HOMEWORK

DS – 4 #W14

1. Mencari Notebook di Kaggle dan memilih salah satu model algoritma menggunakan DBSCAN dan Agglomerative Clustering.

Jawaban : <https://www.kaggle.com/fazilbtopal/popular-unsupervised-clustering-algorithms>

1. Menganalisa notebook menggunakan Data Science Methodology.

Jawaban :

* Business Understanding : Mall Customer Segmentation (me cluster pengunjung mall berdasarkan Pendapatan dan waktu yang di habiskan di dalam mall. )
* Analytic Approach : kita dapat menggunakan Prescriptive analytics.
* Data Requirements : kita akan menggunakan semua feature sebagai analysis kita dan kita akan mecluster feature Pendapatan dan Spending time.
* Data Collecting : menggunakan data dari Kaggle.
* Data Understanding : di tahap ini kita memahami setiap feature pada tabel yang dapat di gunakan pada modelling nantinya.
* Data Preparation : di tahap ini adalah proses pembersihan data dan membuang feature yang tidak digunakan.
* Data Modelling : kita masuk ke tahap modelling yang di mana kita akan menggunakan algoritma Unsupervised Learning di antaranya agglomerative dan DBSCAN.
* Evaluation : setelah proses modelling kita masuk ke tahap evaluation yang dimana kita membandingkan beberapa modelling yang lebih akurat.
* Deployment : bila sudah sesuai modelling mana yang akan di pakai, selanjutnya kita masuk ke tahap deployment yang dimana proses ini akan digunakan oleh stakeholder dan customer.
* Feedback : mengevaluasi permasalahan-permasalahan baru dari modelling yang sudah kita buat sebelumnya dan memperbaruinya.

1. Model algoritma mana yang cocok dengan Notebook tersebut.

Jawaban :

Pada notebook yang kita pakai saat ini dan menurut saya algoritma Unsupervised Learning yang cocok digunakan pada data Mall Customer Segmentation adalah DBSCAN walaupun dataset yang kita gunakan tergolong masih sedikit tapi insight yang kita dapatkan tergolong lebih dari cukup.

1. Hal yang dapat di tingkatkan dari Dataset tersebut.

Jawaban :

Dari proses clustering yang sudah kita lakukan, kita juga dapat menggunakan algoritma Supervised Learning setelah penggunaan DBSCAN yang dimana bertujuan untuk memprediksi customer yang datang ke mall.