

**ANALISIS KLASTER DENGAN ALGORITMA *FUZZY C-MEANS* DALAM  
KLASTERISASI TINGKAT PENJUALAN  
(STUDI KASUS : PINDANG MERANJAT “MAK WAR”)**



**RISKA ARIYANTI MAKKASAU**

**H011201018**



**PROGRAM STUDI MATEMATIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM**

**UNIVERSITAS HASANUDDIN**

**MAKASSAR**

**2024**

**ANALISIS KLASTER DENGAN ALGORITMA *FUZZY C-MEANS* DALAM  
KLASTERISASI TINGKAT PENJUALAN  
(STUDI KASUS : PINDANG MERANJAT “MAK WAR”)**

**RISKA ARIYANTI MAKKASAU  
H011201018**



**PROGRAM STUDI MATEMATIKA DEPARTEMEN MATEMATIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS HASANUDDIN  
MAKASSAR**

**2024**

**ANALISIS KLASTER DENGAN ALGORITMA *FUZZY C-MEANS* DALAM  
KLASTERISASI TINGKAT PENJUALAN  
(STUDI KASUS : PINDANG MERANJAT “MAK WAR”)**

**RISKA ARIYANTI MAKKASAU**

**H011201018**

Skripsi

sebagai salah satu syarat untuk mencapai gelar sarjana

Program Studi Matematika

pada

**PROGRAM STUDI MATEMATIKA  
DEPARTEMEN MATEMATIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS HASANUDDIN  
MAKASSAR**

**2024**

**SKRIPSI**  
**ANALISIS KLASTER DENGAN ALGORITMA FUZZY C-MEANS DALAM**  
**KLASTERISASI TINGKAT PENJUALAN (STUDI KASUS : PINDANG**  
**MERANJAT "MAK WAR")**

**RISKA ARIYANTI MAKKASAU**

**H011201018**

Skripsi,

telah dipertahankan di depan Panitia Ujian Sarjana Sains pada tanggal 7 Juni 2024  
dan dinyatakan telah memenuhi syarat kelulusan  
pada

Program Studi Matematika  
Departemen Matematika  
Fakultas Matematika dan Ilmu Pengetahuan Alam  
Universitas Hasanuddin  
Makassar

Mengesahkan:  
Pembimbing tugas akhir,

Dr. Khaeruddin, M.Sc.  
NIP.19650914 199103 1 003

Mengetahui:  
Ketua Program Studi



Dr. Firman, S.Si., M.Si  
NIP.19680429 200212 1 001

## PERNYATAAN KEASLIAN SKRIPSI DAN PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa, skripsi berjudul "Analisis Kluster Dengan Algoritma *Fuzzy C-Means* Dalam Klusterisasi Tingkat Penjualan (Studi Kasus : Pindang Meranjat "Mak War")" adalah benar karya saya dengan arahan dari Bapak Dr. Khaeruddin, M.Sc. sebagai Pembimbing. Karya ilmiah ini belum diajukan dan tidak sedang diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka skripsi ini. Apabila di kemudian hari terbukti atau dapat dibuktikan bahwa sebagian atau keseluruhan skripsi ini adalah karya orang lain, maka saya bersedia menerima sanksi atas perbuatan tersebut berdasarkan aturan yang berlaku.

Dengan ini saya melimpahkan hak cipta (hak ekonomis) dari karya tulis saya berupa skripsi ini kepada Universitas Hasanuddin.

Makassar, 7 Juni 2024



Riska Ariyanti Makkasau

NIM H011201018

## UCAPAN TERIMA KASIH

Puji dan syukur saya panjatkan kepada Tuhan Yang Maha Esa, yang telah melimpahkan rahmat, hidayah, dan karunia-Nya sehingga penulis dapat menyelesaikan tugas akhir yang berjudul “Analisis Klaster Dengan Algoritma *Fuzzy C-Means* Dalam Klasterisasi Tingkat Penjualan (Studi Kasus : Pindang Meranjat “Mak War”)” dengan baik. Skripsi ini tidak akan terwujud tanpa dukungan, bimbingan, dan bantuan dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis ingin memberikan penghargaan yang sebesar-besarnya kepada orang tua tercinta atas doa, kasih sayang, dukungan moral, dan material yang tiada henti mengalir. Dan dengan segala kerendahan hati, penulis juga ingin mengucapkan terima kasih kepada:

1. Bapak **Prof. Dr. Ir. Jamaluddin Jompa, M.Sc** selaku Rektor Universitas Hasanuddin beserta jajarannya, Bapak **Dr. Eng. Amiruddin, M.Si** selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam, serta Bapak **Dr. Firman, S.Si., M.Si** selaku Ketua Departemen Matematika.
2. **Dosen dan Staf** Departemen Matematika atas ilmu dan bantuan yang selama ini telah diberikan.
3. Bapak **Dr. Khaeruddin, M.Sc.** selaku dosen pembimbing saya, atas semua ilmu yang telah diberikan selama proses perkuliahan dan bimbingan, serta senantiasa memotivasi penulis dalam penulisan skripsi ini.
4. Bapak **Prof. Dr Amir Kamal Amir, M.Sc.** dan Bapak **Prof. Dr. Syamsuddin Toaha M.Sc.** sebagai anggota tim penguji atas segala ilmu yang telah diberikan selama proses perkuliahan, serta telah bersedia meluangkan waktunya untuk memberikan masukan dan kritik yang membangun dalam proses penyusunan skripsi ini.
5. Sahabat **Etang (Nirma, Reza, Lutfiah, Salwa dan Wiwit**, teman sekaligus sepupu **Nack Longka**, teman **Blekpungg**, serta pacar saya **M. Al Habibullah** yang telah menemani, selalu meluangkan waktu, memberikan doa dan dukungan, serta tempat barbagi keluh kesah penulis.
6. Seluruh pihak yang tidak dapat disebutkan satu per satu atas segala bentuk kontribusi, partisipasi, serta motivasi yang diberikan kepada penulis selama ini.

Penulis

Riska Ariyanti Makkasau

## ABSTRAK

Perencanaan menu dalam bisnis kuliner merupakan aspek penting dari strategi penjualan yang efektif. Faktanya setiap menu memiliki tingkat penjualan yang berbeda. Untuk mengoptimalkan efektivitas penjualan dan pengelolaan bahan baku, penting untuk memahami menu mana yang perlu dipertahankan atau sebaliknya. Salah satu cara untuk mengetahui hal tersebut yaitu dengan melakukan analisis pengelompokan tingkat penjualan. Analisis yang bisa dilakukan yaitu analisis dengan teknik data *mining* metode klusterisasi. Salah satu algoritma yang termasuk dalam metode klusterisasi adalah *Fuzzy C-Means*. Dengan menggunakan algoritma *Fuzzy C-Means*, analisis data penjualan menu Pindang Meranjat “Mak War” dari April hingga Agustus 2023 menghasilkan 2 klaster, dengan 13 menu tergabung dalam klaster pertama yang merupakan menu-menu dengan tingkat penjualan yang tinggi dan 77 menu dalam klaster kedua dengan tingkat penjualan yang lebih rendah. Jumlah klaster optimum diperoleh menggunakan 3 uji validitas klaster yaitu *Modified Partition Coefficient*, *Silhouette Coefficient*, dan *Davies Bouldin Index (DBI)*. Berdasarkan ketiga uji validitas yang telah dilakukan, dapat disimpulkan bahwa algoritma *Fuzzy C-Means* sudah cukup baik dalam klusterisasi tingkat penjualan.

**Kata Kunci:** Data *Mining*; Klusterisasi; Algoritma *Fuzzy C-Means*; Validitas Klaster.

## **ABSTRACT**

*Menu planning in the culinary business is an essential aspect of effective sales strategy. In fact, each menu item has different sales levels. To optimize sales effectiveness and raw material management, it is important to understand which menus need to be maintained or otherwise. One way to determine this is by conducting an analysis of sales level grouping. The analysis that can be done is analysis using data mining technique clustering method. One algorithm included in the clustering method is Fuzzy C-Means. Using the Fuzzy C-Means algorithm, analysis of Pindang Meranjat "Mak War" menu sales data from April to August 2023 resulted in 2 clusters, with 13 menus grouped in the first cluster which are menus with high sales levels and 77 menus in the second cluster with lower sales levels. The optimum number of clusters is obtained using 3 cluster validity tests, namely Modified Partition Coefficient, Silhouette Coefficient, and Davies Bouldin Index (DBI). Based on the three validity tests conducted, it can be concluded that the Fuzzy C-Means algorithm is quite good in clustering sales levels.*

**Keywords:** *Data Mining; Clustering; Fuzzy C-Means Algorithm; Cluster Validity.*

**DAFTAR ISI**

	halaman
HALAMAN JUDUL.....	i
PERNYATAAN PENGAJUAN .....	ii
HALAMAN PENGESAHAN.....	iii
PERNYATAAN KEASLIAN SKRIPSI.....	iv
UCAPAN TERIMA KASIH .....	v
ABSTRAK.....	vi
<i>ABSTRACT</i> .....	vii
DAFTAR ISI.....	viii
DAFTAR TABEL .....	x
DAFTAR GAMBAR .....	xi
DAFTAR LAMPIRAN .....	xii
BAB I PENDAHULUAN .....	1
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah.....	3
1.4 Tujuan penelitian .....	3
1.5 Manfaat penelitian .....	3
1.6 Landasan Teori.....	3
1.6.1 Penjualan dan Tingkat Penjualan .....	3
1.6.2 Menu .....	4
1.6.3 Data Mining .....	5

1.6.4	Klasterisasi.....	6
1.6.5	Logika <i>Fuzzy</i> .....	7
1.6.6	<i>Fuzzy C-Means</i> .....	7
1.6.7	Standarisasi Data.....	9
1.6.8	<i>Modified Partition Coefficient (MPC)</i> .....	10
1.6.9	<i>Silhouette Coefficient</i> .....	11
1.6.10	<i>Davies Bouildin Index (DBI)</i> .....	11
BAB II METODOLOGI PENELITIAN.....		13
2.1	Sumber Data.....	13
2.2	Variabel Penelitian.....	13
2.3	Analisis Data.....	13
2.4	Alur Penelitian.....	14
BAB III HASIL DAN PEMBAHASAN.....		15
3.1	Hasil Penelitian.....	15
3.1.1	Data.....	15
3.1.2	Standarisasi Data.....	15
3.1.3	Pengklasteran menggunakan <i>Fuzzy C-Means (FCM)</i> .....	15
3.1.4	Menentukan Jumlah Kluster Optimum.....	25
3.2	Pembahasan.....	29
3.2.1	Hasil Pengklasteran.....	29
3.2.2	Interpretasi Kluster.....	31
BAB IV PENUTUP.....		33
4.1	Kesimpulan.....	33
4.2	Saran.....	33
DAFTAR PUSTAKA.....		34
LAMPIRAN.....		37

**DAFTAR TABEL**

Nomor Urut	halaman
1. Derajat Keanggotaan Iterasi Terakhir (Iterasi ke-25) 2 Klaster .....	22
2. Nilai Hasil Validasi MPC Tiap Klaster .....	26
3. Nilai Silhoutte Coefficient Tiap Klaster .....	27
4. Nilai Davies Bouldin Index Tiap Klaster.....	28
5. Hasil Pengklasteran dengan Fuzzy C-Means .....	29
6. Nilai Minimal dan Maksimal dari Setiap Variabel.....	31
7. Nilai Pusat Klaster pada Klaster Optimum .....	32

**DAFTAR GAMBAR**

Nomor Urut	halaman
1. Tahapan Data Mining .....	5
2. Alur Penelitian.....	14
3. Fungsi Objektif 2 Klaster.....	20
4. Grafik Hubungan antara Fungsi Objektif dan Jumlah Iterasi 2 Klaster ...	21
5. Pusat Klaster Untuk 2 Klaster .....	21
6. Plot Penyebaran Masing-Masing Anggota Klaster Untuk 2 Klaster .....	23
7. Grafik Jumlah Klaster dengan Nilai MPC .....	26
8. Grafik Jumlah Klaster dengan Nilai Silhoutte Coeficient.....	27
9. Grafik Jumlah Klaster dengan Nilai Davies Bouldin Index (DBI).....	28

**DAFTAR LAMPIRAN**

Nomor Urut	halaman
1. Data Penjualan Periode April – Agustus 2023.....	37
2. Matriks dari data penjualan (90x5).....	42
3. Matriks Partisi Awal (90 x 2 ) 2 Klaster.....	44
4. Perhitungan Pusat Klaster Iterasi 1 untuk 2 Klaster .....	46
5. Perhitungan Fungsi Objektif Iterasi 1 untuk 2 Klaster.....	58
6. Perbaikan matriks partisi untuk iterasi pertama (U1) 2 Klaster .....	62
7. Perbaikan matriks partisi untuk iterasi kedua (U2) 2 Klaster .....	64
8. Pengkodean Program pada Matlab .....	66
9. Pusat Klaster .....	67
10. Derajat Keanggotaan.....	69
11. Plot Penyebaran Anggota Klaster.....	118
12. Source Kode Silhoutte Coefficient pada Python.....	121
13. Source Kode Davies Bouldin Index (DBI) pada Python.....	123

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Perencanaan menu dalam bisnis kuliner merupakan aspek penting dari strategi penjualan yang efektif. Faktanya setiap menu memiliki tingkat penjualan yang berbeda. Untuk mengoptimalkan efektivitas penjualan dan pengelolaan bahan baku, penting untuk memahami menu mana yang perlu dipertahankan atau sebaliknya. Oleh karena itu, analisis menu menjadi langkah penting dalam mengembangkan strategi pemasaran yang tepat serta meninjau popularitas dan profitabilitas setiap item. Evaluasi terhadap penjualan memungkinkan untuk mengantisipasi tren pasar dan memperbarui menu secara berkala agar tetap relevan dengan selera pelanggan. Selain itu, dengan pemahaman yang lebih baik tentang pola penjualan, pemilik usaha dapat menyesuaikan strategi harga dan promosi untuk meningkatkan daya tarik menu dan menghasilkan pendapatan yang lebih besar.

Informasi tentang tingkat penjualan dapat diketahui dengan melakukan sebuah analisis pengelompokan menu makanan. Analisis yang bisa dilakukan untuk mengetahui tingkat penjualan adalah dengan teknik data *mining* metode klasterisasi. Data *mining* adalah proses ekstraksi informasi baru dari data dalam jumlah besar yang dapat digunakan untuk mendukung pengambilan keputusan (Megawati, Mukid, & Rahmawati, 2013). Data *mining* memiliki kemampuan untuk meramalkan berdasarkan informasi dari data masa lalu dan sekaligus proses untuk menemukan pola serta kecenderungan yang berguna dalam kumpulan data yang besar (Saifudin, 2018). Di dalam bidang data *mining*, terdapat beragam metode salah satunya metode klasterisasi. Metode klasterisasi atau analisis klaster digunakan untuk mengelompokkan objek-objek menjadi beberapa kelompok berdasarkan indikator yang diamati, sehingga seluruh objek yang masuk dalam kelompok yang sama memiliki kemiripan yang tinggi, dibandingkan dengan objek yang masuk dalam kelompok berbeda (Ananda & Yamani, 2020). Analisis klaster adalah sebuah metode yang digunakan untuk membagi kumpulan data ke dalam kelompok data berdasarkan kemiripannya (Windarto, 2017). Klasterisasi merupakan pilihan yang baik yang bisa digunakan untuk mendeteksi kemiripan dari pola pembelian *customer*.

Salah satu algoritma yang termasuk dalam metode klasterisasi adalah *Fuzzy C-Means* yang merupakan suatu teknik pengelompokan yang mempertimbangkan tingkat keanggotaan dalam himpunan *fuzzy* sebagai dasar pembobotan untuk pengelompokan dan bertujuan untuk mengurangi masalah kegagalan konvergen (Jang, Sun, & Mizutani, 1997). Penelitian sebelumnya menunjukkan bahwa metode *Fuzzy C-Means* memberikan hasil terbaik dalam mengatasi kasus *outliner* dan *overlapping* pada data (Li, Ng, & Cheung, 2008). Salah satu penelitian yang menjelaskan hal tersebut yaitu penelitian yang dilakukan oleh Ahmad Jaini dkk (2020) dengan judul "Perbandingan Algoritma *Fuzzy C-Means* dan *K-Medoids* Untuk

Pengelompokan Data Penjualan Pada 212 Mart”, penelitian ini menunjukkan bahwa pada data penjualan 212 Mart, *Fuzzy C-Means* menghasilkan akurasi 21,59% dan *K-Medoids* 20,18% . Sehingga dari hasil penelitian ini algoritma *Fuzzy C-Means* lebih baik dalam penempatan kluster. Namun, metode ini terdapat kelemahan dalam penentuan jumlah kluster yang optimal. Untuk meminimalisir hal tersebut serta mengoptimalkan hasil pengklasteran, perlu dilakukan validasi dengan menggunakan indeks validitas (Sutoyo & Sumpala, 2015).

Penelitian yang dilakukan oleh Cakra Ramadhana dkk (2013) dengan judul “Data Mining dengan Algoritma *Fuzzy C-Means Klustering* Dalam Kasus Penjualan di PT Sepatu Bata” diperoleh hasil berupa 3 kluster, dimana penentuan jumlah kluster telah ditetapkan oleh peneliti sebelum proses klustering. Penelitian serupa juga dilakukan oleh Ridha Nafila Turrahma dkk (2023), dengan menggunakan 202 data set data harga saham dan berdasarkan distribusi data ditetapkan nilai  $c = 3$  sehingga diperoleh 3 kluster yaitu kluster rendah, kluster sedang, dan kluster tinggi. Pada kedua penelitian tersebut tidak dilakukan pengujian validitas kluster sehingga tidak diketahui bagaimana kualitas kluster dan seberapa baik data ditempatkan pada suatu kluster.

Dalam penelitian yang dilakukan oleh Fajar Agustini (2017) mengenai implementasi algoritma *Fuzzy C-Means* pada penjualan di Sushigroove Restaurant, algoritma *Fuzzy C-Means* digunakan untuk mengelompokkan produk makanan di Restoran *Sushigroove* berdasarkan data transaksi selama bulan Juli hingga September 2016. Hasil proses data *mining* menunjukkan bahwa pada bulan tersebut memiliki kemiripan menu potensial yang tinggi. Selain itu, hasil pengujian validitas kluster juga menunjukkan nilai yang cukup tinggi. Penelitian ini menunjukkan bahwa algoritma *Fuzzy C-Means* dapat digunakan pada analisis penjualan di industri kuliner.

Berdasarkan uraian tersebut, penulis mengusulkan penggunaan teknik data *mining* untuk melakukan pengelompokan atau klusterisasi dengan memproses data historis transaksi penjualan yang telah ada. Dalam hal ini, penulis menggunakan metode klustering dengan algoritma *Fuzzy C-Means* untuk mengelompokkan menu-menu yang telah terjual ke dalam beberapa kategori serta menggunakan uji validitas kluster yaitu *Modified Partition Coefficient* (MPC), *Silhouette Coefficient* dan *Davies Bouldin Index* (DBI) untuk menentukan nilai  $c$  (jumlah kluster) yang optimal. Hasil penelitian akan dituangkan ke dalam tulisan skripsi dengan judul “**Analisis Kluster Dengan Algoritma *Fuzzy C-Means* Dalam Klusterisasi Tingkat Penjualan (Studi Kasus : Pindang Meranjat “Mak War”)**”.

## 1.2 Rumusan Masalah

Rumusan masalah dalam penelitian ini adalah “bagaimana hasil pengklasteran tingkat penjualan melalui analisis kluster dengan algoritma *Fuzzy C-Means*?

## 1.3 Batasan Masalah

Untuk menghindari penyebaran pembahasan yang terlalu luas dan agar fokus lebih terarah, maka dibuat batasan masalah yang mencakup:

1. Data yang akan diolah merupakan data penjualan bulanan untuk setiap menu yang dijual secara online oleh Pindang Meranjat “Mak War” dari periode April hingga Agustus 2023.
2. Harga setiap menu diabaikan.
3. Pengujian kluster menggunakan  $c = 2$  hingga 10.

## 1.4 Tujuan penelitian

Penelitian ini bertujuan memperoleh kluster-kluster tingkat penjualan melalui analisis kluster dengan algoritma *Fuzzy C-Means*.

## 1.5 Manfaat penelitian

Manfaat dari penelitian ini adalah sebagai berikut :

1. Bagi Mahasiswa
  - Menambah pemahaman dan pengetahuan mengenai analisis kluster dengan algoritma *Fuzzy C-Means*.
  - Dapat menjadi referensi bagi peneliti apabila ingin melakukan penelitian mengenai analisis kluster dengan algoritma *Fuzzy C-Means* terutama dalam klusterisasi tingkat penjualan.
2. Bagi Pemilik Bisnis :
  - Dapat mengetahui tingkat penjualan produk.
  - Dengan mengetahui tingkat penjualan produk, pemilik dapat membuat keputusan mengenai pengelolaan dan pemasaran produk.

## 1.6 Landasan Teori

### 1.6.1 Penjualan dan Tingkat Penjualan

Penjualan merupakan suatu metode terencana dan terorganisir untuk mengetahui dan memahami pelanggan demikian baik sehingga produk atau jasa yang ditawarkan sesuai dengan keinginan konsumen. Dengan demikian yang diperlukan adalah membuat produk dan jasa tersedia secara kuantitas dengan tetap

memperhatikan kuantitas barang atau jasa yang dihasilkan (Aslichah, Dwiningwarni, Yulianto, & Supriyadi, 2018).

Tingkat penjualan adalah ukuran yang menunjukkan banyaknya nilai produk yang terjual dalam suatu periode tertentu (Doloksaribu & Wibowo, 2017). Mengetahui tingkat penjualan pada suatu perusahaan sangat penting terutama dalam bisnis makanan. Adapun beberapa manfaat mengetahui tingkat penjualan makanan, terutama dalam bisnis makanan yaitu :

1. Mengukur performa bisnis

Tingkat penjualan dapat digunakan sebagai indikator performa bisnis. Dengan mengetahui tingkat penjualan, pemilik dapat mengevaluasi apakah bisnis mereka mengalami peningkatan atau penurunan.

2. Mengidentifikasi tren

Memantau tingkat penjualan dari waktu ke waktu memungkinkan pemilik bisnis mengidentifikasi tren penjualan, seperti peningkatan penjualan pada hari-hari atau jam-jam tertentu. Data ini berguna untuk perencanaan stok, pengaturan jadwal karyawan, dan pengembangan strategi pemasaran.

3. Mengoptimalkan persediaan

Mengetahui tingkat penjualan memungkinkan pemilik bisnis untuk mengoptimalkan persediaan bahan baku dan bahan makanan. Tujuannya adalah mengurangi biaya pembelian dan pemborosan, sambil memastikan ada cukup stok untuk memenuhi permintaan pelanggan.

4. Menentukan harga

Pemilik bisnis bisa menggunakan pengetahuan tentang tingkat penjualan untuk mengatur harga pasar sesuai dengan permintaan. Jika tingkat penjualan tinggi, mereka mungkin menaikkan harga produk. Namun, jika penjualan rendah, kemungkinan perlu menurunkan harga untuk menarik lebih banyak pelanggan.

5. Mengukur efektivitas promosi

Mengetahui tingkat penjualan sebelum dan setelah pelaksanaan promosi memungkinkan pemilik restoran untuk menilai seberapa efektif promosi tersebut. Informasi ini berguna dalam pengambilan keputusan apakah promosi perlu dipertahankan, dimodifikasi, atau dihentikan.

6. Mengidentifikasi masalah operasional

Jika terjadi penurunan tajam dalam tingkat penjualan, pemilik bisnis dapat mengidentifikasi masalah operasional yang mungkin terjadi. Misalnya, masalah dengan kualitas makanan, pelayanan pelanggan, atau kebersihan tempat. Dengan mengetahui masalah ini, pemilik bisnis dapat segera mengambil tindakan perbaikan yang diperlukan.

## 1.6.2 Menu

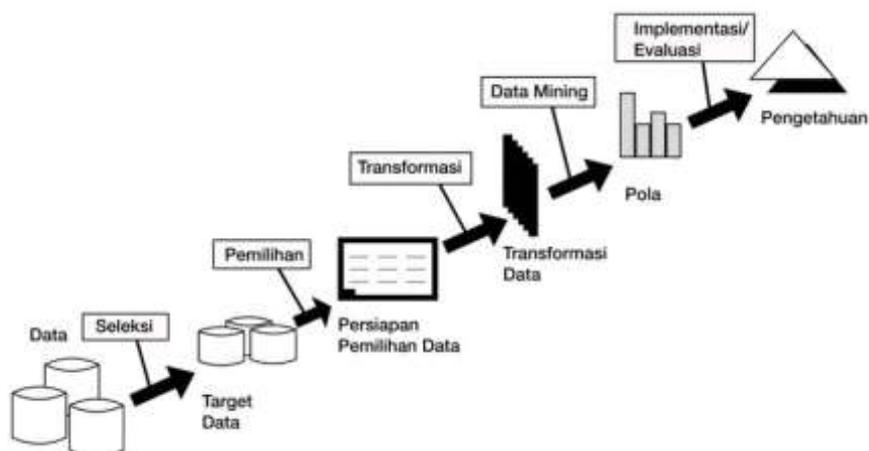
Menu berasal dari bahasa Prancis yaitu *Le Menu* yang berarti daftar makanan yang disajikan kepada tamu di ruang makan. Dalam lingkungan rumah tangga, menu diartikan sebagai susunan makanan atau hidangan tertentu. Menu

disebut juga Bill of Fare oleh orang Inggris. Menu adalah pedoman bagi yang menyiapkan makanan atau hidangan, bahkan merupakan penuntun bagi mereka yang menikmati hidangan tersebut dibuat (Rotua, Siregar, & Agustin, 2015).

### 1.6.3 Data Mining

Data *mining* merupakan suatu proses yang memanfaatkan teknik statistik, kecerdasan buatan, matematika, dan pembelajaran mesin untuk mengekstraksi dan mengidentifikasi informasi berharga serta pengetahuan yang relevan dari beragam basis data besar. Penggunaan metode atau algoritma dalam data *mining* sangat bervariasi, tergantung pada tujuan dan tahapan proses *Knowledge Discovery in Databases* (KDD) yang sangat tergantung pada pemilihan metode atau algoritma tersebut (Wahyuni, 2018). Secara teknis, data *mining* bisa diartikan sebagai langkah-langkah yang diambil untuk mengidentifikasi korelasi atau pola di antara puluhan, ratusan atau bahkan ribuan kolom dalam sebuah basis data relasional yang besar.

*Knowledge Discovery in Database Process* (KDD) adalah salah satu metode yang bisa digunakan dalam melakukan data *mining*. Fayyed dkk.(1996) mendefinisikan KDD sebagai proses dari menggunakan metode data *mining* untuk mencari informasi-informasi yang berharga, pola yang ada di dalam data, yang melibatkan algoritma untuk mengidentifikasi pola pada data. Tahapan yang dilakukan pada proses data *mining* diawali dari seleksi data dari data sumber ke data target, tahap *preprocessing* untuk memperbaiki kualitas data, transformasi, data *mining* serta tahap interpretasi dan evaluasi yang menghasilkan *output* berupa pengetahuan baru yang diharapkan memberikan kontribusi yang lebih baik. Secara detail dijelaskan sebagai berikut (Fayyad, Shapiro, & Smyth, 1996):



**Gambar 1.** Tahapan Data Mining

#### 1. Seleksi data (Selection)

Pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi yang digunakan untuk proses data *mining*, disimpan dalam suatu berkas, terpisah dari basis data operasional.

#### 2. Pemilihan data (*Preprocessing/Cleaning*)

Sebelum proses data *mining* dapat dilaksanakan, perlu dilakukan proses *cleaning* pada data yang menjadi fokus KDD. Proses *cleaning* mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data.

#### 3. Transformasi (*Transformation*)

Pada fase ini yang dilakukan adalah mentransformasi bentuk data yang belum memiliki entitas yang jelas ke dalam bentuk data yang valid atau siap untuk dilakukan proses data *mining*.

#### 4. Data *mining*

Data *mining* adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam data *mining* sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.

#### 5. Interpretasi/Evaluasi (*Interpretation/Evaluation*)

Pola informasi yang dihasilkan dari proses data *mining* perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut *interpretation*. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya.

### 1.6.4 Klasterisasi

Salah satu metode dalam data *mining* yaitu klasterisasi. Klasterisasi atau analisis klaster adalah proses yang bertujuan untuk mengelompokkan objek-objek dengan karakteristik yang mirip ke dalam satu klaster dan objek dengan karakteristik yang berbeda ke klaster yang lain dengan memaksimalkan kesamaan antar objek dalam satu klaster dan meminimalkan kesamaan antar klaster (Nuryani & Darwis, 2021).

Tahapan ini melibatkan pengenalan pola dan struktur dalam data set, baik yang telah diberi label maupun yang belum diberi label. Kualitas hasil klasterisasi sangat dipengaruhi oleh metrik yang digunakan untuk mengukur tingkat kesamaan atau perbedaan antar objek. Terdapat beragam metode yang dapat digunakan untuk mengukur nilai perbedaan antara objek-objek yang dibandingkan, dan salah satunya adalah dengan menggunakan metode Jarak *Euclidean*.

Kualitas kluster juga tergantung metode klusterisasi yang digunakan. Bora & Gupta (2014) menyimpulkan, klusterisasi dapat di klasifikasikan menjadi *soft klustering* dan *hard klustering*. Dalam *soft klustering*, himpunan *fuzzy* digunakan untuk mengelompokkan data sehingga memungkinkan data untuk menjadi anggota dari beberapa kluster berdasarkan derajat keanggotaannya. sedangkan pada teknik *hard klustering*, himpunan *crisp* digunakan untuk mengelompokkan data sehingga jika suatu data masuk ke dalam kluster tertentu maka tidak dapat masuk ke kluster lain. Dalam banyak situasi, *soft klustering* lebih alami daripada *hard klustering*. Teknik *soft klustering* yang sangat populer adalah *Fuzzy C-Means* (FCM).

### 1.6.5 Logika Fuzzy

Logika *fuzzy* pertama kali diperkenalkan oleh Lotfi A. Zadeh pada tahun 1965. Logika *fuzzy* adalah bentuk logika bernilai banyak yang memungkinkan nilai kebenaran variabel dapat berupa bilangan riil antara 0 dan 1. Dasar logika *fuzzy* adalah teori himpunan *fuzzy*. Pada teori himpunan *fuzzy*, peranan nilai keanggotaan sebagai penentu keberadaan elemen dalam suatu himpunan sangatlah penting. Pada himpunan tegas (*crisp*), nilai keanggotaan hanya terdapat dua kemungkinan yaitu 0 dan 1, sedangkan pada himpunan *fuzzy* nilai keanggotaan terletak pada rentang 0 sampai 1. Apabila  $x$  memiliki nilai keanggotaan *fuzzy*  $\mu_A [x] = 0$ , berarti  $x$  tidak menjadi anggota himpunan  $A$ , apabila  $x$  memiliki nilai keanggotaan *fuzzy*  $\mu_A [x] = 1$ , berarti  $x$  menjadi anggota penuh pada himpunan  $A$ , demikian pula apabila  $x$  memiliki nilai keanggotaan *fuzzy*  $\mu_A [x]$  antara 0 – 1, berarti  $x$  menjadi anggota pada himpunan  $A$  secara parsial (Kusumadewi & Purnomo, 2004).

### 1.6.6 Fuzzy C-Means

*Fuzzy C-Means* (FCM) adalah suatu teknik pengklasteran data yang keberadaan tiap-tiap titik data suatu kluster ditentukan oleh nilai keanggotaan. Nilai keanggotaan tersebut akan mencakup bilangan real pada interval 0-1. Pada tahun 1981, Jim Bezdek memperkenalkan metode ini yang memiliki keunggulan dalam penempatan pusat kluster yang lebih akurat daripada metode kluster lainnya. (Ramadhana, Lulu W, & Diah, 2013)

Konsep dasar dari *Fuzzy C-Means* adalah mengidentifikasi pusat-pusat kluster yang mewakili lokasi rata-rata untuk setiap kluster. Dalam metode FCM, pusat-pusat kluster awalnya tidak akurat karena setiap data memiliki tingkat keanggotaan untuk setiap kluster. Oleh karena itu, dilakukan iterasi berulang untuk memperbaiki posisi pusat kluster dan tingkat keanggotaan sehingga pusat kluster dapat bergerak ke lokasi yang lebih tepat. Proses iterasi ini didasarkan pada upaya meminimalkan fungsi objektif yang menggambarkan jarak antara data dengan pusat kluster (Muhardi & Nisar, 2015).

Kusumadewi & Purnomo (2004) menyimpulkan, algoritma yang digunakan pada metode *Fuzzy C-Means* adalah sebagai berikut:

1. Menentukan data yang akan di kluster berupa matriks  $X$  berukuran  $n \times m$  di mana  $n$  adalah jumlah sampel data,  $m$  adalah jumlah variabel yang digunakan dan  $x_{ij}$  adalah nilai data ke- $i$  ( $i = 1, 2, 3, \dots, n$ ), variabel ke- $j$  ( $j = 1, 2, 3, \dots, m$ ).
2. Menentukan nilai parameter awal
  - Jumlah kluster =  $c$  ( $c \geq 2$ )
  - Pangkat pembobot =  $w$   
Berdasarkan jurnal dengan judul "Penentuan Nilai Pangkat Pada Algoritma *Fuzzy C-Means*" menjelaskan bahwa  $w > 1$ , boleh 2,3,4,...  
Saat  $w = 2$ , tingkat akurasi 92% akurat, 8% tidak akurat.  
Saat  $w = 3$ , tingkat akurasi 85% akurat, 15% tidak akurat.  
Saat  $w = 4$ , tingkat akurasi 62% akurat, 38% tidak akurat.  
Sehingga penetapan pangkat sebesar  $w = 2$  akan menghasilkan tingkat akurasi yang lebih tinggi dibandingkan yang lain (Anggraeni, 2015)
  - Maksimum iterasi =  $MaxIter$
  - *Error* terkecil yang diharapkan =  $\xi$
  - Fungsi objek awal =  $P0 = 0$
  - Iterasi awal =  $t = 1$
3. Membangkitkan bilangan acak  $\mu_{ik}$ ,  $i = 1, 2, 3, \dots, n$ ;  $k = 1, 2, 3, \dots, c$ ; sebagai elemen-elemen baris ke- $i$  kolom ke- $k$  dari nilai derajat awal (matriks partisi awal ( $U$ )). Bilangan acak ini bisa dihasilkan dengan bantuan Microsoft *Excel* dan *Matlab*. Nilai derajat awal pada pengklasteran *Fuzzy* memenuhi kondisi sebagai berikut :

$$\mu_{ik} \in [0, 1] \text{ dan } \sum_{k=1}^c (\mu_{ik}) = 1 \quad (1)$$

Matriks partisi awal ( $U$ ) yang sudah terbentuk akan digunakan untuk menentukan pusat kluster. Bentuk matriks partisi awal yaitu :

$$U = \begin{bmatrix} \mu_{11}(x_1) & \cdots & \mu_{1n}(x_n) \\ \vdots & \ddots & \vdots \\ \mu_{c1}(x_c) & \cdots & \mu_{cn}(x_n) \end{bmatrix} \quad (2)$$

4. Menghitung pusat kluster ke- $k$  untuk variabel ke- $j$  ( $v_{kj}$ ),  $k = 1, 2, 3, \dots, c$  dan  $j = 1, 2, 3, \dots, m$ . Setiap pusat kluster didapatkan dengan menghitung jumlah dari hasil pemangkatan kluster dikalikan dengan bobot setiap data. Kemudian dibagi dengan jumlah dari hasil pemangkatan kluster yang dihitung. Matriks  $V$  berukuran yaitu sesuai jumlah kluster yang dibentuk dan banyaknya atribut atau variabel.

$$v_{kj} = \frac{\sum_{i=1}^n (\mu_{ik})^w (x_{ij})}{\sum_{i=1}^n (\mu_{ik})^w} \quad (3)$$

dengan,

$v_{kj}$  = pusat kluster ke-k pada variable ke-j,  
 $n$  = banyak data,  
 $w$  = pangkat pembobot,  
 $x_{ij}$  = nilai data ke-i ( $i = 1, 2, 3, \dots, n$ ), variabel ke-j ( $j = 1, 2, 3, \dots, m$ )

$\mu_{ik}$  = nilai derajat keanggotaan pengamatan ke-i pada kluster ke-k,

5. Menghitung fungsi objektif pada iterasi ke-t ( $P_t$ ). Fungsi objektif digunakan sebagai syarat perulangan untuk mendapatkan pusat kluster yang tepat, sehingga diperoleh hasil kluster pada iterasi terakhir berdasarkan kecenderungan nilai derajat keanggotaan. Kluster dengan nilai derajat keanggotaan tertinggi merupakan kluster dari data tersebut. Adapun formula fungsi objektif yang digunakan yaitu :

$$P_t = \sum_{i=1}^n \sum_{k=1}^c ([\sum_{j=1}^m (x_{ij} - v_{kj})^2] (\mu_{ik})^w) \quad (4)$$

dengan,

$P_t$  = fungsi objektif pada iterasi ke-t,  
 $c$  = banyak kluster yang akan dibentuk,  
 $n$  = banyak data,  
 $w$  = pangkat pembobot,  
 $m$  = banyaknya variable (kriteria), dengan ( $j = 1, 2, 3, \dots, m$ )  
 $x_{ij}$  = nilai data ke-i ( $i = 1, 2, 3, \dots, n$ ), variabel ke-j ( $j = 1, 2, 3, \dots, m$ )

$v_{kj}$  = pusat kluster ke-k pada variable ke-j, dan

$\mu_{ik}$  = nilai derajat keanggotaan pengamatan ke-i pada kluster ke-k,

6. Menghitung perubahan matriks partisi yang digunakan sebagai derajat keanggotaan.

$$\mu_{ik} = \frac{[\sum_{j=1}^m (x_{ij} - v_{kj})^2]^{\frac{-1}{w-1}}}{\sum_{k=1}^c ([\sum_{j=1}^m (x_{ij} - v_{kj})^2]^{\frac{-1}{w-1}})} \quad (5)$$

7. Memeriksa kondisi berhenti. Jika  $|P_t - P_{t-1}| < \xi$  atau  $t >$  maksimum iterasi maka iterasi berhenti, jika tidak maka  $t = t+1$  dan kembali ke langkah 4.

### 1.6.7 Standarisasi Data

Standarisasi data diperlukan saat terdapat perbedaan yang signifikan dalam satuan pengukuran setiap variabel penelitian. Ketidakteraturan atau perbedaan ini bisa menyebabkan ketidakvalidan dalam analisis kluster. Oleh karena itu, diperlukan

standarisasi dengan mengubah data asli (transformasi data) agar memiliki satuan yang seragam (Silvi, 2018). Walpole & Raymond (1995) menyimpulkan, rumus untuk menghitung standarisasi data sebagai berikut :

$$Z = \frac{x - \mu}{\sigma} \quad (6)$$

dengan :

$Z$  = nilai variabel yang sudah distandarisi

$x$  = nilai data sebelum di standarisasi

$\mu$  = nilai rata-rata variabel

$\sigma$  = nilai standar deviasi variable

### 1.6.8 Modified Partition Coefficient (MPC)

Modified partition coefficient (MPC) merupakan metode untuk menentukan jumlah kluster optimum dalam pengelompokan data. MPC sendiri merupakan perkembangan dari metode Partition Coefficient. Partition Coefficient (PC) adalah metode yang mengukur sejauh mana kluster tumpang tindih satu sama lain. Indeks Partition Coefficient memiliki kecenderungan mengalami perubahan yang konsisten ketika  $c$  (jumlah kluster) berubah. Modifikasi dari indeks PC, yang disebut sebagai *Modified Partition Coefficient* (MPC), bertujuan untuk mengatasi perubahan konsisten ini. Nilai MPC memiliki rentang dari 0 hingga 1 (Wu & Yang, 2005). Metode Modified Partition Coefficient (MPC) didefinisikan dengan persamaan :

$$MPC(c) = 1 - \frac{c}{c-1}(1 - PC(c)) \quad (7)$$

dengan:

$c$  = jumlah kluster

$MPC(c)$  = nilai indeks MPC pada kluster ke- $c$

Adapun persamaan nilai indeks PC ini adalah:

$$PC(c) = \frac{1}{n} \sum_{k=1}^c \sum_{i=1}^n \mu_{ik}^2 \quad (8)$$

dengan:

$c$  = jumlah kluster

$n$  = jumlah data

$\mu_{ik}$  = derajat keanggotaan data  $k$ - pada kluster ke- $i$

$PC(c)$  = nilai indeks PC pada kluster ke- $c$

Klaster dikatakan optimal jika nilai MPC mendekati 1 atau yang paling besar. Semakin dekat nilai MPC ke 0 maka semakin kabur keakuratannya.

### 1.6.9 *Silhouette Coefficient*

*Silhouette Coefficient* merupakan salah satu metode validasi untuk menguji kualitas sebuah klaster (Furqon & Muflikhah, 2016). Metode *Silhouette* mengukur seberapa dekat relasi antara objek dalam sebuah klaster dan seberapa jauh sebuah klaster terpisah dari klaster lainnya. Metode ini menggunakan rumus *Eclidean distance* dalam proses perhitungannya. Tahapan perhitungannya adalah sebagai berikut:

$$S(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (9)$$

dengan

$S_i$  = *Silhouette Coefficient*

$a_i$  = rata – rata jarak dari objek  $i$  dengan seluruh objek yang berada pada klaster yang sama

$b_i$  = nilai terkecil dari rata – rata jarak objek  $i$  dengan objek lain pada klaster yang berbeda.

Nilai *Silhouette Coefficient* berada pada rentan (-1) hingga 1. Nilai *Silhouette Coefficient* mendekati 1 menunjukkan bahwa titik data terklaster dengan baik dan secara tepat berada dalam klasternya. Ketika nilai mendekati -1, itu mengindikasikan adanya kemungkinan bahwa titik data ditempatkan di klaster yang tidak sesuai, sedangkan nilai yang mendekati 0 menunjukkan bahwa titik data berada pada daerah yang dekat dengan batas antara klaster. Artinya, semakin tinggi nilainya, maka semakin bagus pula kualitasnya.

### 1.6.10 *Davies Bouldin Index (DBI)*

*Davies Bouldin Index (DBI)* adalah metode evaluasi untuk mengukur validitas klaster hasil metode klasterisasi dengan menguji tingkat maksimal jarak inter-klaster dan pada saat yang sama juga menguji tingkat minimal jarak antar klaster. Mengacu pada penelitian Singh (2020), metode evaluasi DBI memiliki kelebihan dalam hal mengukur evaluasi klaster pada suatu metode pengelompokan disebabkan nilai kohesi dan separasi yang dihasilkannya. Kohesi didefinisikan sebagai jumlah dari kedekatan data terhadap *centroid* dari klaster yang diikuti, dikenal sebagai *Sum of Square Within* klaster. Sedangkan, separasi didasarkan pada jarak antar *centroid* dari klasternya, dikenal sebagai *Sum of Square Between* klaster. Melalui nilai kohesi dan separasi inilah yang menjadikan semakin kecil nilai DBI yang didapatkan (mendekati 0 atau sama dengan 0) maka menunjukkan semakin baik

kluster yang diperoleh dari pengelompokan yang dilakukan oleh algoritma klusterisasi.

Perhitungan DBI diawali dengan mencari nilai melalui perhitungan nilai *Sum of Square Within* kluster (SSW), yaitu untuk mengetahui metrik kohesi dalam sebuah kluster ke- $i$  yang dirumuskan pada Persamaan (10) :

$$SSW_i = \frac{1}{m_i} \sum_{j=1}^{m_i} d(x_j, c_i) \quad (10)$$

dengan

SSW = *Sum of Square Within* kluster  
 $m_i$  = jumlah data dalam kluster ke-  $i$   
 $c_i$  = *centroid* kluster ke- $i$   
 $d(x_j, c_i)$  = jarak dari data ke- $i$  ke titik kluster  $i$

Berikutnya, dilakukan langkah selanjutnya menggunakan Persamaan (11) untuk mengetahui separasi antar kluster.

$$SSB_{ij} = d(i, j) \quad (11)$$

dengan

SSB = *Sum of Square Between* kluster

$d(i, j)$  : Jarak *Euclidence distance* data ke- $i$  dan data ke- $j$

Selanjutnya, yaitu mencari Rasio dari hasil perhitungan SSW dan SSB seperti pada Persamaan:

$$R_{ij} = \frac{SSW_i + SSW_j}{SSB_{ij}} \quad (12)$$

Untuk proses perhitungan tahap akhir, Persamaan (13) digunakan untuk memperoleh nilai DBI.

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_{i,j}) \quad (13)$$

dengan

$k$  = jumlah kluster yang dihitung