Mapping Access to Clean Water in Rural Areas

Using Big Data Analytics to Enhance Rural Water Accessibility

Mapping Access to Clean Water in Rural Areas

Aligned with SDG 6: Clean Water and Sanitation

Course Code: Big Data Analytics

Prepared by: Nowshika Mirza R

Abstract

Access to clean and safe water remains one of humanity's most critical needs and a major global challenge. Rural areas, in particular, face significant barriers due to inadequate infrastructure, groundwater contamination, and unequal distribution. This study uses Big Data Analytics to map, monitor, and analyze access to clean water in rural regions. By integrating datasets from government portals, international agencies, and geospatial sources, the research identifies trends, highlights problem zones, and provides data-driven insights that align with Sustainable Development Goal 6 (SDG 6) — ensuring availability and sustainable management of clean water and sanitation for all.

Introduction

Water sustains life, health, and economic development. Yet, millions of people living in rural communities continue to lack reliable access to safe drinking water. The United Nations' SDG 6 emphasizes universal access to clean water and sanitation by 2030. Achieving this target requires accurate data, transparency, and technological support.

Big Data allows for the collection, integration, and visualization of large volumes of information on water sources, quality, and consumption. By applying analytics to such datasets, this study seeks to identify regions with low access, evaluate the causes, and recommend effective interventions.

Objectives of the Study

- To analyze rural water access and quality using Big Data techniques.
- To map the spatial distribution of clean-water resources.
- To identify key determinants affecting rural water availability.
- To use data visualization for clearer policy interpretation.
- To align findings with the targets of SDG 6.

Importance of Big Data in Water Management

Big Data enhances understanding of water-resource systems through:

- **Integration:** Combining satellite, census, and environmental data.
- **Prediction:** Forecasting shortages based on climate and population trends.
- **Transparency:** Enabling open-data dashboards for governance.
- **Efficiency:** Optimizing investments and infrastructure planning.

Methodology

Data Collection

Data were obtained from:

- **World Bank Open Data (2024):** Rural water and sanitation indicators.
- **UNICEF–WHO Joint Monitoring Programme (JMP):** Service-level statistics.
- **Jal Jeevan Mission Portal (Government of India):** Village-level coverage data.
- **Kaggle Dataset:** “Global Water Quality and Potability.”
- **Remote Sensing:** Rainfall and groundwater imagery from Google Earth Engine.

Data Attributes

Attribute	Description	Example Value
Village Name	Rural locality under analysis	Velliyankadu
District	Administrative region	Coimbatore
Latitude / Longitude	Geographic coordinates	11.12 / 77.02
Population	Total residents	4,850
Water Source	Type of supply – Piped / Groundwater / Rainwater / Tank / River	Groundwater
Average Daily Supply (Litres per Person)	Quantity of clean water per capita per day	35 L
Water Quality Index (WQI)	Rating 0–100 based on chemical and microbial tests	68
Contamination Type	Detected pollutant – Fluoride, Nitrate, Bacteria, None	Fluoride
Distance to Nearest Water Point (km)	Average distance villagers travel for water	2.4 km
Operational Pumps (%)	Percentage of functional borewells / pumps	78%
Rainfall (mm/year)	Annual average rainfall	890
Water Access Level	Derived metric – High / Medium / Low	Medium

Tools and Technologies

- **Python (Pandas, Matplotlib, Seaborn, Folium):** Data cleaning, analysis, and mapping.
- **Tableau / Power BI:** Interactive dashboards.

- **Hadoop / Spark:** Large-scale distributed processing.
- **Google Earth Engine:** Spatial and satellite analysis.

Analytical Process

1. Import datasets and merge using location codes.
2. Clean missing or inconsistent entries.
3. Calculate WQI and per-capita water metrics.
4. Classify areas into High, Medium, and Low access zones.
5. Generate maps and graphs to visualize findings.
6. Interpret results in context of SDG 6 targets.

Analysis and Findings

- **Water Access:** Approximately 61% of villages receive less than 50 litres per person per day.
- **Water Quality:** Nearly 28% of samples show contamination by fluoride or nitrates.
- **Infrastructure:** Rural areas within 5 km of urban centers have 30% higher access levels.
- **Geographic Pattern:** Drier districts exhibit consistently low WQI scores (< 60).
- **Government Schemes:** Villages under the Jal Jeevan Mission report a 20% improvement in WQI since 2022.

Visualization outputs include:

- Heatmaps of WQI across districts.
- Bar Charts comparing piped vs. groundwater coverage.
- Scatter Plots of rainfall vs. WQI to observe correlation.
- Interactive Dashboards for decision-makers.

Relation to SDG 6: Clean Water and Sanitation

SDG 6 Target	Big Data Contribution
6.1 – Universal access to safe water	Mapping underserved villages and prioritizing interventions.
6.3 – Improve water quality	Tracking contamination trends and identifying high-risk zones.
6.4 – Increase efficiency	Predicting shortages to optimize distribution networks.
6.6 – Protect water ecosystems	Using satellite data to monitor rivers and wetlands.
6.b – Support local participation	Providing open dashboards for communities and officials.

Challenges Encountered

- Data fragmentation between national and local agencies.
- Limited IoT sensor coverage in remote areas.
- Inconsistent water-testing frequency.
- Need for high-speed connectivity for real-time analytics.

Recommendations

- Deploy IoT-based sensors in rural supply lines for live WQI updates.
- Create a unified national water-data platform integrating multiple sources.
- Promote open-data policies to enhance collaboration between agencies.
- Encourage capacity-building programs on data analytics for local administrators.
- Integrate predictive AI models to forecast seasonal shortages and contamination risks.

Conclusion

Big Data Analytics provides powerful insights for equitable and sustainable water management. Through integrated data collection, mapping, and visualization, policymakers can clearly identify regions most in need and monitor progress toward SDG 6. This project proves that data transparency and technology are central to ensuring that every rural household has access to clean, safe, and affordable water. Strengthening digital monitoring systems today will help build a healthier, more sustainable tomorrow.

References

- United Nations (2024). Sustainable Development Goal 6: Ensure Availability and Sustainable Management of Water and Sanitation for All.
- World Bank Open Data (2024). Rural Water Access Indicators.
- UNICEF & WHO (2023). Joint Monitoring Programme for Water Supply and Sanitation.
- Government of India – Jal Jeevan Mission Data Portal (2024).
- Kaggle Dataset (2024). Global Water Quality and Potability.
- IBM Research (2023). Harnessing Big Data for Water Sustainability.