Chemical Analysis Report

Classification Model

Petter Lövehagen

2023-05-07

## Brief

The brief is to develop a classification tool to classify chemicals into classes.

In addition, the client is interested in knowing whether all variables are necessary for achieving optimal out-of-sample classification performance, or if a subset of measurements will suffice.

This report will aim to:

* analyse and feedback on the provided dataset
* develop an optimsed classification model
* determine whether all variables are required

## Data

The supplied dataset contains 2500 measurements of 20 numerical variables. We have not been supplied with any additional information or context, for example variable importance, scale, unit of measurement, variable interactions, measurement challenges, etc.

Therefore, variables will be treated equally and without bias when data processing and modelling.

|  |
| --- |
| To consider |
| However, additional contextual information from the client could result in different modelling decisions. |

### Missing data

[analysis] [decision making] [questions for company to consider]

### Exploratory Data Analysis

[histograms] [outliers] [pair plots] [correlation]

### Splitting data

[approach] [decision]