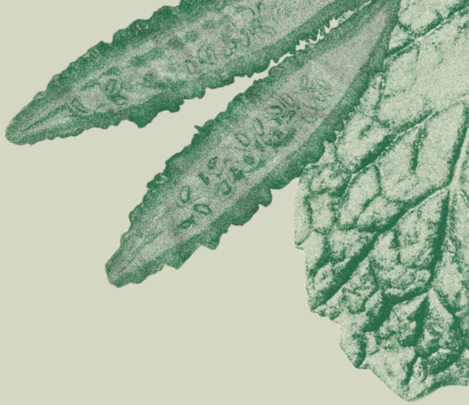
RISK PREDICTION

CSI IIT-BHU

MENTOR: PRAKSHAALE JAIN

MENTEES: BALAJI PRAVINYA, AMAN , VEDIKA,NAVKAR, ADITYAA

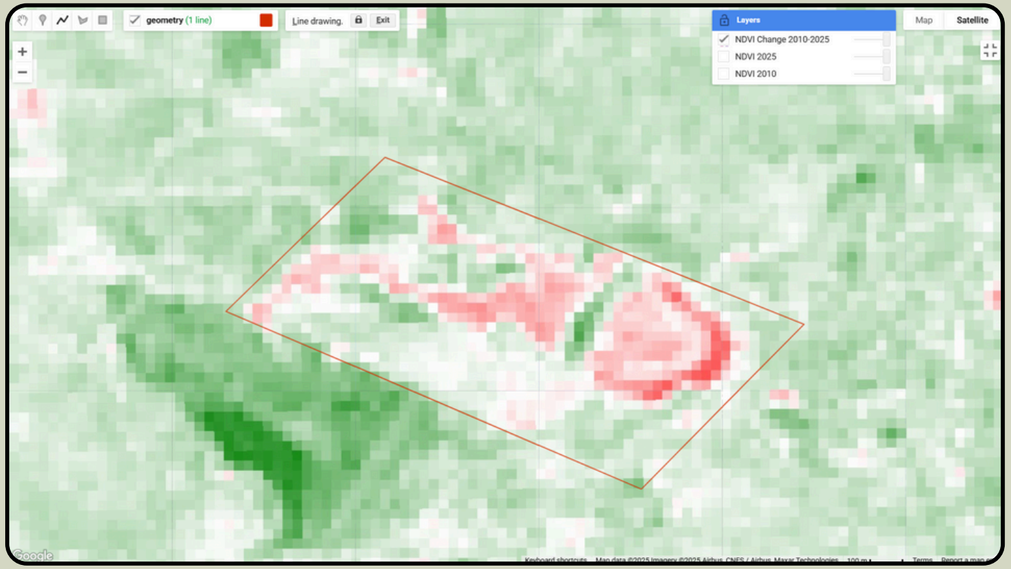


Task 1: Environmental Change Detection via Satellite

Used Google Earth Engine → Detected forest loss between 2010–2025 → Collected NVDI Changes

Key Insights:

- Gradual forest loss observed around UCIL mines
- Possible link to mining expansion or encroachment
- Code and full multi-site results are in the attached file
  LINK: **https://github.com/PrakshaaleJain/SoC - AIML/tree/main/Task%201**
- The files under Task-1 includes the GEE codes and the Maps showing NDVI change (Normalized Difference Vegetation Index) and Vegetation loss
- The red Spots shows loss, green shows gain and white shows no change in vegetation
- for reference we have attached the map for Jadugoda mine, first uranium mine of the country



Task 2: Building a Risk Prediction Model

Objective: To identify Indian districts that are environmentally and medically vulnerable by combining satellite vegetation loss data, proximity to mining areas, and health indicators.

Data Sources and Features:

NDVI_Loss_Percent – Measures the reduction in vegetation (higher = more environmental stress).(NDVI Loss data was collected using satellite-derived vegetation indices, primarily accessed from publicly available remote sensing platforms such as MODIS or Google Earth Engine, and aggregated at the district level to reflect percentage vegetation loss over a defined period.)

Distance to Mine (km) – Distance from the district to the nearest mine (closer = higher potential risk).(All data was programmatically collected using Python by integrating NFHS-5 health statistics, district coordinates from the Nominatim API, and mine location data from OpenStreetMap's Overpass API, followed by geospatial distance calculation to identify the nearest mining impact per district).

Respiratory diseases & Diarrhoea Prevalence (%) – Public health indicators collected from national surveys.

Coordinates – Latitude and longitude of each district for geographical mapping.(collected simultaneously along with distance from mines.)

Hypothesis: Districts that experience both significant vegetation degradation and are near mining activities are more likely to face health risks like ARI and diarrhoea.

Risk Prediction Using Machine Learning

Normalisation and risk calculation

Normalized features to bring them onto the same scale .Averaged features to create a risk score. Districts in the top 30% were labeled as "Risk".

```
#normalisation and risk calculation
normalized = (df[features] - df[features].min()) / (df[features].max() - df[features].min())
df["risk"] = normalized.mean(axis=1)

# Assigning high risk based on a threshold
threshold=df['risk'].quantile(0.70)
df['high_risk']=(df['risk']>threshold).astype(int)
```

Model Training

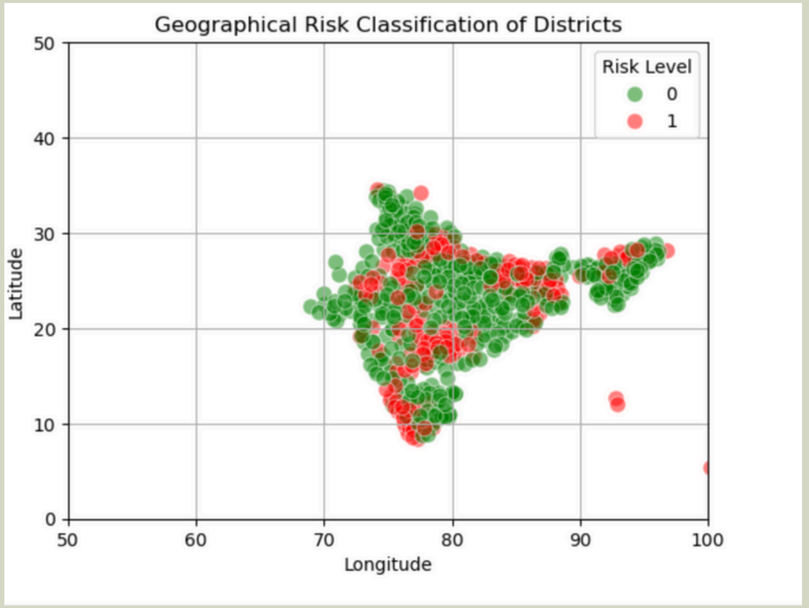
Training the model using random forest classifier, and setting the max depth to 3 to avoid overfitting the data and making the estimators to around 400 for better accuracy

```
clf=RandomForestClassifier(n_estimators= 400 ,max_depth=3,random_state=42)
clf.fit(X_trainscaled,y_train)

# Making predictions on the test set
y_pred = clf.predict(X_testscaled)
```

Visualizations and Risk Maps

Risk HeatMap:



- Each point is color-coded based on the district's composite risk score—ranging from cool (low risk) to warm (high risk) tones.
- The heatmap and scatter plot visualize each district's location using geographic coordinates.
- Red markers indicate high-risk districts, green markers indicate low-risk ones.
- This allows quick visual detection of clusters, showing that mining-intensive states like Odisha and Jharkhand have concentrated high-risk zones.
- These visuals help localize intervention zones and support environmental planning.

Interpretation of the Output from the Model

1. **Composite Score Insight**
- The score combines NDVI loss, proximity to mining, and disease rates into a single risk indicator.
- Higher scores reflect greater environmental degradation and health burden.

2. **Risk Distribution Trends**
- High-risk districts are concentrated in central and eastern India—particularly Jharkhand, Odisha, and Chhattisgarh.
- These areas tend to have both high vegetation loss and close proximity to mining activity.

3. **Model Validation**
- The Random Forest classifier showed good accuracy and recall in identifying high-risk zones.
- The model effectively uses input features to separate low and high-risk areas with minimal misclassification.

4. **Real-World Applications**
- Helps direct health and environmental resources to the most vulnerable districts.
- Can guide district-level planning, emergency preparedness, and sustainable development.
- Serves as a tool for monitoring long-term ecological and health risk patterns.

Conclusion: This model offers a practical and scalable way to prioritize intervention zones using a data-driven, composite risk framework. It empowers decision-makers to act early and allocate resources more effectively.