

Analisis Performa Algoritma K-Means dan DBSCAN Dalam Segmentasi Pelanggan Dengan Pendekatan Model RFM

Furqan Maulana Pranata¹, Satrio Hadi Wijoyo², Nanang Yudi Setiawan³

Program Studi Teknologi Informasi Fakultas Ilmu Komputer Universitas Brawijaya
Email: 1furqanmaulana26@student.ub.ac.id, 2satriowijoyo@ub.ac.id, 3nanang@ub.ac.id

Abstrak

Dalam era globalisasi yang semakin kompetitif, intensifikasi persaingan di dunia bisnis memerlukan strategi yang efektif untuk meraih keunggulan kompetitif. Segmentasi pelanggan merupakan langkah penting untuk memahami pasar dan menargetkan kelompok pelanggan yang tepat. Penelitian ini bertujuan untuk menganalisis performa algoritma K-Means dan DBSCAN dalam segmentasi pelanggan dengan pendekatan model RFM (Recency, Frequency, Monetary). Penelitian ini menggunakan lima dataset retail apparel yang berbeda untuk menguji dan membandingkan performa kedua algoritma clustering tersebut. Evaluasi hasil clustering dilakukan menggunakan matriks Silhouette Score dan Davies-Bouldin Index (DBI). Hasil penelitian menunjukkan bahwa kedua algoritma memiliki kelebihan dan kekurangan masing-masing tergantung pada karakteristik dataset. Kesimpulan dari penelitian ini adalah bahwa tidak ada satu algoritma yang secara universal lebih baik dari yang lain, tetapi pemilihan algoritma yang sesuai sangat bergantung pada karakteristik dataset yang digunakan. Hasil clustering paling baik yang dihasilkan K-Means dengan silhouette score 0.547 dan nilai indeks DBI 0.640 yang merupakan hasil dataset pertama. Sedangkan hasil clustering yang baik dihasilkan DBSCAN dengan silhouette score 0.767 dan nilai indeks DBI 0.731 yang merupakan hasil dataset kedua.

Kata kunci: *segmentasi pelanggan, K-Means Clustering, DBSCAN Clustering, model RFM, retail apparel, Silhouette Score, Davies-Bouldin Index.*

Abstract

In an increasingly competitive era of globalization, the intensification of business competition requires effective strategies to gain a competitive edge. Customer segmentation is a crucial step in understanding the market and targeting the right customer groups. This study aims to analyze the performance of K-Means and DBSCAN algorithms in customer segmentation using the RFM (Recency, Frequency, Monetary) model. This research uses five different retail apparel datasets to test and compare the performance of the two clustering algorithms. The evaluation of clustering results is conducted using Silhouette Score and Davies-Bouldin Index (DBI) metrics. The results show that both algorithms have their strengths and weaknesses depending on the dataset characteristics. The conclusion of this study is that no algorithm is universally superior, but the appropriate algorithm choice heavily depends on the dataset characteristics. The best clustering results were produced by K-Means with a silhouette score of 0.547 and a DBI index value of 0.640 which was the result of the first dataset. Meanwhile, good clustering results were produced by DBSCAN with a silhouette score of 0.767 and a DBI index value of 0.731, which is the result of the second dataset.

Keywords: *customer segmentation, K-Means Clustering, DBSCAN Clustering, RFM model, retail apparel, Silhouette Score, Davies-Bouldin Index.*

1. PENDAHULUAN

Dalam menghadapi persaingan, setiap entitas perusahaan dan aktivitas bisnis harus memiliki strategi yang terencana dengan matang. Menurut Hitt, Ireland dan

Hoskisson (2011) strategi adalah satu set komitmen dan aksi yang terintegrasi dan terkoordinasi yang dirancang untuk digunakan mendapatkan kompetensi inti (*core competencies*) dan keunggulan bersaing (*competitive advantage*).

Untuk menetapkan strategi bersaing

untuk penetapan target pasar, terdapat rekomendasi pendekatan yang paling efektif, yaitu pendekatan yang bersumber dari kebutuhan dan keinginan pelanggan atau biasa disebut pendekatan berorientasi pelanggan. Berorientasi terhadap pelanggan dan pesaing adalah salah satu metode yang dapat digunakan jika suatu perusahaan ingin unggul dalam persaingan (Jusriadi, 2022). Orientasi Pelanggan menurut Avlonitis (2010) adalah sikap konsumen dalam memberikan penjelasan atau pengertian terhadap pelanggan dalam mendahulukan kebutuhan pelanggan, menjadi problem solver bagi pelanggan, hanya merekomendasikan produk yang benar-benar dibutuhkan pelanggan, dan mencari produk yang membantu pelanggan. Dengan memperhatikan keberagaman karakteristik dan kebutuhan pelanggan, serta variasi keinginan yang bersifat individual, penggunaan pendekatan ini dinilai sangat menguntungkan bagi entitas perusahaan.

Dalam menetapkan target pasar berdasarkan kebutuhan dan keinginan konsumen, umumnya dilakukan melalui segmentasi pelanggan. Segmentasi diartikan sebagai kegiatan mempartisi suatu pasar menjadi kelompok pembeli yang mungkin memerlukan produk atau tindakan pemasaran yang berbeda (Lamb, 2006). Dibb (1998) mengemukakan bahwa Segmentasi pelanggan adalah salah satu metode yang efisien untuk mengelola berbagai pelanggan dengan preferensi yang berbeda. Penerapan segmentasi pelanggan dalam strategi bisnis memungkinkan perusahaan untuk menyelaraskan layanan dan produk sesuai dengan kriteria pelanggan, yang pada gilirannya meningkatkan kualitas pemasaran produk.

Segmentasi pelanggan umumnya dapat dilakukan dengan memanfaatkan berbagai data, termasuk di antaranya data transaksi yang tercatat dari aktivitas pelanggan dalam suatu periode waktu tertentu. Implementasi segmentasi pelanggan menggunakan data transaksi di perusahaan dinilai sangat efisien karena berkaitan erat dengan kebiasaan

pelanggan dalam bertransaksi dan sangat relevan dengan penentuan target pasar. Untuk menjalankan proses segmentasi pelanggan dengan optimal, pendekatan model juga harus sejalan. Dengan memanfaatkan data transaksi pelanggan, dapat memanfaatkan untuk menggunakan model pendekatan Recency, Frequency, Monetary (RFM).

Model pendekatan RFM merupakan singkatan dari Recency, Frequency, Monetary, merupakan suatu metode analisis perilaku pelanggan yang mencakup seberapa baru pelanggan melakukan pembelian (Recency), seberapa sering pelanggan melakukan pembelian (Frequency), dan berapa jumlah uang yang dihabiskan pelanggan saat bertransaksi (Monetary) (Berahmana dkk, 2020). Model ini mampu memantau perubahan perilaku pelanggan dari waktu ke waktu dan memungkinkan adaptasi strategi perusahaan sesuai dengan perubahan dalam pola transaksi pelanggan. Penelitian-penelitian sebelumnya yang telah melakukan segmentasi pelanggan menggunakan model RFM banyak memanfaatkan metode clustering untuk mendukung proses segmentasi pelanggan. Metode klusterisasi yang sering diimplementasikan dalam penelitian-penelitian sebelumnya terkait segmentasi pelanggan terdapat K-Means Clustering dan DBSCAN Clustering. K-Means Clustering sering digunakan pada kasus segmentasi pelanggan dikarenakan bentuk metode yang sederhana dan mudah diterapkan pada proses clustering. Di sisi lain juga terdapat metode DBSCAN (Density-Based Spatial Clustering of Applications with Noise) yang sering diterapkan pada studi kasus segmentasi pelanggan karena memiliki kelebihan yang berbeda dalam menghasilkan cluster-cluster yang lebih akurat. Kedua metode ini secara tidak langsung memainkan peran sentral dalam klusterisasi pelanggan berdasarkan karakteristik atau perilaku yang sama dan memberikan wawasan untuk merancang strategi pemasaran yang lebih efektif. Oleh karena itu, kedua metode ini perlu diteliti

lebih lanjut terkait performa kualitasnya ketika diterapkan pada segmentasi pelanggan dengan menggunakan pendekatan model RFM dan pada beberapa dataset retail apparel yang berbeda. Pengujian performa kualitas clustering ini menggunakan dua matriks evaluasi yaitu Silhouette Score dan Davies Bouldin Index. Hal ini bertujuan untuk mengidentifikasi dan mengetahui metode yang paling sesuai berdasarkan karakteristik dataset yang dimiliki untuk menghasilkan kualitas clustering yang paling baik. Dengan demikian, peneliti mengangkat topik "Analisis Performa Algoritma K-Means dan DBSCAN dalam Segmentasi Pelanggan dengan Pendekatan Model RFM" sebagai dasar penelitian. Diharapkan hasil penelitian ini dapat memberikan kontribusi yang berharga bagi industri-industri dan pembaca dalam mengoptimalkan strategi bisnis melalui pemahaman yang lebih mendalam terkait karakteristik dataset dan penerapan metode yang tepat.

2. METODOLOGI PENELITIAN

Penelitian ini menggunakan lima dataset retail apparel yang berbeda untuk menguji dan membandingkan performa kedua algoritma clustering tersebut. Secara garis besar pada penelitian ini menggunakan *Knowledge Discovery in Database* (KDD). Metodologi penelitian meliputi studi literatur, pengumpulan data, preprocessing data, ekstraksi RFM, standarisasi data, dan proses clustering menggunakan algoritma K-Means dan DBSCAN. Evaluasi hasil clustering dilakukan menggunakan matriks Silhouette Score dan Davies-Bouldin Index (DBI).

2.1 Data Selection

Data yang digunakan dalam penelitian ini adalah lima dataset retail apparel yang berbeda, yang diperoleh dari berbagai sumber seperti Kaggle dan lainnya. Data ini kemudian dipilih dan disaring untuk memastikan relevansi dan kualitasnya sebelum digunakan dalam proses clustering.

Data yang digunakan dalam penelitian ini adalah lima dataset retail apparel yang berbeda, yang diperoleh dari berbagai sumber seperti Kaggle dan lainnya. Kelima dataset tersebut adalah:

- a. Different Store Sales - Dataset ini berisi data penjualan atau transaksi pelanggan di retail online dengan total 99.457 data transaksi.
- b. Marketing Insights for E-Commerce Company - Dataset ini berisi data penjualan di retail online dengan total 52.924 data transaksi.
- c. E-Commerce Dataset - Dataset ini berisi data penjualan di e-commerce online di Amerika Serikat dengan total 36.662 data transaksi.
- d. Clothes Retail - Dataset ini berisi data penjualan di retail pakaian dengan total 54.925 data transaksi.
- e. Retail Case Study Data - Dataset ini berisi data penjualan di retail online dengan total 23.053 data transaksi.

2.2 Data Preprocessing

Tahap ini mencakup pembersihan data dan penanganan outlier untuk memastikan kualitas data yang digunakan dalam penelitian. Proses penanganan outlier hanya dilakukan pada data yang dikelompokkan menggunakan K-Means Clustering, sedangkan untuk DBSCAN, outlier dianggap sebagai bagian dari proses clustering.

2.3 Ekstraksi RFM

Dataset yang telah dipilih kemudian dikonversi menjadi model RFM, di mana setiap pelanggan dievaluasi berdasarkan tiga dimensi: Recency, Frequency, dan Monetary. Proses ini dilakukan menggunakan bahasa pemrograman Python dengan library Pandas dan NumPy.

2.4 Standarisasi Data

Data yang telah dikonversi ke model RFM kemudian distandarisasi untuk memastikan bahwa semua data memiliki skala yang serupa. Teknik yang digunakan

adalah Robust Scaler dan Square-root transformation, tergantung pada distribusi data.

2.5 Clustering

Data RFM yang telah distandarisasi kemudian dikelompokkan menggunakan dua metode clustering: K-Means Clustering dan DBSCAN Clustering. Penggunaan dua metode ini bertujuan untuk menguji dan membandingkan performa kinerja kedua metode dalam menghasilkan clustering terbaik.

2.6 Evaluasi Clustering

Evaluasi hasil clustering dilakukan menggunakan matriks Silhouette Score dan Davies-Bouldin Index (DBI). Kedua matriks ini digunakan untuk menentukan metode clustering yang paling baik di antara kedua metode ketika digunakan untuk segmentasi pelanggan menggunakan model RFM pada dataset retail apparel.

3. HASIL DAN PEMBAHASAN

Hasil penelitian menunjukkan bahwa kedua algoritma memiliki kelebihan dan kekurangan masing-masing tergantung pada karakteristik dataset. K-Means Clustering lebih efisien dalam mengelompokkan data dalam jumlah besar namun sensitif terhadap pemilihan centroid awal. DBSCAN Clustering lebih unggul dalam menangani data dengan noise dan dapat membentuk cluster dengan bentuk yang sewenang-wenang namun kesulitan dalam menangani dataset berdimensi tinggi.

3.1 Hasil Clustering

Hasil clustering dari masing-masing dataset menunjukkan variasi dalam efektivitas kedua algoritma. K-Means Clustering cenderung menghasilkan cluster yang lebih kompak namun rentan terhadap outlier, sementara DBSCAN lebih fleksibel dalam menangani bentuk cluster yang tidak beraturan dan noise dalam data.

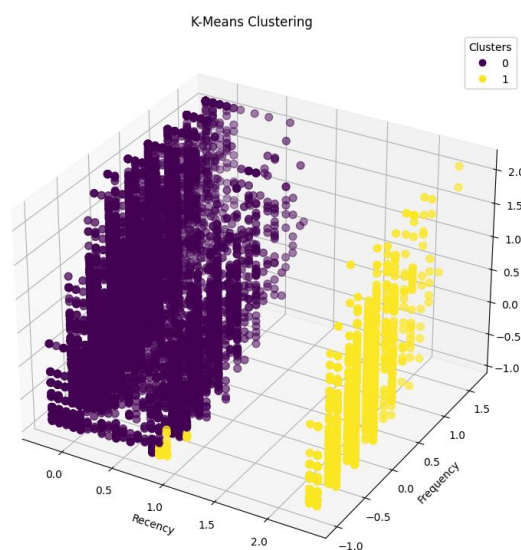
1. K-Means Clustering

Dalam melakukan clustering, langkah awal yang dilakukan penentuan nilai menggunakan elbow method untuk mendapatkan cluster yang optimal. Berikut hasil jumlah cluster yang diperoleh dari elbow method untuk setiap dataset.

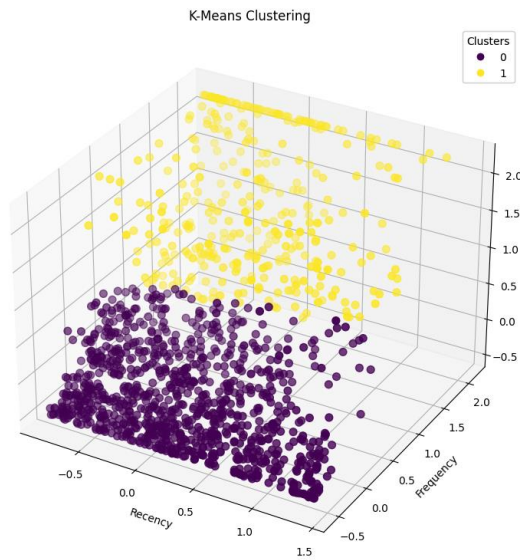
Tabel 1 Jumlah Cluster (K) Optimal K-Means

Dataset 1	2
Dataset 2	2
Dataset 3	3
Dataset 4	3
Dataset 5	2

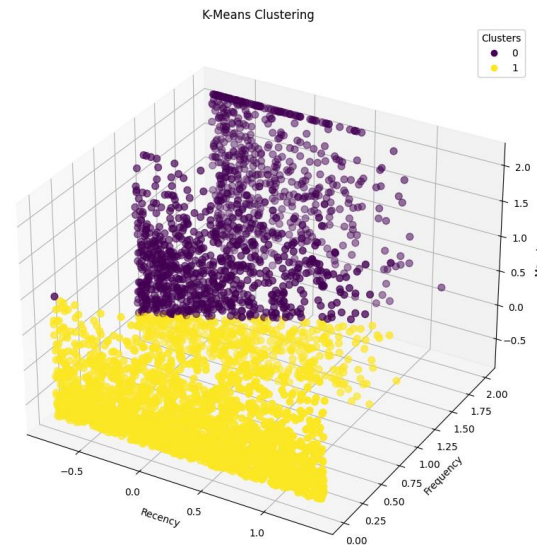
Dari jumlah cluster optimal yang telah didapatkan, dilakukan *clustering* yang menghasilkan visualisasi berbentuk 3d scatterplot untuk setiap dataset.



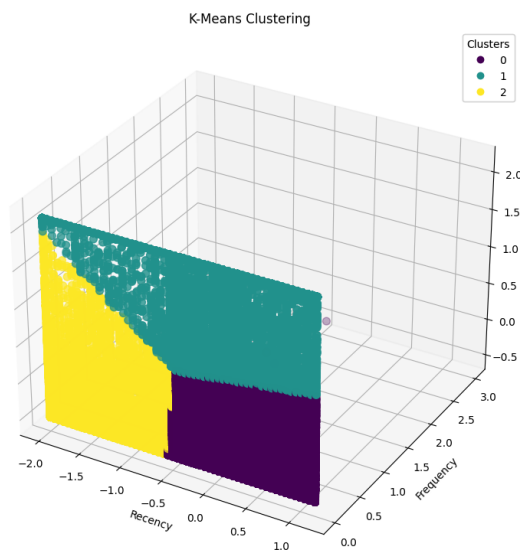
Gambar 1 3D Scatterplot K-Means Dataset 1



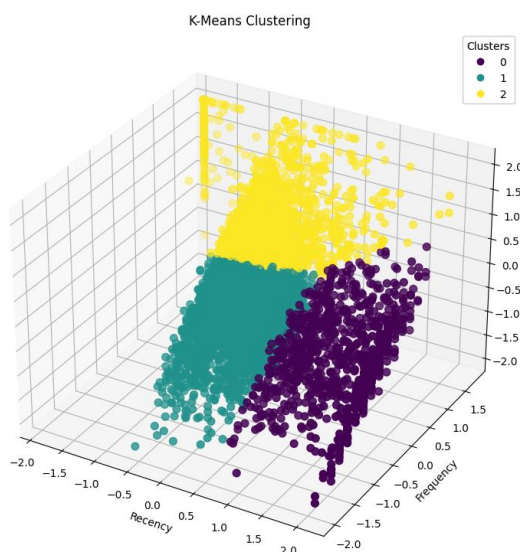
Gambar 2 3D Scatterplot K-Means Dataset 2



Gambar 5 3D Scatterplot K-Means Dataset 5



Gambar 3 3D Scatterplot K-Means Dataset 3



Gambar 4 3D Scatterplot K-Means Dataset 4

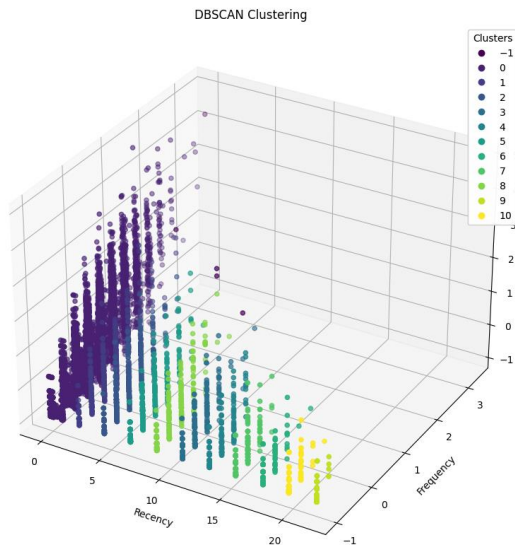
2. DBSCAN Clustering

Dalam melakukan clustering menggunakan algoritma DBSCAN, dilakukan pencarian parameter epsilon dan min samples optimal untuk proses clustering. Untuk memperoleh kedua parameter optimal, dapat memanfaatkan *k-dist graph* dan *grid-search silhouette score*. Berikut merupakan nilai parameter yang diperoleh untuk masing-masing dataset.

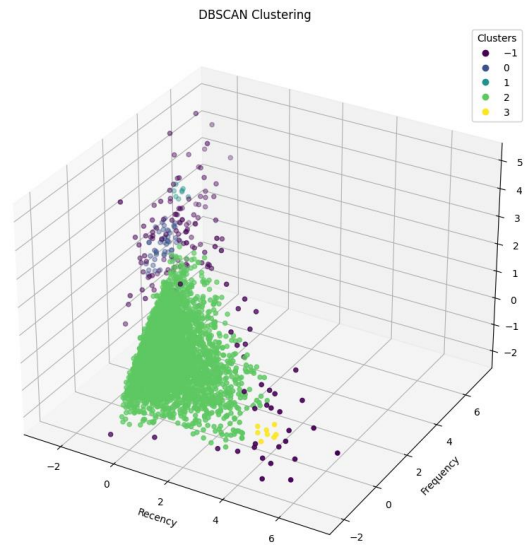
Tabel 2 Epsilon dan Min Samples Optimal

	Epsilon	Min Samples
Dataset 1	0.953	8
Dataset 2	1.510	2
Dataset 3	0.607	4
Dataset 4	0.409	7
Dataset 5	0.370	4

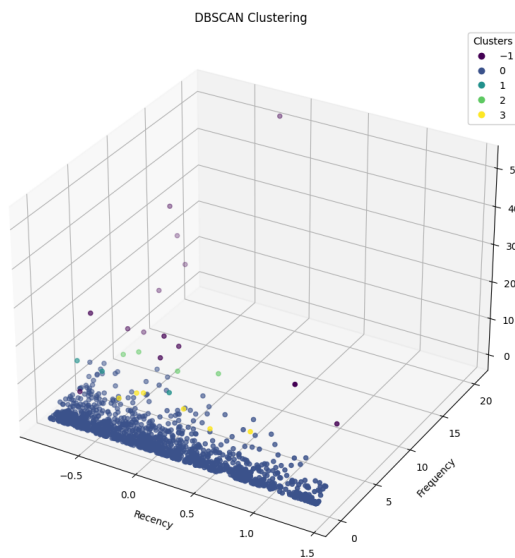
Dari parameter optimal yang telah didapatkan, dilakukan *clustering* yang menghasilkan visualisasi berbentuk 3d scatterplot untuk setiap dataset.



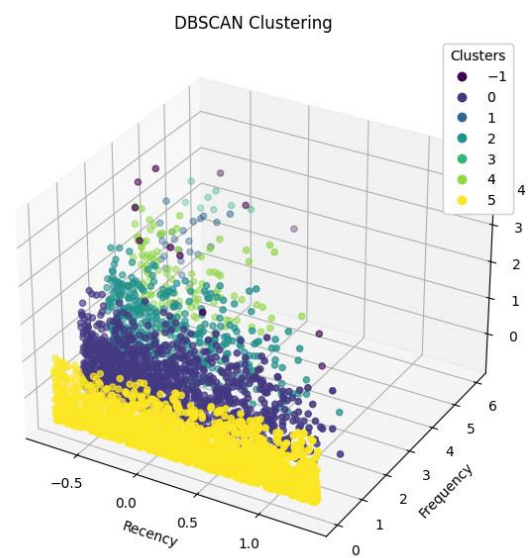
Gambar 6 3D Scatterplot DBSCAN Dataset 1



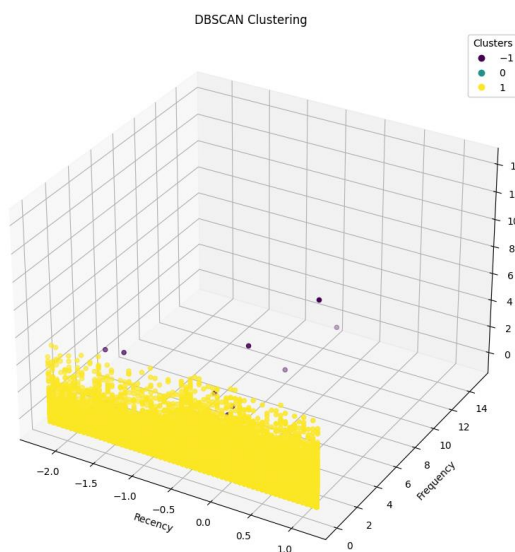
Gambar 9 3D Scatterplot DBSCAN Dataset 4



Gambar 7 3D Scatterplot DBSCAN Dataset 2



Gambar 10 3D Scatterplot DBSCAN Dataset 5



Gambar 8 3D Scatterplot Dataset 3

3.2. Evaluasi Kinerja

Evaluasi kinerja dilakukan dengan menggunakan Silhouette Score dan Davies-Bouldin Index. Hasil menunjukkan bahwa tidak ada satu algoritma yang secara konsisten lebih baik dari yang lain. Pemilihan algoritma yang sesuai sangat bergantung pada karakteristik dataset yang digunakan.

Tabel 3 Hasil Clustering Dataset 1

	<i>Silhouette Score</i>	<i>DBI</i>
K-Means	0.547	0.639
DBSCAN	0.531	0.731

Pada tabel diatas dapat dilihat bahwa nilai

silhouette score dan nilai indeks yang dimiliki oleh K-Means clustering lebih baik dibandingkan dengan nilai yang dimiliki DBSCAN clustering. Sehingga dapat diartikan bahwa performa k-means clustering pada dataset ini lebih baik daripada DBSCAN clustering, baik dari segi keserasian objek dalam cluster, pemisahan antar cluster serta kerapatan setiap objek data pada suatu cluster.

Tabel 4 Hasil Clustering Dataset 2

	<i>Silhouette Score</i>	<i>DBI</i>
K-Means	0.519	0.804
DBSCAN	0.767	0.986

Pada tabel diatas dapat dilihat bahwa nilai silhouette score dari DBSCAN clustering jauh lebih tinggi dibandingkan dengan K-Means clustering, yang berarti cluster yang dihasilkan lebih kompak dan terpisah dengan baik. Tetapi jika dilihat dari nilai davies-bouldin index, K-Means clustering lebih baik dibandingkan dengan DBSCAN clustering dengan perbedaan nilai indeks yang tidak terlalu jauh dan dapat diartikan pemisahan dan kerapatan objek dalam cluster masih hampir sama kualitasnya. Oleh karena itu, pada dataset ini peneliti menentukan bahwa performa clustering menggunakan metode DBSCAN clustering lebih baik daripada K-Means clustering, dari segi kekompakan objek dalam cluster serta pemisahan antar cluster yang cukup jelas.

Tabel 5 Hasil Clustering 3

	<i>Silhouette Score</i>	<i>DBI</i>
K-Means	0.446	0.777
DBSCAN	0.184	1.331

Pada tabel diatas dapat dilihat bahwa nilai silhouette score dan nilai indeks yang dimiliki oleh *K-Means clustering* lebih baik dibandingkan dengan nilai yang dimiliki *DBSCAN clustering*. Sehingga dapat diartikan bahwa performa k-means clustering pada dataset ini lebih baik

daripada DBSCAN clustering, cukup baik dari segi keserasian objek dalam cluster, pemisahan antar cluster serta kerapatan setiap objek data pada suatu cluster.

Tabel 6 Hasil Clustering Dataset 4

	<i>Silhouette Score</i>	<i>DBI</i>
K-Means	0.368	0.927
DBSCAN	0.545	1.239

Pada tabel diatas dapat dilihat bahwa nilai silhouette score dari DBSCAN clustering lebih tinggi dibandingkan dengan K-Means clustering yang berarti cluster yang dihasilkan lebih kompak dan terpisah dengan baik. Tetapi jika dilihat dari nilai davies-bouldin index, K-Means clustering lebih baik dibandingkan dengan DBSCAN clustering dengan perbedaan nilai indeks yang cukup jauh dan dapat diartikan pemisahan dan kerapatan objek dalam cluster cukup berbeda jauh kualitasnya. Dengan mempertimbangkan Silhouette score sebagai matriks utama dan menjadi indikator yang kuat sebagai penentuan kualitas clustering serta tetap mengambil wawasan tambahan yang diperoleh dari hasil matriks davies-boudin index, peneliti menentukan bahwa performa clustering menggunakan metode DBSCAN clustering lebih baik daripada K-Means clustering untuk dataset ini, baik dari segi kekompakan objek data didalam cluster serta baik juga dalam pemisahan antar cluster.

Tabel 7 Hasil Clustering Dataset 5

	<i>Silhouette Score</i>	<i>DBI</i>
K-Means	0.429	0.941
DBSCAN	0.343	1.606

Pada tabel diatas dapat dilihat bahwa nilai silhouette score dan nilai indeks yang dimiliki oleh *K-Means clustering* lebih baik dibandingkan dengan nilai yang dimiliki *DBSCAN clustering*. Sehingga dapat dipastikan bahwa performa k-means clustering pada dataset ini lebih baik daripada *DBSCAN clustering*, baik

dari segi keserasian objek data dalam cluster, pemisahan antar cluster serta kerapatan setiap objek data pada suatu cluster.

Dari seluruh tahap penelitian yang telah dilakukan, dapat diketahui bahwa karakteristik dataset retail apparel yang cocok menggunakan metode k-means clustering dalam segmentasi pelanggan dan model rfm yaitu dataset yang memiliki jumlah data yang besar dan memiliki distribusi yang cukup merata, memiliki sedikit outlier atau tanpa outlier serta harus dilakukan penanganan outlier untuk meningkatkan kualitas hasil clustering karena keterbatasan kemampuan k-means clustering dalam menangani outlier.

Lalu karakteristik dataset retail apparel yang cocok menggunakan DBSCAN clustering dalam hal segmentasi pelanggan dan model rfm yaitu dataset yang memiliki distribusi kepadatan yang variatif dan memiliki noise atau outlier. Karakteristik dataset lainnya yang cocok untuk metode ini yaitu data yang memiliki struktur cluster yang tidak teratur dan penanganan outlier secara langsung.

Dengan melihat performa k-means clustering dan DBSCAN clustering yang diterapkan pada lima dataset pada penelitian ini, 3 dari 5 dataset retail apparel dapat menghasilkan clustering yang baik menggunakan metode k-means clustering. Setiap dataset yang digunakan pada penelitian memiliki total data, karakteristik data, dan periode yang berbeda-beda. Untuk 3 dataset yang menghasilkan clustering yang baik menggunakan k-means clustering memiliki karakteristik dataset dengan distribusi yang cukup merata dan pengelompokan terbentuk dalam kelompok jumlah data yang besar. Di sisi lain, 2 dataset yang menghasilkan cluster yang baik dengan menggunakan DBSCAN clustering memiliki karakteristik dataset dengan kepadatan yang bervariasi. Dengan menggunakan berbagai variasi dataset retail apparel, peneliti dapat menemukan metode clustering yang baik untuk segmentasi pelanggan menggunakan model RFM.

4. KESIMPULAN

Dari seluruh tahap penelitian yang telah dilakukan, dapat diketahui bahwa karakteristik dataset retail apparel yang cocok menggunakan metode k-means clustering dalam segmentasi pelanggan dan

model rfm yaitu dataset yang memiliki jumlah data yang besar dan memiliki distribusi yang cukup merata, memiliki sedikit outlier atau tanpa outlier serta harus dilakukan penanganan outlier untuk meningkatkan kualitas hasil clustering karena keterbatasan kemampuan k-means clustering dalam menangani outlier.

Lalu karakteristik dataset retail apparel yang cocok menggunakan DBSCAN clustering dalam hal segmentasi pelanggan dan model rfm yaitu dataset yang memiliki distribusi kepadatan yang variatif dan memiliki noise atau outlier. Karakteristik dataset lainnya yang cocok untuk metode ini yaitu data yang memiliki struktur cluster yang tidak teratur dan penanganan outlier secara langsung.

Dengan melihat performa k-means clustering dan DBSCAN clustering yang diterapkan pada lima dataset pada penelitian ini, 3 dari 5 dataset retail apparel dapat menghasilkan clustering yang baik menggunakan metode k-means clustering. Setiap dataset yang digunakan pada penelitian memiliki total data, karakteristik data, dan periode yang berbeda-beda. Untuk 3 dataset yang menghasilkan clustering yang baik menggunakan k-means clustering memiliki karakteristik dataset dengan distribusi yang cukup merata dan pengelompokan terbentuk dalam kelompok jumlah data yang besar. Di sisi lain, 2 dataset yang menghasilkan cluster yang baik dengan menggunakan DBSCAN clustering memiliki karakteristik dataset dengan kepadatan yang bervariasi. Dengan menggunakan berbagai variasi dataset retail apparel, peneliti dapat menemukan metode clustering yang baik untuk segmentasi pelanggan menggunakan model RFM.

5. DAFTAR PUSTAKA

Berahmana, R. W., Mohammed, F. A., & Chairuang, K. (2020). Customer Segmentation Based on RFM Model Using K-Means, K-Medoids, and DBSCAN Methods. LONTAR KOMPUTER.

- Christy, A. J., Umamakeswari, A., Priyatharsini, L., & Neyaa, A. (2021). RFM Ranking - An Effective Approach to Customer Segmentation. *Journal of King Saus University - Computer Science and Information Sciences*, 1251-1257.
- Cios, K. J., Pedrycz, W., & Swiniarski, R. W. (1998). *Data Mining Methods for Knowledge Discovery*. New York: Springer Science Business Media, LLC.
- Cochran, W. G. (1997). *Sampling Techniques* third edition. John Wiley & Sons.
- Elbatta, M. N. (2012). An Improvement for DBSCAN Algorithm for Best Result in Varied Densities.
- Elliyana, E., Agustina, T. S., Kadir, H., Imanuddin, B., Yuliani, Mohdari, M., Rachman, S. (2022). *Dasar-Dasar Pemasaran*. Malang: Ahlimedia Press.
- Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (1996). A Density-Based Algorithm for Discovering Clusters. *KDD-96 Proceeding*, 226-231.
- Fayyad, U., Platetsky-Shapiro, G., & Smyth, P. (1996). From Data Mining to Knowledge Discovery in Database. *AI Magazine*, 37-54.
- Liantoni, F. (2022). *Data Mining dan Penerapan Metode*. Purbalingga: Eureka Media Aksara.
- Maimon, O., & Rokach, L. (2010). *Data Mining and Knowledge Discovery Handbook: Second Edition*. New York: Springer Science and Business Media.
- Monko, G. J., & Kimura, M. (2023). Optimized DBSCAN Parameter Selection: Stratified Sampling for Epsilon and GridSearch. *Computer Science & Information Technology (CS & IT)*, 43-61.
- Osborne, J. W. (2008). Best practices in data transformation: the overlooked effect of minimal values.
- Rizan, M., Dimuk, M., Kamal, A. F., Setiabudi, D., Azizy, F. N., Marlina, H., Sari, U. I. (2023). *Strategi Bisnis di Era Digital*. Bandung: Widina.
- Sahinbas, K. (2022). Performance Comparison of K-Means and DBSCAN Methods for Airline Customer Segmentation. *Black Sea Journal of Engineering and Science*, 158-165.
- Saparso. (2021). *Marketing Process: Menciptakan Nilai Bagi Pelanggan*. Jakarta Barat: Ukrida Press.
- Shahapure, K. R., & Nicholas, C. (2020). Cluster Quality Analysis Using Silhouette Score. *IEEE &th International Conference on Data Science and Advanced Analytics (DSAA)*.
- Team DSH. (2023). *Data Cleaning and Processing for Data Science Beginner*. Data Science Horizons.
- Wardani, S. D., Aryanto, A. S., Umroh, M., & Rolliawati, D. (2023). Perbandingan Hasil Metode Clustering K-Means, DBSCANNER & Hierarchical untuk Analisa Segmentasi Pasar. *JKO*, 191-201.
- Wei, J. T., Lee, M. C., Chen, H. K., & Wu, H. H. (2013). Customer Relationship Management in The Hairdressing Industry: An Application of Data Mining Techniques. *Expert System With Application*, 7513-7518.
- Wei, J. T., Lin, S. Y., & Wu, H. H. (2010). A Review of the Application of RFM Model. *African Journal of Business Management*, 4199-4206.