Machine learning-powered Environmental Risk Prediction for Mining Operations

A predictive analytics dashboard designed to monitor and predict environmental risks across multiple mining sites, providing actionable insights for more sustainable practices

by Harshavardhan M

Project Overview & Motivation

Problem:

Mining operations face unpredictable environmental incidents such as landslides, water contamination, and habitat disruption. These risks are difficult to forecast with traditional methods.

Objective :

Develop a real-time risk scoring system using machine learning to monitor and predict environmental threats at mining sites, enabling timely interventions.

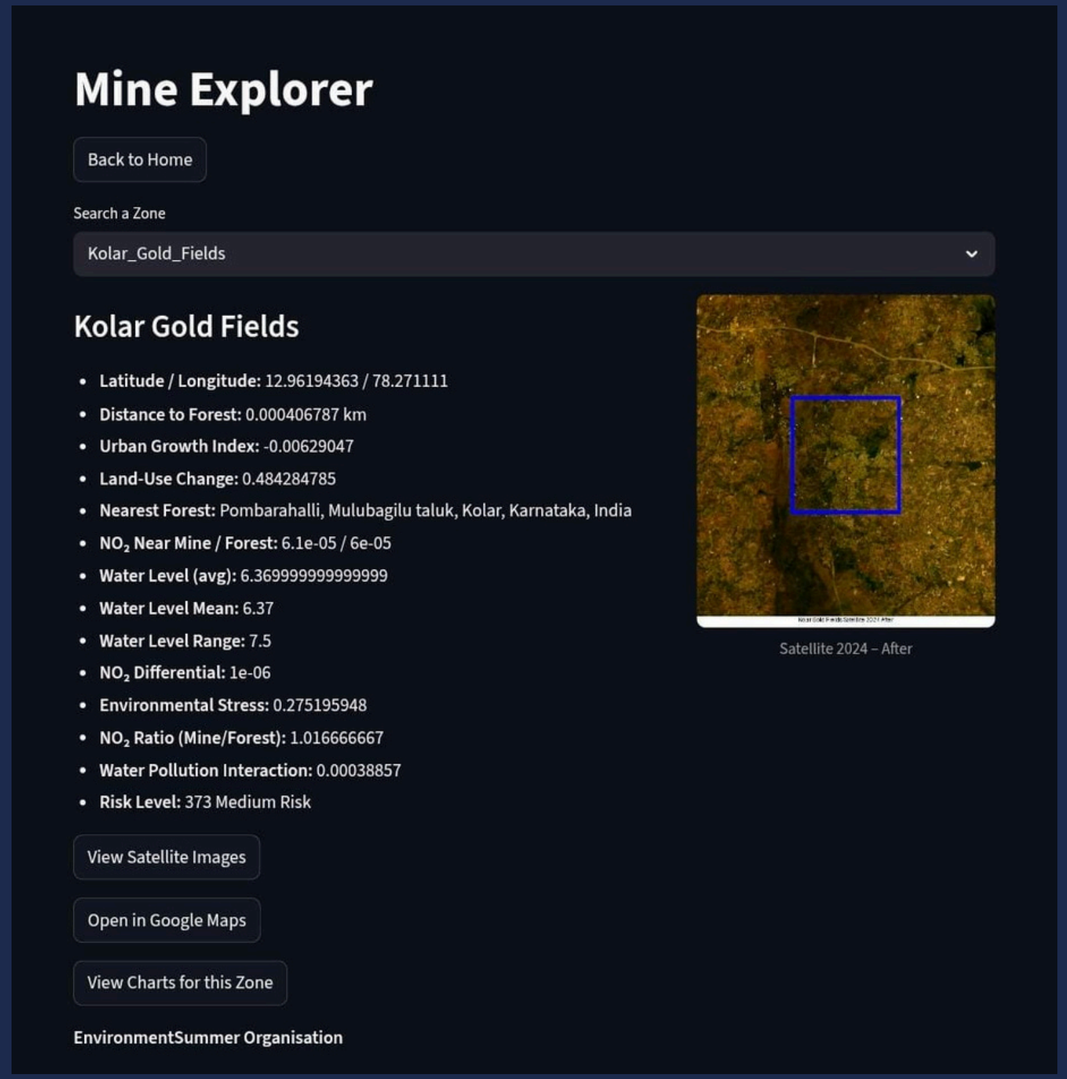
Impact :

Proactive risk management can reduce ecological incidents by up to 30% and improve regulatory compliance.

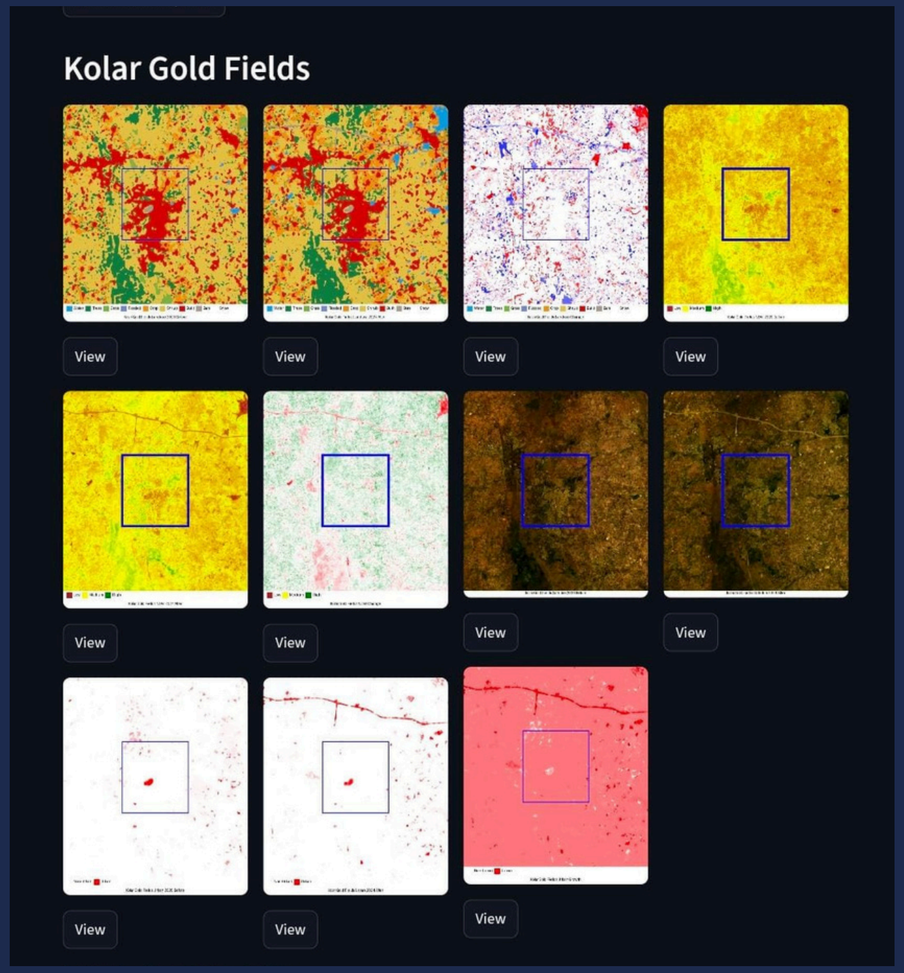
By harnessing advanced analytics, this project aims to transform how mining companies anticipate and respond to environmental risks

Key features of Risk Prediction Dashboard

- Interactive Risk Mapping: Visualize risk scores for all mines on a map or grid, allowing users to quickly identify high-risk sites
- Real-Time Data Updates: Dashboard automatically updates with new data, ensuring users always see the latest risk assessments
- Historical Trend Analysis: Display risk trends over time, enabling users to spot patterns and monitor changes
- User-Friendly Interface: Intuitive controls, filters, and search features make it easy for operators and managers to explore data and make decisions



Map Visualisations in the Dashboard



Satellite Imagery:

- The dashboard integrates high-resolution satellite images to provide a real-world view of mining areas. These images help users visually assess land cover, vegetation health, and changes over time.

Simple Maps:

- Simple maps display basic geographic features, such as mine boundaries and surrounding land use, making it easy to locate and understand the context of each site.

Heat Maps:

- Heat maps use color gradients to represent data intensity, such as areas with higher environmental risk or vegetation loss. In the dashboard, redder zones indicate higher concern, enabling quick identification of critical spots.

Combining satellite imagery with map overlays allows for intuitive monitoring of environmental changes, supporting data-driven decisions and targeted interventions.

Data Collection & Processing Flow

Data Collection

- Satellite Imagery: Use Google Earth Engine to collect NDVI, water levels, Urban Growth, Land use change and other geospatial features
- Operational Data: Gather mine-specific parameters, sensor readings, and historical records

Data Preprocessing

- Handling Missing Values: Impute or remove missing data
- Feature Engineering: Extract relevant features from raw data
- Feature Selection : Select most relevant features for risk prediction
- Export: Save processed data to CSV for next steps

Training the Model

- Model Development: Train machine learning models on processed dataset using Python and Google Colab.
- A Random Forest classifier was implemented to predict environmental risk categories
- The model was trained on a comprehensive dataset and validated using cross-validation, ensuring reliable performance.

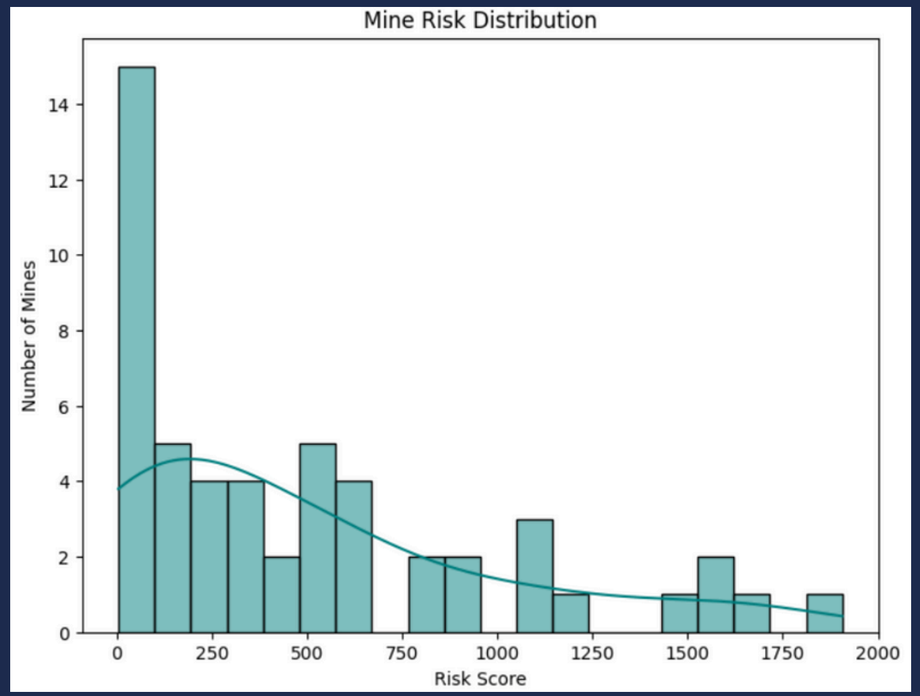
Model & Results

- The model achieved an overall accuracy of 81.84%, demonstrating its effectiveness.
- With an F1 score of 83%, the model maintains a strong balance between precision and recall
- Analysis of feature importance revealed that water level, Environment stress and others are the most influential factors in risk prediction.

Visualising Key Data Insights

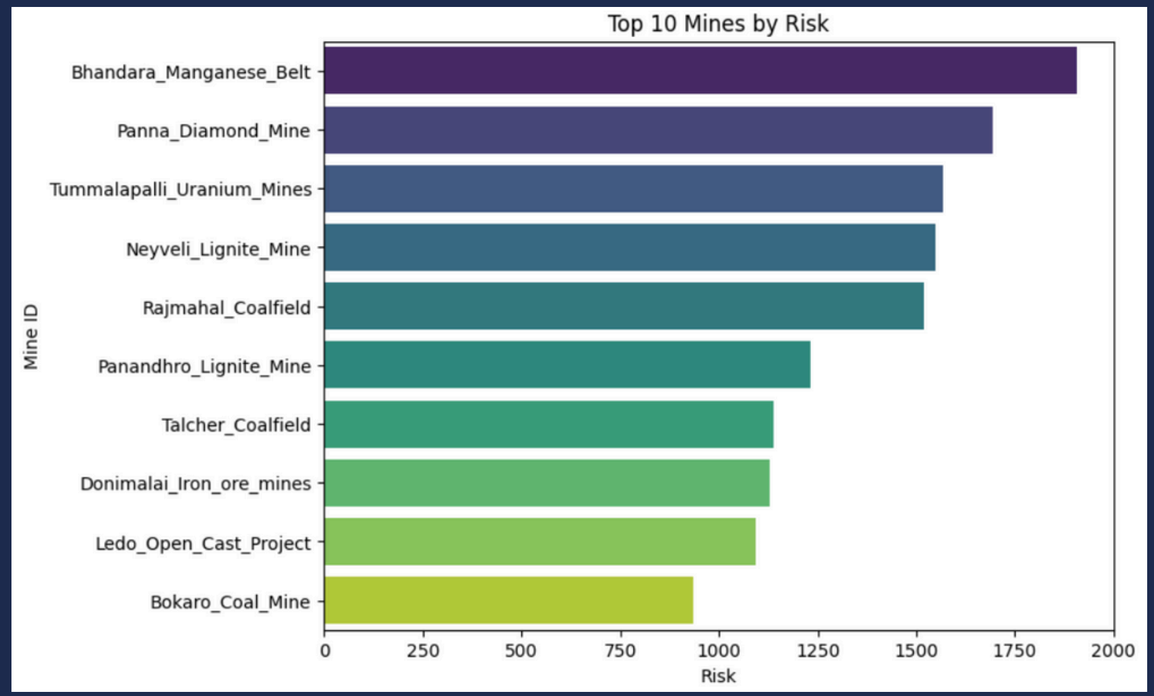
Mine Risk Distribution

This histogram illustrates the distribution of risk scores across all mines in the dataset. Most mines have low to moderate risk scores, while a smaller number exhibit significantly higher risk.



Top 10 Highest-Risk Mines

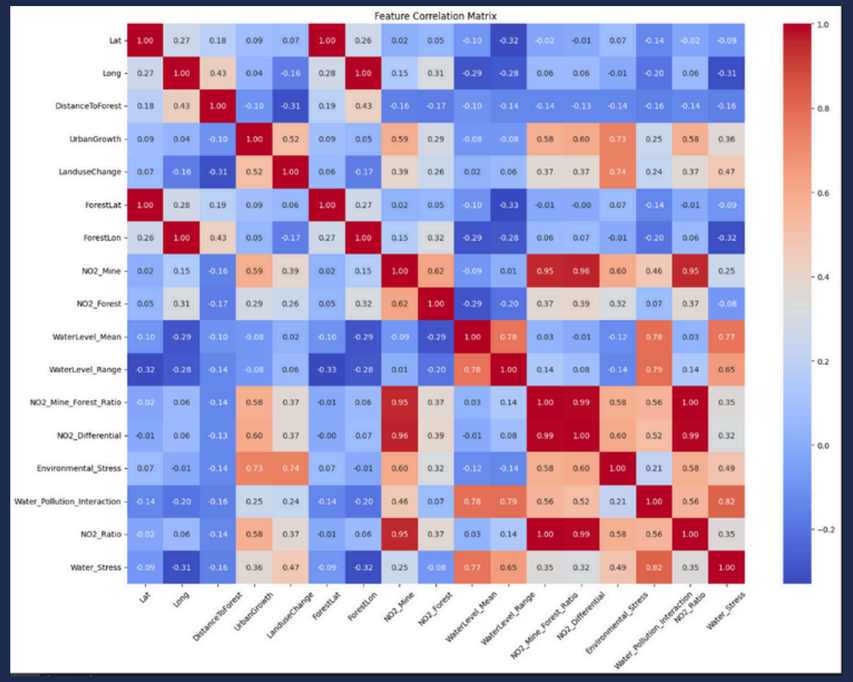
The bar plot ranks mines by their environmental risk scores, clearly identifying those requiring immediate attention and proactive monitoring.



Feature Selection and Importance

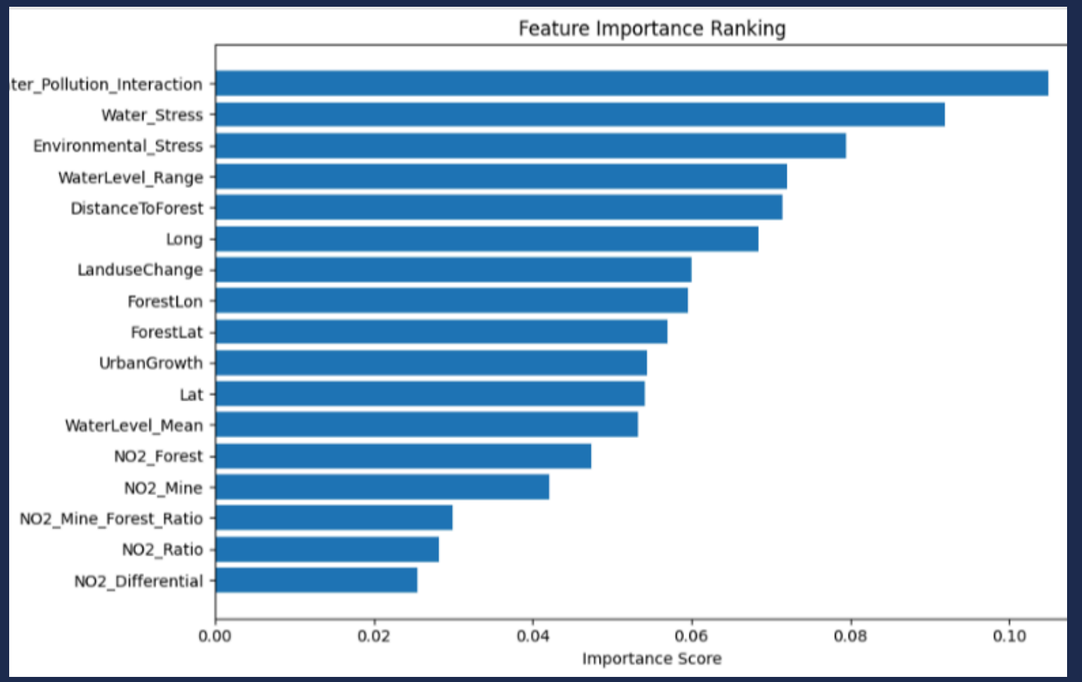
Feature Correlation Matrix

This heatmap visualizes the correlations between key environmental and operational features in the dataset. Strong positive and negative relationships are highlighted, revealing how variables like NO_2 levels, water stress, and land use changes interact. Understanding these correlations helps identify influential factors and guides effective feature selection for risk modeling.



Feature Importance Ranking

This bar chart displays the relative importance of each input feature in predicting environmental risk. Variables such as water pollution interaction, water stress, and environmental stress contribute most significantly to the model's predictions.



Robust Tech Stacks Used

- Python
- Google Colab
- Matplotlib
- Streamlit
- Scikit learn
- Seaborn
- Numpy
- Pandas
- Google Earth Engine

The project leverages a modern, scalable tech stack to ensure efficient data processing, accurate model training, and intuitive user experience. The integration of geospatial tools and machine learning frameworks enables robust risk assessment and actionable insights for mining operations.

Conclusion & Future Roadmap

Conclusion

Our machine learning model and dashboard provide actionable insights into environmental risk for mining operations. By prioritizing high-risk sites and visualizing key risk factors, we enable proactive risk management and improved safety outcomes.

Future Work

- Expand Data Sources: Incorporate real-time sensor data and additional environmental metrics.
- Enhance Model Accuracy: Integrate advanced algorithms and deeper feature engineering.
- Mobile Integration: Develop a mobile app for on-site risk alerts and monitoring.
- Regulatory Compliance: Automate reporting features to streamline regulatory compliance