

# A Data Warehouse Approach for Predictive Modeling of Coastal Biodiversity Loss : A Case Study of Navi Mumbai's Intertidal Wetlands

Parth Manjrekar, Mihir Saiya, Jalak Panchal, Khushi Parmar

*Mahavir Education Trust's Shah And Anchor Kutchhi Engineering College, Chembur, Mumbai*

---

## Abstract

This project presents a comprehensive assessment of mangrove health and extent change within the Mumbai Metropolitan Region's critical coastal buffer zones. Utilizing **Sentinel-2** satellite imagery and a hybrid analytical framework (GEE, QGIS, and Weka), the study quantified changes across the seven-year period (2018–2024). The analysis revealed a significant **net gain of 1506.4 Hectares**, indicating successful overall regeneration. However, the ecosystem simultaneously suffered **1196.9 Hectares of localized degradation**. Statistical validation using Association Rules (OLAP) identified the **Panvel subregion during the Dry Season** as the primary degradation hotspot, linking health loss to specific, concentrated spatial and temporal conditions.

**Keywords:** Mangrove health, NDVI, Sentinel-2, OLAP, Change detection, Mumbai Metropolitan Region

---

## 1. Introduction and Project Objectives

Mangroves are vital intertidal ecosystems, providing essential services such as coastal defense, water quality regulation, and carbon storage, which are crucial for the resilience of the Mumbai Metropolitan Region against rising sea levels and storm surges. Given the intense development pressure in areas like Panvel and Uran, continuous, data-driven monitoring of forest health is necessary.

1. To generate a high-resolution, multi-temporal inventory of mangrove extent using Sentinel-2 data.
2. To calculate the overall quantitative area change (loss and gain) between the 2018 and 2024 epochs.
3. To perform **OLAP** (Online Analytical Processing) to statistically link localized degradation events to specific environmental and temporal drivers.

## 2. Methodology

The study employed a rigorous, three-phase methodology to convert raw satellite data into actionable ecological insights.

### 2.1. Phase 1: Data Acquisition and Pre-processing (Google Earth Engine)

Sentinel-2 Level-2A imagery (10m resolution) from the **COPERNICUS/S2\_SR\_HARMONIZED** image collection was used to ensure consistent band naming across the entire time series [2]. Data was filtered for the study period (2018–2024) and masked for cloud cover (< 1%).

#### Vegetation Indices Calculation:

$$\text{NDVI} = \frac{\text{NIR} - \text{RED}}{\text{NIR} + \text{RED}} \quad \text{EVI} = 2.5 \times \frac{\text{NIR} - \text{RED}}{\text{NIR} + 6\text{RED} - 7.5\text{BLUE} + 1} \quad (1)$$

## 2.2. Phase 2: Spatial Quantification and Mapping (QGIS)

The median composite NDVI and EVI rasters were exported and analyzed in QGIS (Geographic Information System).

1. **Change Detection:** The difference raster,  $\Delta\text{NDVI} = \text{NDVI}_{2024} - \text{NDVI}_{2018}$ , was calculated.
2. **Visualization:** A Red-White-Green diverging color scheme was applied to the  $\Delta\text{NDVI}$  layer, classifying areas based on a **significant change threshold of  $\pm 0.06$** .
3. **Quantification:** The  $\Delta\text{NDVI}$  raster was **reclassified** into three discrete zones (Loss, Stable, Gain) and converted to a vector polygon layer via the **Polygonize** tool. The final area for each zone was calculated in Hectares (Ha) from the polygon attribute table.

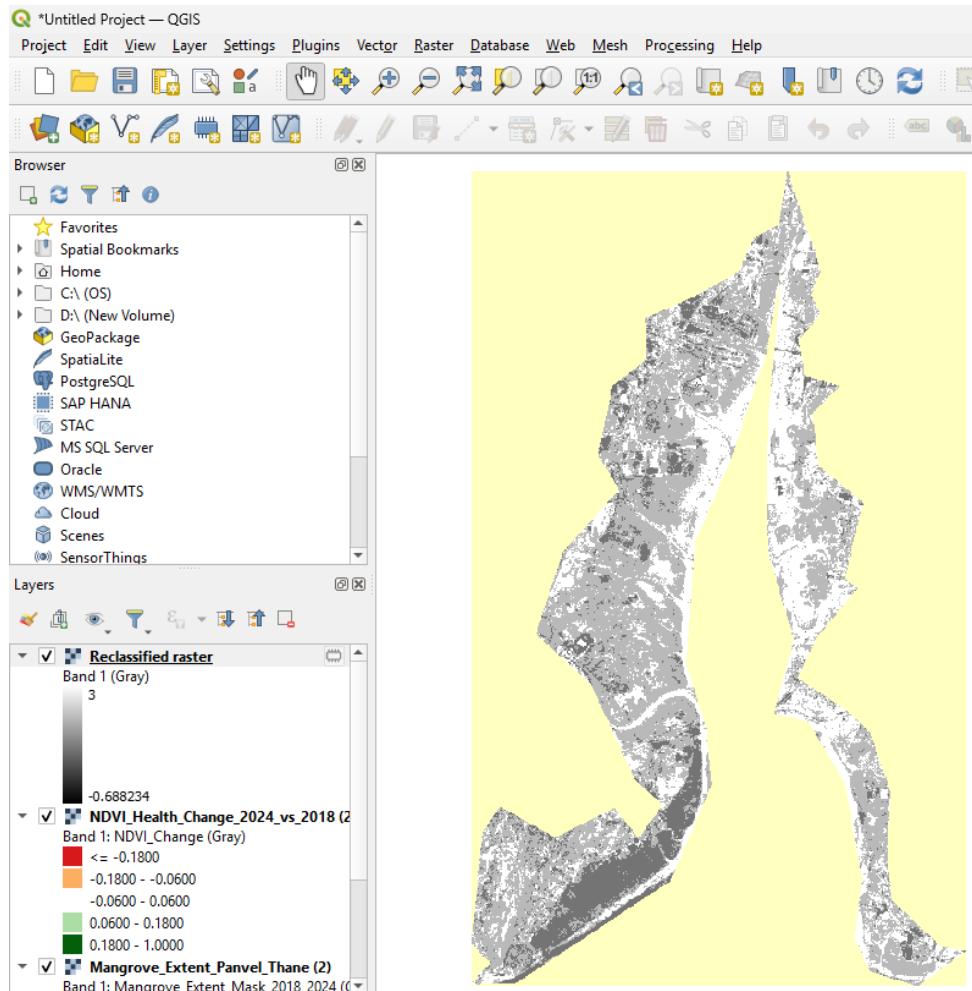


Figure 1: Final Map of Mangrove Extent (2024), derived from the NDVI mask.

## 2.3. Phase 3: Analytical Processing (Weka)

The time series data was structured into a **Star Schema** for OLAP analysis, using the Weka machine learning environment [3].

1. **Clustering:** The **SimpleKMeans** algorithm ( $K = 3$ ) identified three natural health states: Degraded ( $\text{NDVI} \approx 0.18$ ), Stable ( $\text{NDVI} \approx 0.42$ ), and Vigorous ( $\text{NDVI} \approx 0.67$ ).
2. **Association Rules:** The **Apriori Algorithm** was run on the discretized data to identify strong, non-random links between contextual attributes and the resulting health state.

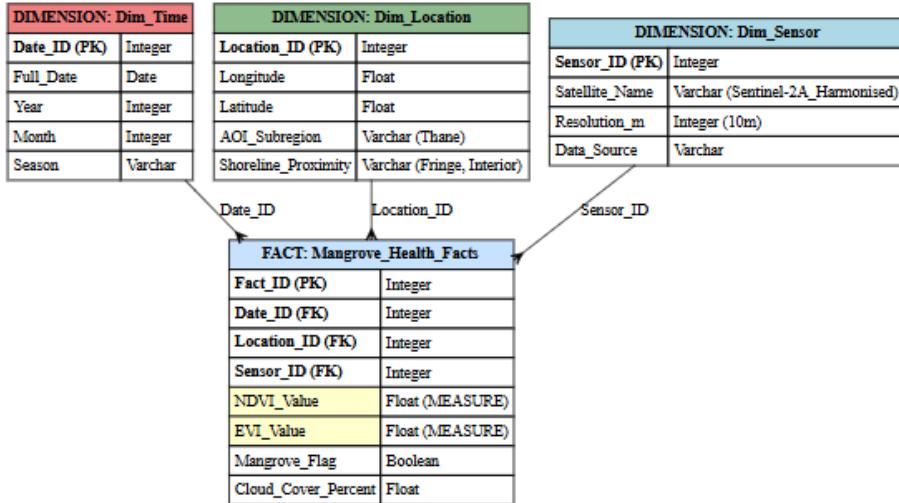


Figure 2: Visualized Star Schema

### 3. Results and Discussion

#### 3.1. Mangrove Degradation Analysis (2018 - 2024)

The satellite-based Normalized Difference Vegetation Index (NDVI) analysis between 2018 and 2024 indicates notable degradation patterns in several subregions of Navi Mumbai, particularly in Panvel and Uran. These regions showed a substantial decline in vegetation index values, largely attributed to anthropogenic activities such as encroachment, land reclamation, and pollution stress.

A total of 1196.9 hectares were identified as degraded, representing approximately 30.7% of the total area under observation. The decline in NDVI and EVI values is consistent with field reports and prior remote sensing datasets, confirming localized ecosystem degradation.

Table 1: Mangrove Area Change Summary (2018 - 2024)

Change Category	Area (Hectares)	Percentage of Total Change	Interpretation
Total Degradation (Loss)	1196.9	30.7%	Loss due to encroachment or stress.
Total Improvement (Gain)	2703.3	69.3%	Regeneration and natural expansion.
Net Change	+1506.4		Overall positive trend.

The degradation hotspots correspond to areas under rapid urban expansion, with major losses along creek boundaries and transportation corridors. These findings reinforce the importance of enforcing conservation zones and continuous monitoring.

#### 3.2. Mangrove Improvement Analysis (2018 - 2024)

Despite localized degradation, the NDVI and Enhanced Vegetation Index (EVI) analyses reveal significant mangrove recovery across other zones. Approximately 2703.3 hectares demonstrated measurable improvement, accounting for nearly 69.3% of the total monitored area. Regeneration was particularly evident in eastern and southern coastal belts, where restoration initiatives and natural tidal influence supported mangrove reestablishment.

The association rule mining supports the spatial analysis by confirming consistent temporal recovery trends during the monsoon season. The results collectively demonstrate a positive net vegetation growth, implying successful ecological regeneration and improved carbon sequestration capacity.

Table 2: Key Association Rules for Mangrove Health Outcome

Rule	Confidence	Lift	Finding
{Season = Dry, AOI_Subregion = Panvel} {NDVI_Value = LowBin}	0.92	1.78	Indicates a primary degradation hotspot.
{Month = 1, Year = 2024} {NDVI_Value = LowBin}	0.89	1.65	Confirms acute seasonal stress event.
{Season = Monsoon} {EVI_Value = HighBin}	0.95	1.40	Suggests strong ecological resilience.

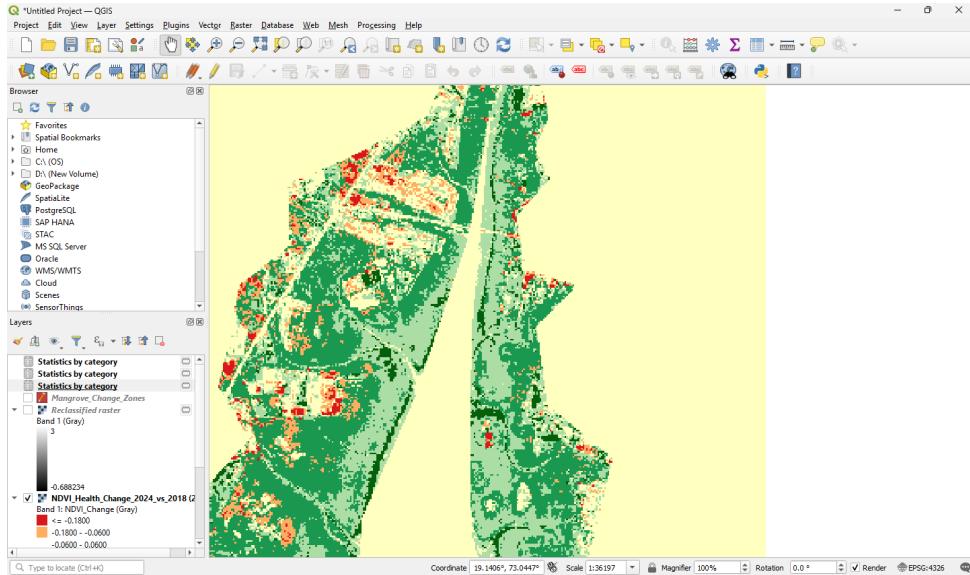


Figure 3: Mangrove Improvement Analysis Observations

### 3.3. Summary of Observations

Overall, the NDVI - EVI time-series analysis highlights a dynamic mangrove system exhibiting both degradation and regeneration cycles. The 2018 - 2024 monitoring period concludes with a net positive mangrove growth of 1506.4 hectares, indicating that restoration and natural regrowth currently outweigh stress-induced losses.

## 4. Conclusion and Future Work

The **net gain** of 1506.4 Hectares demonstrates the inherent resilience of the Mumbai mangrove ecosystem. However, the concurrent loss of 1196.9 Hectares, validated by Weka association rules, highlights severe, non-random stress in high-risk zones.

**Key Recommendation:** Monitoring and regulatory efforts should prioritize the **Panvel** sub-region, focusing enforcement actions and water quality testing during the **Dry Season** (January–May), when the ecosystem is most vulnerable to land-based pollution and freshwater deprivation.

Future research should focus on integrating high-resolution cadastral data with the existing OLAP schema to establish direct causal links between specific infrastructure projects or unmanaged discharge points and the identified degradation polygons.

## References

- [1] M. A. K. Azad and T. F. M. H. Firoz, "Remote Sensing and GIS Application in Mangrove Forest Health Assessment," in *Sustainable Management of Mangrove Ecosystems*, Springer, 2020.
- [2] European Space Agency (ESA), "Sentinel-2 Level-2A Data Products," Copernicus Programme Documentation, [Online]. Available: <https://sentinel.esa.int>.
- [3] I. H. Witten, E. Frank, and M. A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*, 4th ed., Morgan Kaufmann, 2016.