Evaluating Clustering Algorithms for Prediction of Rock Type for Oil and Gas Applications Using United Kingdom Core Data

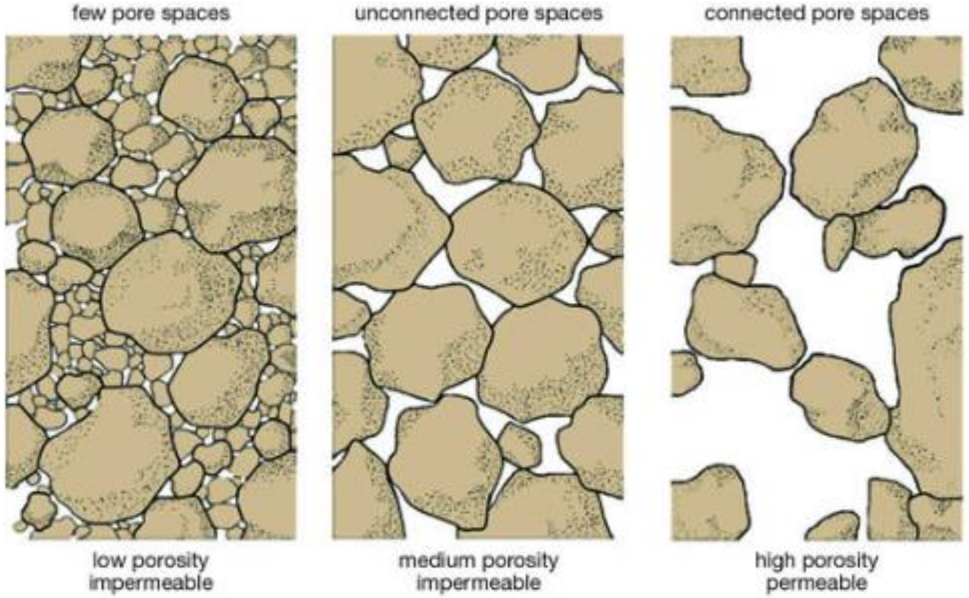
Tan Wen Xuan

Z1981074

Presentation date: 200922

Introduction

- Rock typing is the process of grouping rocks based on their characteristics
- Background
  - Oil & Gas are acquired by drilling wells in sedimentary basins
  - Rock samples are extracted to discover flow capacity and storage
  - Generalised for future drilling
- Business Problem
  - Traditional methods are time consuming
  - Require domain knowledge
  - Can be costly
- Objective
  - Compare unsupervised algorithms against a traditional method



Note. Department of Mines, Industry Regulation and Safety (n.d.).

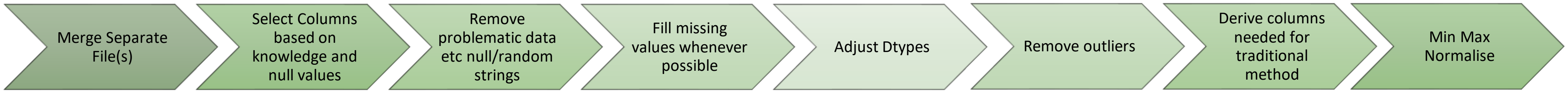
Previous Studies

- Revolves around usage of Indices for rock typing
  - Flow Zone Indicator (FZI)
    - $FZI = \frac{RQI}{\varphi_z}$
      - RQI – Rock Quality Index: $0.0314 \times \sqrt{\frac{k}{\varphi_z}}$
      - k – Permeability
      - φ – Porosity
      - φ_z - Normalized Porosity: $\frac{\varphi}{1-\varphi}$
- Machine learning with indices
- Rock Typing with Machine learning (standalone)
- What I aim to do:
  - Compare the first method against the third
  - Add value due time savings

Data

- Publicly available core data from three regions (UK, USA and North Sea)
- Columns needed were derived first
- All data were min-max normalized
- Cleaned data’s columns are mainly porosity and permeability
- Training of models done on UK first
- End Goal: Complete training on 2-3 regions

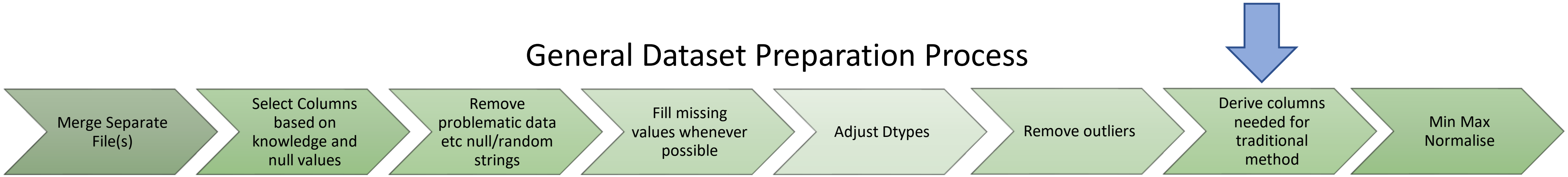
General Dataset Preparation Process



Pre-Modelling Data Preparation

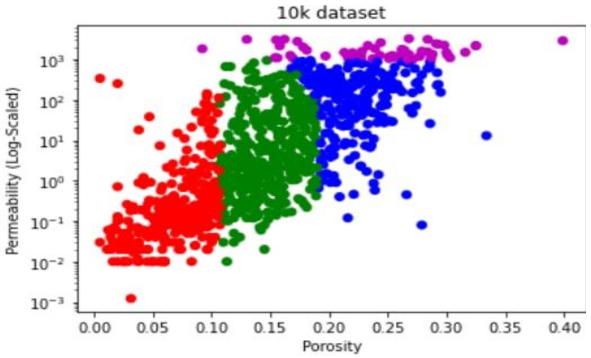
- Data preparation needed for the traditional method (using FZI)
- Calculation of RQI, Porosity Index, FZI, Log RQI and Log Porosity Index
- Recap:
  - $FZI = \frac{RQI}{\varphi_z}$
  - RQI – Rock Quality Index: $0.0314 \times \sqrt{\frac{k}{\varphi}}$
  - k – Permeability
  - φ – Porosity
  - φ_z - Normalized Porosity: $\frac{\varphi}{1-\varphi}$ (Porosity Index)

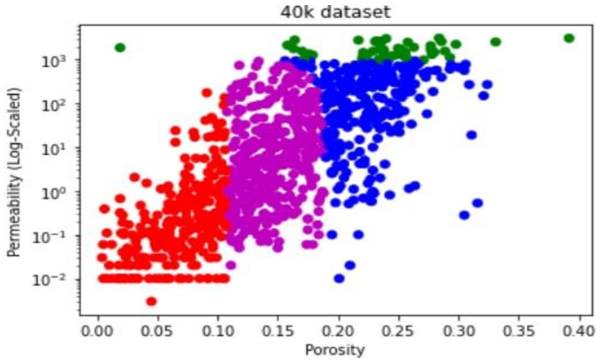
General Dataset Preparation Process

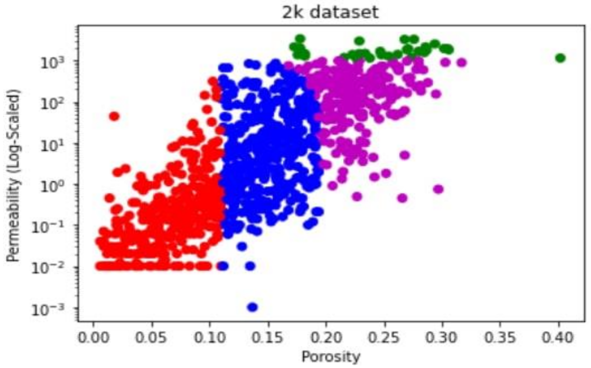


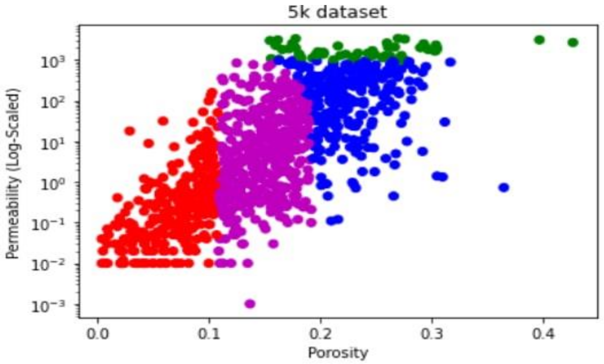
Choosing Size of Dataset

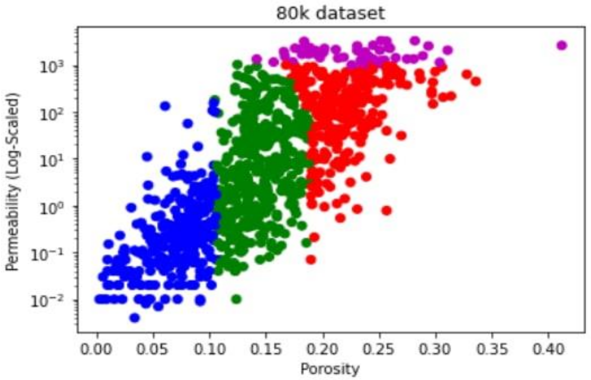
- Cleaned dataset consisted of 104,832 data points
- For computation efficiency, explore possibility to use smaller set of data
- Use Kmeans to perform clustering on different sizes of dataset
  - Random state is the same
  - Optimal cluster set to 4 using inertia plot
- Compare porosity vs log-scaled permeability plot
  - To observe cluster distribution
- Sizes include 2k, 5k, 10k, 40k, 80k, and full dataset
- 2k sized dataset selected

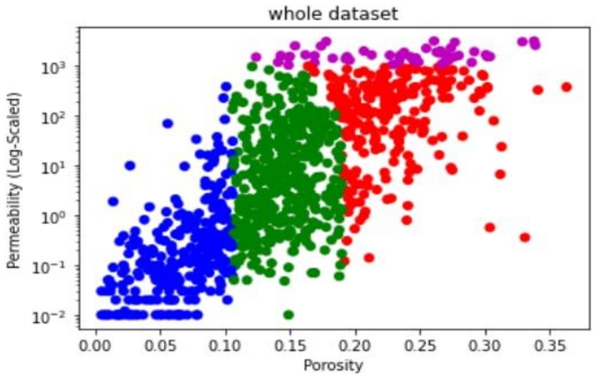






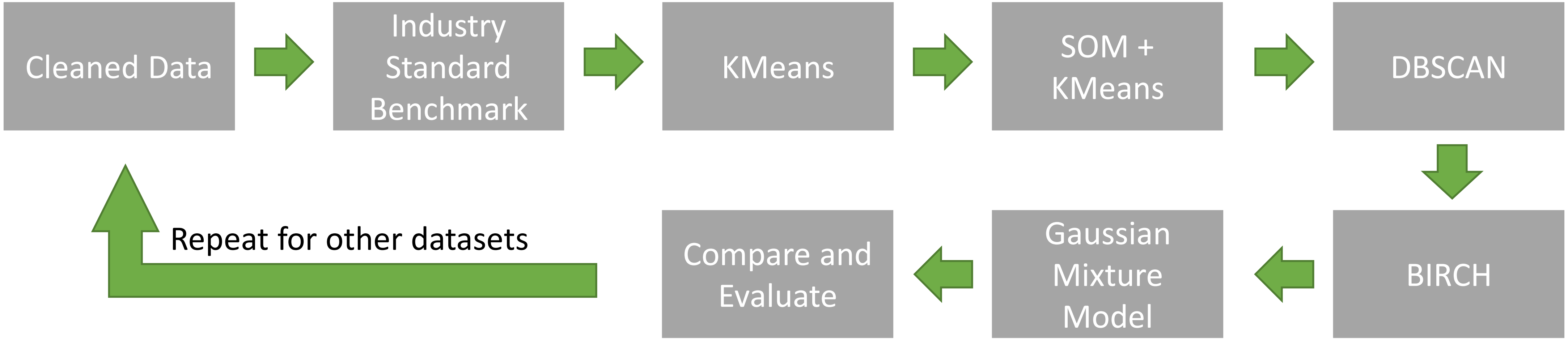






Modelling

- 1 Industry standard and 5 unsupervised algorithms
- Only UK core data was trained
- 2,000 samples (rows), 2 features
- Results and comparison
- Repeat the process for other datasets (as many as possible)

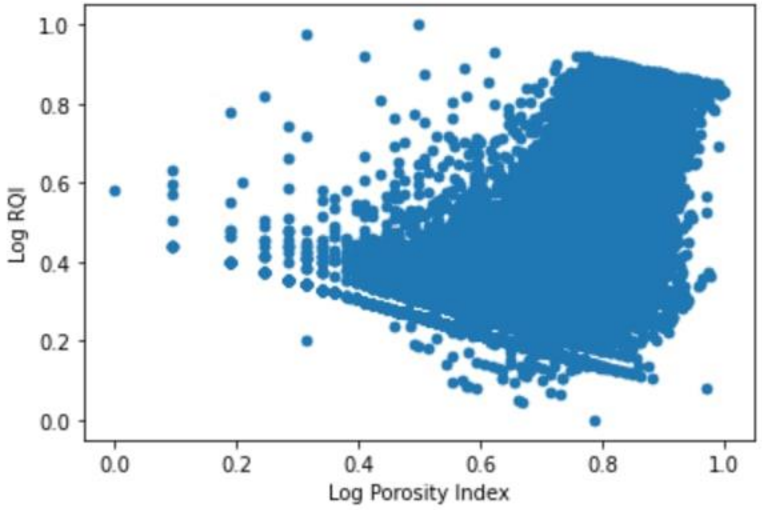


Traditional Method Selected

- Method – Iterative multi-linear regression (IMLR) (Khalid et. al., 2019)
  1. Plot scatter plot log normalized porosity (x-axis) against log RQI (y-axis)
  2. Guess intercept value based on mean FZI for each visible straight line
  3. Lines selected based on the points
  4. Line all have slope = 1
  5. Allocate samples to closest line
  6. Calculate new intercept using least square regression
  7. Re-adjust lines if difference is too big
- Other Methods
  - Histogram Analysis
    - Log fzi
  - Normal Probability Plot
    - Log fzi

Failure of Traditional Method

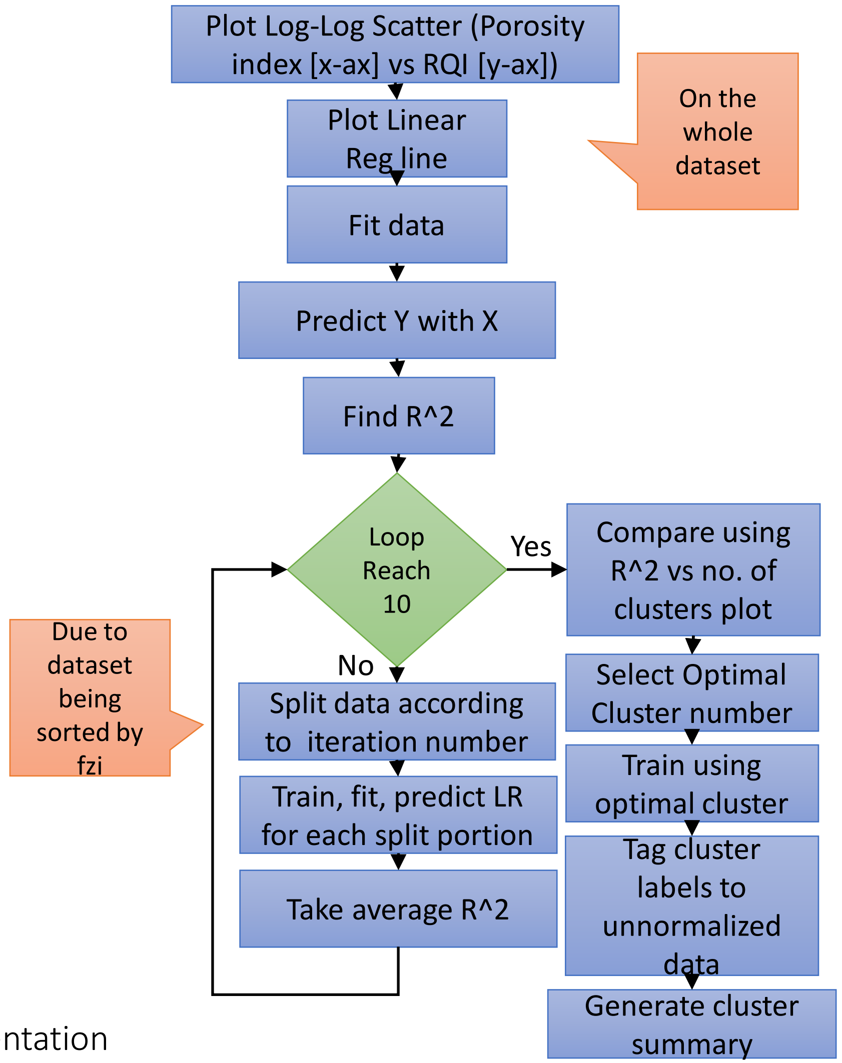
- Method Requirement
  - there must be clear separation in the log-log plot to select initial lines
- Failure – method not feasible due to data too close
- Failure -> limited traditional method
- Proposed method – own method (adapted from different sources)



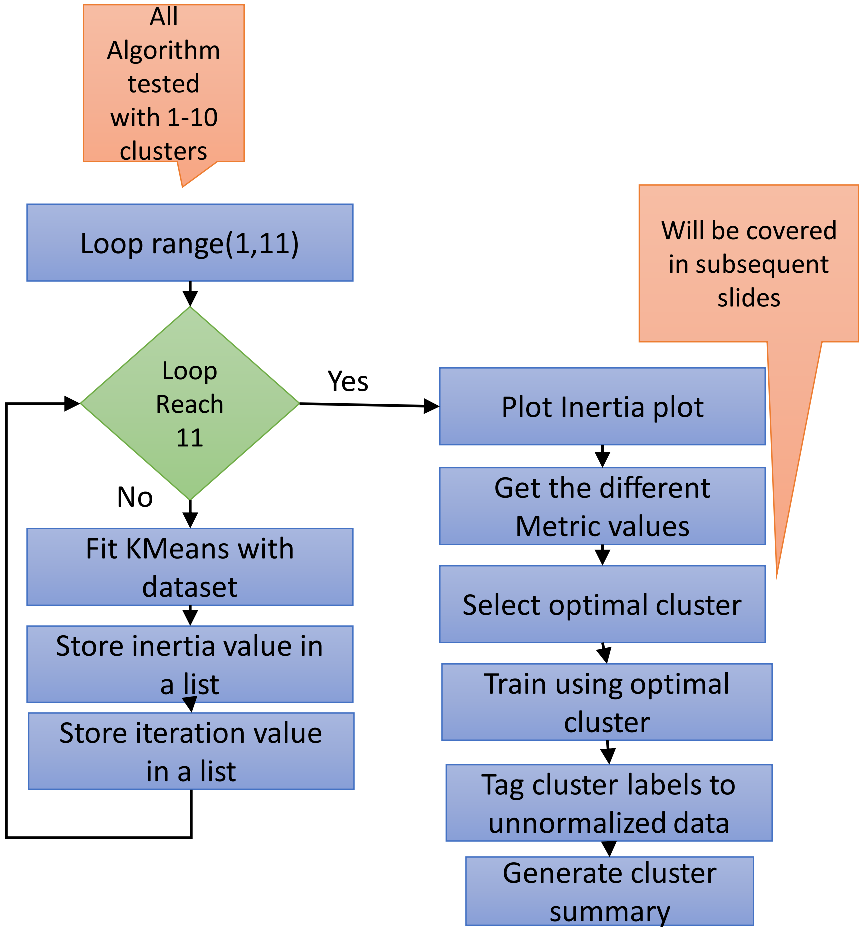
Methods Employed

SINGAPORE UNIVERSITY OF SOCIAL SCIENCES

Flow Zone Inzdicator (FZI) [Own Method]

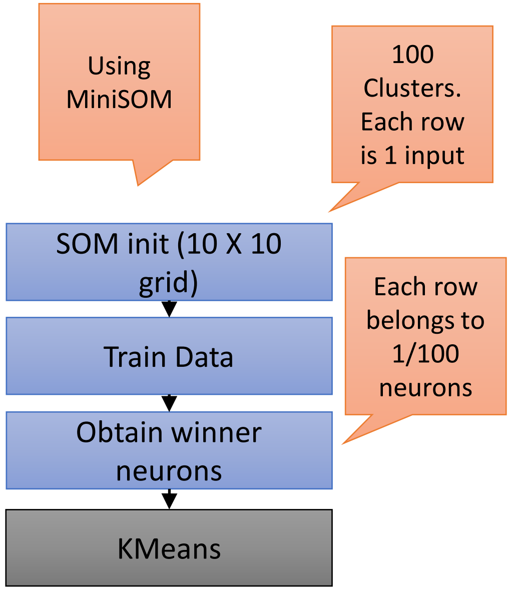


KMeans

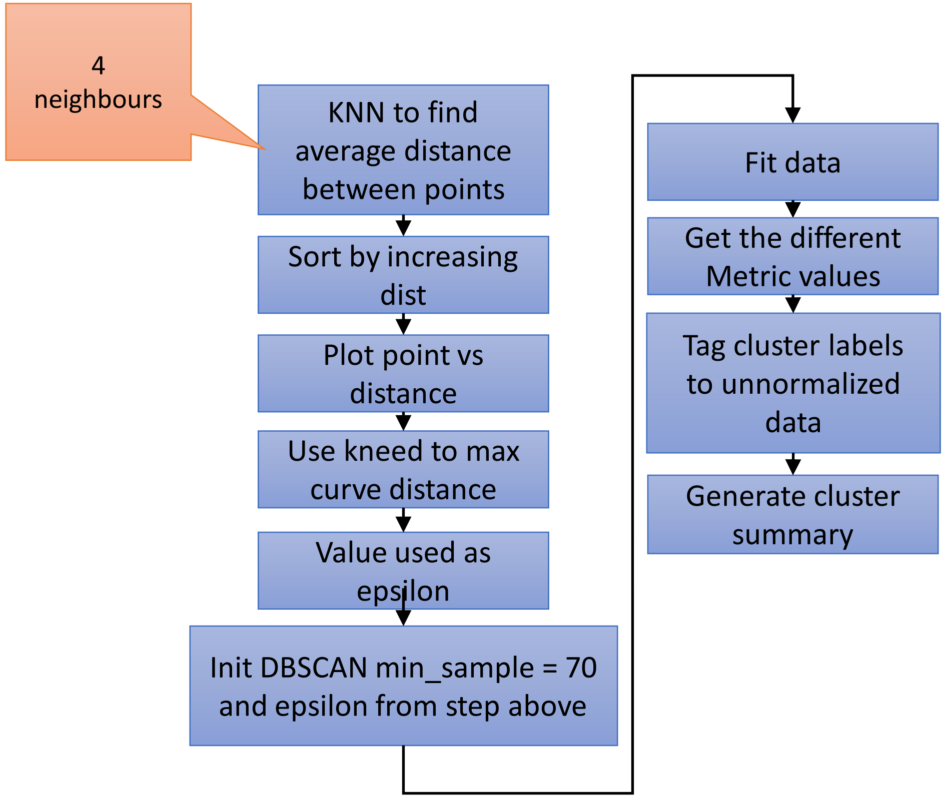


Methods Employed

Self Organising Maps (SOM) + KMeans

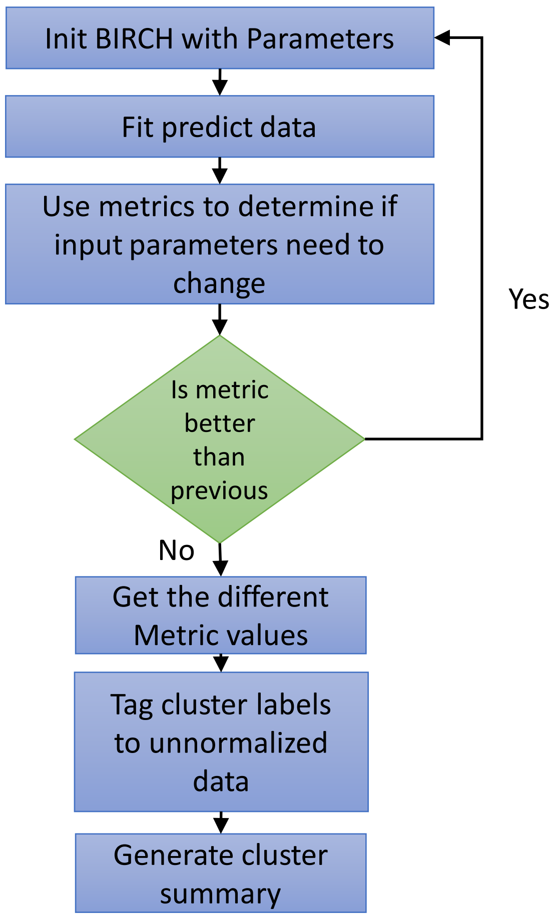


DBSCAN

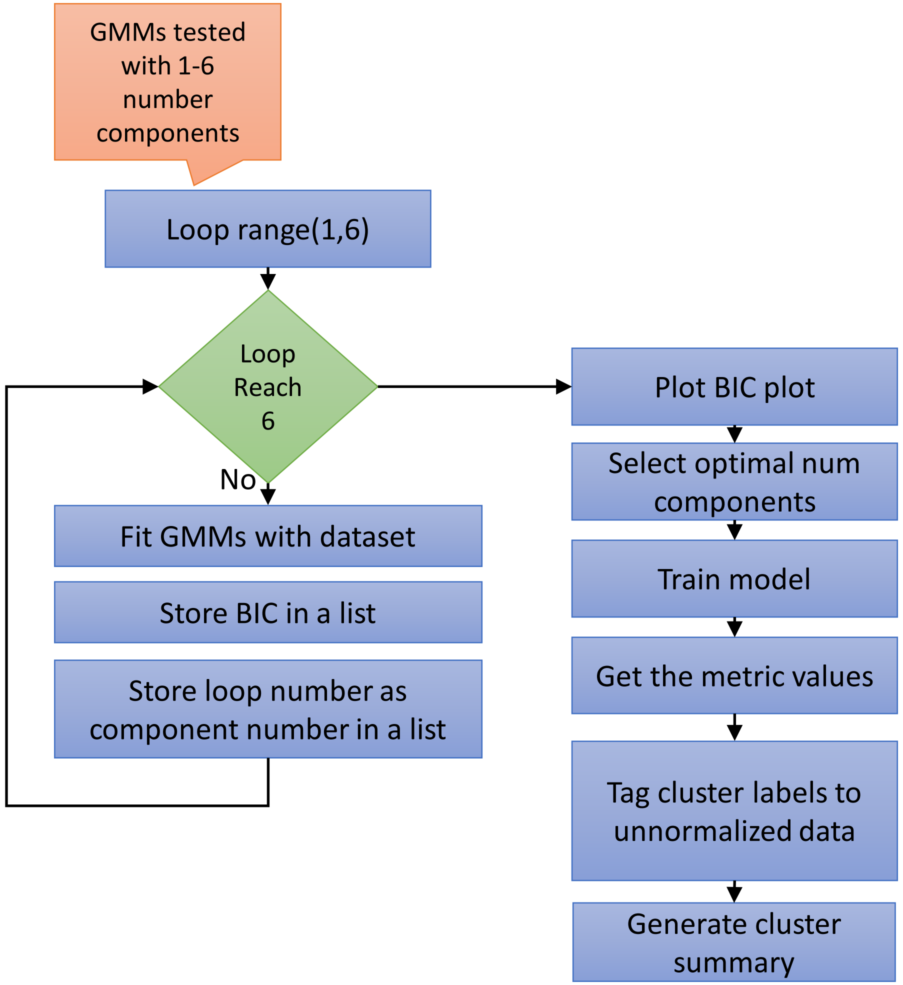


Methods Employed

BIRCH



Gaussian Mixture Model



Training and Visualisation

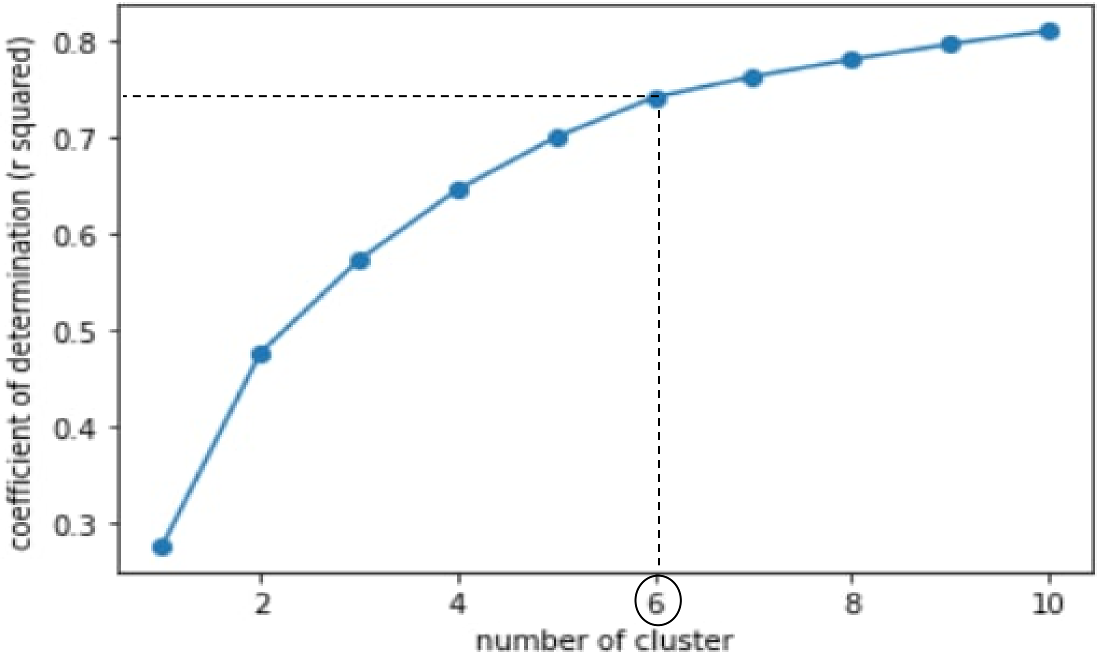
- Different method's modelling and visualisation
- Scatterplots done for visualisation
- Cluster labelling are fixed with a specific colour

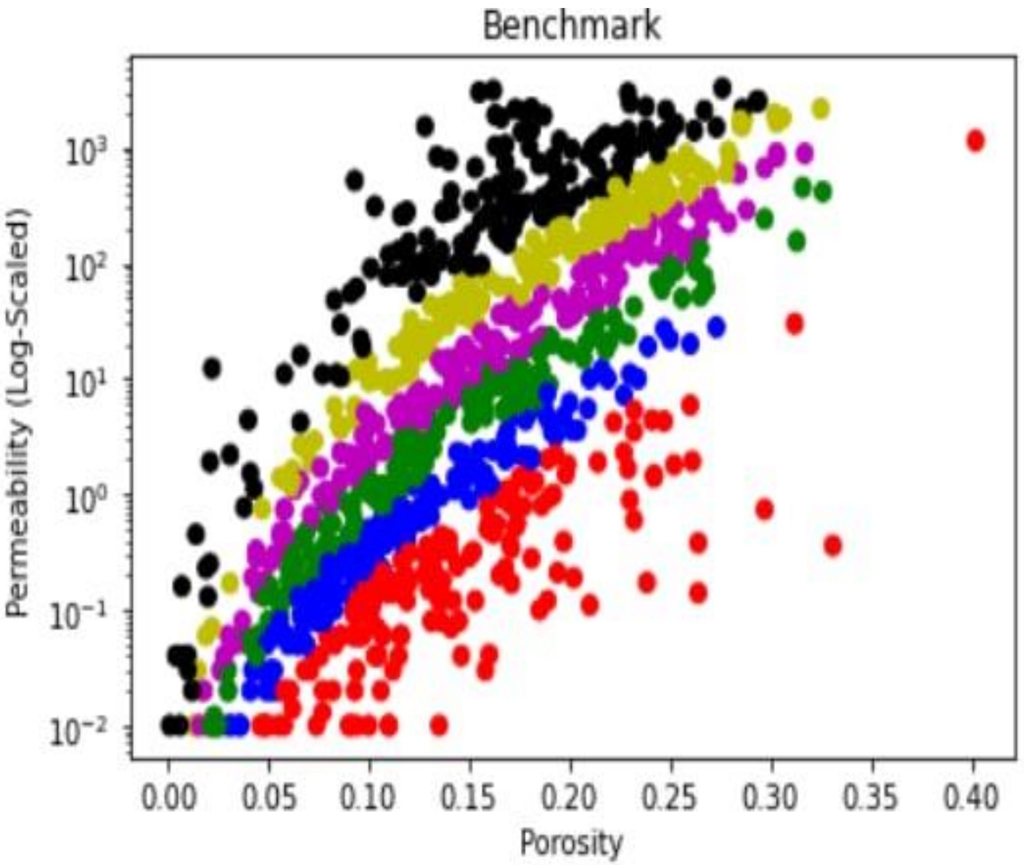
Cluster	1	2	3	4	5	6
Colour	Red	Blue	Green	Magenta	Yellow	Black

- Cluster 0 is fixed with gold for DBSCAN
- Clusters re-ordering was attempted for comparison
- Visualisation with 1000 sample data
- Y-axis (permeability) log scaled to remove skew

Own Implementation of Traditional Method

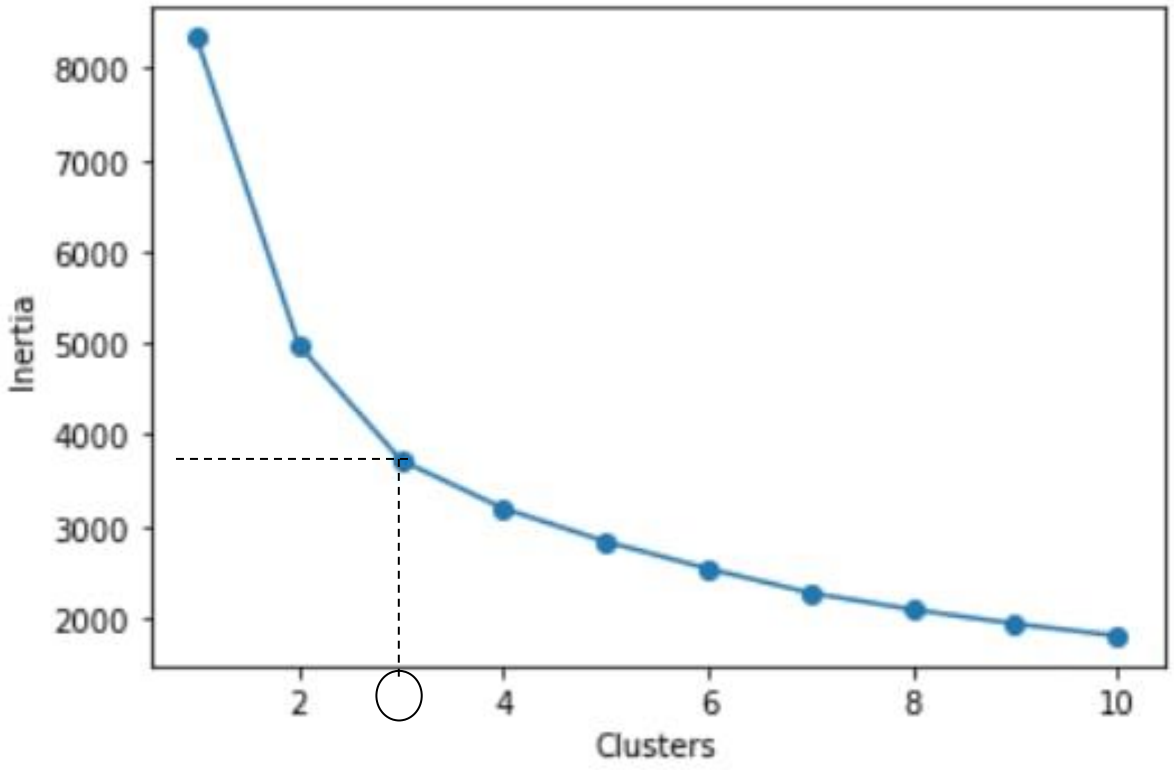
- Iteration 1-10
- Plot corresponding number of LR line
- Get average R^2
- Compare using R^2 vs Cluster plot
- Optimal Cluster of 6 chosen

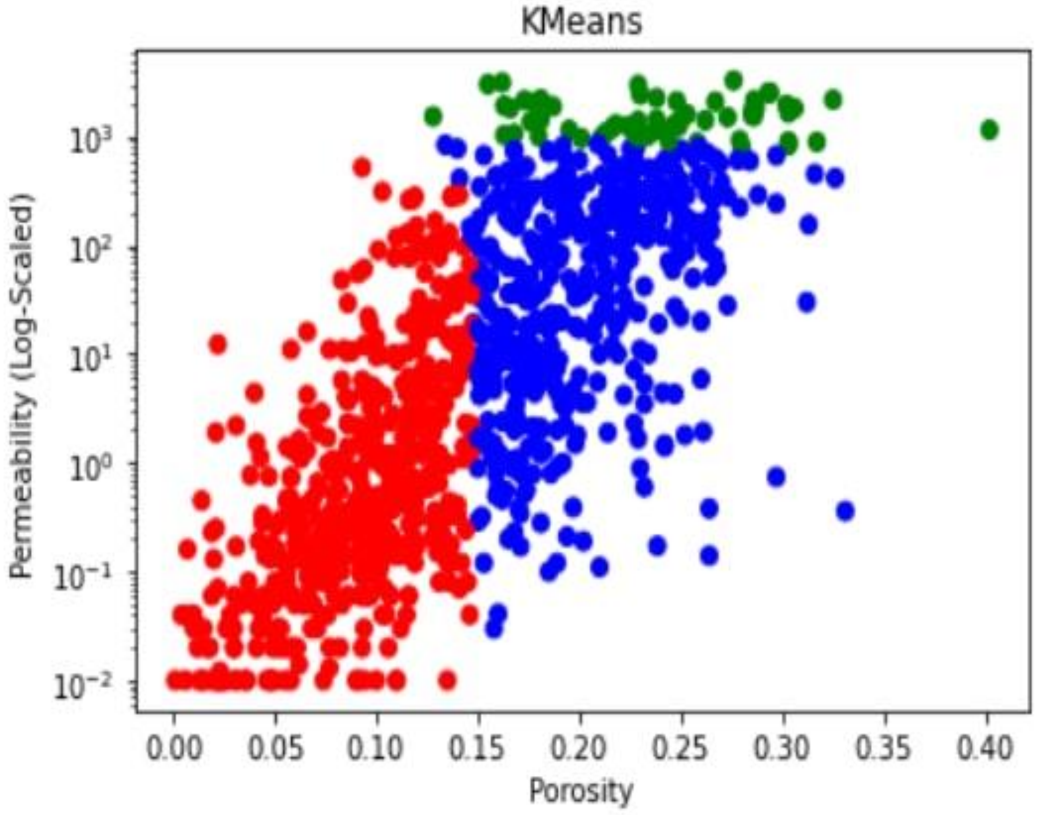




KMeans

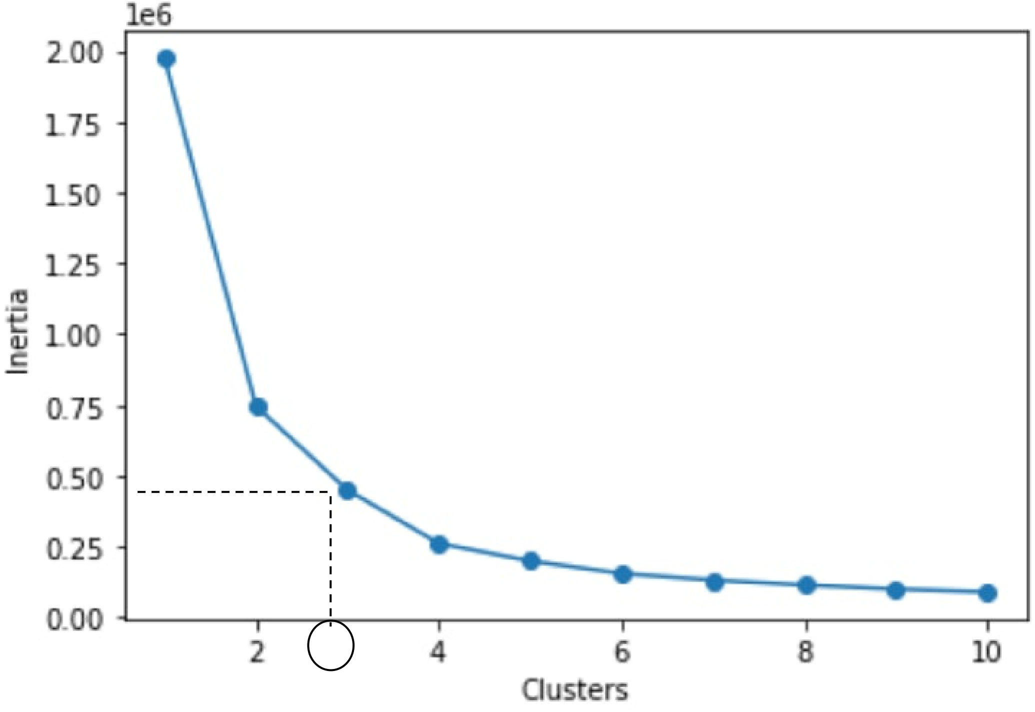
- Used Elbow plot to determine optimal clusters
- Range trained was from 1 to 11
- Based on elbow plot, optimal cluster of 6 is chosen

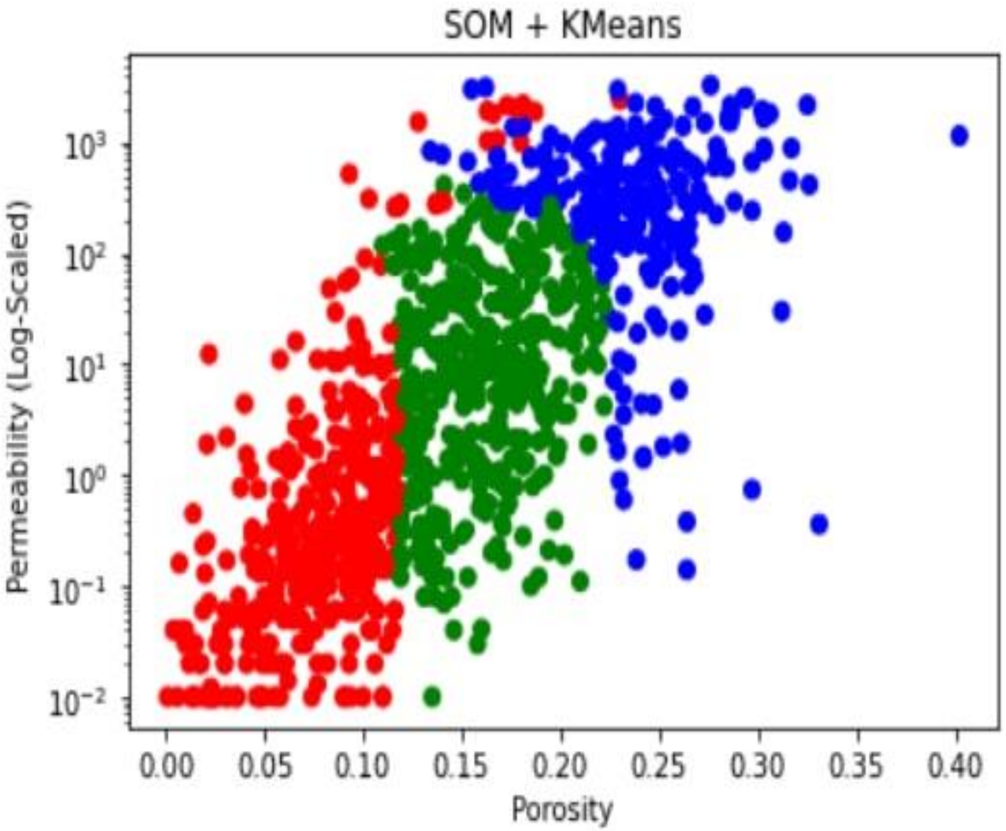




Self Organizing Map (SOM) + KMeans

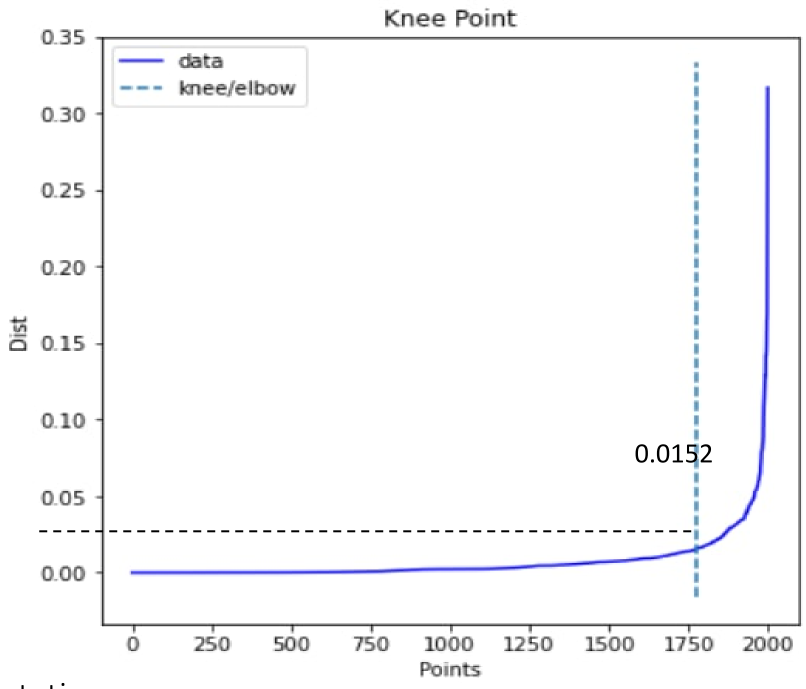
- Trained UK data using SOM with 10 X 10 grid
- Output is 100 clusters
- Used KMeans to combine the clusters
- Optimal clusters chosen using elbow plot
- Optimal clusters of 3 was chosen

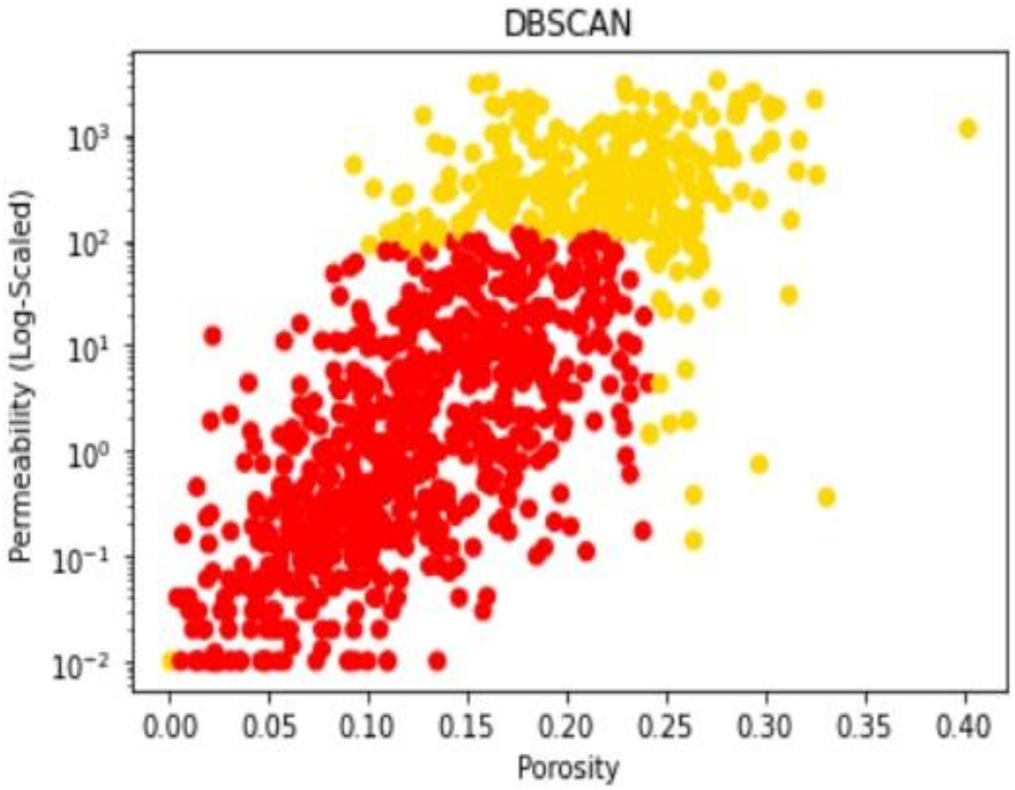




DBSCAN

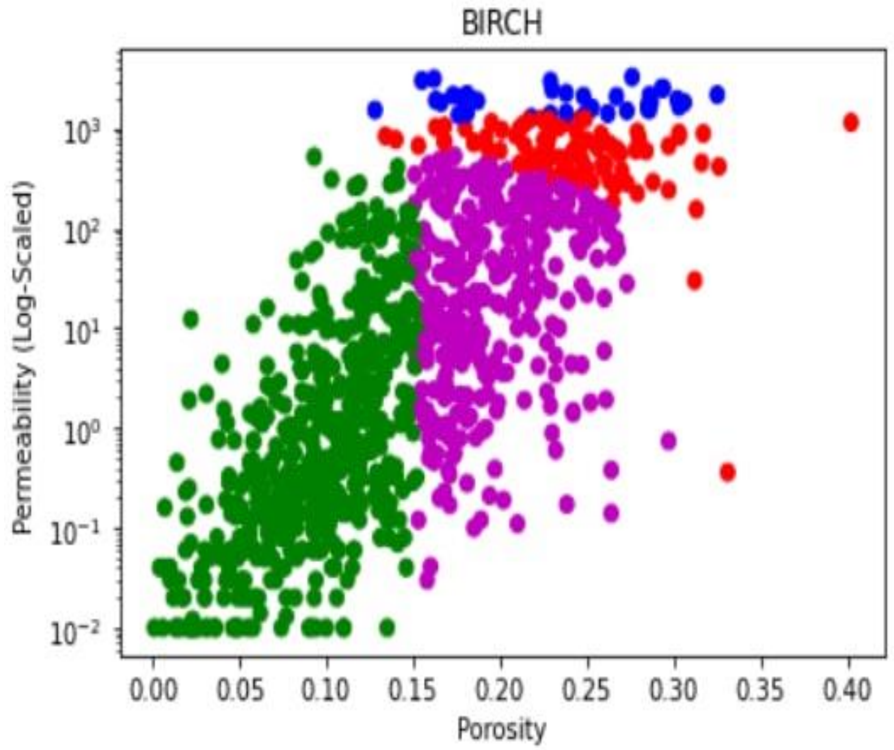
- Find Optimal eps (epsilon) using avg distance of k-nearest neighbour
- K set to 4 at random
- Maximum curvature used as epsilon
- Dbscan initialised and fitted
- min_samples set at 25





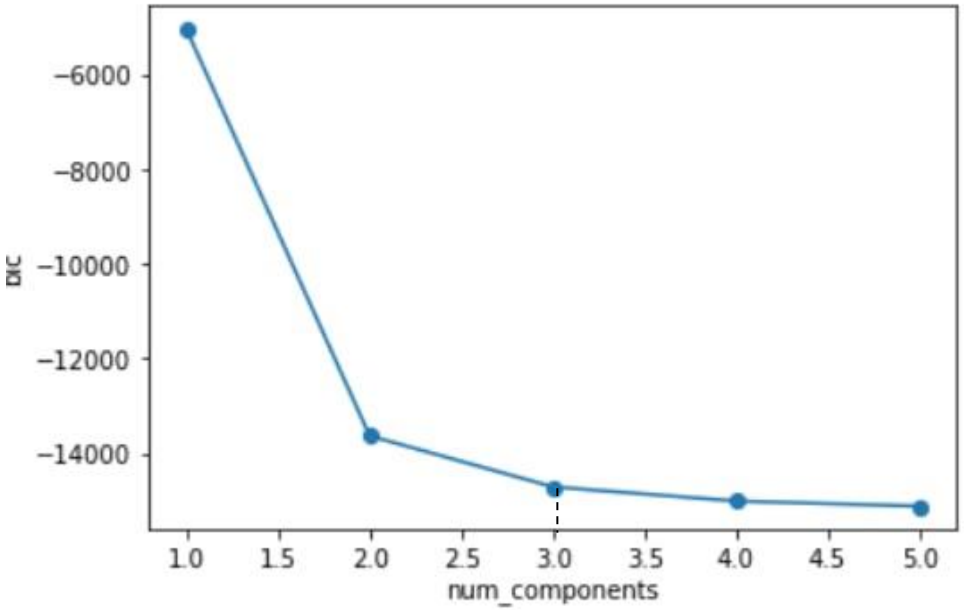
BIRCH

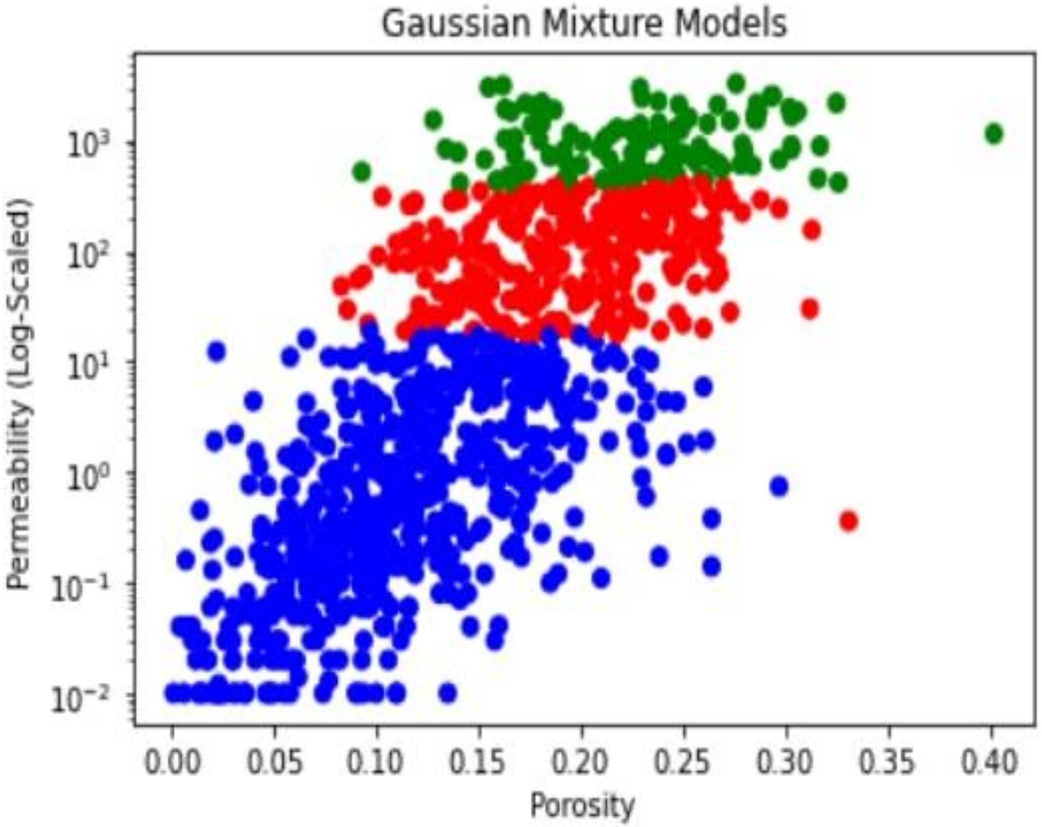
- Initialise BIRCH
- Set n_cluster as 4
- Trial for values of branching_factor and threshold]
- Fit predict data
- Use metrics to evaluate the braching_factor and threshold



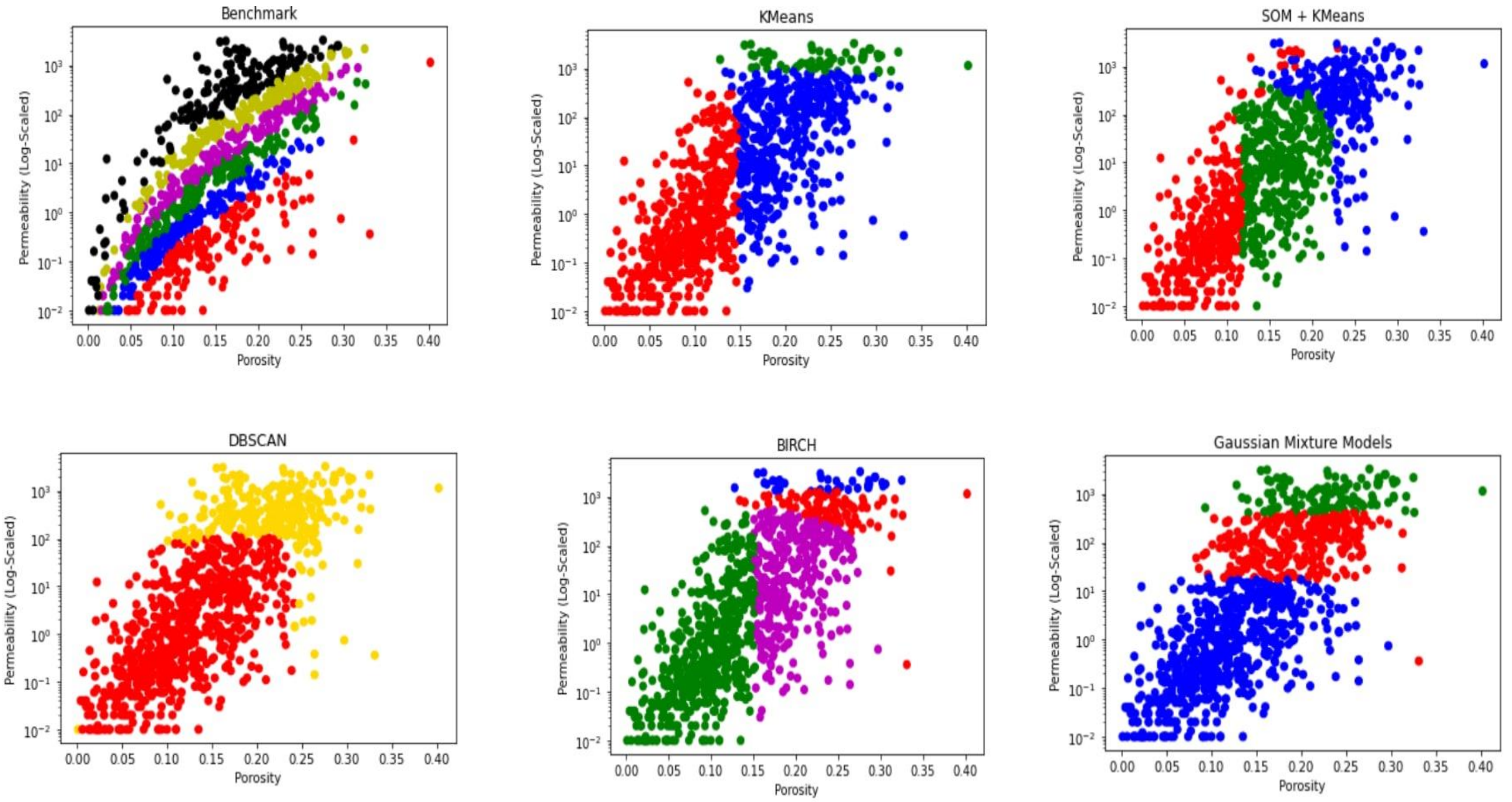
GMM

- Use BIC to find optimal num_components
- Train GMM for num_components 1-6
- Get BIC
- Plot BIC Chart
- Train using optimal n_components = 3





Comparison of the Cluster Distribution



Comparison of the Cluster Porosity and Permeability Averages

Benchmark

Cluster	Porosity	Permeability
1	0.120	1.888
2	0.138	12.175
3	0.140	16.426
4	0.154	591.318
5	0.157	68.599
6	0.181	278.419

KMeans

Cluster	Porosity	Permeability
1	0.093	15.225
2	0.200	152.504
3	0.243	1596.954

SOM+KMeans

Cluster	Porosity	Permeability
1	0.079	50.100
2	0.161	34.684
3	0.234	550.760

DBSCAN

Cluster	Porosity	Permeability
1	0.122	10.720
2	0.215	549.689

BIRCH

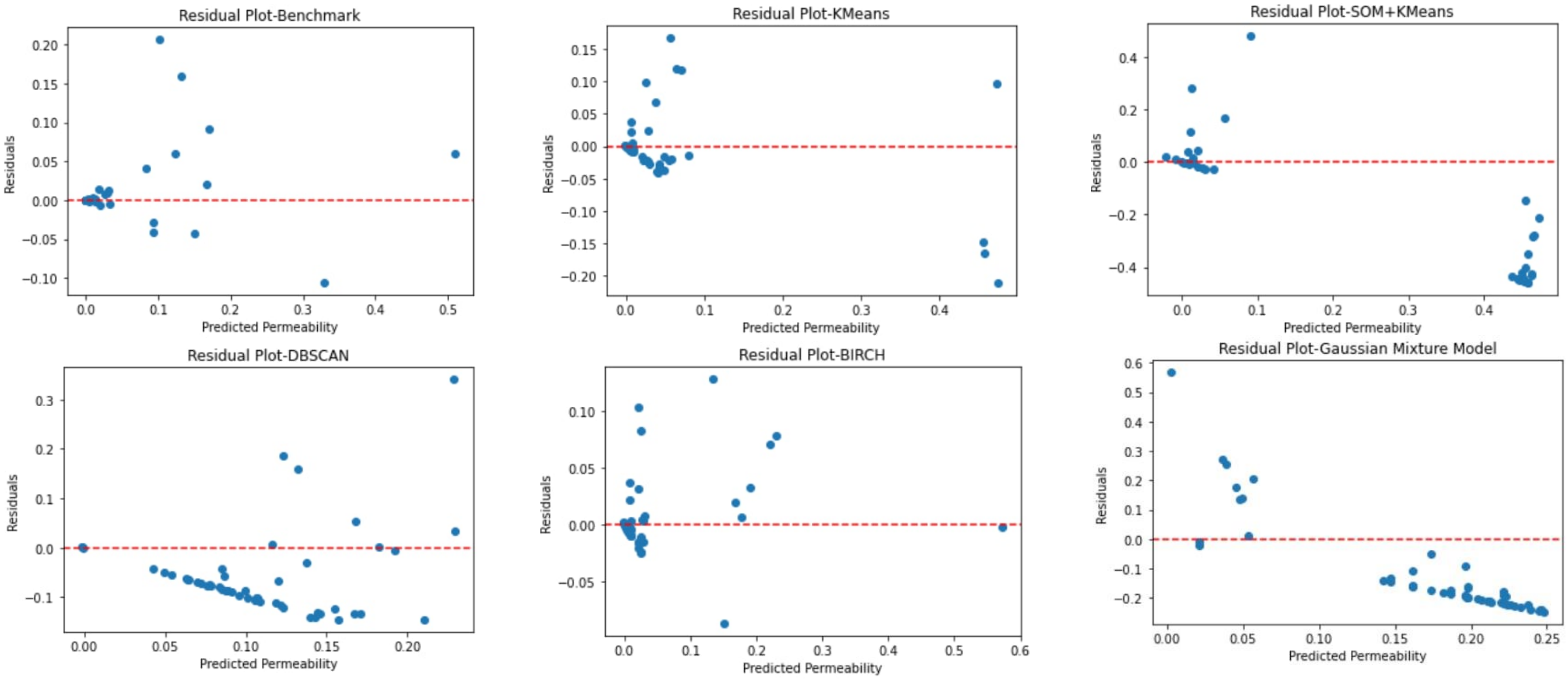
Cluster	Porosity	Permeability
1	0.095	16.812
2	0.196	87.472
3	0.238	620.094
4	0.244	1928.485

GMM

Cluster	Porosity	Permeability
1	0.113	2.424
2	0.192	138.307
3	0.227	1087.203

Comparison of the Residual Charts from Permeability Prediction

SUSS
SINGAPORE UNIVERSITY OF SOCIAL SCIENCES



Comparison Between Clustering Algorithms Using Metrics

- Metrics use: Silhouette Coefficient, Calinski-Harabaz Index, Davies-Bouldin Index
- The metrics evaluates uniqueness of clusters
- For Silhouette Coefficient and Calinski-Harabaz Index, higher values are better
- For Davies-Bouldin Index, the lower the value the better
- KMeans and BIRCH has a better performance than SOM+Kmeans, DBSCAN and GMM

Algorithm	Silhouette Coefficient	Calinski-Harabaz Index	Davis-Bouldin Index
KMeans	0.529	2556.740	0.642
SOM+KMeans	0.412	1505.616	0.859
DBSCAN	0.404	1180.614	1.041
BIRCH	0.474	2170.326	0.839
GMMs	0.263	1090.622	1.032

Future Works and Conclusion

- There are further areas that can definitely be included given time
  - Application of more algorithms
  - More sets of data to account for variability
  - Benchmarking against one or more industry standards
  - Using different methods for benchmarking
- Comparing to benchmark
  - Clustering algorithms are not similar to benchmark
- Comparing between clustering algorithms
  - KMeans and BIRCH has a better performance than SOM+Kmeans, DBSCAN and GMM
- Kmeans and BIRCH performed as well as the benchmark
- More research can be done for a better comparison

Thank You

References

Department of Mines, Industry Regulation and Safety. (n.d.). *Introduction to unconventional resources.* https://www.dmp.wa.gov.au/Petroleum/Introduction-to-unconventional-25621.aspx

Khalid, M. Saad, E. D., Desouky, S., Rashed, M., Shazly, T., & Sediek, K. (2019). Application of hydraulic flow units' approach for improving reservoir characterization and predicting permeability. *Journal of Petroleum Exploration and Production Technology, 10*(2). https://doi.org/10.1007/s13202-019-00758-7