Machine Learning–Driven Prediction and Risk Assessment of Heavy Metal Contamination in Coastal Soil Regions

Introduction:

Coastal regions are more prone to heavy metal accumulation because of industrial and agricultural pollution. These metals enter soil and affect plants, water sources, and humans. Therefore, identifying heavy metal concentration in coastal soil regions is very important to reduce environmental and health impacts. Machine Learning helps us predict contamination levels without physically collecting soil samples.

Abstract:

In this project we use Sentinel remote-sensing images and machine learning algorithms to predict the concentration of heavy metals like Cadmium (Cd), Chromium (Cr), Lead (Pb), and Zinc (Zn) in coastal soils. We mainly implement Random Forest (RF), Support Vector Machine (SVM) and XGBoost algorithms to calculate prediction accuracy. The model gives predicted metal concentration and also generates a risk flag. XGBoost gives improved accuracy compared to the existing RF and SVM techniques.

Existing System:

The existing system uses Random Forest and Support Vector Machine algorithms for predicting heavy-metal values. These models give good accuracy but have limitations in large datasets and non-linear relations.

Proposed System:

The proposed method adds XGBoost algorithm, which improves accuracy, handles complex data well, and gives faster results. The model compares performance of RF, SVM and XGBoost and chooses the best prediction.

Process of Data Collection:

- We used Sentinel EO Browser (remote-sensing platform)
- Downloaded coastal area images
- Selected useful spectral bands
- Converted features into CSV dataset
- No physical soil samples were collected

Code Details:

- Data reading (CSV)
- Libraries import
- Preprocessing
- Feature selection
- Train–Test split
- Model training
- Prediction
- Evaluation

Python libraries used: pandas, numpy, sklearn, xgboost

Model Details:

Models used:

- Random Forest
- Support Vector Machine

- XGBoost (proposed)

For each metal (Cd, Cr, Pb, Zn) the model gives:

- predicted value
- accuracy
- risk level

Output:

The project outputs:

- Predicted metal concentration for each metal
- Accuracy of each algorithm
- Final best algorithm
- Risk Flags (safe/unsafe)

Final Model Total Accuracy:

In this project, three machine-learning models—Random Forest, Support Vector Machine, and XGBoost—were trained for predicting heavy-metal concentration in coastal soil. The final model accuracy was calculated by taking the **average accuracy of all the three algorithms**. After comparing individual metal-wise performances, XGBoost obtained the highest accuracy and therefore it is selected as the **best performing model** for predicting heavy-metal concentration in coastal regions.