# Portfolio Submission

Week 2

Studio 1-5

**Dataset selected in Studio 1**: Liquid Battery Electrolyte Formulation in Electronics

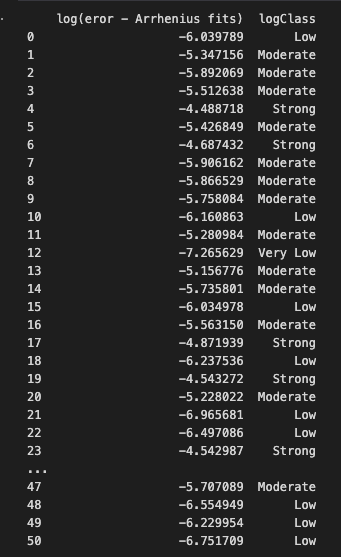
**Reason**: All of the dataset at Studio 1 are not familiar from myself because I only study Computer Science, majoring in AI. However, as I have learned Computer System, which let me understand quite a bit about hardware system, so electronics is the kind of what I think they have a connection in each other.

Summary of exploratory data analysis (EDA) in studio 1:

* ﻿﻿ ﻿﻿﻿Features like Salt, VC, EC are having weak significantly in terms of relationship with log (error - Arrhenius fits) feature and does not account for making statistical decision (of correlation)
* ﻿﻿﻿ChemRiv Salt feature is having Low Positive Correlation with VC and C6-50C features, perhaps we can create additional features such as (Salt + VC) and (Salt + C6-50C) to predict the concrete strength
* ﻿﻿﻿Range of clusters in the dataset from 2 to 6

Class labelling for target variable / developing ground truth data

-8, -7, -6, -5, -4, -3 Equal as 'Very Low', 'Low', 'Moderate', 'Strong', 'Very Strong'



Feature engineering – column VC

A black background with white numbers

Description automatically generated

Feature selection

I choose these columns:

'C5-40C', 'C7-60C', 'VC', 'EMC', 'saltEC', 'ecEMC', 'c1C2', 's1S2'

Training and decision tree model development

It still loading for a long time

# Appendix

1. Source-code: https://github.com/Phonginhere/cos40007/tree/main/week1/practicalWeek1/homework