

# Computer Science Capstone Topic Approval Form

The purpose of this document is to help you clearly explain your capstone topic, project scope, and timeline. Identify each of these areas so that you will have a complete and realistic overview of your project. Your course instructor cannot sign off on your project topic without this information.

*Note: You must fill out and submit this form. Space beneath each number will expand as needed.*

*Any cost associated with developing the application will be the responsibility of the student.*

## INFORM INSTRUCTOR:

Potential use of proprietary company information: (Y/**N**)

## ANALYSIS:

1. Project topic AND description:

The project topic is the classification of Water Potability – which is the determination of whether a water sample is safe to drink or not. Water potability is a growing concern among many populations across the globe. In a world with diminishing water sources, increasing global temperatures, and rising sea levels, finding a water source is only half of the problem – the other half is to determine whether it is safe to drink.

**Scenario:** I am part of a machine learning team at a fictitious company called EcoBrain. In this hypothetical scenario, EcoBrain is contracted by a committee representing various African countries. This committee has requested an application that will analyze whether water sources are safe to drink for their population.

2. Project purpose/goals: Water potability is determined by various factors such as pH, hardness, chloramine, turbidity, and more. By analyzing a water sample and its composition of these factors, we can determine whether a provided water sample is safe to drink. The value of this application has a great impact on the well-being and health of the affected populations. By confirming water sources as potable or not, we can both bring confidence to a water source and avoid perilous situations in which water is unsafe to drink.
3. Descriptive method: To describe the data, I can use various graphics to illustrate how a water sample's properties measures against potable/ideal values. First, I can use a bar chart or histogram to illustrate the quantity of each property within a given sample. Second, I can use a scatter plot to represent the distribution of property data. Finally, I can use box plot to illustrate the quartets of distributions of a property in the a data set.
4. Predictive/Prescriptive method: Since the machine learning algorithm will be focusing on determining whether a given sample is either potable or not, we will use a classification model. Specifically, we will use a Random Forest Classification model that will work by constructing a multitude of decision trees during training; decisions are made in rank order of importance that will ultimately congregate to define the resulting accuracy. Random Forests is a supervised learning algorithm and is flexible and easy to use with Sci-Kit Learn.

## DESIGN and DEVELOPMENT:

1. Computer science application type:
- **WEB**
2. Programming/development language(s) you will use: Python for the backend server/API and JavaScript for the frontend.
3. Operating System(s)/Platform(s) you will use: Windows 10.

4. Database Management System you will use: N/A. Source will be a CSV file.
5. Estimated number of hours for the following:
  - i. Planning and Design: 20
  - ii. Development: 60
  - iii. Documentation: 20
  - iv. Total: 100
6. Projected completion date: 7/23

**IMPLEMENTATION and EVALUATION:**

1. Describe how you will approach the execution of your project:
  - a. Acquire data set from Kaggle
  - b. Analyze data for null values to replace
  - c. Build model with processed data
  - d. Process and evaluate results and reevaluate if necessary
  - e. Create documentation

**X This project does not involve human subjects research and is exempt from WGU IRB review.**

**STUDENT SIGNATURE S**

Alexander Le

**By signing and submitting this form, you acknowledge** any cost associated with development and execution of the application will be your (the student) responsibility.

**COURSE INSTRUCTOR'S NAME:**



**COURSE INSTRUCTOR APPROVAL DATE: July 9, 2021**

---