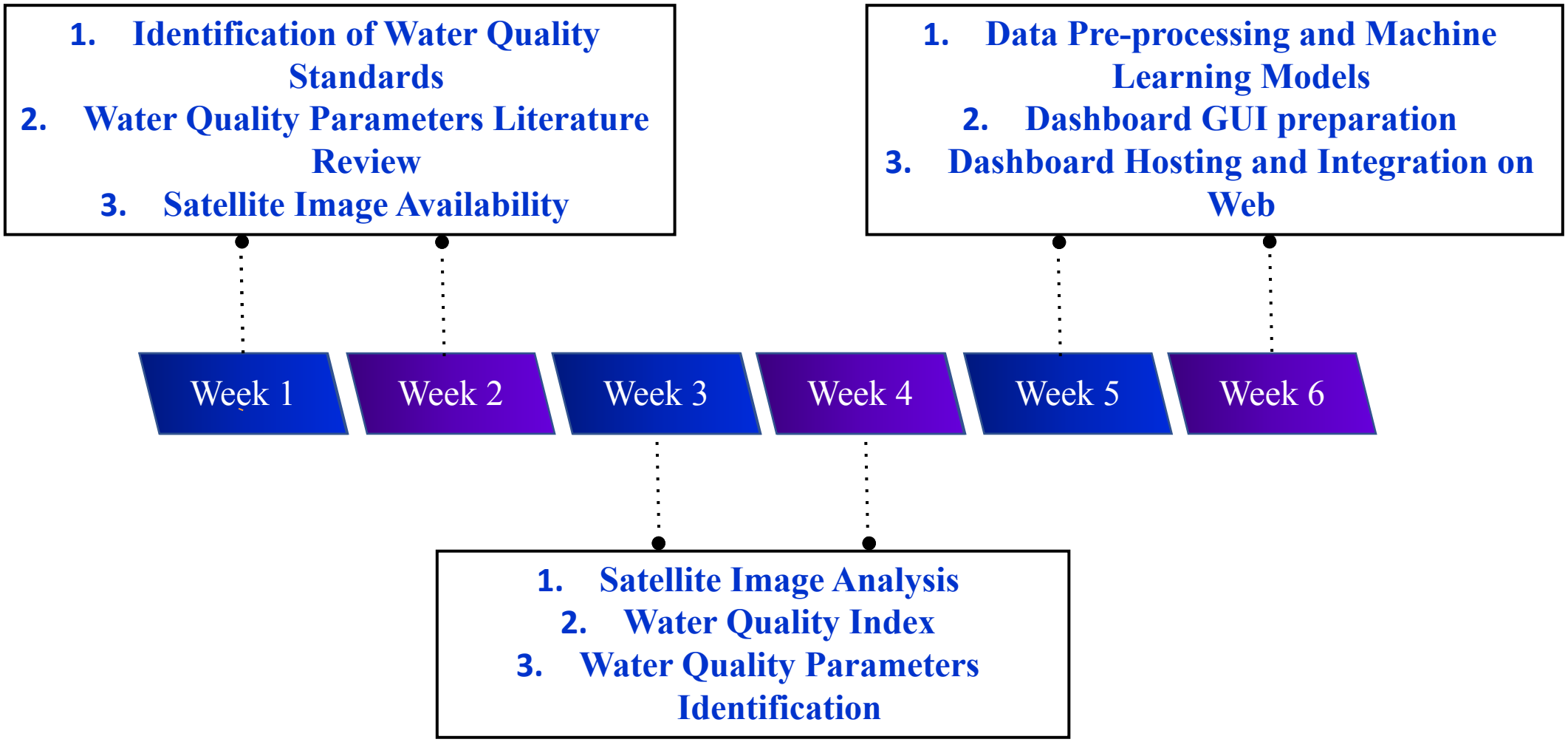
Water Quality Monitoring Dashboard for Kutch Region

By: Kutch, Gujarat India Chapter
https://omdena.com/omdena-chapter-page-india/

CONTENTS

Timeline


1. Identification of Water Quality Standards
2. Water Quality Parameters Literature Review
3. Satellite Image Availability
1. Data Pre-processing and Machine Learning Models
2. Dashboard GUI preparation
3. Dashboard Hosting and Integration on Web
Week 1
Week 2
Week 3
Week 4
Week 5
Week 6
1. Satellite Image Analysis
2. Water Quality Index
3. Water Quality Parameters Identification

Water Quality Standards

PHYSICAL

- pH
- Temperature (deg Celsius)
- Turbidity (NTU)
- T.S.S.(Total Suspended Solids)
- T.D.S.(Total Dissolved Solids)
- Conductivity (mS/cm)

INORGANIC

- Magnesium
- Alkalinity
- Total hardness
- Calcium hardness
- Sodium Chloride(salinity)
- Sulphate
- Sodium
- Nitrate(NO3)

ORGANIC / NUTRIENTS

- Phosphate(PO4)
- DO(Dissolved Oxygen)
- COD(Chemical Oxygen Demand)
- BOD(BioChemical Oxygen Demand)

BIOLOGICAL

- Total Coliform(CFU)
- Chlorophyll

Literature Review

Numerous remote sensing data incorporated with machine learning were evaluated and key points were noted, and different satellite sources were effectively adjusted.

A handful of the most notable sources are listed here: -

1. https://earthdata.nasa.gov/learn/pathfinders/water-quality-data-pathfinder
2. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5017463/
3. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6111878/

GITHUB -

robmarkcole/satellite-image-deep-learning: Resources for deep learning with satellite & aerial imagery

Exploring the Data

1. https://satellites.pro/
2. https://developers.google.com/earth-engine

Satellite Data

1. **Sentinel-2 can be used to map changes in land cover and to monitor the world's forests. It also provides information on pollution in lakes and coastal waters.**
2. **The main objective of the Sentinel-3 mission is to measure sea surface topography, sea and land surface temperature, and ocean and land surface colour with high accuracy and reliability to support ocean forecasting systems, environmental monitoring and climate monitoring.**

Satellites & Sensors for Water Quality Monitoring

Satellites	Sensors	Resolution
Landsat 7	Enhanced Thematic Mapper (ETM+)	185 km Swath; 15 m, 30 m, 60 m; 16-Day Revisit
Landsat 8	Operational Land Imager (OLI)	185 km Swath; 15 m, 30 m, 60 m; 16-Day Revisit
Terra & Aqua	MODerate Resolution Imaging Spectroradiometer (MODIS)	2330 km Swath; 250 m, 500 m, 1 km; 1–2-Day Revisit
SNPP[1] and JPSS[2]	Visible Infrared Imaging Radiometer Suite (VIIRS)	3040 km Swath; 375 m – 750 m; 1–2-Day Revisit
Sentinel 2A and 2B	Multi Spectral Imager (MSI)	290 km Swath; 10 m, 20 m, 60 m; 5-Day Revisit
Sentinel 3A and 3B	Ocean and Land Color Instrument (OLCI)	1270 km Swath; 300 m; 27-Day Revisit

Satellite Image Analysis

pH	pH = 8.339-0.827*(B1/B8)
SALINITY	Salinity = (B11-B12)/(B11+B12)
TURBIDITY	Turbidity= (B4-B3)/(B4+B3)
LAND SURFACE TEMPERATURE	Land Surface Temperature = ST_B10* 0.00341802+149.0 – 273.15
CHLOROPHYLL	Chlorophyll = (B5-B4)/(B5+B4)
SUSPENDED MATTER	Suspended_matter = Oa08_radiance/Oa06_radiance
DISSOLVED OXYGEN MATTER	Dissolved Organic Matter = Oa08_radiance/Oa04_radiance
DISSOLVED OXYGEN	Dissolved Oxygen = -0.0167*B8+0.0067*B9+0.0083*B11+9.577

Water Quality Parameters Identification

1. pH
2. Salinity
3. Turbidity
4. Land Surface Temperature (LST)
5. Chlorophyll
6. Suspended matter
7. Dissolved Organic Matter (DOM)
8. Dissolved Oxygen

Machine Learning

DATA PRE-PROCESSING AND EXPLORATION

First we concatenated the individual data files.

Null Values :- nearly 60% of the data had null values so we decided to drop them.

Outliers :- Only dissolved oxygen showed data points outside of the IQR however to it's high number, they couldn't be considered as clear outliers.

No significant correlation between parameters were found except one relation between DOM and Suspended Matter.

TRAINING AND TESTING DATA PREPARATION

Due to the lack of in-situ data in India for training, we applied the research based thresholds for labelling the records into 'good', 'poor' and 'Needs treatment' classes and generate our own training and testing data.

We normalized the data using Min-Max scalar.

Since the data was imbalanced, we used SMOTE to up-sample it, so there were balanced observations for each class.

MACHINE LEARNING MODELS

Various Machine learning models were applied on the final dataframe and the metrics were analysed and the best model was chosen with having a good validation accuracy.

Among all the models we evaluated Random Forest Classifier performed best and was used for the final deployment.

Dashboard

1. **Problem Statement** - Water Quality Centralized Dashboard for Better Decision Making
2. **Project Goals** - Analysis, Interpretation and Visualization with comparison of Water Quality Limits
3. **Location Chosen** - Kutch Region (Harmisar, Shinai and Tappar Lake)
4. **Developments Made** - Identified Parameters, Selected Sources, Analyzing and Applying Formulas, Finalized Parameters along with Band, ML model with Good Accuracy, Interactive Dashboard
5. **Project Endorsements** - Projecting the water quality, Monitoring and Analyzing existing conditions, Identification of Parameters with Threshold Values
6. **Select AOI Data Parameters** - Water Body, Parameters, Latitude, Longitude, Dates
7. **Visualizations** - Lake Satellite Imaging using Sentinel
8. **Project Summary** - Water Quality Crisis, Ability of Decision Making, Real-Time Enforcement
9. **Conclusion** - Centralized Dashboard to check on with the Water Conditions visually in Real-Time.

Potential Next Steps

- Decide airborne sensor or spaceborne sensor based on the size of the study area.
- Choose the remote sensor with the suitable spatial, temporal, and spectral scales according to the problem to be solved.
- Choose close dates of both remote sensor images and the in-situ samples to reach acceptable results.
- Apply the proposed model on different in-situ sampling datasets in different regions to ensure model applicability and robustness.
- Solve the problem of temporal scale (revisit times) by data fusion between over one remote sensing data to fill the gaps of missing dates.
- Use deep learning as a powerful machine to improve the prediction accuracy and try to solve the problem of extra bands of the hyperspectral sensors by using band selection and also generalize the model to predict the quality of water all-over the country

Thank You

PROJECT MANAGER: -

CHANCY SHAH

- Tanisha Banik
- Renju Zachariah
- Sairam Kannan
- Ishita Kheria
- Bharati Panigrahi
- Himanshu Mishra
- Prathima Kadari
- Deepali Bidwai
- Drij Chudasama
- Kiran Ryakala

THANK YOU!