### GUJARAT TECHNOLOGICAL UNIVERSITY

Chandkheda, Ahmedabad Affiliated

|  |  |  |
| --- | --- | --- |
| A picture containing text  Description automatically generated | **ADANI INSTITUTE OF INFRASTRUCTURE ENGINEERING** | A picture containing company name  Description automatically generated |

A

Project Report On

## MACHINE LEARNING MODEL FOR WATERBORNE DISEASE OUTBREAKS

Under Subject of

**Summer Internship - (3170001)**

**B.E, Semester - VII**

(Information Communication Technology Branch) Academic Year (2024-2025)

**Submitted By:**

Group No:

|  |  |
| --- | --- |
| **Name of Student** | **Enrollment No.** |
| Patel Kavy Rashkinkumar | 211310132081 |

## INTERNSHIP OFFER LETTER

## 

## CERTIFICATE

****

**TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| **Sr. No.** | **Title** | **Page No.** |
| **1** | Acknowledgement | 1 |
| **2** | About the Project | 2 |
| **3** | Dataset Description | 3 |
| **4** | Features | 4 |
| **5** | Methodology | 5 |
| **6** | Mind Map | 6 |
| **7** | Model Information | 7 |
| **8** | Conclusion | 9 |
| **9** | Future scope | 10 |
| **10** | Plagiarism Certificate | 11 |

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| **Sr. No.** | **Title** | **Page No.** |
| **1** | Mind Map | 6 |
| **2** | Model graphs and statistics | 7,8 |

**ACKNOWLEDGEMENT**

Every work owes its success to many people. Likewise, the successful completion of my project could not have been possible without the co-ordination and support of the faculties of AIIE-ICT.

I am grateful to **Dr. Ajay Kumar Vyas (H.O.D, AIIE-ICT)**, for providing me with all the facilities required for my project. My sincere thanks to **Asst. Prof. (Dr.) Mani Shekhar Gupta(Faculty Guide)** for imparting valuable guidance throughout the project.

Thanks to the Gujarat Technological University for introducing **Summer Internship** as a significant aspect of IT program providing an environment to students for practical application of their knowledge.

|  |  |
| --- | --- |
| **Name of Student** | **Enrollment No.** |
| Patel Kavy Rashkinkumar | 211310132081 |

# MACHINE LEARNING MODEL FOR WATERBORNE DISEASE OUTBREAKS

### Concept of the Project

Waterborne diseases pose significant public health challenges worldwide, particularly in regions with inadequate access to clean water and sanitation facilities. Diseases such as cholera, typhoid, hepatitis, and giardiasis are prevalent in these areas, leading to severe illness, mortality, and economic burden. The prevention and management of waterborne disease outbreaks require proactive strategies that leverage advanced technologies and data-driven approaches. Traditional methods for monitoring and responding to disease outbreaks often rely on retrospective data analysis and reactive measures, which can result in delayed interventions and increased transmission rates. This project addresses these shortcomings by developing Machine Learning Model for Waterborne Disease Outbreaks powered by machine learning. The model aims to predict and mitigate the impact of waterborne disease outbreaks through timely detection and targeted interventions. This project directly contributes to **SDG 3: Good Health and Well-being** by improving disease prevention strategies and enhancing public health responses.

### Problem Statement

Waterborne diseases are a significant public health challenge, particularly in regions with limited access to clean water and proper sanitation. These diseases can result in severe health consequences and high mortality rates, especially among vulnerable populations. The complexity of predicting and preventing these outbreaks stems from the intricate interactions between environmental, climatic, and social factors.

### Objective of the Project

This project aims to achieve the following objectives:

**Predict Outbreaks:** Forecast the likelihood and potential severity of waterborne disease outbreaks.

**Inform Public Health Actions:** Provide actionable insights for targeted interventions and resource allocation.

**Enhance Community Awareness:** Improve public awareness and preparedness for potential outbreaks.

* **Overview**

# DATASET DESCRIPTION

We have developed a machine learning model to predict waterborne disease outbreaks using data sourced from **Kaggle**. This data includes various parameters related to health records, environmental quality measures, climate data, and socio-economic indicators. We utilized **IBM Watson Machine Learning** to analyze, train, and validate our model, ensuring robust and accurate predictions for four specific diseases: typhoid, cholera, hepatitis, and giardiasis.

### Content of dataset

The dataset utilized for this project was sourced from Kaggle, containing the following features relevant to water quality and potential disease predictors:

* + **pH:** Measure of acidity or alkalinity of water.
  + **Hardness:** Concentration of calcium and magnesium salts in water.
  + **Solids:** Total dissolved solids (TDS) in water.
  + **Chloramines:** Chemical compounds used as disinfectants in water treatment.
  + **Sulfate:** Concentration of sulfate ions in water.
  + **Conductivity:** Ability of water to conduct electrical current.
  + **Organic\_carbon:** Presence of carbon-based organic matter in water.
  + **Trihalomethanes:** By-products of water chlorination.
  + **Turbidity:** Cloudiness or haziness of water due to suspended particles.

### Disease Predicted:

The project aimed to predict the likelihood of four waterborne diseases based on these water quality parameters:

* Typhoid
* Cholera
* Hepatitis
* Giardia

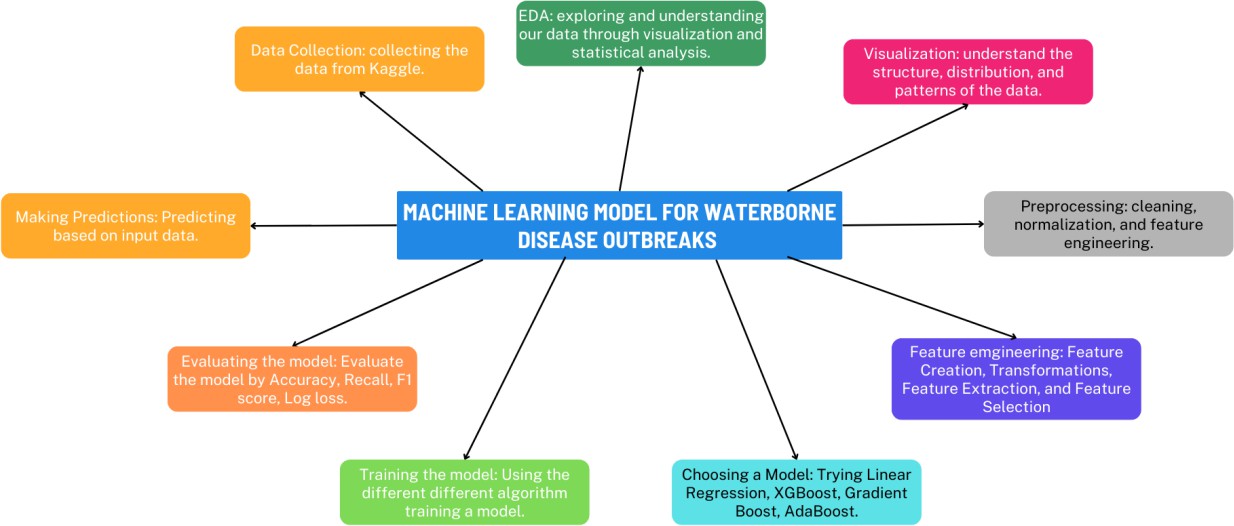
# FEATURES

* **Disease Prediction:** The model predicts the likelihood of waterborne disease outbreaks based on environmental and health data.
* **Risk Assessment:** It assesses the potential severity of outbreaks, providing insights into high-risk areas.
* **Data Integration:** Comprehensive aggregation of health, environmental, climatic, and social data.
* **Predictive Accuracy:** High accuracy in predicting outbreaks for typhoid, cholera, hepatitis, and giardiasis.
* **Scalability:** Capable of handling large datasets and scalable for future data integration.
* **Automated Analysis:** Automated preprocessing and feature engineering to streamline the workflow.

# METHODOLOGY

* **Data Collection and Preprocessing:** Data sourced from Kaggle was cleaned, normalized, and structured for analysis.
* **Feature Engineering:** Relevant features were extracted and engineered to enhance model performance.
* **Model Development:** Multiple machine learning algorithms were tested, with the best-performing model selected for final deployment.
* **Training and Validation:** The model was trained and validated using IBM Watson Machine Learning, ensuring it met the required performance criteria.
* **Evaluation:** The model’s performance was evaluated using standard metrics and cross-validation techniques.

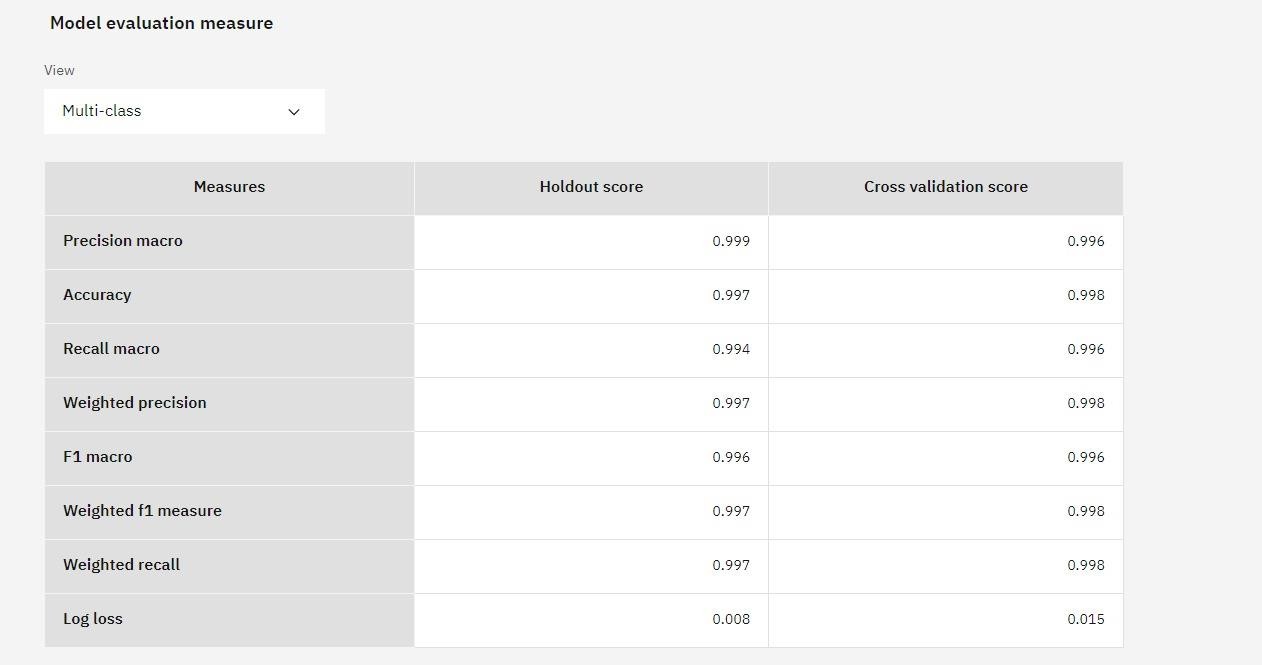
# MINDMAP



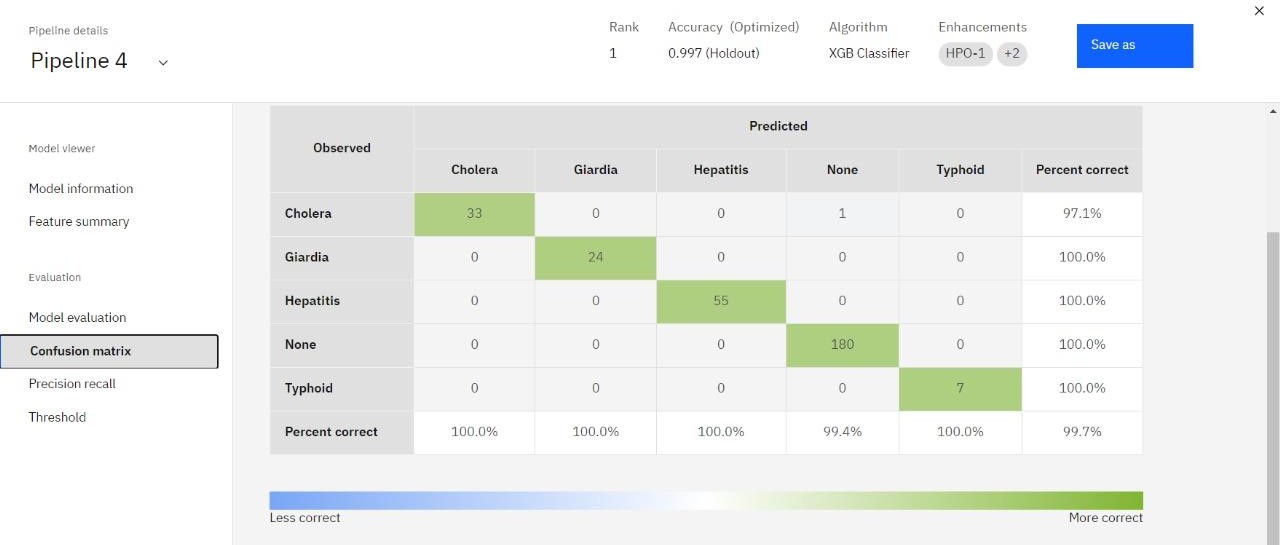
**MODEL INFORMATION**



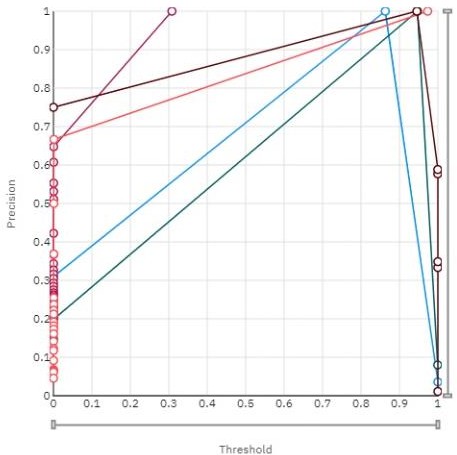
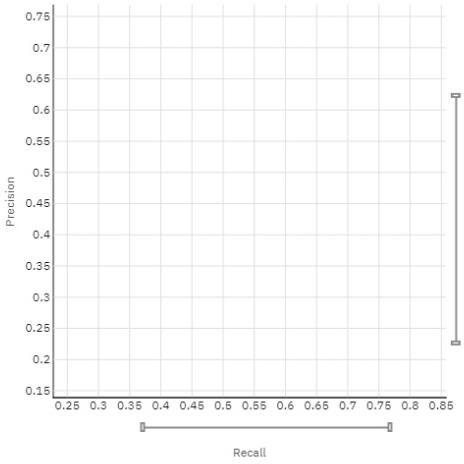
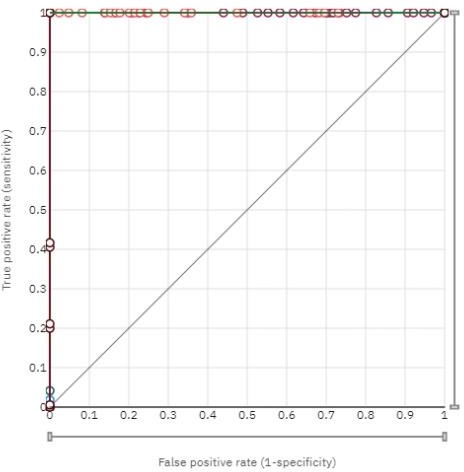
**Fig.1: Experiment parameters**



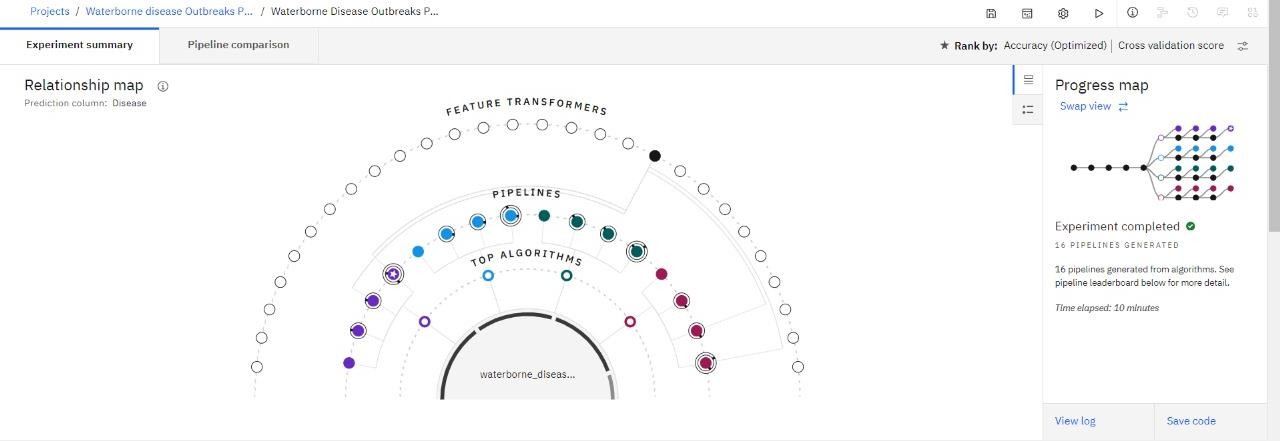
**Fig.2: Model evaluation measure**



**Fig.3: Confusion matrix**



**Fig.4: ROC curve Fig.5: Precision Recall curve Fig.6: Threshold chart**



**Fig.7: Relationship map**

# CONCLUSION

The development of this machine learning model represents a significant step forward in predicting waterborne disease outbreaks. By leveraging data from Kaggle and utilizing IBM Watson Machine Learning, we have created a robust, scalable, and accurate model that can aid public health authorities in preventing and mitigating the impacts of these diseases. This project not only aligns with global health initiatives but also demonstrates the power of machine learning in addressing critical public health challenges.

# FUTURE GOALS

* Enhanced Model Accuracy
* Expansion to Additional Diseases
* Geographical Scalability
* Integration with Public Health Systems
* User-Friendly Interface Development
* AI-Powered Early Warning System
* Collaboration with Global Health Organizations
* Publication and Knowledge Sharing

**PLAGIARISM CERTIFICATE**

