

# **WATER QUALITY PREDICTION USING MACHINE LEARNING**

*A Project Report submitted in the partial fulfillment of the  
Requirements for the award of the degree*

## **BACHELOR OF TECHNOLOGY In COMPUTER SCIENCE AND ENGINEERING**

Submitted by

**Nelluri Joshna (19471A05H1)**

**Chandana Ramoji (19571A05D7)**

**Kasu Pravallika (19471A05F4)**

Under the esteemed guidance of

**Dr. B. JHANSI VAZRAM, M.tech., Ph.D.**

**Professor**



**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING  
NARASARAOPETA ENGINEERING COLLEGE: NARASARAOPET  
(AUTONOMOUS)**

Accredited by NAAC with A+ Grade and NBA under Cycle -1  
NIRF rank in the band of 251-320 and an ISO 9001:2015 Certified  
Approved by AICTE, New Delhi, Permanently Affiliated to JNTUK, Kakinada  
KOTAPPAKONDA ROAD, YALAMANDA VILLAGE, NARASARAOPET-522601  
2022-2023

**NARASARAOPETA ENGINEERING COLLEGE  
(AUTONOMOUS)**

**DEPARTMENT OF COMPUTER SCIENCE AND  
ENGINEERING**



**CERTIFICATE**

This is to certify that the project that is entitled with the name **“Water Quality Prediction using Machine Learning”** is a bonafide work done by the team **Nelluri Joshna (19471A05H1), Chandana Ramoji (19571A05D7), Kasu Pravallika (19471A05F4)** in partial fulfilment of the requirements for the award of the degree of **BACHELOR OF TECHNOLOGY** in the Department of **COMPUTER SCIENCE AND ENGINEERING** during **2022-2023**.

**PROJECT GUIDE**

**Dr. B. Jhansi Vazram, M.tech., Ph.D.**  
**Professor**

**PROJECT CO-ORDINATOR**

**Ms. Sireesha, M.Tech., Ph.D.**  
**Professor**

**HEAD OF THE DEPARTMENT**

**Dr. S. N. Tirumala Rao, M.Tech., Ph.D.,**  
**Professor**

**EXTERNAL EXAMINER**

## ACKNOWLEDGEMENT

We wish to express our thanks to the various personalities who are responsible for the completion of the project. We are extremely thankful to our beloved chairman Sri **M.V. Koteswara Rao**, B.Sc., who took an interest in us in every effort throughout this course. We owe sincere return to our beloved principal **Dr.M. Sreenivasa Kumar**, M.Tech., Ph.D., MISTE., FIE(I), for showing his kind attention and valuable guidance throughout the course.

We express our deep felt gratitude towards **Dr. S.N. Tirumala Rao**, M.Tech., Ph.D. HOD of the CSE department, and also to our guide **Dr. B. Jhansi Vazram**, M.Tech., Ph.D. Professor of the CSE department whose valuable guidance and unstinting encouragement enable us to accomplish our project successfully in time.

We extend our sincere thanks to **Ms. Sireesha .M**, M.Tech., Ph.D. Associate professor & Project coordinator of the project for extending her encouragement. Their profound knowledge and willingness have been a constant source of inspiration for us throughout this project work.

We extend our sincere thanks to all other teaching and non-teaching staff to the department for their cooperation and encouragement during our B.Tech degree. We have no words to acknowledge the warm affection, constant inspiration, and encouragement that we received from our parents.

We affectionately acknowledge the encouragement received from our friends and those who involved in giving valuable suggestions had clarifying out doubts which had really helped us in successfully completing our project.

By

Nelluri Joshna (19471A05H1)

Chandana Ramoji (19471A05D7)

Kasu Pravallika (19471A05F4)

## **ABSTRACT**

Water quality evaluation approaches have been proposed for the purpose of analysis of the quality of water to survive more number of peoples from several diseases caused by lower quality of water. Since by several different algorithms calculate pH values by using attributes related to water quality. And the accuracy should be high. By taking the observations from several water resources then we may perform the machine learning algorithms. Horton produced the first Water Quality Index (WQI) in the 1960s two indices for determining the general state of drinking source water quality are straightforward, adaptable, and stable, with little sensitivity to input data.

Similarly, to give water quality information, we employed the weighted arithmetic WQI approach. WQIs convert a huge number of variables into a digital number and aid in the comprehension of water quality, making them the most widely used water quality assessment tool, despite significant flaws. Recent water quality assessments used matter element extension analysis (MEEA) and entropy TOPSIS in a wastewater irrigation area and a rapidly urbanizing area, respectively. Both approaches are mathematical, but they are accurate in estimating overall water quality. Water quality evaluation methods, on the other hand, rely on water quality standards for classification.



## **INSTITUTE VISION AND MISSION**

### **INSTITUTION VISION**

To emerge as a Centre of excellence in technical education with a blend of effective student centric teaching learning practices as well as research for the transformation of lives and community.

### **INSTITUTION MISSION**

M1: Provide the best class infra-structure to explore the field of engineering and research

M2: Build a passionate and a determined team of faculty with student centric teaching, imbibing experiential, innovative skills

M3: Imbibe lifelong learning skills, entrepreneurial skills and ethical values in students for addressing societal problems



## **DEPARTMENT OF COMPUTER SCIENCE And ENGINEERING**

### **VISION OF THE DEPARTMENT**

To become a center of excellence in nurturing the quality Computer Science & Engineering professionals embedded with software knowledge, aptitude for research and ethical values to cater to the needs of industry and society.

### **MISSION OF THE DEPARTMENT**

The department of Computer Science and Engineering is committed to

**M1:** Mould the students to become Software Professionals, Researchers and Entrepreneurs by providing advanced laboratories.

**M2:** Impart high quality professional training to get expertise in modern software tools and technologies to cater to the real time requirements of the Industry.

**M3:** Inculcate team work and lifelong learning among students with a sense of societal and ethical responsibilities.

### **Program Specific Outcomes (PSO's)**

**PSO1:** Apply mathematical and scientific skills in numerous areas of Computer Science and Engineering to design and develop software-based systems.

**PSO2:** Acquaint module knowledge on emerging trends of the modern era in Computer Science and Engineering

**PSO3:** Promote novel applications that meet the needs of entrepreneur, environmental and social issues.



## **Program Outcomes**

- 1. Engineering knowledge:** Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.
- 2. Problem analysis:** Identify, formulate, research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.
- 3. Design/development of solutions:** Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.
- 4. Conduct investigations of complex problems:** Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.
- 5. Modern tool usage:** Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.
- 6. The engineer and society:** Apply reasoning informed by the contextual knowledge to



assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.

**7. Environment and sustainability:** Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.

**8. Ethics:** Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.

**9. Individual and team work:** Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.

**10. Communication:** Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.

**11. Project management and finance:** Demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.

**12. Life-long learning:** Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

### Project Course Outcomes (CO'S):

**CO425.1:** Analyse the System of Examinations and identify the problem.

**CO425.2:** Identify and classify the

requirements. **CO425.3:** Review the

Related Literature **CO425.4:** Design and

Modularize the project

**CO425.5:** Construct, Integrate, Test and Implement the Project.

**CO425.6:** Prepare the project Documentation and present the  
Report using appropriate method.

### Course Outcomes – Program Outcomes mapping

|               | PO1 | PO2 | PO3 | PO4 | PO5 | PO6 | PO7 | PO8 | PO9 | PO10 | PO11 | PO12 | PSO1 | PSO2 | PSO3 |
|---------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|------|------|------|------|------|
| <b>C425.1</b> |     | ✓   |     |     |     |     |     |     |     |      |      |      | ✓    |      |      |
| <b>C425.2</b> | ✓   |     | ✓   |     | ✓   |     |     |     |     |      |      |      | ✓    |      |      |
| <b>C425.3</b> |     |     |     | ✓   |     | ✓   | ✓   | ✓   |     |      |      |      | ✓    |      |      |
| <b>C425.4</b> |     |     | ✓   |     |     | ✓   | ✓   | ✓   |     |      |      |      | ✓    | ✓    |      |
| <b>C425.5</b> |     |     |     |     | ✓   | ✓   | ✓   | ✓   | ✓   | ✓    | ✓    | ✓    | ✓    | ✓    | ✓    |
| <b>C425.6</b> |     |     |     |     |     |     |     |     | ✓   | ✓    | ✓    |      | ✓    | ✓    |      |

## Course Outcomes – Program Outcome correlation

|               | PO1 | PO2 | PO3 | PO4 | PO5 | PO6 | PO7 | PO8 | PO9 | PO10 | PO11 | PO12 | PSO1 | PSO2 | PSO3 |
|---------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|------|------|------|------|------|
| <b>C425.1</b> | 2   | 3   |     |     |     |     |     |     |     |      |      |      | 2    |      |      |
| <b>C425.2</b> |     |     | 2   |     | 3   |     |     |     |     |      |      |      | 2    |      |      |
| <b>C425.3</b> |     |     |     | 2   |     | 2   | 3   | 3   |     |      |      |      | 2    |      |      |
| <b>C425.4</b> |     |     | 2   |     |     | 1   | 1   | 2   |     |      |      |      | 3    | 2    |      |
| <b>C425.5</b> |     |     |     |     | 3   | 3   | 3   | 2   | 3   | 2    | 2    | 1    | 3    | 2    | 1    |
| <b>C425.6</b> |     |     |     |     |     |     |     |     | 3   | 2    | 1    |      | 2    | 3    |      |

Note: The values in the above table represent the level of correlation between CO's and PO's:

**1.** Low level

**2.** Medium level

**3.** High level

### Project mapping with various courses of Curriculum with AttainedPO's:

| Name of the course from which principles are applied in this project | Description of the device   | Attained PO   |
|--|---|---------------|
| C3.2.4, C3.2.5   | Gathering the requirements and defining the problem, plan to develop a <b>smart bottle for health care using sensors.</b>   | PO1, PO3      |
| CC4.2.5  | Each and every requirement is critically analyzed, the process model is identified and divided into <b>five modules</b>   | PO2, PO3      |
| CC4.2.5  | Logical design is done by using the unified modelling language which involves individual team work  | PO3, PO5, PO9 |
| CC4.2.5  | Each and every module is tested, integrated, and evaluated in our project   | PO1, PO5      |
| CC4.2.5  | Documentation is done by all our four members in the form of a group  | PO10          |
| CC4.2.5  | Each and every phase of the work in group is presented periodically   | PO10, PO11    |
| CC4.2.5  | Implementation is done and the project will be handled by the <b>hospital management and in future updates in our project can be done based on air bubbles occurring in liquid in saline.</b> | PO4, PO7      |
| CC4.2.8 CC4.2.   | <b>The physical design includes hardware components like sensors, gsm module, software and Arduino.</b>   | PO5, PO6      |

# INDEX

| S. No.   | CONTENT  | PAGE NO |
|----------|--|---------|
|          | <b>LIST OF FIGURES</b>                               | XV      |
| <b>1</b> | <b>INTRODUCTION</b>                                  |         |
|          | 1.1 Introduction                                     | 1       |
|          | 1.2 Existing System                                  | 2       |
|          | 1.3 Proposed System                                  | 2       |
|          | 1.4 System Requirements                              | 3       |
|          | 1.4.1 Hardware Requirements                          | 3       |
|          | 1.4.2 Software Requirements                          | 3       |
| <b>2</b> | <b>LITERATURE SURVEY</b>                             |         |
|          | 2.1 Literature Survey                                | 4       |
|          | 2.2 Some Machine Learning Methods                    | 5       |
|          | 2.3 Applications of Machine Learning                 | 6       |
|          | 2.4 Advantages of Machine Learning                   | 6       |
|          | 2.5 Libraries in Machine Learning                    | 7       |
| <b>3</b> | <b>SYSTEM ANALYSIS</b>                               |         |
|          | 3.1 Importance of Machine Learning in Python         | 9       |
|          | 3.2 Implementation of Machine Learning using Python  | 9       |
|          | 3.3 Scope of the Project                             | 10      |
|          | 3.4 Data Set Analysis                                | 10      |
|          | 3.5 Methodology                                      | 13      |
|          | 3.5.1 Missing Values                                 | 14      |
|          | 3.5.2 Correlation Coefficient Method                 | 15      |
|          | 3.5.3 Cross Validation                               | 17      |
|          | 3.6 Classification                                   | 17      |
|          | 3.6.1 Machine Learning Algorithms for Classification | 18      |
|          | 3.7 Implementation Code                              | 20      |
|          | 3.8 Confusion Matrix                                 | 30      |

|          |                       |           |
|----------|-----------------------|-----------|
| <b>4</b> | <b>OUTPUT SCREENS</b> | <b>31</b> |
| <b>5</b> | <b>CONCLUSION</b>     | <b>34</b> |
| <b>6</b> | <b>FUTURE SCOPE</b>   | <b>35</b> |
| <b>7</b> | <b>REFERENCES</b>     | <b>36</b> |

## LIST OF FIGURES

| S.NO. | FIG NO. | FIGURE CAPTION                    | PAGE NO |
|-------|---------|-----------------------------------|---------|
| 1     | 3.1     | Dataset                           | 12      |
| 2     | 3.2     | Work-flow diagram of the proposal | 13      |
| 3     | 3.3     | Before missing data               | 14      |
| 4     | 3.4     | After filling missing data        | 14      |
| 5     | 3.5     | Correlation                       | 16      |
| 6     | 3.6     | Histogram                         | 27      |
| 7     | 3.7     | Boxplot                           | 28      |
| 8     | 3.8     | Bar plot                          | 29      |
| 9     | 3.9     | Scatter plot                      | 30      |
| 10    | 4.1     | Water Quality Prediction          | 31      |
| 11    | 4.2     | Water is safe                     | 32      |
| 12    | 4.3     | Water is unsafe                   | 33      |