**“Perbandingan penggunaan K Means dan DBSCAN pada**

**Clustering Mall Customer Data”**

Christopher Darren - 00000054804

Program Studi Sistem Informasi, Fakultas Teknik dan Informatika,

Universitas Multimedia Nusantara

Jl.Scientia Boulevard,Gading Serpong, Tangerang, Banten – 15811 Indonesia

e-mail: [christopher.darren@student.umn.ac.id](mailto:christopher.darren@student.umn.ac.id)

Abstrak

Zaman sekarang teknologi berkembang cepat dan banyak kegiatan bisnis dilakukan dengan menggunakan teknologi. Contohnya saja seperti merekap data customer yang berada di mall. Salah satu cara yang bisa menelaah data lebih dalam adalah dengan melakukan segmentasi data.bagaimana caranya? Segmentasi dapat dilakukan dengan cara clustering. Clustering bisa dilakukan banyak cara yaitu dengan cara Dbscan maupun K-Means. Dengan adanya segmentasi maka pemilik mall bisa membuat strategi baru dalam meningkatkan efisiennya dalam proses kegiatan yang ada di dalam mall serta merancang sesuatu yang baru sehingga bisa meningkatkan aktivititas pengunjung di mall.

**Kata kunci :Customer, Mall K-Means, DBScan, Clustering**

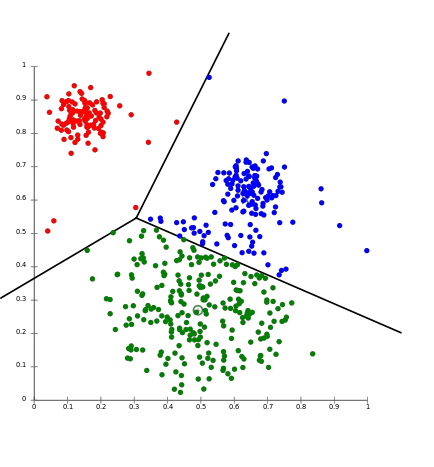
1. **Pendahuluan**

Kmeans adalah salah satu algoritma pembelajaran tanpa pengawasan. K-Means memiliki fungsi untuk mengelompokkan data ke dalam kelompok data. Cluster sampling adalah teknik pengambilan sampel yang secara acak memilih unit populasi dari kelompok yang ada. Kemudian definisi dari DBscan adalah teknik pengolahan data yang menggunakan clustering berbasis kerapatan atau density. Metode clustering adalah suatu proses yang bertujuan untuk membagi data menjadi kelompok-kelompok tertentu berdasarkan kategori atau tingkat kesamaan data. DBSCAN juga merupakan teknik ML yang menggunakan unsupervised machine learning. Disini penulis akan membandingkan cara kerja DBscan dengan kmeans dalam dataset mall customer beserta dengan visualisasinya.

1. **Metode Algoritma**

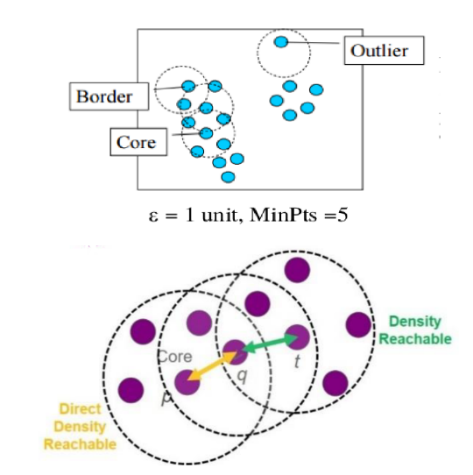
**\*KMeans**

Metode analisis data atau metode penambangan data yang melakukan proses pemodelan pembelajaran tanpa pengawasan dan menggunakan metode mengelompokkan data ke dalam bagian-bagian terpisah.



**Gambar 1. Contoh Kmeans**

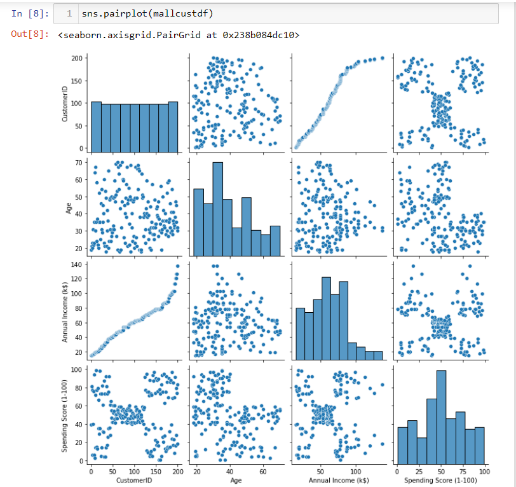
**\*DBscan**

Metode DBscan ini menggunakan algoritma berbasis kepadatan. Dan ada beberapa parameter yang harus diperhatikan yaitu epsilon (eps) dan minimum points (minPts). Epsilon adalah jarak terbesar antara dua data dalam sebuah cluster. Sedangkan titik minimum adalah jumlah data terkecil dalam jarak epsilon untuk membentuk cluster.  


**Gambar 2. Dbscan.**

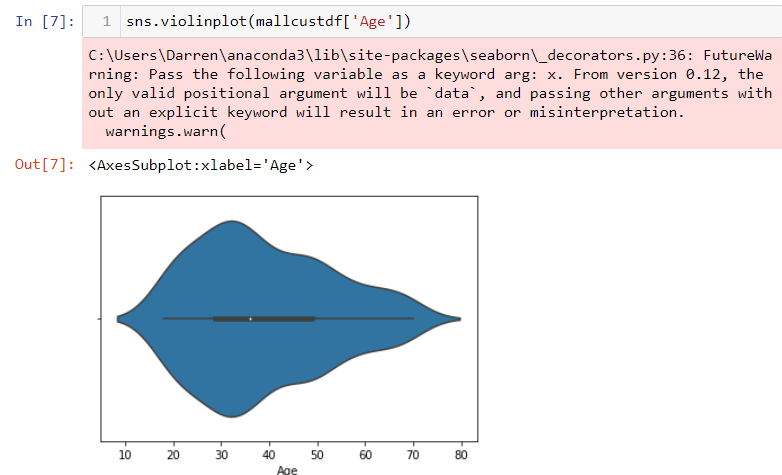
1. **Hasil & Pembahasan**

Karena importing datanya terlalu kecil gambarnya maka penulis langsung skip menuju hasil EDA beserta pembahasannya untuk cara kerja Kmeans.

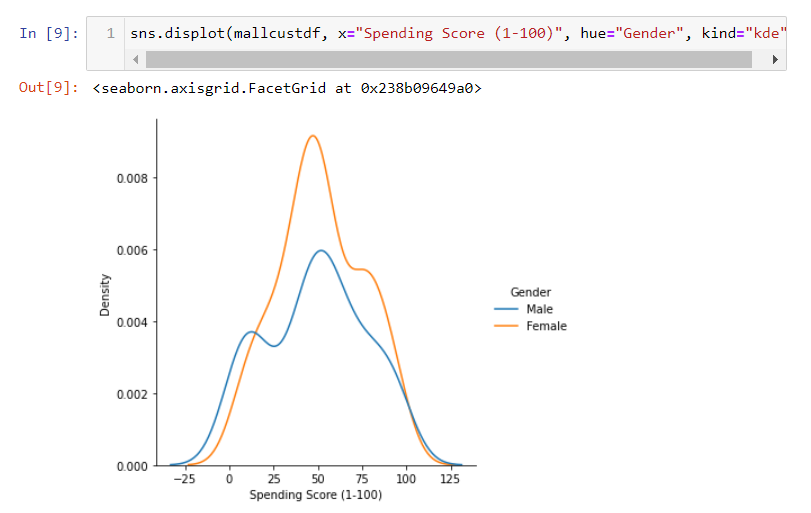


**Gambar 3. Sns pairplot**

Gambar diatas merupakan visualisasi secara keseluruhan yang bisa dilakukan oleh sns seaborn pada jupyter notebook.

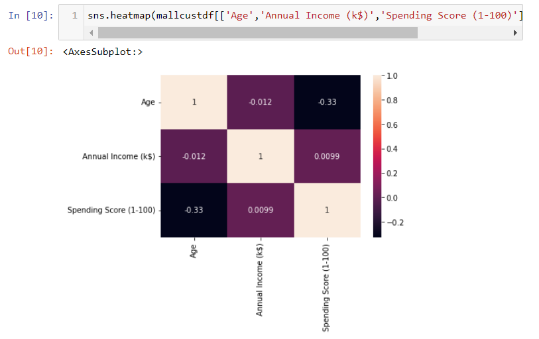


**Gambar 4. Sns violin plot untuk age.**

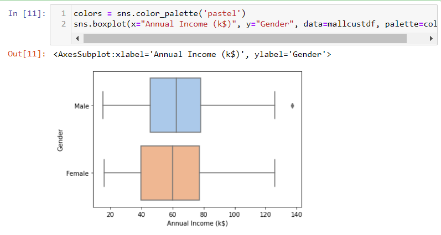
****

**Gambar 5. Distribusi antara spending score dengan Age**

Untuk distribusi diatas tidak gaussian karena lengkungan antara ekor kanan dan kiri tidak simetris.

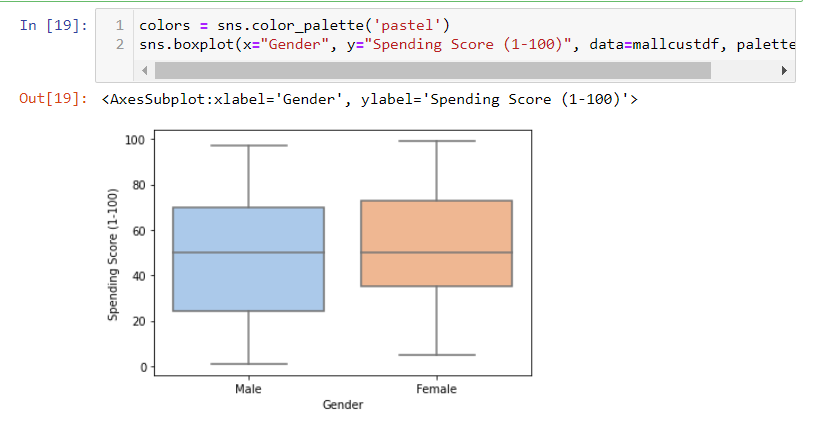
****

**Gambar 6.heatmap korelasi antara age, annual income dan spending score.**



**Gambar 7. Boxplot antara annual income dengan gender.**

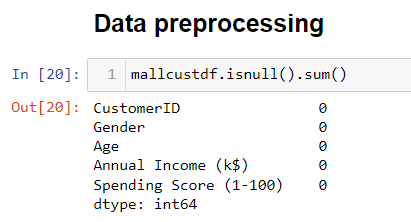
Pada gambar diatas terdapat 1 outlier pada boxplot untuk Male, sedangkan female tidak ada.



**Gambar 8. Boxplot antara spending score dengan gender**

Terlihat bahwa wanita lebih banyak untuk spending dibandingkan dengan pria.

**Proses K-Means**

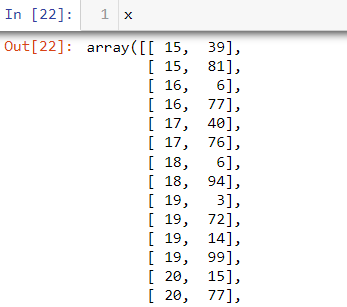
****

**Gambar 9. Data preprocessing**

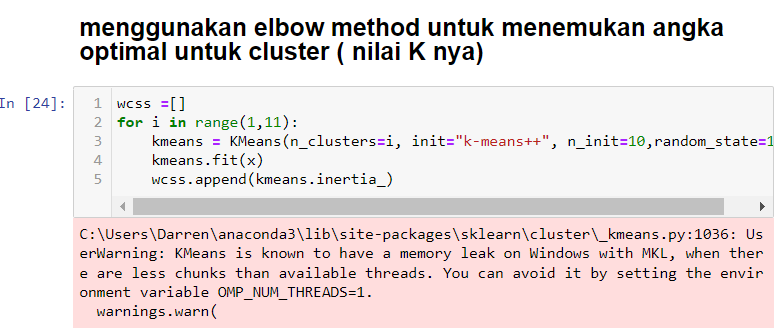
Pertama tama harus dilakukan langkah diatas untuk mengecek apakah ada angka yang non null values atau tidak.

****

**Gambar 10. Membuat train model**

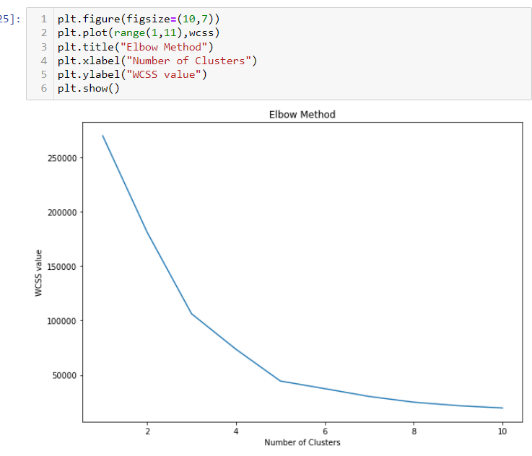


**Gambar 11. Beberapa hasil array dari training model.**

****

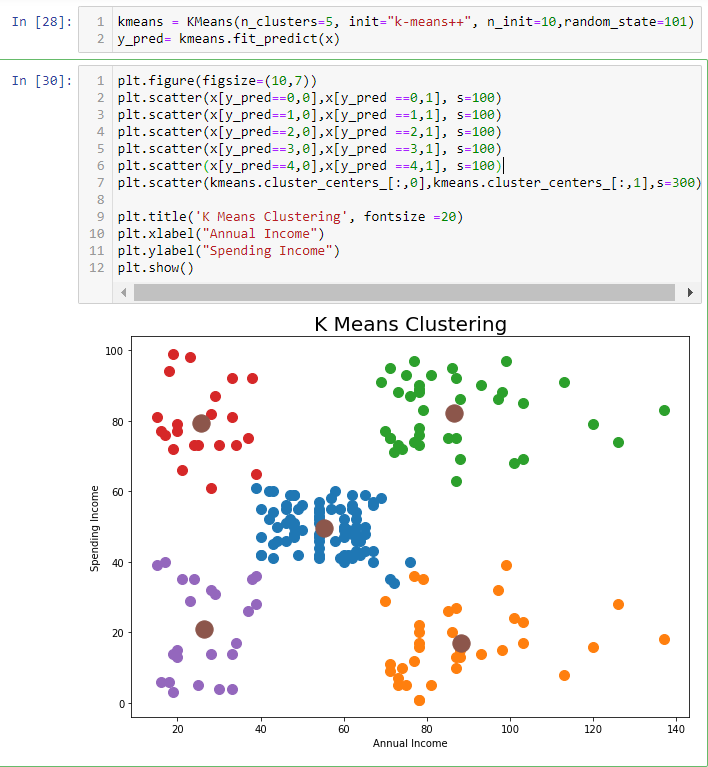
**Gambar 12. Elbow method**

Elbow method menggunakan WCSS untuk menggambarkan plot, untuk memploting values pada wcss di y axis , number cluster di x axis.



**Gambar 13. Visualisasi elbow method**

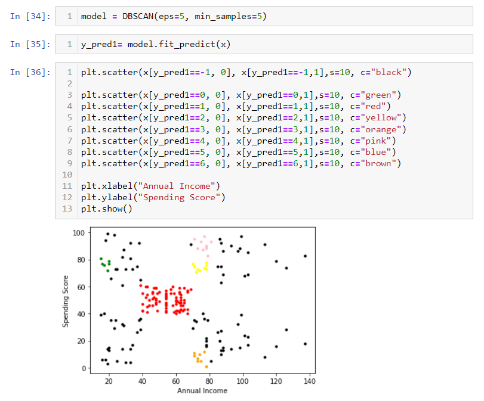
Outputnya elbow pointya di K = 5, kita ambil K nya 5.



**Gambar 14. Visualisasi K-Means**

Karena sesuai kodingan maka clusternya sama dengan 5.

Lalu apa perbedaan antara Kmeans dengan DBscan.. menurut saya sendiri lebih kearah visualisasi beserta cara clusteringnya.Kalau ***untuk Kmeans lebih kearah clustering cluster berdasarkakn grouping(data yang bulet bulet), sedangkan Dbscan lebih kearah cluster mana saja yang tingkat kerapatannya(cluster) tinggi maupun rendah***. Contoh untuk perbandingan DBSCAN.



**Gambar 15. DBSCAN.**

Dbscan lebih jelas dalam pembagian cluster, dan untuk setiap kerapatan akan dilabeli dan diwarnai sehingga data analyst tahu akan kerapatan sebuah data yang ada pada dataset.

1. **Kesimpulan dan saran**

Dari semua proses yang sudah penulis lakukan berakhir dengan 2 kesimpulan yakni:

1. Kmeans

Kmeans dalam penerapannya adalah si Kmeans ini menentukan jumlah cluster yang ada pada data cluster. Kmeans mempunyai tujuan untuk meminimalkan variasi data yang ada pada data cluster. Dan Kmeans penerapannya lebih mudah ketimbang dbscsan karena clustering penerapannya lebih diterapkan pada metode K-Means dan bukan berarti DBscan tidak bisa diterapkan, Dbscan bisa diterapkan tapi rada sulit namun untuk hasilnya lebih bagus untuk DBSCAN(contoh visualisasi DB scan pada gambar 15).

1. DBScan

Dbscan dalam penerapan algoritma clustering dilihat berdasarkan tingkat kerapatan data(density) sehingga bisa menampilkan kepadatan data yan tinggi maupun tingkat kerapatan data yang rendah.

Penulis berharap dengan adanya penelitian ini bisa membuka wawasan dan ilmu pengetahuan baru bagi para pembaca dalam mempelajari ilmu data , dan bisa saja sebagai calon data analyst kedepannya dengan cara memahami apa itu algoritma Dbscan dan Kmeans dalam clustering sebuah segmentasi sebuah data pada dataset.

1. **Daftar Pustaka**

Nursyafitri. (2022, August 12). *K-Means Clustering, Salah Satu Contoh Teknik Analisis Data Populer*. dqlab.id. Retrieved December 20, 2022, from <https://www.dqlab.id/k-means-clustering-salah-satu-contoh-teknik-analisis-data-populer>

Agusta. (n.d.). *A Weblog A web log For Data Mining and Clustering As Well As Other General Ideas and Interests*. wordpress.com. Retrieved December 20, 2022, from <https://yudiagusta.wordpress.com/k-means/https:/geospasialis.com/k-means-clustering/>

Firdaus. (2020, March 1). *Kmeans Clustering dan Implementasinya*. medium.com. Retrieved December 20, 2022, from <https://medium.com/@afrizalfir/kmeans-clustering-dan-implementasinya-5e967dc604cf>

KETUTRARE. (n.d.). *lgoritma K-Means Clustering dan Contoh Soal*. https://www.ketutrare.com/. Retrieved December 20, 2022, from <https://www.ketutrare.com/2018/11/algoritma-k-means-clustering-dan-contoh.html>

*Clustering Algoritma (K-Means)*. (2022, January 31). sis.binus.ac.id. Retrieved December 20, 2022, from <https://sis.binus.ac.id/2022/01/31/clustering-algoritma-k-means/>

algoritma. (2022, March 22). *Cara Kerja K Means Clustering*. algorit.ma. Retrieved December 20, 2022, from <https://algorit.ma/blog/k-means-clustering-2022/>

Education Ecosystem (LEDU). (2018, September 13). *Understanding K-means Clustering in Machine Learning*. medium.com. Retrieved December 20, 2022, from <https://towardsdatascience.com/understanding-k-means-clustering-in-machine-learning-6a6e67336aa1>