

BAB II

LANDASAN TEORI

2.1 Tinjauan Teori

2.1.1 Customer

Customer juga dikenal sebagai pelanggan, adalah kunci kemajuan bisnis, dan perusahaan berusaha untuk mempertahankan pelanggan mereka dan menghasilkan keuntungan yang signifikan bagi mereka., namun kesulitan dalam memperoleh pelanggan yang *loyal* kerana terkadang sifat pelanggan yang dapat berubah-ubah [6]. *Customer* atau pelanggan merupakan individu atau kelompok yang menjadi pelaku dalam pembelian produk atau jasa berdasarkan keperluan mereka yang berdasarkan pertimbangan dari cara pandang dalam memanfaatkan dan harga produk atau jasa tersebut [7].

2.1.2 Segmentasi

Segmentasi adalah proses dimana berbagai pasar dikelompokkan menjadi pelanggan potensial, kebutuhan dan preferensi yang sebanding saat berbelanja. Ini memungkinkan bisnis untuk berkonsentrasi pada pembagian sumber daya dan mendapatkan pemahaman yang jelas tentang elemen strategis. Berdasarkan karakteristik pelanggan, segmentasi membagi pasar menjadi empat jenis :

1. Segmentasi Geografis, yang membagi pasar berdasarkan karakteristik pelanggan seperti rumah tangga, lingkungan, provinsi, wilayah, negara bagian & negara.
2. Demografis, ini adalah pembagian pasar yang didasarkan pada kebangsaan, ras, status pernikahan, agama, pendidikan, pendapatan, pekerjaan, kelamin, jenis, dan usia. Variabel ini populer karena terkait dengan kebutuhan dan keinginan pelanggan. Selain itu, ketika segmentasi dilakukan dengan variabel non-demografis, seperti yang didasarkan pada jenis kepribadian, ukuran pasar masih dihitung dengan variabel demografis.
3. Psikografis, yang membagi pelanggan berdasarkan kepribadian, status sosial, atau gaya hidup.

4. Tingkah Laku, yang membagi konsumen ke dalam segmen berdasarkan sikap, status sosial, atau karakteristik kepribadian. Bagaimana sikap pelanggan terhadap produk, merek, keuntungan, atau loyalitas dapat diamati dengan menggunakan model berbasis perilaku yang disebut model RFM.

Skema segmentasi yang berbeda dapat dikembangkan menurut tujuan bisnis yang spesifik dari organisasi. Segmentasi umumnya digunakan melalui riset data pasar untuk mendapatkan wawasan tentang sikap pelanggan, keinginan, pandangan, preferensi, dan opini tentang perusahaan dan kompetisi.

2.1.3 Customer Segmentation (Segmentasi Pelanggan)

Segmentasi pelanggan yaitu untuk membuat strategi pemasaran yang menghasilkan pemasukan yang lebih besar, segmentasi pelanggan adalah proses membagi pelanggan ke dalam beberapa segmen berdasarkan kategori kesetiaan pelanggan mereka [8]. Diperlukan segmentasi untuk mengidentifikasi *customer* yang memiliki karakteristik sama agar untuk mengidentifikasi perilaku mereka. [9].

Segmentasi pelanggan berdasarkan pada riset pasar dan demografi seringkali membutuhkan pemabahan karakteristik semua pelanggan agar lebih efektif mengetahui segmen apa yang menjadi menarik pelanggan. Penggalan data dapat mengembangkan segmentasi pelanggan yang juga mengidentifikasi segmentasi pada perilaku pelanggan. Selain data penelitian eksternal atau pasar, data transaksi dan pembayaran pelanggan juga dapat digunakan untuk mendapatkan wawasan tentang perilaku pelanggan. Segmentasi dengan cara tersebut, dapat mengalokasikan pelanggan untuk membentuk kelompok berdasarkan jumlah pengeluaran mereka. Hal ini dapat digunakan untuk mengidentifikasi pelanggan yang bernilai tinggi dan memprioritaskan pelayanan [10]. Berikut pada Tabel 2.1 merupakan karakteristik pelanggan.

Tabel 2. 1 Karakteristik Pelanggan

Kelas Pelanggan	Karakteristik
<i>Superstar</i>	a. Pelanggan dengan loyalti yang tinggi b. Mempunyai nilai <i>monetary</i> yang tinggi

	<ul style="list-style-type: none"> c. Mempunyai frekuensi yang tinggi d. Mempunyai transaksi paling tinggi
<i>Golden customer</i>	<ul style="list-style-type: none"> a. Mempunyai nilai <i>monetary</i> tertinggi yang kedua b. <i>Frequency</i> yang tinggi c. Mempunyai rata-rata transaksi
<i>Typical customer</i>	Mempunyai rata-rata nilai <i>monetary</i> dan rata-rata transaksi
<i>Occasional customer</i>	<ul style="list-style-type: none"> a. Nilai <i>monetary</i> terendah kedua setelah <i>dormant customer</i> b. Nilai <i>recency</i> paling rendah c. Transaksi paling tinggi
<i>Everyday shopper</i>	<ul style="list-style-type: none"> a. Memiliki peningkatan transaksi b. Transaksi yang rendah c. Mempunyai nilai <i>monetary</i> sedang sampai dengan rendah
<i>Dormant customer</i>	<ul style="list-style-type: none"> a. Mempunyai <i>frequency</i> dan <i>monetary</i> paling rendah b. Nilai <i>recency</i> yang paling rendah

2.1.4 Data Mining

Menambang data yaitu proses menemukan data yang sangat besar yang disimpan di *warehouse data* untuk menemukan pola, korelasi, dan tren baru dengan menggunakan statistik, pembelajaran mesin, kecerdasan buatan (AI), dan teknik visualisasi data dikenal sebagai *data mining* [11]. Merupakan proses mengumpulkan dan menemukan informasi berguna dengan menggunakan matematika, statistik, AI, dan pembelajaran mesin. [12]. Berdasarkan definisi tersebut, dapat disimpulkan bahwa data mining adalah sekumpulan proses *knowledge discovery in database* (KDD) yang menggali nilai berupa informasi yang diperoleh dari basis data. Informasi yang dihasilkan diperoleh dengan mengekstraksi dan mengidentifikasi pola-pola penting dari database besar.

Perkembangan pesat data mining, baik dari segi perangkat maupun algoritma, memungkinkan implementasi data mining dalam bentuk bidang.

Kemajuan luar biasa yang terus berlanjut pada data mining didorong oleh beberapa faktor, antara lain [10]:

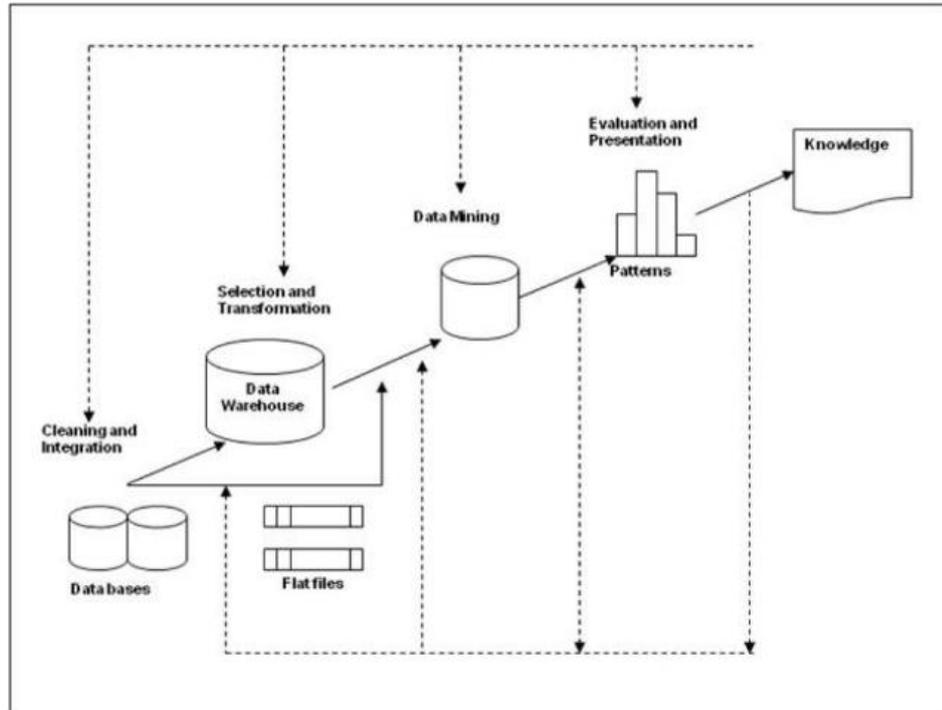
1. Mempunyai pertumbuhan yang cepat dalam kumpulan data
2. Penyimpanan data, sehingga seluruh perusahaan memiliki akses ke dalam database yang andal
3. Peningkatan akses data melalui navigasi web dan intranet
4. Adanya tekanan kompetisi bisnis untuk meningkatkan penguasaan pasar dalam globalisasi ekonomi
5. Perkembangan teknologi dari perangkat lunak untuk mendorong proses *data mining*
6. Perkembangan yang hebat dalam kemampuan komputasi dan pengembangan kapasitas media penyimpanan.

Terdapat beberapa hal penting yang perlu diperhatikan terkait dengan *data mining* diantaranya adalah sebagai berikut:

1. *Data mining* adalah suatu proses otomatis terhadap data yang sudah ada.
2. Data yang digunakan pada proses *data mining* merupakan data dalam jumlah besar
3. Tujuan *data mining* adalah mendapatkan hubungan atau pola yang mungkin memberikan indikasi yang bermanfaat.

Berikut pada Gambar 2.1 merupakan proses *data mining*





Gambar 2. 1 Proses *Data Mining*
 Sumber: Syaiful Zuhri Harahap [13]

Konsep *data mining* menurut Randi Farmana Putra, dkk [14] merupakan suatu proses kerangka kerja yang komprehensif untuk memahami dan menerapkan teknik-teknik *data mining* dengan efektif. Kerangka kerja ini terdiri dari beberapa tahap yang saling terkait dan saling mendukung, mulai dari pra-pemrosesan data hingga penggunaan hasil *data mining*. Tahap pertama dalam konsep *data mining* adalah pra-pemrosesan data. Pra-pemrosesan data melibatkan langkah-langkah penting seperti pembersihan data, integrasi data, transformasi data, dan reduksi dimensi data. Pembersihan data bertujuan untuk mengatasi masalah dimensi data. Pembersihan data bertujuan untuk mengatasi masalah data yang tidak lengkap, duplikat, atau tidak valid. Selanjutnya integrasi data digunakan untuk menggabungkan data dari berbagai sumber yang berbeda menjadi satu kesatuan yang koheren. Transformasi data melibatkan konversi data ke format yang sesuai untuk analisis *data mining*, sedangkan reduksi dimensi data dilakukan untuk mengurangi kompleksitas data dengan memilih fitur yang relevan atau mengekstraksi fitur yang lebih penting.

Setelah pra-pemrosesan data, tahap berikutnya adalah penambangan pola. Penambangan pola mencakup teknik-teknik seperti klasifikasi, pengelompokan, asosiasi, dan analisis urutan. Klasifikasi merupakan sebuah proses pengelompokan entitas ke dalam kategori atau kelas yang telah ditentukan sebelumnya berdasarkan atribut-atribut yang relevan. Pengelompokan disisi lain, melibatkan pengelompokan entitas yang serupa menjadi kelompok-kelompok yang saling berdekatan berdasarkan kesamaan atribut. Asosiasi adalah teknik yang digunakan untuk menemukan hubungan dan ketergantungan antara item-item dalam dataset, misalnya hubungan antara produk yang sering dibeli bersama. Sedangkan analisis urutan digunakan untuk mengidentifikasi pola urutan dalam data, seperti pola pembelian pelanggan dari waktu ke waktu.

Setelah tahap penambangan pola, evaluasi model dilakukan untuk mengukur kinerja dan validitas model yang telah dibangun. Beberapa metrik evaluasi yang umum digunakan termasuk akurasi, presisi, *recall*, *f1-score*, dan kurva ROC. Metrik-metrik ini membantu untuk menilai sejauh mana model dapat menghasilkan prediksi yang akurat dan relevan. Evaluasi model juga membantu dalam memahami sejauh mana model dapat menggeneralisasi pada data baru yang tidak terlihat sebelumnya.

Tahap terakhir dalam konsep *data mining* adalah interpretasi. Pada tahap ini, pola-pola atau penemuan yang ditemukan melalui penambangan pola dianalisis secara mendalam. Ahli *data mining* menganalisis hubungan, keterikatan, dan implikasi pola dalam konteks aplikasi yang relevan. Interpretasi bertujuan untuk menghasilkan pemahaman yang lebih mendalam tentang pola-pola tersebut, sehingga dapat digunakan untuk pengambilan keputusan atau Tindakan yang lebih baik.

Selain itu, hasil dari *data mining* juga harus digunakan secara efektif dalam pengambilan keputusan dan penerapan praktis. Penggunaan hasil *data mining* melibatkan penerapan pengetahuan yang dihasilkan dalam konteks yang relevan. Pengetahuan ini dapat digunakan untuk mengoptimalkan proses bisnis, mengidentifikasi peluang baru, menginformasikan keputusan bisnis, atau mengembangkan strategi yang lebih efektif.

Dalam keseluruhan konsep *data mining*, setiap tahap memiliki peran penting dalam memahami data dan menghasilkan wawasan berharga. Tahap pra-pemrosesan data memastikan kebersihan dan integritas data sebelum dilakukan analisis lebih lanjut. Penambangan pola membantu mengungkap pola-pola tersembunyi dan tren dalam data. Evaluasi model memastikan bahwa model yang dibangun memiliki kinerja yang baik dan dapat diandalkan. Interpretasi memungkinkan pemahaman mendalam tentang makna pola yang ditemukan. Dan akhirnya penggunaan hasil *data mining* menghubungkan pengetahuan yang dihasilkan dengan pengambilan keputusan dan Tindakan nyata dalam dunia bisnis. Berikut merupakan beberapa tantangan penggunaan *data mining* dalam era digital yang semakin maju.

1. Variasi dan kompleksitas data

Data mining harus menghadapi beragam jenis data seperti teks, gambar, suara, dan video, serta perbedaan dan tingginya dimensi sumber data. Variasi dan kompleksitas ini menyulitkan analisis dan pemodelan data, dan seringkali membutuhkan pendekatan algoritma yang lebih canggih serta pemrosesan yang lebih intensif.

2. Kualitas data dan pra-pemrosesan

Kualitas data yang buruk dapat signifikan mempengaruhi akurasi dan keandalan hasil *data mining*. Data yang tidak lengkap, tidak akurat, atau tidak konsisten dapat menghasilkan hasil yang tidak dapat diandalkan dalam *data mining*. Oleh karena itu, pra-pemrosesan data menjadi krusial untuk mempersiapkan data sebelum dilakukan analisis. Tantangan dalam memperbaiki kualitas data melibatkan deteksi dan penanganan nilai yang hilang, *outlier*, atau *noise* dalam data, serta menangani duplikasi atau inkonsistensi.

3. Privasi dan keamanan

Perlindungan privasi dan keamanan data menjadi penting dalam *data mining*, terutama ketika berurusan dengan *data sensitive*. Memastikan privasi data dan mencegah akses yang tidak sah merupakan tantangan yang krusial. Perlindungan privasi dan keamanan data dapat membatasi akses dan

penggunaan data, serta memerlukan pertimbangan tambahan dalam teknik *data mining* yang digunakan, seperti teknik penghapusan rincian identitas atau pengenkripsian data.

4. Interpretabilitas dan kemampuan interpretasi

Kemampuan untuk menginterpretasi hasil *data mining* dengan cara yang dapat dimengerti adalah kunci untuk pengambilan keputusan dan adopsi teknik *data mining*. Kompleksitas model dan algoritma yang digunakan dalam *data mining* seringkali sulit dipahami oleh pengguna akhir. Oleh karena itu, penting untuk mengembangkan cara yang efektif.

2.1.5 Clustering

Clustering merupakan teknik penggalian data dalam *data mining*. Fungsi dari *clustering* adalah untuk mengelompokkan data berdasarkan tingkat kemiripan dari data yang dapat pada kelompok tertentu dan meminimalkan kemiripan pada kelompok lain. *Clustering* dapat diartikan sebagai proses untuk melakukan pengamatan, mengelompokkan *record* dan kelas objek yang terdapat kemiripan [15].

Pengklusteran berbeda dengan klasifikasi yang tidak adanya variabel target dalam pengklusteran. Dalam pengklusteran tidak dilakukan proses klasifikasi, estimasi, atau prediksi nilai dari variabel target. Tetapi dalam algoritma pengklusteran terdapat pembagian data ke dalam masing-masing kelompok yang berdasarkan kemiripan (homogen). Berikut merupakan beberapa contoh pengklusteran dalam dunia bisnis maupun penelitian adalah [10]:

1. Mendapatkan kelompok-kelompok pelanggan untuk target pemasaran dari suatu produk bagi perusahaan yang tidak memiliki dana pemasaran yang besar.
2. Untuk tujuan audit akuntansi, yaitu melakukan pemisahan terhadap perilaku finansial dalam baik dan mencurigakan.
3. Melakukan pengklusteran terhadap ekspresi dari sebuah gen yang berfungsi untuk mendapatkan kemiripan perilaku dari gen dalam jumlah besar.

Algoritma *clustering* membangun sebuah model dengan melakukan serangkaian pengulangan dan berhenti ketika model tersebut telah terpusat dan batasan dari segmentasi telah stabil. Hasil dari *clustering* yang bagus tergantung dengan ukuran kesamaan dan metode yang digunakan. Pendekatan dalam *cluster* berdasarkan saran dari Fraley dan Raftery, membagi metode pengelompokan menjadi dua kelompok utama yaitu metode hirarki dan metode partisi [16].

1. Metode hirarki yaitu metode yang membentuk *cluster* dengan mempartisi secara berulang-ulang dari atas ke bawah atau sebaliknya. Hasil dari metode hirarki berupa dendogram yang mewakili kelompok objek dan tingkat kesamaan di mana terdapat perubahan pengelompokan. Sebuah pengelompokan objek data diperoleh dengan memotong dendogram pada tingkat kemiripan yang diinginkan.
2. Metode partisi yaitu metode yang membuat inisial partisi k di awal dimana parameter k merupakan jumlah partisi untuk membentuk. Kemudian secara *iterative* menggunakan teknik relokasi dengan mencoba berulang-ulang memindahkan objek dari satu kelompok ke kelompok lain untuk mendapatkan partisi yang optimal.

Beberapa model algoritma pada metode *clustering* yang digunakan secara umum [14].

1. *K-means*

Algoritma ini merupakan salah satu metode *clustering* yang paling populer. Algoritma *k-means* mengelompokkan data ke dalam k kelompok berdasarkan jarak *Euclidean* antara data dan pusat *cluster*.

2. *Hierarchical clustering*

Metode *hierarchical clustering* yang menghasilkan dendogram yang merepresentasikan struktur hierarki dari kelompok data. Metode ini dapat berupa *agglomerative* (pemisahan) atau *divisive* (pembagian) dalam pendekatan *clustering*.

3. DBSCAN

Algoritma DBSCAN (*density-based spatial clustering of applications with noise*) yang mengelompokkan data berdasarkan kepadatan data dalam

ruang. Algoritma ini dapat mengidentifikasi kelompok yang memiliki kepadatan yang tinggi dan menemukan outlier dalam data.

Parameter yang perlu diperhatikan dalam algoritma *clustering* antara lain meliputi jumlah *cluster* yang diinginkan (k), metode pengukuran jarak, dan metode inialisasi pusat *cluster*. Pentingnya pemilihan parameter yang tepat dapat mempengaruhi hasil *clustering* yang akurat dan relevan.

2.1.6 Model *Recency, Frequency, dan Monetary* (RFM)

RFM Model yaitu digunakan untuk mengidentifikasi perilaku atau karakteristik pelanggan [4]. Model RFM didasarkan pada tiga variabel: *recency* pembelian terakhir, *frequency* transaksi, dan *monetary* yang berasal dari jumlah transaksi setiap pelanggan. Dalam penelitian ini, terdapat tiga atribut model RFM dianalisis yaitu adalah *Recency* (R), *Frequency* (F), lalu *Monetary* (M). Kebaruan pembelian terakhir (R) adalah *Recency*, ialah jarak antar ke waktu pembelian terakhir dan waktu yang saat ini, yang semakin pendek jarak semakin besar R. *Frequency* pembelian (F) adalah banyaknya transaksi selama periode waktu tertentu, misalnya satu tahun dua kali atau satu bulan dua kali. Semakin banyak *frequency*, semakin besar F. Nilai *monetary* dari pembelian (M). M adalah *monetary*, yang menunjukkan uang yang dipakai belanja selama periode waktu tertentu. Semakin besar pengeluaran uang, maka semakin besar M [14].

2.1.7 *K-Means Clustering*

K-Means Clustering Algorithm merupakan algoritma *clustering* data nonhirarki yang mengelompokkan data menjadi satu atau lebih *cluster* atau kelompok. *Clustering* menggabungkan data dengan atribut yang sama dan kemudian menggabungkannya dengan atribut yang berbeda sehingga ada sedikit variasi antara *cluster* atau kelompok data yang berbeda, maka itu data yang terdapat dalam satu *cluster* atau kelompok memiliki tingkat variasi yang rendah. [17]. *Clustering* merupakan metode algoritma dalam materi *data mining* yang bertujuan sebagai pengelompokan beberapa objek dalam *cluster-cluster* dengan karakteristik yang sama, pengelompokkan *item* data kedalam jumlah *cluster*

kecil. Cluster yang sama dapat disesuaikan dengan objek-objek yang berbeda [18].

Langkah-langkah *clustering K-means* yaitu [18]:

1. Menetapkan jumlah *cluster* yang diinginkan untuk dibentuk, dengan nilai K adalah jumlah *cluster* atau banyaknya *cluster*.
2. Selanjutnya dapat ditentukan *centroid* awal, atau pusat *cluster* awal, dengan mengambil *centroid* awal secara acak dari data saat ini, dan jumlah *centroid* awal sama dengan jumlah *cluster*.
3. Sesudah menentukan *centroid* awal, data akan menemukan *centroid* yang dekat dengan menghitung jarak masing-masing data ke *centroid* tersebut memakai rumus korelasi antar dua objek yaitu rumus *Euclidean Distance*.
4. Selanjutnya setelah jarak data ke *centroid* dihitung, lalu tahap selanjutnya yaitu data dikelompokkan berdasar minimum jaraknya. Sebuah data menjadi bagian anggota dari sebuah *cluster* yang mempunyai jarak terdekat (terkecil) dari pusat *cluster*-nya.
5. Tahap berikutnya yaitu dengan dicari *centroid* baru berdasarkan *membership* dari masing-masing *cluster* dimana masing-masing *cluster* dihitung rata-rata datanya.
6. Proses kembali pada tahap 3.
7. Selanjutnya, perulangan akan *stop* jika tidak ada data yang pindah.

Untuk tentukan nilai pusat (*centroid*) pada tahap iterasi digunakan rumus seperti berikut :

$$V_{ij} = \frac{1}{N_i} = \sum_{k=0}^{N_i} X_{ki} \quad (2.1)$$

Keterangan:

V_{ij} = *centroid* rata-rata *cluster* ke-i untuk *variable* k-j

N_i = jumlah anggota *cluster* ke-i

i, k = indeks dari *cluster*

j = indeks dari *variable*

X_{kj} = nilai data ke-k *variable* ke-j

Selain itu untuk menentukan korelasi antar dua obyek yaitu dengan menggunakan rumus *Euclidean Distance* berikut:

$$d_{Euclidean}(x,y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2.2)$$

Keterangan:

$d(x,y)$ = jarak data ke x ke pusat *cluster* y

X_i = data ke-i pada atribut data ke n

Y_j = data ke-j pada atribut data ke n.

Metode *k-means* memiliki kelebihan yaitu dapat mengelompokkan data yang besar dengan cepat, mengimplementasikan dengan mudah untuk dilanjutkan dan dijalankan, prinsip yang digunakan sederhana sehingga dapat dijelaskan dalam non-statistik, dan sangat umum digunakan. Sedangkan kelemahan dari metode *k-means* adalah jumlah *cluster* sebanyak k harus ditentukan sebelum dilakukan perhitungan, kontribusi dari atribut dalam pengelompokan tidak diketahui karena semua atribut dianggap mempunyai bobot yang sama, dan tidak pernah mengetahui *real cluster* dengan menggunakan data yang sama, tetapi dimasukkan dengan cara berbeda, mungkin dapat diproduksi *cluster* yang berbeda jika jumlah datanya sedikit.

2.1.8 Fuzzy C-Means Clustering

Fuzzy C-Means clustering algorithm merupakan algoritma *unsupervised* yang kuat untuk analisis data serta konstruksi model serta objek, algoritma *Fuzzy C-Means* sangat efektif itu membuat model dan objek antar beberapa kelas tidak harus untuk dimiliki oleh satu kelas sebaliknya, itu menunjukkan keanggotaan objek dengan derajat keanggotaan antara 0 sampai 1 [19]. Pengelompokan data menjadi sejumlah besar sampel yang diubah menjadi sejumlah kecil perwakilan prototipe atau *cluster* dikenal sebagai *clustering*. Dalam kompresi data, kelompokan membagi sejumlah besar sampel menjadi representasi kelas atau *cluster* yang lebih kecil, sehingga data dalam satu kelompok sangat berbeda dari kelompok lain. [20].

Langkah-langkah melakukan *clustering fuzzy c-means* [21]:

1. Menentukan matriks c berukuran $n \times p$, dengan n = jumlah data yang akan di*cluster*, dan p = parameter/jumlah variabel. Kemudian jumlah

cluster yang akan dibentuk = c (2), serta untuk eksponen pemboot, m . Dan iterasi awal, $t = 1$

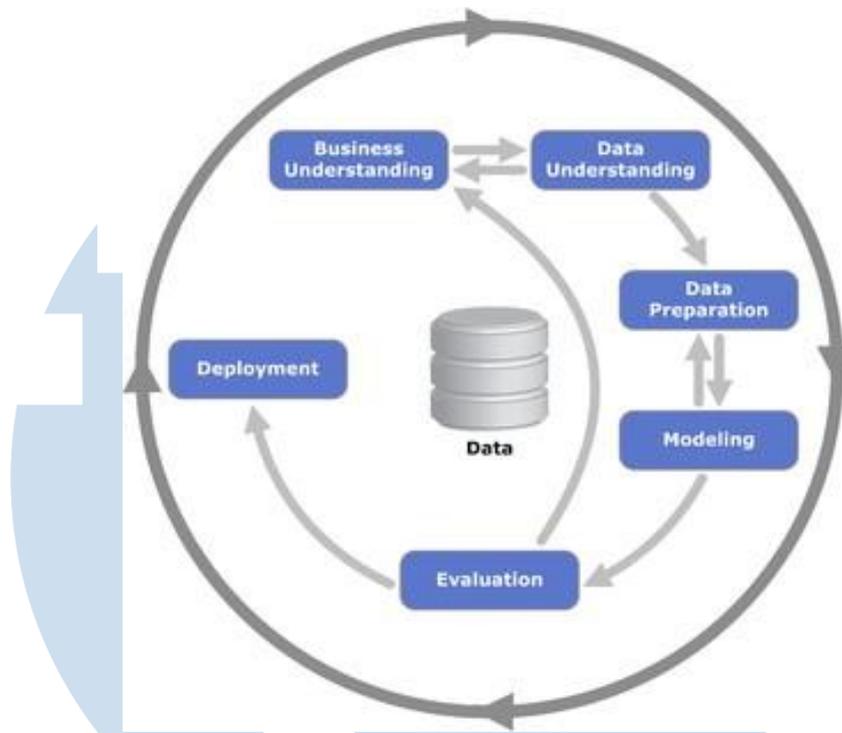
2. Setelah itu bentuk matriks keanggotaan awal dengan bilangan acak untuk $i = 1, 2, 3, \dots, n$; $j = 1, 2, 3, \dots, c$.
3. Kemudian cari nilai pusat klaster untuk setiap klaster dengan $i = 1, 2, 3, \dots, c$ dan $j = 1, 2, 3, \dots, p$.
4. Selanjutnya derajat keanggotaan seluruh data masing-masing klaster diperbaiki, dengan catatan menghitung fungsi objek (formula (1)) terlebih dahulu, yang kemudian menggunakan persamaan sebagai berikut:

$$u_{ij} = \frac{[\sum_{j=1}^p (x_{j-})^2]_{m-1}^{-1}}{\sum_{i=1}^c [\sum_{j=1}^p (x_{j-})^2]_{m-1}^{-1}} \quad (2.3)$$

5. Tentukan kriteria berhenti atau konvergen dari perulangan, yaitu dengan melihat perubahan nilai fungsi objektif iterasi saat dengan iterasi sebelumnya atau iterasi kurang dari maksimum iterasi.

2.1.9 CRISP-DM

Pada tahun 1996 analisis dari beberapa *industry* seperti DiaimlerChrysler SPSS dan NCR mengembangkan *cross industry standard process for data mining* (CRISPM-DM). CRISP-DM adalah standar pemrosesan *data mining* yang telah dikembangkan. Metode ini memungkinkan data saat ini diproses melewati fase terstruktur secara jelas dan terdefinisi, dan telah banyak digunakan untuk memecahkan masalah [10]. 6 (enam) tahapan dapat dilihat pada Gambar 2.2. keseluruhan tahap berurutan yang ada tersebut bersifat adaptif. Fase berikutnya dalam urutan bergantung kepada keluaran dari fase sebelumnya. Hubungan penting antar fase yang digambarkan oleh panah. Misalkan, jika proses berada fase *modeling*, dengan melihat perilaku serta karakteristik dari model memungkinkan proses harus kembali kepada fase *data preparation*.



Gambar 2.2 Tahapan CRISP-DM
Sumber: Yogasetya Suhanda [22]

Berdasarkan Gambar 2.1 dapat dijelaskan tahapan-tahapan CRISP-DM sebagai berikut [23]:

1. *Business Understanding*

Business understanding atau pemahaman bisnis adalah proses menetapkan apa yang dituju oleh perusahaan, memahami situasi dan kondisi saat penelitian, dan menetapkan tujuan penelitian untuk masalah yang diselesaikan melalui data mining.

2. *Data Understanding*

Data understanding atau pemahaman data yaitu proses melakukan pengecekan terhadap data yang digunakan, mengumpulkan data awal, dan mengidentifikasi kualitas data. Dalam pemahaman data, data yang digunakan akan melalui proses untuk mendeskripsikan setiap fiturnya.

3. *Data Preparation*

Data preparation atau persiapan data merupakan proses yang dilakukan setelah data dikumpulkan. Pada tahap ini, data akan diidentifikasi, dipilih,

dibersihkan, dan diubah.

4. *Modelling*

Modelling atau pemodelan yaitu fase implementasi algoritma yang akan digunakan untuk pencarian, identifikasi, dan pembuatan pola yang akan diterapkan pada data penelitian.

5. *Evaluation*

Evaluation atau evaluasi merupakan proses dimana melakukan pengukuran hasil evaluasi dari model yang telah diimplementasikan pada tahap modeling sebelumnya. Hasil evaluasi tersebut menggambarkan proses data mining yang dilakukan dan menentukan model mana yang paling cocok untuk digunakan.

6. *Deployment*

Penyebaran atau *deployment* merupakan proses menggunakan model yang dihasilkan sebelumnya. Ada dua jenis kegiatan penyebaran yaitu perencanaan dan pengawasan hasil penyebaran. Tugas terakhir adalah menyelesaikan tugas penutup dengan membuat laporan akhir dan melakukan tinjauan proyek. Kedua kegiatan ini dapat dilakukan secara bersamaan atau hanya menyelesaikan salah satunya.

2.1.10 DBI (*Davies Bouldin Index*)

Davies bouldin index adalah sebuah evaluasi yang memiliki rencana untuk evaluasi *cluster internal*. Ini memungkinkan untuk mengetahui seberapa baik hasil *cluster* dari kuantitas antar data. Metode pengukuran DBI melibatkan mengurangi jarak intra *cluster* dan meningkatkan jarak inter *cluster*. Nilai DBI yang lebih rendah menunjukkan skema *cluster* terbaik [23]. Adapun perhitungan *Davies Bouldin Index* (DBI) dapat memakai persamaan 2.4 dan juga persamaan 2.5.

$$R_{j,k} = \frac{MAE_j + MAE_k}{d(c_j, c_k)} \quad (2.4)$$

$$DBI = \frac{1}{M} \sum_{j=1}^M \max_{j \neq k} R_{j,k} \quad (2.5)$$

2.1.11 Python

Python merupakan salah satu bahasa pemrograman yang banyak digunakan oleh perusahaan besar maupun para *developer* untuk mengembangkan berbagai macam aplikasi berbasis *desktop*, *web* dan *mobile*. *Python* diciptakan oleh Guido van Rossum di Belanda pada tahun 1990 dan namanya diambil dari acara televisi kesukaan Guido *Monty Python's Flying Circus*. Van Rossum mengembangkan *Python* sebagai hobi, kemudian *Python* menjadi bahasa pemrograman yang dipakai secara luas dalam industri dan pendidikan karena sederhana, ringkas, sintak intuitif dan memiliki pustaka yang luas [24].

Python tergolong bahasa pemrograman yang berlevel tinggi, namun *python* dirancang sedemikian rupa sehingga mudah dipelajari dan dipahami. Kelebihan program *python* ini yaitu mudah untuk dipelajari, dapat menjalankan program dengan banyak fungsi kompleks didalamnya dengan mudah, lebih sedikit menggunakan kode, dan bahkan mampu mengubah program dengan tingkat kerumitan tinggi menjadi mudah [25]. Berikut merupakan beberapa kekuatan dan kelemahan *Python* [26].

1. Kekuatan

Menurut penelitian dan survei, *Python* adalah bahasa terpenting kelima serta bahasa paling populer untuk pembelajaran mesin dan ilmu data. Karena kekuatan berikut yang dimiliki *Python*.

- a. *Syntax* bahasa *python* telah disimplikasi sehingga cenderung mudah dimengerti. *Python* banyak memberi fokus yang tertanam pada *natural language* sehingga tidak rumit bila dibandingkan bahasa pemrograman lain dan karena ini bahasa pemrograman *python* lebih mudah ditulis serta pengekseskuan programnya lebih cepat.
- b. Bahasa pemrograman *python* dinamis karena bahasa pemrograman *python* dapat menggunakan berbagai paradigma dalam melakukan proses *coding*, dari mulai secara *functional*, *procedural*, maupun OOP.

- c. Bahasa *python* dianggap *powerful* karena dapat diaplikasikan multiplatform diantara penggunaannya seperti untuk *web development*, *mobile apps*, *desktop apps*, hingga pada *devops*.
 - d. Dukungan komunitas *open source* sebagai bahasa pemrograman *open source*, *python* didukung oleh komunitas pengembang yang sangat besar. Karena ini, *bug* mudah diperbaiki oleh komunitas *python*. Karakteristik ini membuat *python* sangat kuat dan adaptif.
 - e. Skalabilitas *python* adalah bahasa pemrograman yang skalabel karena menyediakan struktur yang lebih baik untuk mendukung program besar daripada skrip *shell*.
2. Kelemahan
- a. Kurang ideal digunakan untuk melakukan proses *memory intensive task* dikarenakan memakan memori yang cenderung tinggi serta kurang adanya dukungan dari multiprosesor sehingga terkadang terjadi pembatas dalam penulisan kode.
 - b. *Python* kurang populer dibandingkan *java* dan *kotlin* dalam pengembangan aplikasi *mobile* karena terdapat beberapa batasan desain serta memiliki kecepatan yang cenderung lebih rendah dari bahasa pemrograman lainnya.

Dalam pemrograman *python*, dapat digunakan untuk berbagai macam modul yang kemudian disatukan ke dalam suatu folder atau disebut juga *package*. Berbagai *package* kemudian dapat dihimpun ke dalam sebuah kesatuan bernama *library*. *Python* memiliki *library* yang berjumlah lebih dari 140.000 dan akan terus bertambah. Adapun *library* pada *python* yang dikembangkan bersifat *open-source project* sehingga kita dapat dengan mudah mengaksesnya secara gratis. Menggunakan *library* yang terdapat pada *python*, dapat dihasilkan kode-kode yang secara sistematis dianggap lebih efisien serta dapat menghemat waktu karena tidak wajib menulis seluruh skrip yang diperlukan. Kelebihan lainnya adalah *library* yang terdapat pada *python* bersifat *reusable* sehingga pengguna dapat menggunakannya berulang dan berkali-kali, kapan saja dan dimana saja secara praktis [27].

2.1.12 *Pandas*

Pandas adalah pustaka *open source* untuk *Python* untuk spesialisasi analisis data dan timbul dari kebutuhan untuk memiliki pustaka yang terspesifikasi dalam menganalisa data dengan menyediakan cara tersimpel untuk alat *data processing*, *data extraction*, dan *data manipulation* [28].

Pandas merupakan *library Python* untuk mengolah dan menganalisis data-data yang terstruktur. Untuk mengambil fungsi *pandas*, dilakukan dengan menuliskan “*import pandas as pd*” [29]. Dalam *pandas* terdapat dua tipe data yaitu *Series* dan *DataFrame*. *DataFrame* merupakan sebuah data yang diatur dalam baris dan kolom, sedangkan *Series* adalah satu kolom pada *DataFrame* [30]. Dengan bantuan *pandas*, dalam pemrosesan data dapat menyelesaikan lima langkah seperti memuat, mempersiapkan, memanipulasi, model, menganalisa. Seluruh representasi data dalam *pandas* dilakukan dengan bantuan tiga struktur data berikut [26]:

1. *Series*

Pada dasarnya adalah *ndarray* satu dimensi dengan label sumbu yang berarti seperti *array* sederhana dengan data homogen. Misalnya deret berikut adalah himpunan bilangan bulat 1,5,10,15,24,25.

2. *Data frame*

Data frame adalah struktur data yang paling berguna dan digunakan untuk hampir semua jenis representasi dan manipulasi data di *pandas*. Ini pada dasarnya adalah struktur data dua dimensi yang dapat berisi *data heterogeny*. Umumnya data tabular direpresentasikan dengan menggunakan bingkai data. Sebagai contoh data siswa yang memiliki nama, *roll number*, usia, dan jenis kelamin.

3. *Panel*

Panel adalah struktur data 3 dimensi yang berisi *data heterogeny*. Sangat sulit untuk merepresentasikan panel dalam representasi grafis. Tetapi dapat diilustrasikan sebagai wadah *dataframe*.

2.1.13 Numpy

Numerical Python atau biasa disebut *NumPy* merupakan sebuah *library Python* untuk *scientific computing* terutama pada analisis data. Pada *library* ini sudah lebih baik dibandingkan *library Python* yang standar karena *library Python* yang standar lebih sederhana atau tidak memadai untuk analisis data. *Numpy* ini biasa digunakan pada perhitungan *array* multidimensi atau *array* dengan jumlah besar [30].

Kunci dalam tipe data *Numpy* yaitu *array n-dimensi (ndarray)*. *Ndarray* disebut *N-dimensi* karena mereka dapat memiliki sejumlah dimensi. *Array* satu dimensi kira-kira sama dengan *list Python*. Pada *ndarray* ini mirip dengan *list Python* tapi memiliki keunggulan yaitu teknik manipulasi data. *Array* ini harus bertipe homogen atau semua *item* pada *array* harus memiliki tipe yang sama. [30]. Dengan menggunakan *numpy*, kita dapat melakukan operasi penting seperti berikut

4. Operasi matematika dan logika pada *array*.
5. Transformasi *fourier*
6. Operasi yang terkait dengan aljabar linier

Berikut merupakan contoh pembuatan *array* menggunakan *numpy* [27].

```
#pembuatan array menggunakan numpy
#bernama arr dan berisi 10,11,13,14

import numpy as np
contoh_array = np.array([10,11,13,14])
print (contoh_array)
```

```
Import numpy as np
Panjang = np.random.uniform(low = 10, high = 50, size
= 300)
Lebar = np.random.uniform(low = 5, high = 10, size =
100)
```

```
Tinggi = np.random.uniform(low = 5, high = 10, size  
= 200)
```

Selain membuat dan memanipulasi *array*, *numpy* juga dapat digunakan untuk melakukan *generate* angka *random*. Kode diatas melakukukan perintah *generate random number* memiliki distribusi seragam. Pada distribusi seragam, *low*, dan *high* merupakan batas angka yang akan degenerate, sementara *size* memiliki arti jumlah data yang degenerate (200 data). Beberapa perintah *generate random number* dengan *numpy*:

```
#distribusi normal  
Numpy.random.normal(loc=0.0, scale=2.0, size=None)  
#distribusi seragam  
Numpy.random.uniform(loc=0.0, high=5.0, size=None)  
#sequence number  
Numpy.arange([start, ]  
stop, [step, ]dtype=None, *, like=None)
```

2.1.14 Library Matplotlib

Matplotlib adalah salah satu *library* yang banyak digunakan untuk keperluan visualisasi data multiplatform yang dibangun di atas *array numpy*. Pada awal peluncurannya, *library* ini dirancang agar dapat menciptakan sebuah plot grafik yang sesuai dengan artikel ilmiah. Pada umumnya *matplotlib* bekerja pada skrip *python*, *IPython shell*, dan server aplikasi web. Namun ternyata *library* ini juga dapat berjalan pada *toolkit graphical user interface* (GUI) [27].

Output yang dihasilkan *library* ini berupa visualisasi gambar grafik yang memiliki sumbu-sumbu. Sumbu ini kemudian disebut sebagai sumbu kartesius yang terbagi ke dalam sumbu horizontal (x) serta sumbu vertical (y). grafik *matplotlib* juga dapat disesuaikan dengan pemberian label serta judul. Adapun data-data yang tergambar di dalam grafik direpresentasikan menjadi warna-

warna serta *glyphs* mirip seperti marker (seperti lingkaran) atau garis-garis serta bentuk polygon.

Untuk dapat menggunakan *matplotlib* kita dapat menginstallnya melalui *command prompt* dengan perintah.

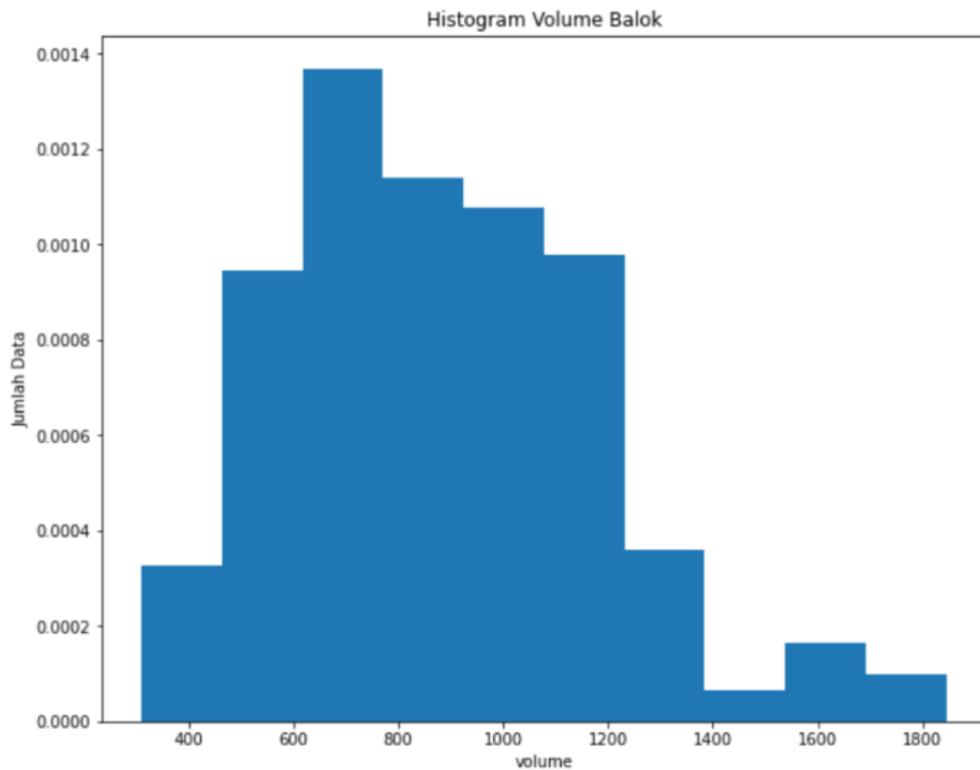
```
Pip install matplotlib
```

Berikut merupakan contoh penggunaan *matplotlib* untuk membuat sebuah histogram.

```
#membuat histogram dari kolom volume
Nilai_volume = df["volume"]

#membuat frame
Fig, ax = plt.subplots(figsize = (10,8))
Ax.set_xlabel('volume')
Ax.set_ylabel('jumlah data')
Ax.set_title('histogram volume balok')
Plt.hist(x=nilai_volume, bins=10, density=true)
Plt.show()
```

Dari kode diatas, telah dilakukan pengimporan *library matplotlib* dengan menggunakan variabel *plt*. Untuk membuat *frame* dapat menggunakan perintah *plt.subplots* yang dimasukkan kedalam variabel *ax*, dengan diberikan nama sumbu *x* dan *y* serta judul dari histogram menggunakan perintah *set_(x,y)label*. Setelah *frame* terbentuk maka dapat digunakan untuk membuat histogram dengan perintah *plt.hist*. sementara *plt.show* berfungsi untuk menampilkan histogram yang sudah dibuat [27]. Berikut pada Gambar 2.3 merupakan contoh hasil histogram menggunakan *matplotlib*.



Gambar 2. 3 Contoh Histogram Matplotlib

2.1.15 Scikit Learn

Scikit-learn merupakan salah satu modul *Python* yang merupakan penyedia berbagai macam algoritma *machine learning*. *Scikit-Learn* tersedia dalam bentuk *library Python*. *Scikit-learn* menggunakan *task-oriented interface* yang konsisten sehingga mudah dalam melakukan perbandingan antar metode [31].

Penggunaan *scikit-learn* berfokus pada pemodelan data. Namun keunggulan dari *library* tersebut adalah dapat diintegrasikan dengan berbagai *library*, contohnya untuk visualisasi data dapat digabungkan dengan *library matplotlib*, kemudian untuk proses analisis *dataframe* yang akan digunakan untuk model dapat diintegrasikan dengan *library pandas*. Berikut merupakan beberapa pilihan algoritma yang dapat digunakan pada *scikit-learn* [27].

1. *Regresi* : berfungsi untuk membuat model linear maupun model non-linear.

2. *Classification* : melakukan proses klasifikasi *unsupervised learning*.
3. *Decision tree* : pembuatan pohon keputusan yang digunakan untuk proses klasifikasi serta regresi.
4. SVM : untuk mempelajari mengenai batasan dari suatu keputusan.
5. *Naïve bayes* : Termasuk ke dalam pemodelan *probabilistic* yang umumnya digunakan pada klasifikasi
6. Metode *ensemble* : Dapat melakukan metode *ensemble* dengan *bagging, boosting, voting*, hingga *random forest*
7. Manipulasi fitur : Merupakan proses pengurangan dari suatu dimensi, pemilihan serta analisis pada fitur *machine learning*
8. Deteksi *outlier* : Memiliki peran dalam pendeteksian *data outlier* sehingga mengurangi *noise*

2.2 Penelitian Terdahulu

Berikut pada Tabel 2.2 merupakan beberapa penelitian terdahulu yang digunakan dalam penelitian ini

Tabel 2.2 Penelitian Terdahulu

No	Jurnal/ Volume/Tahun	Judul Penelitian	Penulis	Hasil	Kesimpulan
1	Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer/3/2019 [4].	Analisis Segmentasi Pelanggan Dengan RFM Model Pada Pt. Arthamas Citra Mandiri Menggunakan Metode	Wiratama Ahsani Taqwim, Nanang Yudi Setiawan, Fitra A. Bachtiar.	Hasil dari penelitian dengan menggunakan algoritma <i>Fuzzy C-Means</i> , menghasilkan nilai <i>within sum of squares by</i>	Hasil dari model <i>recency</i> , <i>frekuensi</i> , dan <i>monetary</i> dapat digambarkan dengan karakteristik pelanggan individu yang

No	Jurnal/ Volume/Tahun	Judul Penelitian	Penulis	Hasil	Kesimpulan
		Fuzzy C-Means Clustering.		<i>cluster</i> dari tiga <i>cluster</i> adalah 58,36%, dan nilai untuk <i>cluster</i> empat <i>cluster</i> adalah 72,95%.	melakukan transaksi dengan PT. Arthamas Citra Mandiri. Ini dapat diperhitungkan untuk setiap model.
2	Jurnal Optimalisasi/ 8/2022, [32]	Segmentasi Pelanggan Distributor Daging Sapi Menggunakan Pendekatan Recency, Frequency, Monetary (RFM) dan Fuzzy C-Means Clustering	Paduloh, Murwan Widyantoro, Jasan Supratman.	Hasil penelitian ini yaitu pelanggan dibagi menjadi tiga pemilihan berdasarkan data pelanggan selama tiga tahun. <i>cluster</i> 2 terdiri dari kelompok pelanggan yang potensial atau baik, sedangkan <i>cluster</i> 3 terdiri dari kelompok pelanggan yang kurang loyal terhadap perusahaan.	Hasil pengolahan data menunjukkan bahwa distributor daging sapi membagi produknya menjadi tiga <i>cluster</i> . Dari 1.430 pelanggan selama tiga tahun, 544 anggota, atau sekitar 38% dari total anggota, masuk ke dalam <i>cluster</i>

No	Jurnal/ Volume/Tahun	Judul Penelitian	Penulis	Hasil	Kesimpulan
					1, hasil dari uji validasi yang dilakukan menggunakan <i>Scatterplot SEE</i> dan <i>Scatterplot DB</i> .
3	Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer/3/2019, [33]	Segmentasi Pelanggan Menggunakan Metode Fuzzy C-Means Clustering Berdasarkan LRFM Model Pada Toko Sepatu (Studi Kasus: Ride Inc Kota Malang)	Muhammad Taufik Dharmawan, Nanang Yudi Setiawan, Fitra Abdurrachman Bachtiar	Hasil penelitian menggunakan metode <i>fuzzy c-means clustering</i> dalam segmentasi pelanggan berdasarkan model LRFM adalah dua <i>cluster</i> dan tiga <i>cluster</i> berdasarkan <i>elbow method</i>	Visualisasi <i>dashboard</i> menjadi <i>output</i> penelitian berdasarkan nilai LRFM yang diberikan ke Ride Inc. hasil rata-rata dari uji <i>usability</i> dari 2 responden adalah 65 yang berarti pihak Ride Inc. menerima visualisasi <i>dashboard</i> tersebut

No	Jurnal/ Volume/Tahun	Judul Penelitian	Penulis	Hasil	Kesimpulan
4	Techno.CO M/19/2020, [34]	Implementasi Fuzzy C- Means Clustering dalam Pengelompo- kan UKM Di Kabupaten Rokan Hulu	Erni Rouza, Luth Fimawahib	Dengan implementasi <i>fuzzy c-means clustering</i> didapatkan pengelompokan jenis UKM dicluster menjadi 3 jenis, yaitu usaha menengah, usaha kecil, dan usaha mikro	Penelitian ini menghasilkan tingkat akurasi yang tinggi sebesar 80.90% yang diperoleh dari proses menggunakan metode <i>fuzzy c- means</i>
5	Sebatik/26/ 2022, [35]	Analisis segmentasi pelanggan menggunakan K-means clustering studi kasus aplikasi Alfagift	Satria Ardi Perdana, Sara Famayla Florentin , dan Agus Santoso.	Hasil dari kelompok pelanggan menjadi tiga <i>cluster, cluster</i> pertama berjumlah 7.219, <i>cluster</i> kedua sebanyak 6.902, dan <i>cluster</i> ketiga sebanyak 5.371.	Segmentasi pelanggan menggunakan metode <i>clustering</i> <i>K-Means</i> untuk <i>clustering</i> pelanggan. <i>Clustering</i> pelanggan ditentukan dari tiga kriteria yaitu umur, jenis kelamin, frekuensi

No	Jurnal/ Volume/Tahun	Judul Penelitian	Penulis	Hasil	Kesimpulan
					pesanan, tipe pembayaran dan kota pembelian barang selama satu bulan.
6	Jurnal ilmiah teknologi informasi terapan/8/2022, [36]	Analisis Segmentasi Pelanggan Bank Berdasarkan Pengambilan Kredit Dengan Kmeans Clustering	Nuril Huda Ahsina, Fety Fatimah, Fitri Rachmawati	Hasil dari <i>k-means clustering</i> dari 1000 data yang didapat jumlah dari <i>cluster</i> 1 dengan jumlah 286 klien dengan presentase 28,6%, <i>cluster</i> 2 dengan jumlah 130 klien dengan presentase 13%, dan <i>cluster</i> 3 dengan jumlah terbesar yaitu 542 klien dengan presentase 54,2%.	Hasil dari <i>clustering</i> dapat mempengaruhi nilai <i>centroid</i> awal dan jumlah data yang dipakai, perbedaan data pusat <i>centroid</i> awal dapat mempengaruhi hasil dari <i>centroid</i> akhir.

No	Jurnal/ Volume/Tahun	Judul Penelitian	Penulis	Hasil	Kesimpulan
7	Jurnal Informatika/ 7/2019, [37]	Clustering Data Penjualan pada Toko Perlengkapan Outdoor Menggunakan Metode K-Means	Fintri Indriyani, Eni Irfiani	Dengan menggunakan <i>k-means</i> didapatkan hasil akhir berupa tiga <i>cluster</i> , dimana terdapat 2 jenis barang paling laris, 8 jenis barang cukup laris, dan 18 jenis barang yang kurang laris	Data yang dijadikan rujukan bagi manajemen untuk mengatur stok barang agar toko tidak mengecewakan pelanggan karena barang yang ingin dibeli tidak tersedia
8	Prosiding Seminar Nasional komputer/1/2021 [38]	Analisis <i>Clustering</i> Pada Pengguna <i>Brand Hp</i> Menggunakan Metode <i>K-Means</i>	Indah Nuryani, Dedi Darwis	Dengan menggunakan metode <i>k-means</i> dihasilkan 3 <i>cluster</i> yaitu sangat baik, baik, dan cukup	Terdapat 120 mahasiswa memiliki HP dengan kategori sangat baik, 148 memiliki HP dengan kategori baik, dan 54 mahasiswa memiliki HP dengan kategori cukup
9	Jurnal Informatika	Penerapan Data Mining	Haditsah Annur	Dengan dilakukan	Kelompok barang dengan

No	Jurnal/ Volume/Tahun	Judul Penelitian	Penulis	Hasil	Kesimpulan
	UPGRIS/5/2019, [39]	Menentukan Strategi Penjualan Variasi Mobil Menggunakan Metode K-Means Clustering (Studi Kasus Toko Luxor Variasi Gorontalo)		<i>clustering</i> menggunakan metode <i>k-means</i> untuk menentukan strategi penjualan variasi mobil didapatkan 3 <i>cluster</i>	harga yang rendah dan diminati konsumen yaitu pada <i>cluster</i> 1, <i>cluster</i> 2 kelompok barang dengan harga yang sedang dan diminati konsumen, sedangkan <i>cluster</i> 3 kelompok barang dengan harga tinggi dan tidak terlalu sering dibeli konsumen
10	Ultima Computing: Jurnal Sistem Komputer/11/2020, [40]	<i>K-Means Clustering</i> Video Trending di Youtube Amerika Serikat	Kevin Wijaya, Raymond Sunardi Oetama	Hasil dari penelitian ini terbentuk 3 <i>cluster</i> menggunakan algoritma <i>K-Means</i> dengan ciri-cirinya	Pola dan wawasan yang ditemukan dalam penelitian ini dapat bermanfaat bagi calon

No	Jurnal/ Volume/Tahun	Judul Penelitian	Penulis	Hasil	Kesimpulan
				masing-masing. <i>cluster</i> dengan <i>view, likes, dan</i> <i>dislikes</i>	pembuat video yang ingin meraih kesuksesan sebagai Youtuber

Berdasarkan penelitian terdahulu yang telah dijelaskan pada Tabel 2.2, penelitian ini mengacu pada penelitian yang dilakukan oleh Wiratama Ahsani Taqwim, Nanang Yudi Setiawan, Fitra A. Bachtiar [4] yang menerapkan metode *Fuzzy C-Means Clustering* untuk segmentasi pelanggan dengan menggunakan model RFM.

Perbedaan penelitian yang telah dilakukan oleh Wiratama Ahsani Taqwim, Nanang Yudi Setiawan, Fitra A. Bachtiar [4], serta penelitian yang telah dilakukan oleh Satria Ardi Perdana, Sara Famayla Florentin, dan Agus Santoso [35] adalah hanya menggunakan satu metode *Fuzzy C-Means Clustering* atau metode *K-Means Clustering*, sehingga tidak mengetahui hasil terbaik dari metode lainnya. Sedangkan pada penelitian yang akan dilakukan menggunakan metode *Fuzzy C-Means Clustering* dan metode *K-Means Clustering* untuk melihat performa mana yang terbaik dari metode tersebut. Metode tersebut akan diterapkan untuk segmentasi pelanggan pada Perusahaan bisnis ritel XYZ dengan menggunakan model RFM, dan melakukan evaluasi uji kedua algoritma menggunakan *Davies Bouldin Index (DBI)*.

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A