

# Supporting Information for “A framework for an AI pipeline for borehole data”

John M. Aiken<sup>1,2,3</sup>, Elliot Dufronet<sup>1</sup>, Hamed Amiri<sup>3</sup>, Lotta Ternieten<sup>3,4</sup>,

Oliver Plümper<sup>3</sup>

<sup>1</sup>Njord Centre, Departments of Physics and Geosciences, University of Oslo, PO BOX 1048, Blindern, Oslo, 0316, Oslo, Norway

<sup>2</sup>Expert Analytics, Oslo, Norway

<sup>3</sup>Department of Earth Sciences, Utrecht University, Utrecht, The Netherlands

<sup>4</sup>Department of Ocean Systems, Royal Netherlands Institute for Sea Research, The Netherlands

## Contents of this file

1. Text S1

## Introduction

This supplemental provides the responses to prompts given to ChatGPT with regards to the selection of the catboost model features.

### 1. ChatGPT designing catboost models

Below is the response from ChatGPT in reply to the prompt asking it to categorize the non-keyword columns in the dataset. These were used to produce the almost identical feature groupings and catboost comparison models (Figure ??) as the expert groupings (see Figure ??).

---

## 1.1. ChatGPT reply

Thank you for providing the column names. I'll define each column, categorize them, and explain the rationale behind the categorization.

### 1. Geological Composition and Properties:

- **Columns:** SiO<sub>2</sub>, TiO<sub>2</sub>, Al<sub>2</sub>O<sub>3</sub>, Fe<sub>2</sub>O<sub>3t</sub>, MnO, MgO, CaO, Na<sub>2</sub>O, K<sub>2</sub>O, P<sub>2</sub>O<sub>5</sub>, 100\*Fe(III)/FeT, Vrecal, Crrecal, Co, Nirecal, Curecal, Znrecal, Srrecal.

- **Definition:** These columns represent the chemical composition of the borehole core sample, measured in terms of the concentration of various elements and compounds.

- **Rationale:** These features are crucial for understanding the mineralogical composition of the rock, which is directly related to its alteration characteristics.

### 2. Physical Properties

- **Columns:** Mean dry electrical Resistivity (ohmm), Bulk density (g/cm<sup>3</sup>), AMS bulk susceptibility, LOI wt%, CO<sub>2</sub> wt%, H<sub>2</sub>O wt%, CaCO<sub>3</sub> calc.

- **Definition:** These columns represent physical properties of the core samples, such as resistivity, density, magnetic susceptibility, and content of specific compounds.

- **Rationale:** Physical properties can influence how a rock reacts to geological processes, including alteration.

### 3. Biological Influence

- **Column:** Cell abundance (cells/g).

- **Definition:** This measures the abundance of microbial life per gram of the sample.

- **Rationale:** Microbial activity can significantly impact mineral alteration processes.

#### 4. Fracture Analysis

- **Columns:** % of fractures, PnS2\_sum, PnL\_sum, PnP3V\_sum, PnP3H\_sum, PnP4\_sum, PnP6V\_sum, FnS2\_sum, FnL\_sum, FnP3V\_sum, FnP3H\_sum, FnP4\_sum, FnP6V\_sum.

- **Definition:** These columns quantify the presence and characteristics of fractures in the sample.

- **Rationale:** Fractures can significantly affect alteration processes, potentially facilitating fluid flow and chemical reactions.

#### 5. Rock Type and Unit Classification

- **Columns:** UNIT\_TYPE\_Dunite, UNIT\_TYPE\_Fault rock, UNIT\_TYPE\_Gabbro, UNIT\_TYPE\_Harzburgite, UNIT\_TYPE\_Metagabbro, UNIT\_TYPE\_Other, UNIT\_CLASS\_OPHIO, UNIT\_CLASS\_UND.

- **Definition:** These columns categorize the rock types and geological units of the samples.

- **Rationale:** Different rock types have different susceptibilities to alteration.

#### 6. Textural Features

- **Columns:** TEXTURES\_Brecciated, TEXTURES\_Sheared, GRAINSIZE\_Cryptocrystalline, GRAINSIZE\_Fine grained, GRAINSIZE\_Medium grained, GRAINSIZE\_Microcrystalline, GRAINSIZE2\_Coarse grained, GRAINSIZE2\_Cryptocrystalline, GRAINSIZE2\_Fine grained, GRAINSIZE2\_Medium grained, GRAINSIZE2\_Pegmatitic.

- **Definition:** These columns describe the textural characteristics of the rock samples.

- **Rationale:** Textural features can influence the rate and extent of alteration processes.

## 7. Color and Visual Properties

- **Columns:** Redness, Greenness, Blueness, Y (luminance).
- **Definition:** These columns represent the colorimetric properties of the samples.
- **Rationale:** Color and luminance can be indicative of certain mineralogical changes and alteration processes.

This categorization helps in understanding the various aspects that influence peridotite alteration in borehole cores, especially in the context of reaction-driven cracking and the impact of fractures. Each category represents a different aspect of the geological, biological, physical, and chemical environment of the rock samples.