10.1590/S0102-311X2007000400015) | [Google Scholar](https://scholar.google.com/scholar_lookup?title=GIS%20applications%20for%20mapping%20and%20spatial%20modeling%20of%20urban-use%20water%20quality:%20a%20case%20study%20in%20District%20of%20Cuiab%C3%A1,%20Mato%20Grosso,%20Brazil&author=P.%20Zeilhofer&author=L.%20V.%20A.%20C.%20Zeilhofer&author=E.%20L.%20Hardoim&author=Z.%20M.%20.%20Lima&author=&author=C.%20S.%20Oliveira&publication_year=2007)
* M. A. Kahlown, M. A. Tahir, and H. Rasheed, *National Water Quality Monitoring Programme, Fifth Monitoring Report (2005–2006)*, Pakistan Council of Research in Water Resources Islamabad, Islamabad, Pakistan, 2007, <http://www.pcrwr.gov.pk/Publications/Water%20Quality%20Reports/Water%20Quality%20Monitoring%20Report%202005-06.pdf>.
* UN water, “Clean water for a healthy world,” Tech. Rep., Development, 2010, <https://www.undp.org/content/undp/en/home/presscenter/articles/2010/03/22/clean-water-for-a-healthy-world.html>.View at: [Google Scholar](https://scholar.google.com/scholar_lookup?title=Clean%20water%20for%20a%20healthy%20world&author=UN%20water&publication_year=2010)
* K. Farrell-Poe, W. Payne, and R. Emanuel, *Water Quality & Monitoring*, University of Arizona Repository, 2000, <http://hdl.handle.net/10150/146901>.

**SOURCE CODE:**

<https://www.kaggle.com/anbarivan/indian-water-quality-analysis-and-prediction>

CODE FOR THE SOLUTION WE BUILT:

