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'
```

```
      log(eror - Arrhenius fits)  logClass
0                     -6.039789        Low
1                     -5.347156   Moderate
2                     -5.892069   Moderate
3                     -5.512638   Moderate
4                     -4.488718     Strong
5                     -5.426849   Moderate
6                     -4.687432     Strong
7                     -5.906162   Moderate
8                     -5.866529   Moderate
9                     -5.758084   Moderate
10                    -6.160863        Low
11                    -5.280984   Moderate
12                    -7.265629   Very Low
13                    -5.156776   Moderate
14                    -5.735801   Moderate
15                    -6.034978        Low
16                    -5.563150   Moderate
17                    -4.871939     Strong
18                    -6.237536        Low
19                    -4.543272     Strong
20                    -5.228022   Moderate
21                    -6.965681        Low
22                    -6.497086        Low
23                    -4.542987     Strong
...
47                    -5.707089   Moderate
48                    -6.554949        Low
49                    -6.229954        Low
50                    -6.751709        Low
```

Feature engineering – column VC

```
How many 24
0.0
0.108546
0.103225
0.100083
0.307606
0.289704
0.273225
0.101654
0.096633
0.082602
0.310562
0.286128
0.268895
0.103901
0.099193
0.296476
0.282732
0.267502
0.114292
0.105917
0.108919
0.322698
0.323792
0.308579
```

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