

BAB II

LANDASAN TEORI

2.1 Pengelompokan Pelanggan

Pelanggan adalah konsumen yang secara rutin membeli atau menggunakan suatu produk atau jasa, yang dimana konsumen merupakan individu yang membeli atau memanfaatkan produk atau jasa tersebut. Pelanggan merupakan faktor utama dalam keberhasilan suatu bisnis dan perusahaan, karena mereka menjadi penentu kinerja penjualan, tingkat keuntungan, dan pangsa pasar. Keberhasilan operasional sangat bergantung pada pelanggan yang datang atau melakukan pembelian ulang[2].

Segmentasi pelanggan adalah strategi membagi pelanggan ke dalam kelompok-kelompok berbeda berdasarkan karakteristik dan perilaku yang serupa. Tujuannya adalah untuk memahami perilaku pelanggan, menganalisis preferensi mereka, dan mengidentifikasi kelompok pelanggan dengan kebutuhan serupa[3].

Manajemen perusahaan mencakup prinsip-prinsip berikut:

- a. Pelanggan tidak bergantung pada kita, melainkan kita yang bergantung pada mereka.
- b. Pelanggan adalah individu yang menghadirkan kebutuhan mereka kepada kita.
- c. Tidak ada yang pernah menang dalam perdebatan dengan pelanggan
- d. Pelanggan adalah pihak yang sangat penting dan harus mendapatkan kepuasan.

Dalam dunia bisnis, terdapat tiga jenis pelanggan yaitu [4]:

1. Pelanggan internal (*internal customer*)

Merupakan individu yang berada di dalam perusahaan dan memengaruhi kinerja serta hasil kerja perusahaan. Contohnya meliputi bagian pembelian, produksi, penjualan, penggajian, rekrutmen, dan karyawan lainnya.

2. Pelanggan antara (*intermediate customer*)

Adalah pihak yang berperan sebagai perantara dan bukan pengguna akhir dari suatu produk. Contohnya adalah distributor yang menyalurkan produk kepada konsumen dan agen perjalanan yang memesan kamar hotel untuk pengguna akhir.

3. Pelanggan eksternal (*external customer*)

Merupakan pembeli atau pengguna akhir produk, yang sering disebut sebagai pelanggan nyata (*real customer*). Mereka adalah pihak yang membayar untuk menggunakan produk tersebut. Dalam beberapa kasus, pihak yang membayar dan yang menggunakan produk bisa berbeda.

2.2 Preferensi Belanja

Kotler (2000:154) mendefinisikan preferensi konsumen sebagai menunjukkan kesukaan konsumen dari berbagai pilihan produk jasa yang ada sementara Poerwadaminta (2006:769) dalam Kamus Besar Bahasa Indonesia mendefinisikan preferensi yaitu merupakan kesukaan (kecenderungan hati) kepada sesuatu. Preferensi juga diartikan sebagai pilihan suka atau tidak suka oleh seseorang terhadap suatu produk, barang atau jasa yang dikonsumsi[5].

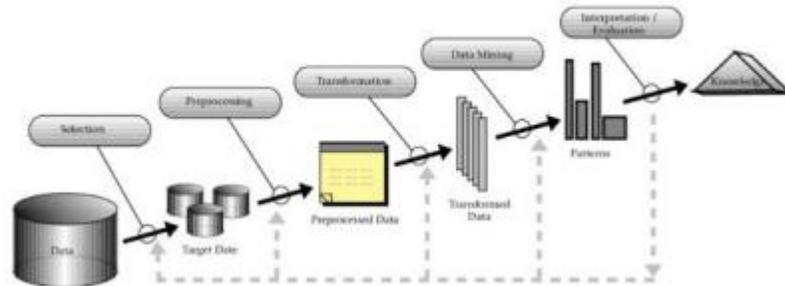
Preferensi belanja pelanggan muncul ketika pelanggan melakukan perbandingan, evaluasi, dan penilaian terhadap berbagai pilihan yang tersedia, kemudian mengambil keputusan. Preferensi ini merupakan hasil dari perilaku yang ditunjukkan konsumen selama proses mencari, membeli, hingga membuang produk. Jadi preferensi belanja pelanggan adalah proses evaluasi berdasarkan pemilihan, minat, perbandingan, penilaian yang dilakukan oleh pelanggan sebelum memutuskan membeli/menggunakan suatu barang/jasa [6].

2.3 Data Mining

Data mining itu sendiri adalah proses pencarian pola-pola yang tersembunyi (*hidden patern*) berupa pengetahuan (*knowledge*) yang tidak diketahui sebelumnya dari suatu sekumpulan data yang mana data tersebut dapat berada didalam *database*, *warehouse* data, atau media penyimpanan informasi yang lain [7].

Data mining mulai ada sejak 1990-an sebagai cara yang benar dan tepat untuk mengambil pola dan informasi yang digunakan untuk menemukan hubungan antara data untuk melakukan pengelompokkan ke dalam satu atau lebih *cluster* sehingga objek-objek yang berada dalam satu *cluster* akan mempunyai kesamaan yang tinggi antara satu dengan lainnya. *Data mining* merupakan bagian dari proses penemuan pengetahuan dari basis data *Knowledge Discovery in Databases* (KDD) [8].

Ada lima tahapan dalam proses KDD yaitu *selection*, *preprocessing*, *transformation*, *data mining*, *interpretation/evolution* yang dapat dilihat pada gambar di bawah ini.



Gambar 2. 1 Proses *Knowledge Discovery in Database (KDD)*

Sumber: [8]

Berikut penjelasan dari proses *Knowledge Discovery in Database (KDD)*:

1. *Selection*

Selection merupakan proses seleksi data dari sekumpulan data yang ada dan proses penentuan jenis serta kategori data yang akan digunakan pada penelitian.

2. *Preprocessing*

Preprocessing merupakan tahapan dalam melakukan pembersihan data, *missing value* atau yang bernilai (*null*), memeriksa data yang tidak konsisten, dan memperbaiki kesalahan pada data.

3. *Transformation*

Tahap ini melibatkan penyesuaian data agar sesuai dengan algoritma *K-Means*. Proses ini dilanjutkan dengan memilih atribut-atribut yang paling tepat untuk mendukung analisis *K-Means*.

4. *Data Mining*

Proses *data mining* pada penelitian ini menggunakan *K-means Clustering* untuk mengelompokkan pelanggan berdasarkan jenis sepatu yang dibeli, frekuensi pembelian, dan merek sepatu yang paling diminati. Selanjutnya, menentukan jumlah kelompok atau *cluster* yang optimal untuk dianalisis.

5. *Interpretation / Evaluation*

Setelah data melalui proses pengolahan dengan algoritma *machine learning* menggunakan *Python* dan *Google Colab*, tahapan selanjutnya adalah evaluasi. Evaluasi ini merupakan tahapan akhir dari proses *data mining* untuk menilai kualitas hasil pengolahan data.

2.4 *Clustering*

Clustering adalah proses pengelompokan data ke dalam cluster berdasarkan kemiripan. Sebagai bagian dari *unsupervised learning*, *clustering* tidak memerlukan label untuk setiap *cluster*. Tujuannya adalah agar data dalam satu *cluster* lebih mirip satu sama lain daripada dengan data di *cluster* lainnya. *Clustering* merupakan pendekatan yang digunakan untuk mengelompokkan data ke dalam beberapa kelompok atau *cluster*. Pengelompokan ini didasarkan pada kemiripan atribut yang dimiliki oleh data-data dalam *cluster* tersebut[9]. *Clustering* juga dikenal sebagai pengelompokan data tanpa pengawasan dengan teknik pembelajaran mesin tak berlabel yang membagi data ke dalam kelompok-kelompok serupa tanpa memerlukan fase pembelajaran sebelumnya. Beberapa metode *clustering* yang umum digunakan antara lain *K-Means*, *K-Medoid*, *K-Mode*, dan *Hierarchical Clustering*[10].

Metode *clustering* sangat bergantung pada konsep kemiripan dan kesamaan antar data. Pada dasarnya, clustering bertujuan untuk mengelompokkan data ke dalam kelompok-kelompok yang memiliki tingkat kemiripan internal yang tinggi dan tingkat kemiripan eksternal yang rendah. Dalam konteks ini, kemiripan mengacu pada sejauh mana dua objek memiliki karakteristik atau atribut yang serupa. Sementara itu, kesamaan mengacu pada sejauh mana dua objek dapat dianggap identik atau memiliki karakteristik yang sama persis. Konsep yang mendasari metode *clustering* adalah bahwa semakin tinggi tingkat kemiripan antara dua objek, semakin tinggi pula tingkat kesamaan mereka. Hal ini berarti bahwa objek-objek yang memiliki banyak kesamaan akan cenderung dikelompokkan bersama dalam *cluster* yang sama[11].

2.5 K-Means Clustering

2.5.1 Pengertian K-Means Clustering

Algoritma *K-Means* adalah salah satu metode *clustering* dalam *data mining* yang digunakan untuk mengelompokkan data. Algoritma ini bekerja dengan membagi data menjadi dua kelompok atau lebih berdasarkan kemiripan karakteristiknya. Data dengan karakteristik serupa akan dikelompokkan bersama, sementara data dengan karakteristik berbeda akan dipisahkan ke kelompok lain[12].

Algoritma K-Means bekerja dengan cara mengelompokkan data ke dalam sejumlah kelompok berdasarkan kedekatan titik data dengan pusat kelompok. Prosesnya dimulai dengan memilih titik-titik pusat secara acak, lalu memasukkan setiap titik data ke kelompok terdekat. Selanjutnya, pusat kelompok dihitung ulang

berdasarkan rata-rata titik data dalam kelompok tersebut. Proses ini diulang hingga pusat kelompok stabil atau mencapai batas iterasi[13].

2.5.2 Langkah-Langkah *K-Means Clustering*

Berikut langkah-langkah *K-Means Clustering*[14]:

1. Tentukan jumlah *cluster* (k) yang diinginkan.
2. Tentukan k titik pusat *cluster* (centroid) awal secara acak. Hitung *centroid cluster*

berikutnya menggunakan rumus berikut:

$$D(ij) = \sqrt{(x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2 + \dots + (x_{ki} - x_{kj})^2}$$

Keterangan:

D(ij) = Jarak data ke i ke pusat cluster j

Xki = Data ke i atribut data ke j

Xkj = Titik pusat ke j pada atribut ke k

3. Hitung jarak setiap data ke masing-masing *centroid* menggunakan rumus jarak

Euclidean (d).

$$d = \sum_{i=1}^n (xi - yi)^2$$

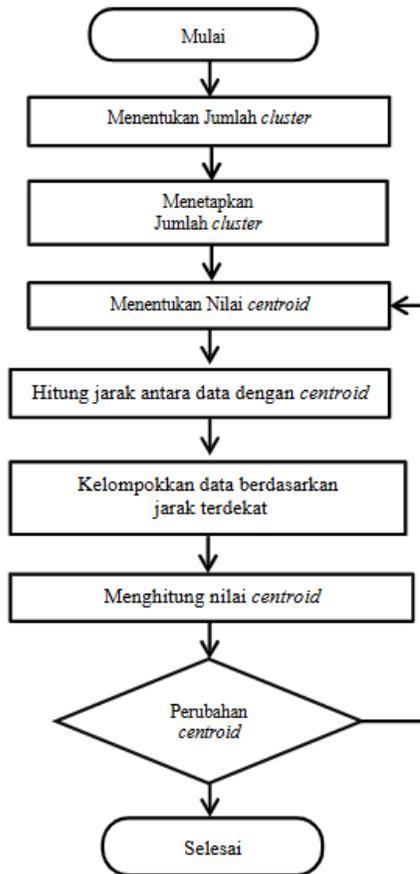
Keterangan:

xi = Objek pengamatan ke i

yi = *Centroid* ke i

n = Banyaknya obyek yang menjadi anggota klaster

4. Kelompokkan setiap data ke *centroid* terdekat.
5. Hitung *centroid* baru dengan mencari rata-rata dari data dalam *cluster* yang sama.



Gambar 2. 2 Flowchart K-Means Clustering

Sumber:[15]

6. Ulangi langkah 3-5 hingga posisi centroid baru sama dengan *centroid* lama atau kriteria berhenti lainnya terpenuhi.

2.6 Evaluasi Model Clustering

2.6.1 Elbow Method

Setelah proses pengelompokan data menggunakan algoritma *K-Means* selesai, hasil pengelompokan untuk setiap nilai *k* (jumlah *cluster*) akan divalidasi menggunakan dua metode, yaitu *Elbow Method* dan *SSE (Sum of Square Error)*.

Elbow Method digunakan untuk menentukan jumlah *cluster* yang optimal dengan menganalisis grafik perbandingan antara jumlah *cluster* (k) dan penurunan error. Titik siku pada grafik tersebut menunjukkan jumlah *cluster* yang terbaik. Sementara itu, SSE digunakan untuk mengukur perbedaan antara data aktual dengan hasil prediksi model. Nilai SSE yang rendah menunjukkan bahwa model pengelompokan yang dihasilkan baik[16].

2.6.2 Davies Bouldin Index (DBI)

Indeks Davies-Bouldin adalah metrik penting dalam analisis data yang digunakan untuk mengukur kualitas *clustering*. Metrik ini memberikan informasi tentang seberapa baik algoritma *clustering* dapat memisahkan data ke dalam kelompok-kelompok yang berbeda dan seberapa dekat data dalam satu *cluster* dengan pusat *clusternya*. Nilai *Indeks Davies-Bouldin* yang rendah menunjukkan bahwa semakin baik kualitas *clustering* yang dihasilkan, arena menunjukkan bahwa klaster-klasternya lebih terpisah dan lebih dekat dengan pusat klasternya[17].

2.6.3 Silhoutte Coefficient (SC)

Dalam proses *clustering*, algoritma akan terus beriterasi hingga mencapai kondisi konvergensi. Konvergensi terjadi ketika hasil pengelompokan tidak lagi mengalami perubahan signifikan, yang ditandai dengan tidak adanya data yang berpindah antar *cluster*. Setelah konvergensi tercapai, hasil pengelompokan tersebut perlu divalidasi untuk memastikan kualitasnya. Salah satu metode validasi yang umum digunakan adalah *Silhouette Coefficient*. Metode ini mengukur seberapa baik setiap data dikelompokkan dengan menghitung jarak rata-rata data tersebut ke data

lain dalam *cluster* yang berbeda. Validasi ini penting untuk memastikan bahwa pengelompokan yang dihasilkan tidak hanya stabil, tetapi juga akurat dan bermakna[18]. Nilai *Silhouette Coefficient* dapat diinterpretasikan seperti yang tercantum dalam tabel berikut.

Tabel 2. 1 Interpretasi Nilai *Silhouette Coefficient*

Rentang Nilai	Interpretasi
0.71 – 1.0	Struktur Kuat
0.51 – 0.70	Struktur yang beralasan
0.26 – 0.50	Struktur Lemah
<0.25	Tidak ditemukan struktur yang substansial

2.7 Machine Learning

Machine Learning adalah cabang AI yang berfokus pada pengembangan sistem yang dapat belajar dari data. Data menjadi kunci utama dalam *Machine Learning*, karena melalui data inilah komputer dapat mengenali pola, membuat prediksi, dan meningkatkan kinerjanya. Proses pembelajaran ini tidak memerlukan 17 campur tangan manusia secara terus-menerus, karena komputer dapat belajar secara mandiri berdasarkan data yang diberikan[19].

Dalam *Machine Learning*, data adalah fondasi utama. Tanpa data, algoritma *Machine Learning* tidak dapat berfungsi. Data yang digunakan biasanya dibagi menjadi dua bagian: data training dan data testing. Data training berperan untuk melatih algoritma agar dapat belajar dan mengenali pola dari data. Setelah dilatih, algoritma akan diuji menggunakan data testing untuk mengukur seberapa baik

kinerjanya dalam memprediksi atau mengklasifikasikan data baru yang belum pernah dilihat sebelumnya[20].

2.8 Alat Bantu Pemrograman dan *Tools* Pendukung

2.8.1 *Python*

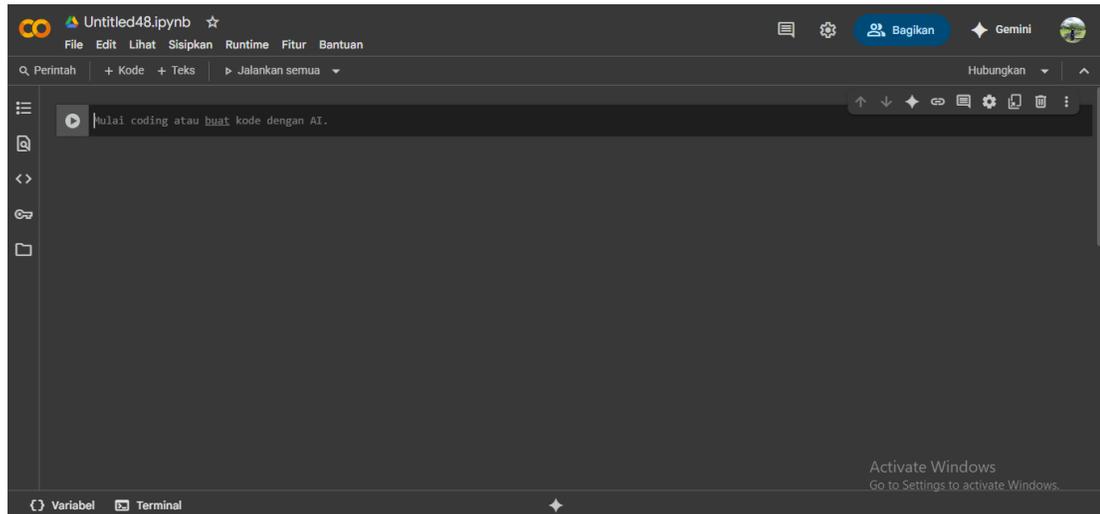
Python merupakan bahasa pemrograman yang menggunakan interpreter untuk mengeksekusi kode program secara langsung. Dengan sifatnya tersebut, *Python* dapat dijalankan di berbagai platform seperti Windows, Linux, dan lainnya. Bahasa ini mengadopsi beragam paradigma pemrograman dari bahasa lain, seperti pemrograman prosedural ala C, pemrograman berorientasi objek seperti Java, serta pemrograman fungsional seperti Lisp. Perpaduan berbagai paradigma tersebut memberikan kemudahan bagi programmer dalam mengembangkan beragam proyek menggunakan *Python*[21].

Python dikenal sebagai bahasa pemrograman yang sangat mudah dibaca. Sebagai *interpreted language* (bahasa pemrograman yang tidak memerlukan proses kompilasi), *Python* menjalankan kode tanpa terlebih dahulu mengubahnya menjadi bentuk yang dapat dibaca komputer. Bahasa ini bersifat tujuan umum dan berada pada tingkat tinggi, dirancang sebagai “bunglon” di dunia pemrograman karena fleksibilitasnya. *Python* bertujuan menghasilkan kode yang jelas dan logis, baik untuk proyek skala kecil maupun besar. Ibarat kubus Rubik, *Python* memiliki banyak sisi yang dapat dieksplorasi dan dimanfaatkan. Kemampuannya memungkinkan penerapan beragam prosedur komputasi untuk menciptakan teknologi yang sering kali memberi kejutan[22].

2.8.2 Google Collaboratory

Colab atau *Google Colaboratory*, adalah platform yang dikembangkan oleh *Google Research* yang memudahkan siapa saja untuk menulis dan menjalankan kode *Python* melalui browser. Platform ini sangat ideal untuk *Machine Learning*, analisis data, dan tujuan pendidikan. *Colab* pada dasarnya adalah layanan *notebook Jupyter* yang dihosting, sehingga pengguna tidak perlu melakukan pengaturan apapun. Selain itu, *Colab* juga menyediakan akses gratis ke sumber daya komputasi, termasuk GPU, menjadikannya pilihan yang menarik bagi para pengembang dan peneliti[23].

Google Colab adalah salah satu platform yang dirancang khusus untuk mempermudah berbagai pekerjaan yang berkaitan dengan bidang *data science*. Platform ini menyediakan lingkungan pemrograman berbasis *cloud* yang memungkinkan pengguna untuk menulis, menjalankan, dan membagikan kode *Python* tanpa perlu melakukan instalasi perangkat lunak tambahan di komputer lokal. Keunggulan lainnya, *Google Colab* mendukung penggunaan secara kolaboratif, di mana beberapa pengembang aplikasi dapat bekerja bersama pada proyek yang sama secara *real-time*. Fitur ini sangat membantu dalam proses pengembangan, pengujian, maupun analisis data, karena setiap anggota tim dapat saling berkontribusi, memeriksa, serta memperbarui kode atau data secara langsung. Dengan demikian, *Google Colab* menjadi solusi yang efektif untuk mendukung kerja sama tim, meningkatkan produktivitas, dan mempercepat penyelesaian proyek yang memerlukan integrasi pengetahuan serta keahlian dari berbagai anggota tim[24].



Gambar 2. 3 Tampilan *Google Collaboratory*

2.9 Penelitian Terdahulu

No.	Judul	Nama Penulis	Tahun	Hasil
1.	Segmentasi Pelanggan Menggunakan <i>K-Means Clustering</i> Di Toko Retail	Syifa Lathifah Achmad, Ahmad Fauzi, Rahmat, Jamaludin Indra [25].	2024	Penelitian ini melakukan segmentasi pelanggan menggunakan <i>K-Means clustering</i> dan PCA, menghasilkan 6 cluster dengan karakteristik unik. Model <i>K-Means</i> mencapai <i>Silhouette score</i> 0.666. Analisis umur, daya beli, pekerjaan, dan status pernikahan menunjukkan perbedaan antar kluster. Evaluasi dengan Metode <i>Elbow</i> dan <i>Silhouette Score</i> mendukung 6 cluster sebagai jumlah optimal, memberikan wawasan untuk strategi pemasaran yang lebih efektif.

No.	Judul	Nama Penulis	Tahun	Hasil
2.	Analisis Segmentasi Pelanggan Menggunakan <i>K-Means Clustering</i> Studi Kasus Aplikasi Alfagift	Satria Ardi Perdana, Sara Famayla Florentin, Agus Santoso [26].	2022	Segmentasi pelanggan menggunakan metode <i>clustering K-Means</i> berhasil mengelompokkan pelanggan berdasarkan umur, jenis kelamin, frekuensi pesanan, tipe pembayaran, dan kota pembelian selama satu bulan. Hasilnya, terbentuk tiga <i>cluster</i> , yaitu <i>cluster 1</i> , <i>cluster 2</i> , dan <i>cluster 3</i> . Informasi ini diharapkan dapat membantu perusahaan memahami kebutuhan pelanggan dengan lebih baik dan menentukan strategi pemasaran yang tepat.
3.	Distribusi Spasial <i>Unmet Need</i> Pelayanan Kesehatan dengan Algoritma <i>K-Means</i> untuk Pemetaan Provinsi di Indonesia	Kusmanto, Samsir, Ronal Watrianthos, Sudi Suryadi [13].	2023	Penelitian ini memetakan tingkat <i>unmet need</i> pelayanan kesehatan di Indonesia menggunakan <i>K-Means clustering</i> , menghasilkan tiga klaster provinsi. Klaster dengan <i>unmet need</i> terendah (2,47%) didominasi provinsi di Jawa, sementara klaster dengan <i>unmet need</i> tertinggi (7,35%) umumnya berada di luar Jawa.

No.	Judul	Nama Penulis	Tahun	Hasil
4.	Analisis Segmentasi Pelanggan Berbasis Model <i>Recency Frequency</i> , dan <i>Monetary</i> (RFM) Menggunakan Algoritma <i>K-Means</i>	Panji Indra Pangestu, Teguh Iman Hermanto, Dede Irmayanti [27].	2023	Analisis pengelompokan data transaksi e-commerce menggunakan K-Means menghasilkan tiga cluster. Cluster terbesar (2395 pelanggan) adalah <i>Highest Loyalty</i> . Penelitian ini menunjukkan bahwa K-Means dan <i>Silhouette Coefficient</i> efektif untuk segmentasi pelanggan, membantu perusahaan memahami karakteristik pelanggan dan menerapkan strategi yang tepat. Kekurangannya adalah kebutuhan akan data pelanggan terbaru.
5.	Segmentasi Konsumen Berdasarkan Model <i>Recency, Frequency, Monetary</i> dengan Metode <i>K-Means</i>	Atik Febriani, Syahfara Ashari Putri[28].	2020	Penelitian ini menggunakan K-means dan SPSS untuk mengelompokkan pelanggan berdasarkan perilaku pembelian. Tiga klaster terbentuk, dengan Klaster 1 (pelanggan potensial dengan R rendah, F dan M tinggi) perlu dipertahankan. Klaster 2 memerlukan peningkatan strategi, dan penelitian lebih lanjut disarankan untuk memahami profil pelanggan mereka. Penelitian selanjutnya disarankan untuk membandingkan dengan metode klasterisasi lain.

No.	Judul	Nama Penulis	Tahun	Hasil
6.	Analisis Performa Algoritma <i>K-Means</i> dan DBSCAN Dalam Segmentasi Pelanggan Dengan Pendekatan Model RFM	Furqan Maulana Pranata, Satrio Hadi Wijoyo, dan Nanang Yudi Setiawan[29]	2024	Hasil clustering dari masing-masing dataset menunjukkan variasi dalam efektivitas kedua algoritma. <i>K-Means Clustering</i> 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.
7.	Penerapan Algoritma <i>K-Medoids</i> Untuk Menentukan Segmentasi Pelanggan	Anggi Ayu Dwi Sulistyawati, Mujino Sadikin[30]	2021	Dari hasil penelitian yang telah dilakukan, segmentasi pelanggan dengan menerapkan algoritma <i>K-Medoids</i> menghasilkan jumlah cluster optimal adalah 3 (tiga). Secara umum dapat disimpulkan bahwa segmentasi paling baik untuk kasus dataset Perum BULOG ini adalah 3 (tiga) segmen yaitu Lost Customer, Core Customer, dan New Customer. Segmentasi pelanggan dapat mengidentifikasi tipe dan karakteristik pelanggan berdasarkan perilaku atau kebiasaan yang sama. Hasil segmentasi pelanggan dapat menjadi acuan

No.	Judul	Nama Penulis	Tahun	Hasil
				dalam menyusun rekomendasi strategi pemasaran. Potensi riset masa depan yang bisa dikembangkan adalah memprediksi Customer Lifetime Value (CLV) dan memprediksi customer churn. Hal tersebut bertujuan untuk penargetan pelanggan agar strategi pemasaran yang disusun lebih tepat sasaran.

2.10 Kelebihan Penelitian

Penelitian ini memiliki sejumlah kelebihan baik dari segi metode maupun manfaat praktis. Penggunaan algoritma *K-Means Clustering* memungkinkan proses pengelompokan pelanggan dilakukan secara efisien dan cepat berdasarkan data preferensi belanja seperti jumlah item yang dibeli, frekuensi pembelian, dan jumlah transaksi. Metode ini juga mudah diimplementasikan dan hasilnya dapat langsung dimanfaatkan oleh pemilik toko untuk memahami pola belanja pelanggan. Dengan segmentasi tersebut, toko sepatu TRIKO dapat menyusun strategi pemasaran yang lebih tepat sasaran, seperti memberikan promo khusus kepada pelanggan loyal.

Selain itu, pendekatan penelitian ini berbasis data aktual, sehingga hasilnya bersifat objektif dan dapat diandalkan. Informasi yang dihasilkan membantu menyederhanakan pengambilan keputusan dan dapat menjadi dasar pengembangan sistem berbasis teknologi seperti CRM (*Customer Relationship Management*) atau

sistem rekomendasi. Penelitian ini juga memberikan kontribusi ilmiah sebagai dasar untuk studi lanjutan di bidang data *mining* dan segmentasi pelanggan. Secara keseluruhan, penelitian ini tidak hanya bermanfaat secara akademis, tetapi juga memberikan dampak nyata dalam praktik bisnis ritel.