

Project Idea Proposal

Group 17: Nurselam Hussen Adam, Rejeb Dendir, Rahmet Abdella Ismael

1. Project Idea Title: Predicting Access to Clean Water Using Machine Learning

Problem Statement: Access to clean water remains a critical global challenge, with 2 billion people lacking access to safely managed drinking water services (UNICEF, 2023). This issue is particularly acute in Sub-Saharan Africa, where water scarcity and poor sanitation infrastructure exacerbate health risks and hinder socioeconomic development.

Goal: Develop a predictive machine learning model to identify areas at risk of water scarcity using socioeconomic, environmental, and infrastructure data.

2. Relevance to Sustainable Development Goals (SDGs): This project aligns with SDG 6: Clean Water and Sanitation, which aims to ensure the availability and sustainable management of water for all. By predicting water scarcity, this project supports targeted policy interventions to improve water access and reduce inequalities. It also indirectly contributes to SDG 3 (Good Health and Well-being): Addressing health risks associated with unsafe water, such as waterborne diseases. The focus on Sub-Saharan Africa underscores the urgency of addressing water scarcity in one of the world's most vulnerable regions.

3. Literature Examples:

"Machine Learning for Water Resource Management" (Smith et al., 2022): This study demonstrated the potential of machine learning models to predict water quality based on environmental factors like rainfall patterns, soil composition, and temperature. The authors highlighted the importance of integrating diverse datasets to improve prediction accuracy.

"Predictive Analytics for Water Scarcity in Sub-Saharan Africa" (Ogundele & Adeyemi, 2021): The authors developed a regression model to forecast water scarcity using socioeconomic indicators such as GDP per capita, population density, and access to sanitation infrastructure. These studies provide a strong foundation for leveraging machine learning to address water scarcity challenges, particularly in Sub-Saharan Africa.

4. Describe Your Data Source:

World Bank Open Data: Provides socioeconomic indicators such as GDP per capita, population density, and poverty rates. **UNICEF Water, Sanitation, and Hygiene (WASH) datasets:** Offers detailed information on water access, sanitation infrastructure, and hygiene practices. **Climate datasets (e.g., CHIRPS):** Include rainfall levels, temperature, and other environmental variables relevant to water resource management.

Data Format: CSV files containing structured data on socioeconomic, environmental, and infrastructure indicators. **Size:** Approximately 10,000 rows of country-level and regional data, with a focus on Sub-Saharan African countries.

5. Approach (Machine Learning or Deep Learning)

Approach: Machine Learning

Given the structured nature of the dataset and the goal of predicting a continuous variable (water access percentage), machine learning models such as Random Forest Regression or Gradient Boosting are suitable. These models are interpretable, efficient for medium-sized datasets, and capable of handling nonlinear relationships between variables.