Submitted By:
Shivani Gupta
2430521

SURGE: Mid Term Report

In this mid-term research project, the objective was to extract, integrate, and prepare water quality data from the India-WRIS website for further analysis using machine learning techniques. In the first week, I explored the India-WRIS website to understand the available water quality data. During the second and third weeks, I employed various web scraping methods to retrieve this data, utilizing Python libraries such as Beautiful Soup and Selenium. These tools facilitated the extraction of the necessary water quality information from the website.

The focus of weeks four and five was on merging the extracted water quality data with geographical information. Specifically, I worked with two Excel workbooks: one containing water quality data from different regions in India and the other comprising the longitude and latitude coordinates of the monitoring stations. Initial attempts to combine these datasets using Python scripts were somewhat cumbersome and inefficient. Consequently, I opted for the Excel "Index-Match" function, which significantly simplified the data integration process, allowing for a quick and accurate merge of the datasets.

With the data successfully extracted and integrated, the next phase of the project will involve applying machine learning techniques to this comprehensive dataset. The goal is to analyze and derive meaningful insights regarding water quality trends and patterns across various regions in India. This project showcases the effective use of web scraping tools and data management techniques, and sets the stage for advanced data analysis using machine learning in the coming weeks.

Approved.
Shankar Pradesh