Project 2

03 January 2025 21:02

Revise (P1) ✓

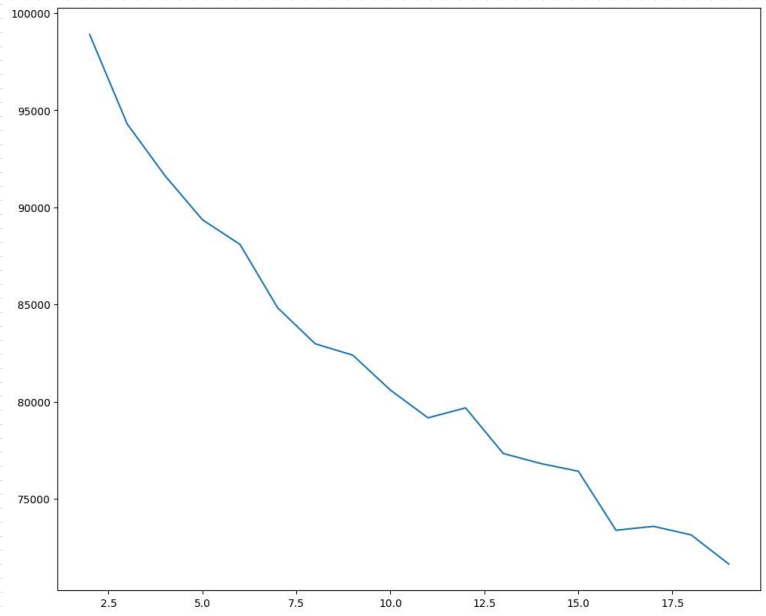
Agenda

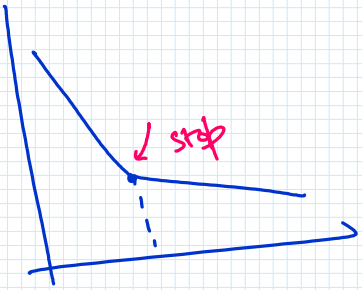
Apply Clustering on my data

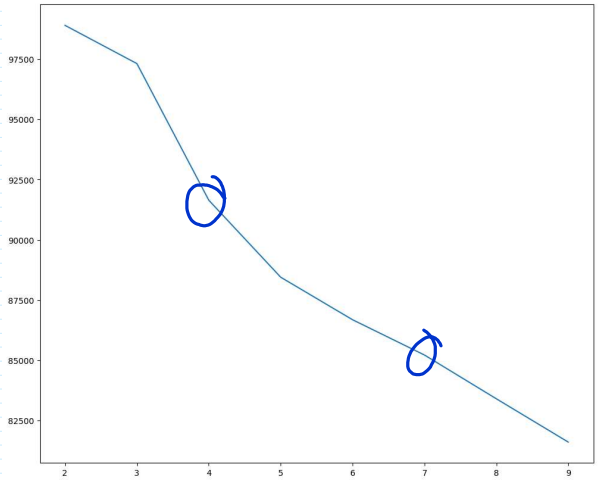
EDA clusters

Technique

Further Improvements







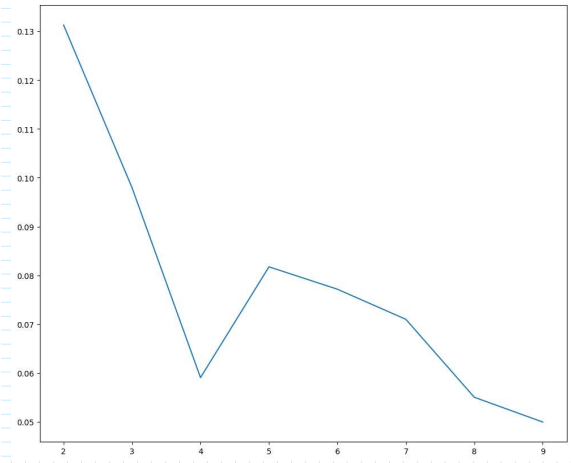
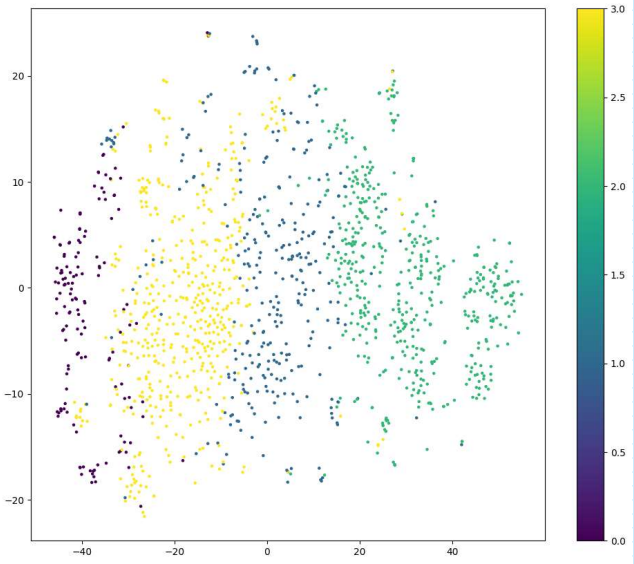
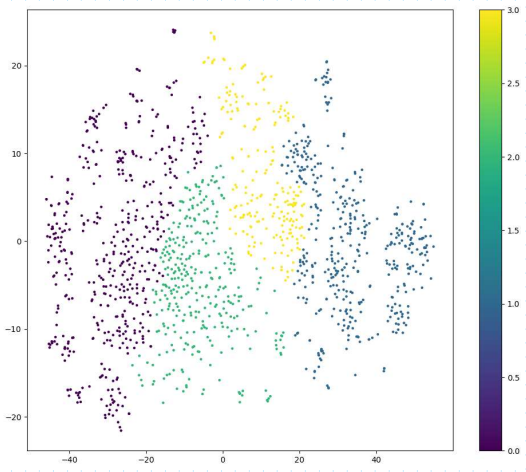
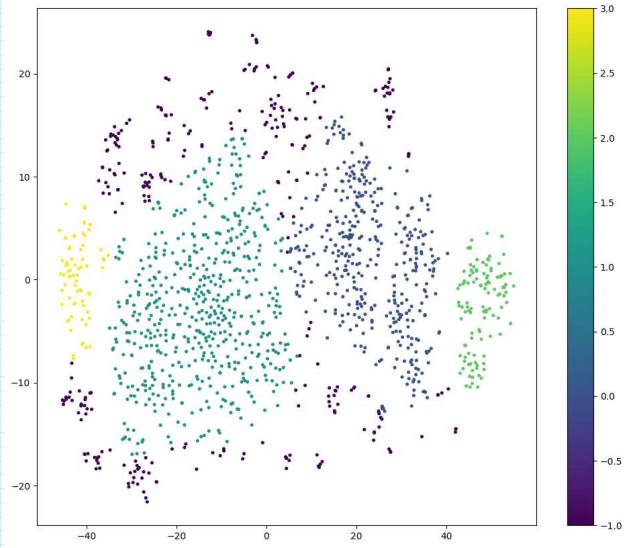


Table · 4 clusters

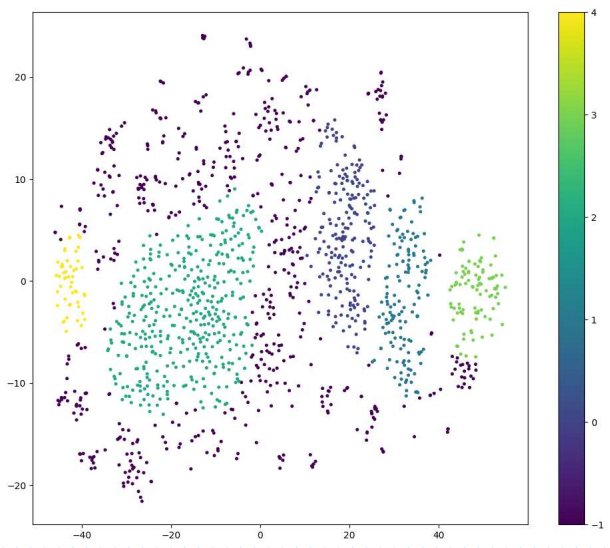




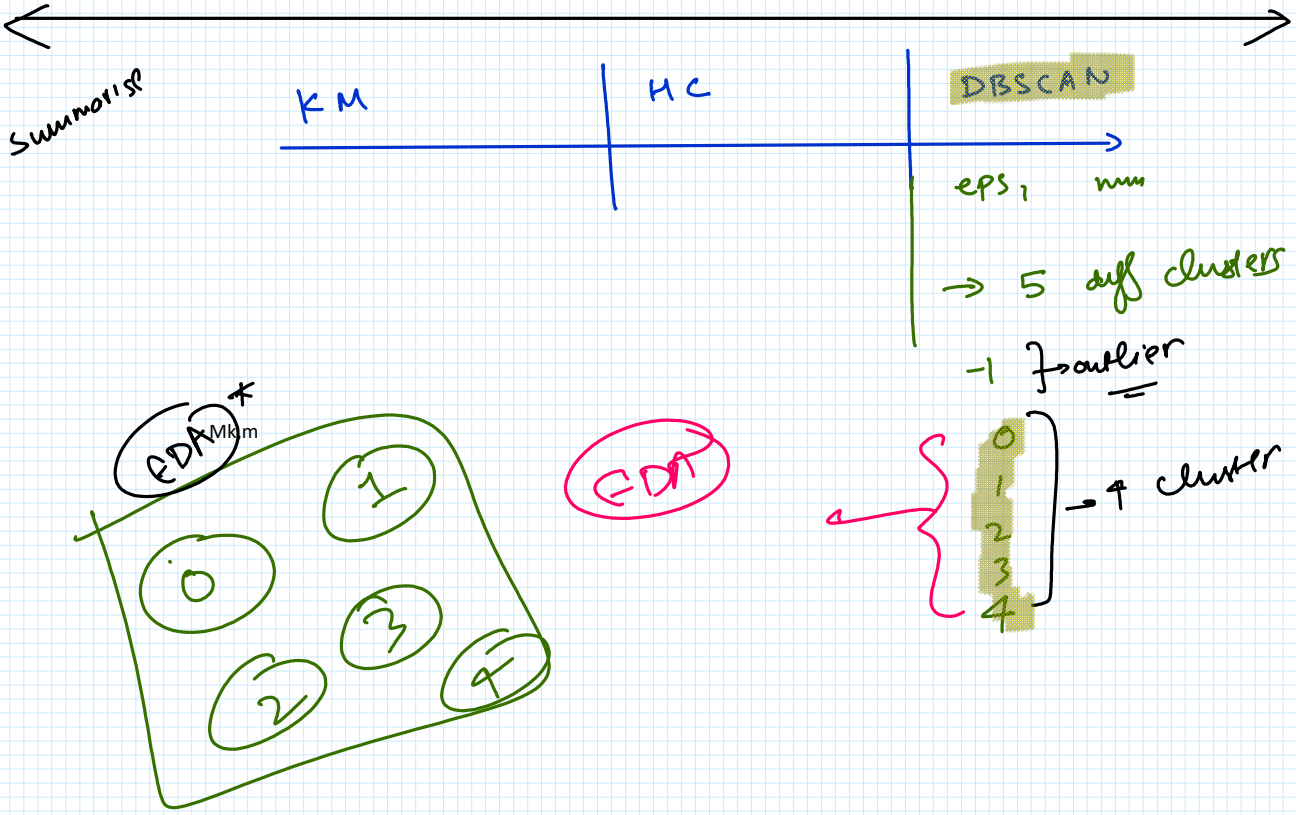


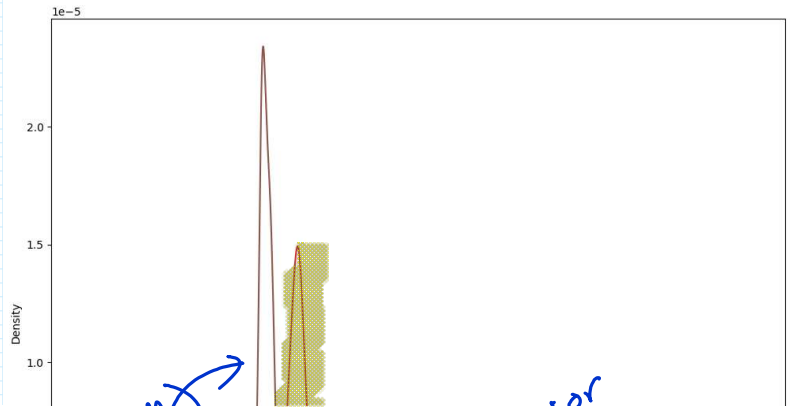
DBSCAN

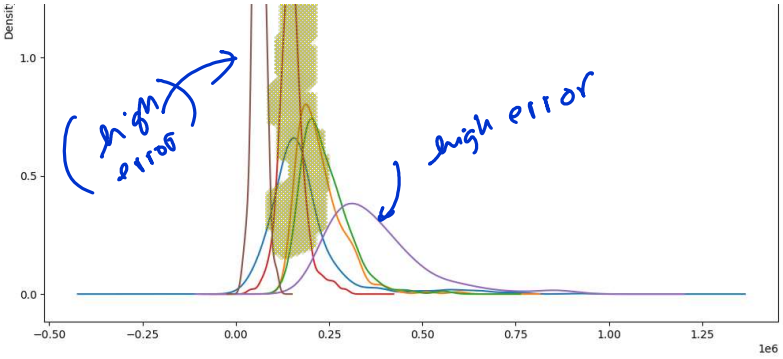
(very how to Tune)


20
10
0
−10
−20
−40
−20
0
20
40
4
3
2
1
0
−1

,


Summarise
KM
HC
DBSCAN
eps, num
→ 5 diff clusters
-1 → outlier
→ 4 cluster
EDA
EDA
0
1
2
3
4


1e−5
2.0
1.5
1.0
Density



original df

All the rows

Model (10.18% error)

used DBSCAN to remove outliers

outliers

(16% error)

8.9%

X

hell

y

X

X

found outliers using DBSCAN

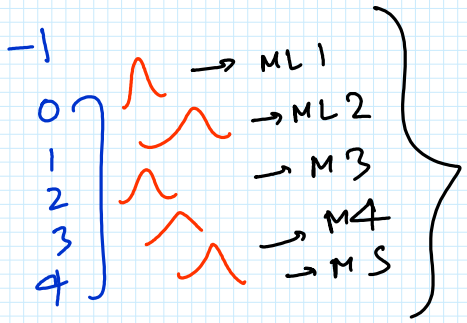
=> Removed

22:12

10:12 pm

Baseline Models

Cluster_wise Baseline Models



Go home and fit each model on each cluster

5 models

```
0 mse : 22467.674531822915 mape: 0.06589067784448051
1 mse : 24574.950851701356 mape: 0.06338192633284288
2 mse : 16308.418622526266 mape: 0.08394762397351883
3 mse : 61321.06152425323 mape: 0.10326860518877166
4 mse : 13033.677322225501 mape: 0.1735627516723023
64576340.91122606 0.08373568049975669
```

3 and 4

Total MSE

Avg MAPE

ARI / NMI (Not used)

Training individual models for each cluster

Architecture

↓

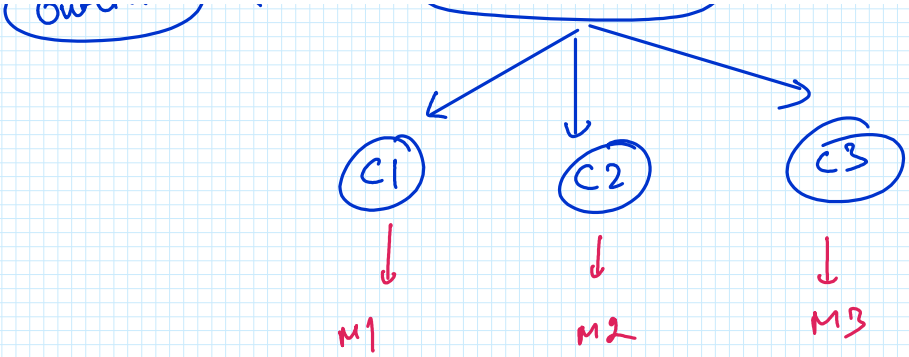
Dimens Redn

↓

Clustering + outlier detection

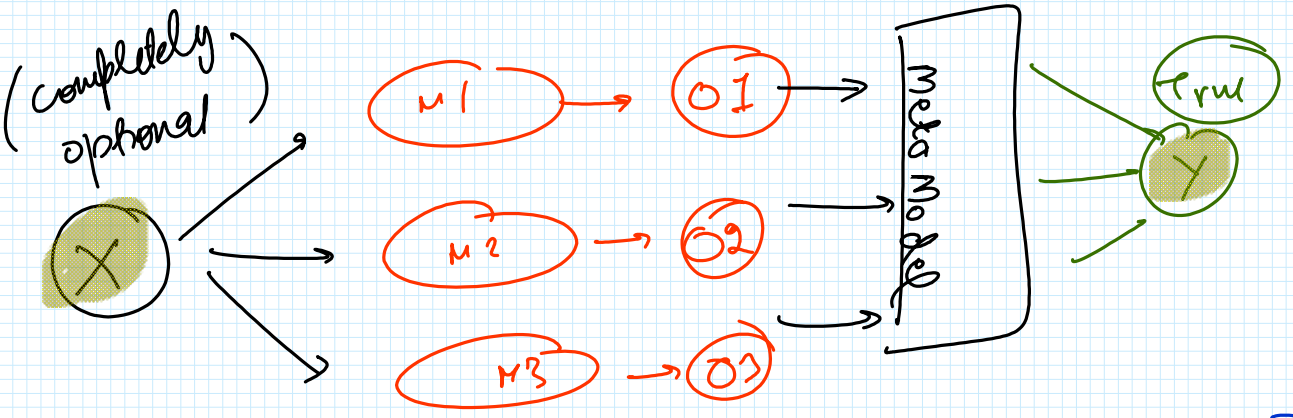
↓

Outlier (handle separately)



Class Concepts
OOP
→ Implement of ensemble Algo

Meta Models (ensemble)



estimators = [(M1) , (M2) , (M3)]

= [('rf', RandomFor) , ('gbt', Gradu)]

Doubts *=

"Development"

4 hours → Project

ML-2