# Merged Rockfall Prediction Dataset

Rows: 400, Columns: 28

## Layman's explanation

This dataset is a snapshot 'health report' for slopes in open-pit mines. Each row represents a location (tagged by a unique location\_id) with the most recent measurements collected from sensors, field surveys, drones, or geotechnical tests. The fields describe the steepness of the slope, the amount of recent rain, small ground movements, soil and rock strength, and whether a rockfall event has been observed. These combine to give a probability of a rockfall and allow for early warning alerts.

## Why it matters

Steeper slopes, high rainfall, rising pore pressure, and sudden displacements often precede rockfalls. By using this dataset, machine learning models can learn the patterns that tend to occur before a rockfall and trigger warnings for people and equipment at risk.

## Column dictionary (technical)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| column\_name | type | unit | technical meaning | example |
| location\_id | string | - | Unique ID for the site | LOC\_001 |
| data\_timestamp | datetime | ISO8601 | Time of most recent measurement | 2025-05-12T10:23:00 |
| slope\_angle\_deg | float | degrees | Slope steepness measured in degrees; higher = steeper | 34.5 |
| slope\_height\_km | float | km | Vertical height of the slope in kilometers | 0.12 |
| surface\_roughness | categorical | Low/Medium/High | Surface texture - influences how blocks detach and runout | Medium |
| rock\_size\_kg | float | kg | Representative mass of free blocks observed on slope | 1200 |
| runout\_distance\_km | float | km | Estimated distance a rock would travel after detachment | 0.02 |
| impact\_velocity\_mps | float | m/s | Estimated impact velocity of falling rock | 12.3 |
| displacement\_mm | float | mm | Measured movement of the slope face | 8.5 |
| displacement\_rate\_mm\_day | float | mm/day | Rate of movement per day | 1.2 |
| pore\_pressure\_kpa | float | kPa | Water pressure in pores; higher weakens slope | 45.2 |
| pore\_pressure\_ratio | float | - | Normalized pore pressure (0-1) | 0.12 |
| unit\_weight\_kn\_m3 | float | kN/m3 | Unit weight of rock/soil | 25.1 |
| cohesion\_kpa | float | kPa | Cohesive strength of material | 18.0 |
| internal\_friction\_deg | float | degrees | Friction angle of material | 34.0 |
| factor\_of\_safety | float | - | Computed stability factor; <1 indicates failure potential | 1.15 |
| reinforcement\_type | string | - | Type of reinforcement (bolt, mesh, none) | bolt |
| reinforcement\_numeric | int | - | Quantified reinforcement index | 3 |
| rainfall\_mm\_year | float | mm/year | Converted rainfall intensity to yearly equivalent | 2000 |
| cumulative\_rainfall\_mm | float | mm | Recent cumulative rainfall measured on site | 45 |
| seismic\_vibration\_g | float | g | Peak ground acceleration from local seismic/vibration sensors | 0.02 |
| crack\_length\_km | float | km | Length of visible cracks on slope face | 0.005 |
| crack\_width\_km | float | km | Width of cracks (in km, usually small) | 0.00002 |
| time\_since\_last\_blast\_hr | float | hours | Hours since most recent blasting event nearby | 72 |
| truck\_activity | categorical | Yes/No | Recent heavy truck traffic near the face | Yes |
| rockfall\_event | categorical | Yes/No | Recorded rockfall event at the location | No |
| alert\_level | categorical | High/Medium/Low | Combined alert level assigned by source systems | Medium |
| data\_conflict | categorical | Yes/No | Indicates whether datasets had conflicting values | No |

## Example row

**location\_id:** LOC\_219  
**data\_timestamp:** 2024-11-14 20:51:05  
**slope\_angle\_deg:** 68.31069832316169  
**slope\_height\_km:** 0.14036663879567654  
**surface\_roughness:** Low  
**rock\_size\_kg:** nan  
**runout\_distance\_km:** nan  
**impact\_velocity\_mps:** nan  
**displacement\_mm:** 0.15746670789081804  
**displacement\_rate\_mm\_day:** 0.3369222698361119  
**pore\_pressure\_kpa:** 408.393407021318  
**pore\_pressure\_ratio:** 0.9881774412544156  
**unit\_weight\_kn\_m3:** 24.89083209670861  
**cohesion\_kpa:** 34.46559075250027  
**internal\_friction\_deg:** 24.991025098811992  
**factor\_of\_safety:** 3.0121602653458477  
**reinforcement\_type:** Retaining Wall  
**reinforcement\_numeric:** 0  
**rainfall\_mm\_year:** 661684.078543772  
**cumulative\_rainfall\_mm:** 160.3541723888226  
**seismic\_vibration\_g:** 0.30683558309162073  
**crack\_length\_km:** 0.00412327723573715  
**crack\_width\_km:** 5.831463128093322e-05  
**time\_since\_last\_blast\_hr:** 73.80587932003911  
**truck\_activity:** No  
**rockfall\_event:** nan  
**alert\_level:** Red  
**data\_conflict:** yes

## Notes on cleaning & merging

- Column names standardized to short, lowercase, underscores.  
- Distances converted to kilometers. Rainfall intensities converted to mm/year.  
- When the same variable existed in multiple sources, the most recent timestamped value was chosen. data\_conflict is 'Yes' if different sources disagreed by more than 5% (numeric) or differed (categorical).  
- Some critical fields kept as NA where data missing; non-essential fields were kept or removed as described.  
- Small random noise (~±3%) added to numeric values to avoid synthetic bias and improve model robustness.