

## **BAB II**

### **TINJAUAN PUSTAKA**

#### **2.1 Kajian Pustaka**

Pada tahapan ini, dilakukan studi literatur untuk mengumpulkan informasi dan data terkait dengan topik penelitian. Studi literatur digunakan untuk memberikan dasar teori terkait metode yang akan digunakan dalam penelitian. Hal ini penting untuk memperkirakan keberhasilan penelitian, memperkuat alasan pentingnya penelitian, dan memahami metode dan teknik yang digunakan dalam penelitian sebelumnya[12]. Informasi dan data yang dikumpulkan digunakan sebagai pendukung dan pembandingan dalam penelitian. Penelitian-penelitian sebelumnya yang bertujuan membantu pengolahan dan pengelompokan data telah banyak dilakukan dan ditetapkan di berbagai bidang. Algoritma *K-Means Clustering* merupakan salah satu algoritma yang telah banyak digunakan dalam penelitian sebelumnya. Penelitian-penelitian tersebut menghasilkan berbagai temuan yang berbeda. Algoritma *K-Means Clustering* merupakan salah satu algoritma dalam *Data Mining* yang populer untuk mengelompokkan data menjadi beberapa kelompok yang berbeda[13]. Kajian dilakukan untuk memberikan dasar teoritis terkait metode yang akan diterapkan dalam penelitian ini. Penelitian sebelumnya juga memiliki tujuan sebagai bahan perbandingan dan acuan untuk memperkuat teori yang digunakan dalam penelitian yang sedang dilakukan. Bagian ini mencakup rangkuman berbagai hasil penelitian terdahulu yang sama dengan topik penelitian yang akan dilakukan. Berdasarkan hasil eksplorasi, maka ditemukan beberapa penelitian sebelumnya yang sama dengan penelitian yang akan dilakukan, antara lain :

Tabel 2. 1 Penelitian Sebelumnya

No	Judul	Compare	Contrast	Criticize	Synthesize	Summarize
1.	<i>Enhancing Thyroid Patient Dietary Management with an Optimized Recommender System based on PSO and K-means</i> [14].	Dalam penelitian ini memberikan rekomendasi makanan sehat sama dengan penelitian yang dilakukan menggunakan metode <i>K-Means</i> .	Penelitian ini memberikan rekomendasi makanan yang sesuai untuk individu dengan kondisi Tiroid.	Pada penelitian membuat sebuah sistem rekomendasi akan tetapi belum terdapat jelas keadaan sistem yang telah dibuat.	Metode yang diterapkan dalam penelitian ini yaitu <i>Data Mining</i> selaras dengan penelitian yang akan dilakukan.	Pada penelitian ini akan digunakan metode PSO dan <i>K-Means</i> untuk menganalisis makanan yang cocok untuk kondisi individu Teroid. Hasil yang didapatkan setelah melakukan pengujian diperoleh rekomendasi makanan yang sesuai dengan individu yang berkondisi Teroid.
2.	<i>A new approach to signal filtering method using K-means clustering and distance-based Kalman filtering</i> [15]	Dalam penelitian ini menggunakan teknik <i>clustering</i> sama seperti penelitian yang dilakukan.	Penelitian ini membahas tentang penggunaan pendekatan baru dalam pengurangan noise pada data yang dihasilkan oleh alat	Pada penelitian ini metode yang diusulkan membutuhkan waktu yang lebih lama untuk menghitung karena Kalman	Teknik yang diterapkan dalam penelitian ini adalah <i>clustering</i> selaras dengan penelitian yang akan dilakukan.	Penelitian ini menunjukkan bahwa metode yang diusulkan mampu meningkatkan akurasi pengolahan data sebesar 18,95% dan akurasi seimbang sebesar 11,87% dibandingkan dengan

No	Judul	Compare	Contrast	Criticize	Synthesize	Summarize
			pengindera bau elektronik (e-nose) dengan menggunakan metode K-means.	Filtering bekerja secara iteratif dan semakin banyak data referensi yang digunakan, semakin lama waktu yang dibutuhkan.		data asli sebelum dilakukan filtering.
3.	<i>Data Mining</i> Pengelompokan Barang Makanan Ringan/ <i>Snack</i> Menggunakan Metode <i>Clustering</i> [16]	Penelitian ini melakukan metodologi yang sama yaitu <i>Clustering</i> dan <i>K-Means</i>	Penelitian ini melakukan pengelompokan makanan ringan/ <i>snack</i> sedangkan pada penelitian yang akan dilakukan adalah mengelompokkan menu makanan katering diet sehat.	Penelitian ini masih menggunakan variable merk makanan yang sedikit dari banyaknya merk makanan ringan/ <i>snack</i>	Penelitian yang dilakukan mengenai pengelompokan data dengan penelitian ini untuk penelitian tujuan yang sama	Penelitian ini mengelompokkan makanan ringan/ <i>snack</i> menggunakan algoritma <i>K-Means</i> pada PT. Dalam penelitian ini, Mitra Distrindo mendapatkan informasi tentang makanan yang paling diminati oleh pelanggan, serta jumlah persediaan yang perlu diproduksi, dipertahankan, atau ditingkatkan. Temuan

No	Judul	Compare	Contrast	Criticize	Synthesize	Summarize
						ini membantu perusahaan dalam menentukan perubahan dan penambahan yang perlu dilakukan.
4.	Implementasi <i>Data Mining</i> untuk <i>Clustering</i> Makanan dan Minuman Favorit dengan Menggunakan Algoritma <i>K-Means</i> [17].	Penelitian ini menggunakan metodologi yang sama yaitu <i>Clustering</i> dan <i>K-Means</i>	Pada penelitian ini mengelompokkan menu makanan favorit, sedangkan pada penelitian yang akan di lakukan adalah mengelompokkan menu makanan sehat	Penelitian ini hanya menggunakan 2 atribut, yang mana seharusnya bisa lebih banyak agar ke validitasan penelitian lebih baik	Penelitian yang dilakukan mengenai pengelompokan data dengan penelitian ini untuk penelitian tujuan yang sama	Penelitian ini dilakukan untuk mengidentifikasi menu yang paling disukai dengan menggunakan metode <i>Data Mining</i> menggunakan <i>K-Means</i> . Atribut yang digunakan dalam penelitian ini menghasilkan tiga kelompok ( <i>cluster</i> ) dengan menggunakan 40 data sampel.
5.	Implementasi <i>Data Mining</i> Dalam Menentukan Destinasi Unggulan Berdasarkan	Penelitian ini menggunakan metodologi yang sama yaitu <i>Clustering</i> dan <i>K-Means</i>	Penelitian ini dilakukan untuk menentukan destinasi unggulan berdasarkan	Penelitian ini hanya dilakukan pada satu perusahaan	Penelitian selanjutnya disarankan untuk melakukan penelitian di tempat lain	Informasi perjalanan dapat lebih baik dan pengetahuan baru terkait destinasi utama dapat dihasilkan berdasarkan nilai klaster yang diperoleh

No	Judul	Compare	Contrast	Criticize	Synthesize	Summarize
	Online <i>Reviews</i> <i>Tripadvisor</i> Menggunakan Algoritma <i>K-Means</i> [18]		online <i>reviews</i> <i>tripadvisor</i>			dalam implementasi di simulator <i>RapidMiner</i>
6.	Penerapan <i>Data Mining</i> Pengelompokan Menu Makanan dan Minuman Berdasarkan Tingkat Penjualan Menggunakan Metode <i>K-Means</i> [19]	Penelitian ini membahas <i>Data Mining</i> dengan menggunakan metode algoritma <i>K-Means Clustering</i> untuk mengelompokkan beberapa objek	Dalam penelitian ini menggunakan objek kuliner untuk mengelompokkan menu makanan dan minuman di <i>Dpom Coffee</i>	Data yang digunakan hanya menggunakan tiga perbandingan	Dengan menggunakan lebih banyak perbandingan akan menghasilkan data pengelompokan yang lebih akurat	Dalam pengolahan dataset penjualan di <i>Dpom Coffee</i> menggunakan alat bernama <i>Rapidminer</i> , terbentuk 3 <i>cluster</i> . <i>Cluster</i> pertama memiliki 8 menu dengan tingkat penjualan yang rendah. <i>Cluster</i> kedua terdiri dari 40 menu dengan tingkat penjualan sedang. Sedangkan <i>cluster</i> ketiga memiliki 7 menu dengan tingkat penjualan yang tinggi.
7.	Penerapan Metode <i>Clustering</i> Untuk	Melakukan penelitian menggunakan	Penelitian ini dilakukan untuk memperoleh objek wisata	Penelitian ini terdapat ketidaktepatan format file saat	Penelitian selanjutnya disarankan	Penelitian ini dilakukan untuk membangun sistem kelompok prioritas objek wisata

No	Judul	Compare	Contrast	Criticize	Synthesize	Summarize
	Pengelompokan Potensi Wisata Di Kabupaten Sumedang [20]	Metode <i>K-Means Clustering</i>	potensial dan yang cocok untuk dikembangkan.	<i>import</i> ke dalam sistem	memperbaharui format file	yang akan dikembangkan. Dari perhitungan <i>clustering</i> yang dilakukan mendapatkan 3 <i>cluser</i> , <i>cluster 1</i> sebanyak 12 tempat wisata, <i>cluster 2</i> sebanyak 18 tempat wisata, dan <i>cluster 3</i> sebanyak 7 tempay wisata.
8.	<i>Data Mining in Tourism Data Analysis : Inbound Visitors to Japan</i> [21]	Melakukan penelitian menggunakan metode <i>Data Mining</i>	Penelitian ini menggunakan metode pohon keputusan dan metode statistik lanjutan dari <i>Data Mining</i>	Dalam penelitian ini hanya menggunakan satu data dari perusahaan <i>Japan Travel Bureau (JTB) Foundation</i>	Penelitian selanjutnya disarankan untuk menggunakan data dari beberapa perusahaan	Hasil dari pohon keputusan menunjukkan bahwa terdapat dua kelompok berbeda yaitu wisatawan Asia dan non-Asia.
9.	Implementasi <i>K-Means</i> Untuk Pengelompokan Status Gizi Balita (Studi	Melakukan penelitian menggunakan Metode <i>K-Means Clustering</i>	Pada penelitian ini mengelompokan status Gizi balita sedangkan pada penelitian yang	Dalam penelitian ini seharusnya bisa menambahkan lebih banyak	Meode yang dilakukan menggunakan <i>K-Means</i> ini sejalan dengan	Penelitian ini menunjukkan bahwa gizi pada balita Banjar Titih dapat dikelompokkan dengan metode <i>K-Means</i>

No	Judul	Compare	Contrast	Criticize	Synthesize	Summarize
	Kasus Banjar Titih)[22].		akan dilakukan mengelompokan menu makanan sehat pada ketring diet sehat	atribut lagi seperti tinggi badan ataupun dalam bentuk kuisioner agar mendapat hasil yang lebih baik.	penelitian yang akan dilakukan	menggunakan 3 atribut yaitu umur, jenis kelamin, dan berat badan. Kategori gizi dibagi menjadi tiga kelompok, yaitu gizi kurang, gizi normal, dan gizi lebih. Hasil pengelompokan menunjukkan terdapat tiga <i>cluster</i> , dengan gizi buruk sebesar 30,43%, gizi normal sebesar 47,83%, dan gizi lebih sebesar 21,74%.
10.	Analisis Dan Implementasi <i>Data Mining</i> untuk Menentukan Gaji Karyawan Tetap Serta Honorer Memakai Prosedur	Melakukan penelitian menggunakan Metode <i>K-Means Clustering</i>	Pada penelitian ini menggunakan dua metode sedangkan pada penelitian yang akan dilakukan hanya menggunakan satu metode.	Pada penelitian ini tidak terdapat cara penentuan jumlah klaster.	Meode yang dilakukan menggunakan <i>K-Means</i> ini sejalan dengan penelitian yang akan dilakukan	Penelitian ini, dilakukan penggunaan <i>Data Mining</i> untuk menganalisis dan mengimplementasikan penentuan gaji bagi karyawan tetap dan honorer. Teknik <i>clustering K-Means</i> dan algoritma C4.5

No	Judul	Compare	Contrast	Criticize	Synthesize	Summarize
	Algoritma <i>K-Means Clustering</i> Dan C4.5[23].					<p>digunakan dalam penelitian ini, dan terbukti berhasil dalam menentukan gaji untuk kedua jenis karyawan tersebut.</p> <p>Implementasinya dilakukan melalui sebuah <i>website</i> yang berfungsi sebagai platform untuk memasukkan dan memilih data penggajian berdasarkan status karyawan, dan hasilnya terbukti sangat efektif bagi perusahaan.</p>



Studi literatur ini dilakukan untuk menambah informasi atau menambah landasan teori sebagai bahan untuk menyelesaikan penelitian ini. Hasil dari studi literatur yang diterapkan terhadap beberapa jurnal seperti pada Tabel 2.1, terdapat beberapa metode yang ada dalam *Data Mining* termasuk *Apriori*, *K-Means Clustering*, DBSCAN, dan C4.5. Beberapa jurnal menjelaskan perbandingan antara algoritma *K-Means* dengan algoritma lainnya. *K-Means Clustering* banyak diterapkan untuk mengelompokkan sebuah data besar menjadi beberapa kelompok sedangkan, *Apriori* berfokus pada penemuan pola asosiasi antar item-item dalam transaksi, DBSCAN merupakan algoritma kluster berbasis kepadatan, ini mengelompokkan titik-titik data yang berdekatan dan memiliki kepadatan yang tinggi ke dalam klaster, algoritma C4.5 digunakan untuk pembelajaran mesin dan pengambilan keputusan untuk membangun model prediksi berdasarkan pohon keputusan. Berdasarkan beberapa jurnal yang digunakan dalam studi literatur terdapat beberapa metode dan tema yang sama dengan penelitian yang akan dilakukan.

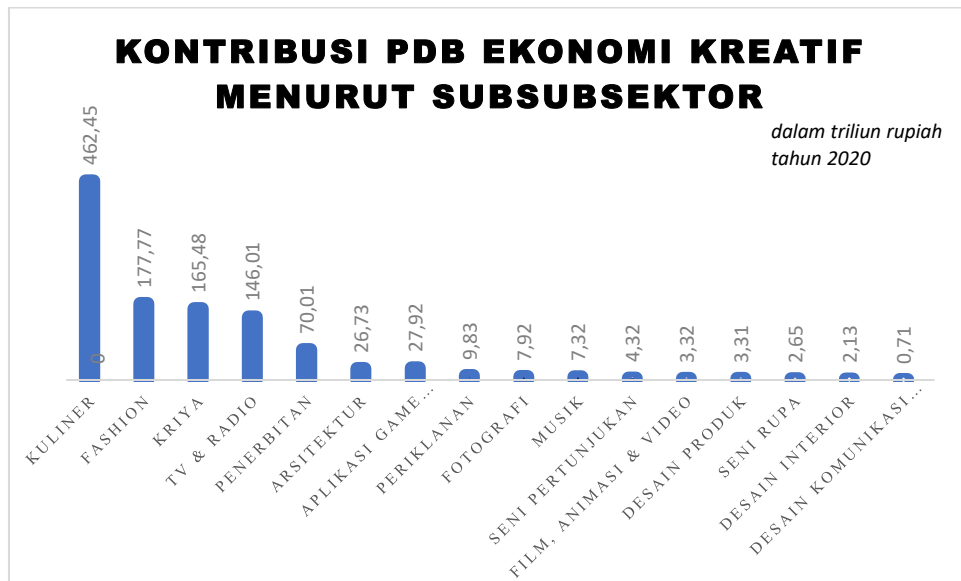
## **2.2 Dasar Teori**

Bagian ini mencakup pembahasan terkait dasar teori yang relevan dengan topik penelitian yang dilakukan. Dasar teori yang digunakan sebagai acuan dalam penelitian yang dilakukan yaitu:

### **2.2.1 Kuliner**

Kuliner adalah aktivitas yang meliputi pembuatan, pengolahan, dan penyajian makanan dan minuman, yang memperhatikan aspek kreativitas, estetika, tradisional, dan kebudayaan setempat untuk menambah nilai dan memperkuat cita rasa produk serta meningkatkan daya tarik konsumen dan memberikan kesan yang menyenangkan bagi mereka[24]. Memiliki usaha milik sendiri saat ini menjadi salah satu pilihan para pencari pekerjaan. Salah satu jenis usaha yang menjanjikan atau memiliki prospek kerja yang baik adalah usaha di bidang kuliner. Usaha kuliner saat ini banyak yang bermunculan dengan beragam inovasi menarik, di iringi dengan budaya konsumtif masyarakat saat ini yang lebih gemar menempatkan diri mereka menjadi konsumen. Potensi kuliner Indonesia sangat berlimpah. Indonesia memiliki reputasi sebagai sumber rempah-rempah yang sangat baraneka ragam, yang memungkinkan untuk menciptakan berbagai jenis masakan dengan cita rasa yang kaya dan beragam. Hal ini membuat kuliner menjadi salah satu bidang dalam rencana pengembangan ekonomi kreatif di Indonesia[24].

Indonesia terdapat 4 bidang atau sektor bisnis ekonomi kreatif. Dari 4 bidang tersebut, ada tiga sektor yang memberikan kontribusi terbesar, yaitu kuliner, fashion, dan kriya. Subsektor kuliner, fashion, dan kriya merupakan subsektor yang memberikan kontribusi terbesar terhadap Produk Domestik Bruto (PDB) ekonomi kreatif, dengan total kontribusi mencapai sekitar 75% setiap tahun. Secara rata-rata, subsektor kuliner menyumbang sekitar 42% setiap tahun, sementara subsektor fashion menyumbang sekitar 18%, dan subsektor kriya sebesar 15% dari total PDB ekonomi kreatif[25].



Gambar 2. 1 Kontribusi PDB per-Subsektor ekonomi kreatif[25]

### 2.2.2 Makanan Sehat

Makanan adalah sesuatu yang penting bagi Kesehatan tubuh saat melakukan aktivitas. Mengonsumsi satu jenis makanan tanpa mempertimbangkan keseimbangan dengan jenis makanan lain dapat menyebabkan konsekuensi buruk bagi kesehatan tubuh. Pola hidup yang tidak sehat, seperti makan makanan yang tidak sehat seperti makanan instan atau *junk food*, dapat meningkatkan resiko berbagai penyakit. Banyak orang lebih memilih makanan yang mengenyangkan dan lezat dengan mengabaikan nutrisi yang terkandung dalam makanan yang dimakan[26].

Makanan yang sehat yaitu makanan yang memiliki kandungan gizi seimbang, kaya serat dan mencakup semua nutrisi yang berguna untuk perkembangan tubuh[27]. Menurut Atmojo, unsur gizi adalah salah satu aspek penting untuk menghasilkan sumber daya manusia yang sehat, cerdas, dan produktif. Pola makan yang sehat yaitu pola makan yang memperhitungkan asupan nutrisi yang baik. Ini dapat dicapai dengan mengonsumsi makanan yang sederhana dan alami, seperti makanan dari rantai makanan yang paling rendah. Dengan demikian, kandungan vitamin, mineral, dan nutrisi penting lainnya akan lebih tinggi

dan terpenuhi dengan baik. Pedoman Umum makanan dengan gizi seimbang sebagai berikut[27]:

1. Memenuhi kebutuhan energi dengan makan yang benar.
2. Memenuhi setengah dari kebutuhan energi dengan makan makanan sumber karbohidrat
3. Menggunakan garam beryodium
4. Mengonsumsi makanan yang kaya zat besi
5. Minum air putih yang cukup
6. Menjalani kegiatan fisik dan olahraga dengan rutin
7. Tidak meminum minuman keras

### **2.2.3 Diet sehat**

Secara umum, kebanyakan orang menginginkan berat badan yang ideal, terutama Wanita. Orang yang memiliki berat badan berlebih dianggap oleh masyarakat seringkali tidak mempunyai keinginan yang besar untuk mengurangi berat badan. Penyebab dari kelebihan berat badan dipengaruhi oleh beberapa faktor yaitu termasuk di dalamnya pola makan, kurang olahraga, kelainan metabolisme, dan faktor genetik. Akan tetapi tidak sedikit masyarakat khususnya wanita saat ingin menurunkan berat badan menerapkan diet yang tidak sehat atau dengan cara yang instan, sehingga menyebabkan tidak sedikit orang salah caranya menurunkan berat badan dan sakit. Maka dari itu untuk melakukan diet disarankan untuk melakukan diet sehat[28].

Diet sehat adalah pola makan yang seimbang dan sehat yang terdiri dari beberapa jenis makanan seperti buah-buahan, sayuran, bahan pangan nabati, bahan pangan hewani seperti daging, ikan, telur, dan produk susu, serta bahan pangan bernutrisi lainnya seperti kacang-kacangan, biji-bijian, dan minyak nabati. Dalam diet sehat juga diperlakukan asupan cukup protein, lemak sehat, karbohidrat kompleks, vitamin, mineral dan serat yang cukup sesuai dengan kebutuhan tubuh. Sebaliknya dihindari atau dibatasi makanan yang tinggi kalori, lemak jenuh, natrium, dan gula serta makanan olahan. Juga diperlukan asupan air yang cukup dan aktivitas fisik yang rutin untuk menjaga Kesehatan.

Diet sehat adalah mengatur asupan nutrisi yang diterima oleh tubuh agar dapat mencapai dan mempertahankan kesehatan yang baik. Prinsip dasar yang perlu diperhatikan dalam menjalani diet sehat, diantaranya:

1. Konsumsi makanan yang bergizi seimbang: Asupan nutrisi yang diterima harus seimbang antara karbohidrat, protein, lemak, vitamin, mineral, dan serat.
2. Konsumsi makanan yang bervariasi: Konsumsi beragam jenis makanan akan memastikan bahwa tubuh mendapatkan berbagai jenis nutrisi yang dibutuhkan.
3. Membatasi konsumsi makanan yang tinggi lemak, gula, dan garam: Makanan yang tinggi lemak, gula, dan garam dapat beresiko terkena berbagai penyakit kronis seperti obesitas, hipertensi, dan diabetes.
4. Konsumsi makanan yang rendah kalori: Membatasi asupan kalori akan membantu dalam mempertahankan berat badan dan memperoleh berat badan yang sehat.
5. Cukupi kebutuhan cairan: Asupan cairan yang cukup sangat penting bagi tubuh dan dapat membantu dalam menjaga kesehatan.
6. Konsumsi makanan organik dan/atau bersih: Makanan yang dibuat dengan cara yang baik dan tanpa bahan kimia berbahaya akan memastikan kesehatan tubuh.

#### **2.2.4 Z-Score Normalization**

Proses normalisasi data merupakan tahapan yang esensial dalam praproses data. Tujuannya adalah untuk mengurangi skala nilai-nilai pada *dataset* sehingga memudahkan proses pengolahan. Normalisasi data diperlukan karena seringkali terdapat variasi rentang nilai yang berbeda pada setiap atribut pada *dataset*, sehingga dapat mempengaruhi efektivitas atribut dalam *dataset*[29]. *Z-Score Normalization* adalah metode yang didasarkan pada nilai rata-rata (mean) dan deviasi standar (*standard deviation*) dari data, dan dapat digunakan jika ada Informasi mengenai nilai minimum dan maksimum yang sebenarnya dari data. Penggunaan metode ini sangat membantu untuk normalisasi data[30]. Teknik *Z-Score* melakukan transformasi data dengan membuat rentang nilai baru yang berdasarkan pada nilai-nilai yang terdapat pada *dataset*. Persamaan untuk menghitung *Z-Score Normalization* dapat ditemukan pada persamaan berikut[29].

$$X_{baru} = \frac{X_{lama} - \bar{X}}{\sigma} \quad (2.1)$$

**Keterangan:**

$\bar{X}$  = rata – rata

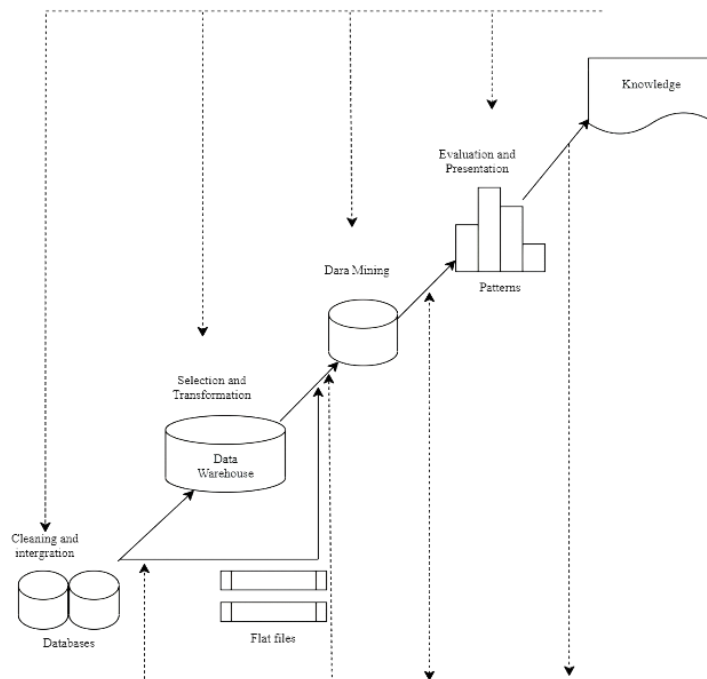
$\sigma$  = standar deviasi

### 2.2.5 Data Mining

*Data Mining* yaitu teknik pembelajaran mesin, diterapkan dalam *Data Mining* untuk mengambil Informasi dan pengetahuan secara otomatis dari data. Definisi lainnya adalah pembelajaran berbasis induksi, yang merupakan proses membentuk pemahaman umum suatu konsep dengan memperhatikan contoh-contoh tertentu dari konsep tersebut[31]. *Data Mining*, juga dapat diartikan sebagai *pattern recognition* adalah algoritma yang dimanfaatkan untuk mengola data dengan tujuan menemukan pola yang tersembunyi di dalamnya. Hasil dari pengolahan data ini kemudian menghasilkan pengetahuan baru yang berasal dari data lama, yang dapat diterapkan untuk membuat keputusan di masa depan[32].

*Data Mining* memiliki tahapan yang diantaranya yaitu[33]:

1. Pembersihan Data (membersihkan data yang tidak stabil).
2. Penggabungan Data (menggabungkan data dari berbagai sumber).
3. Seleksi Data (data yang sesuai dengan tujuan analisis dimasukan Kembali ke dalam database).
4. Transformasi Data (data diperoleh dan diorganisir untuk membuat laporan performa atau operasi agresif yang singkat).
5. *Data Mining* (proses mencari paduan dan pola yang bermakna dalam data besar dengan metodologi dan Teknik statistik dan matematis).
6. Evolusi Pola (menganalisis pola penting yang merepresentasikan pengetahuan berdasarkan beberapa perilaku menarik).
7. Presentasi Pengetahuan (menampilkan Informasi yang diambil dari *Data Mining* dengan menggunakan visualisasi untuk memberikan wawasan kepada pengguna).



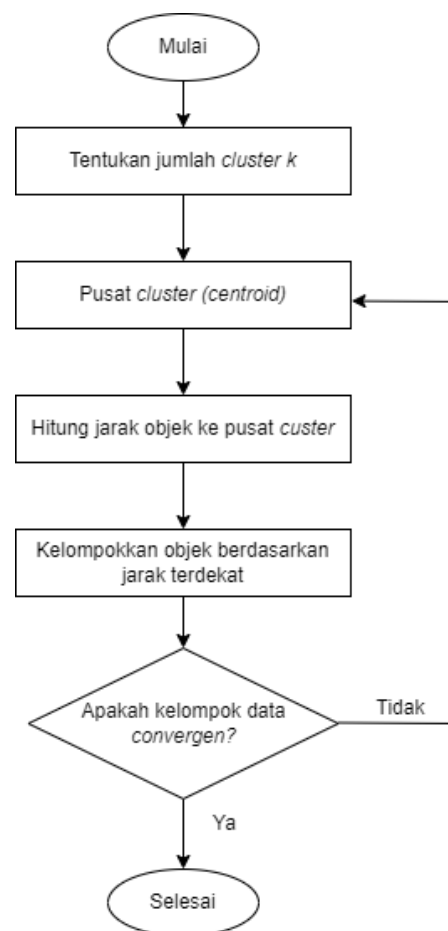
Gambar 2. 2 Tahapan Data Mining[33]

### 2.2.6 Clustering

*Clustering* atau pengelompokan merupakan suatu cara untuk membagikan banyak data atau objek menjadi beberapa kelompok berdasarkan kesamaan atribut-atribut mereka. Metode *clustering* adalah sebuah teknik yang tidak bergantung pada arahan atau supervisi, di mana karakteristik dari setiap kluster tidak ditentukan sebelumnya, tetapi ditentukan berdasarkan kemiripan atribut dalam suatu kelompok[19]. Proses *clustering* dapat membantu dalam menemukan tata letak data yang bisa digunakan untuk berbagai keperluan seperti klasifikasi, pengolahan gambar, dan identifikasi pola[32]. *Clustering* dalam proses *Data Mining* dapat membantu dalam menentukan pola pembagian dalam sebuah kumpulan data, yang bermanfaat untuk analisis. Objek yang mirip didapat dari kesamaan nilai atribut yang mendeskripsikan dari objek-objek data, dan biasanya diwakili sebagai titik dalam ruang multidimensi[34].

*K-Means clustering* merupakan metode *clustering non-hiyerarkis* yang mulai dengan memilih sejumlah titik dari populasi sebagai pusat awal *cluster*. Pusat *cluster* dipilih secara acak dari kumpulan data populasi. Lalu, melalui proses *K-Means*, setiap unsur dalam data populasi diperiksa dan dikategorikan dalam salah

satu dari pusat *cluster* yang sudah ditentukan[19]. Menerapkan dan menjalankan algoritma *K-Means clustering* sangatlah mudah, proses *clustering* relatif cepat, mudah beradaptasi, dan banyak digunakan dalam praktik. Ide atau alur dari algoritma ini cukup sederhana. Pada awalnya, jumlah kelompok atau *cluster* yang akan digunakan ditentukan terlebih dahulu. Kemudian, dokumen atau elemen pertama dipilih sebagai pusat *cluster*. Selanjutnya, dilakukan iterasi untuk menghitung jarak antara dokumen atau objek dengan pusat *cluster*, hingga mencapai *konvergensi*[35].



Gambar 2. 3 Tahapan *K-Means Clustering*[35]

Berikut penjelasan mengenai tahapan-tahapan yang terjadi dalam pengelompokan data menggunakan algoritma *K-Means Clustering*[35]:

- 1) Melakukan penentuan nilai jumlah *cluster*(*k*).



- 2) Melakukan inisialisasi *Centroid* awal (*initial Centroid*) secara acak. Dalam algoritma *K-Means* pusat *cluster* baru ditentukan dengan menghitung rata-rata dari nilai total objek dalam suatu *cluster*. Persamaan yang diterapkan untuk menghitung pusat *cluster* adalah persamaan (2.2) sebagai berikut[36]:

$$C_i = \frac{\sum d_i}{n} \quad (2.2)$$

**Keterangan:**

$C_i$  = *Centroid* baru ke  $i$

$d_i$  = jumlah nilai jarak yang masuk dalam *cluster*

$n$  = jumlah data pada tiap *cluster*

- 3) Menghitung jarak pada masing-masing data setiap *Centroid* menggunakan rumus jarak *Euclidean Distance* berikut:

$$d(i, j) = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \dots + (X_{ki} - X_{kj})^2} \quad (2.3)$$

**Keterangan:**

$d(i, j)$  = jarak data ke  $i$  ke pusat *cluster*  $j$

$X_{ki}$  = data ke  $i$  pada atribut data ke  $k$

$X_{kj}$  = titik pusat ke  $j$  pada atribut data ke  $k$

- 4) Melakukan pengelompokkan dengan menghitung jarak antara objek dengan *Centroid* dan memilih *Centroid* terdekat sebagai representasi *cluster*
- 5) Tahap terakhir dalam proses ini yaitu menguji penyesuaian antara kelompok data yang baru dibentuk dengan kelompok data pada iterasi sebelumnya. Namun, jika kedua kelompok data masih berbeda, maka akan dilakukan iterasi dan menemukan *cluster* baru.

### 2.2.7 Metode *Elbow*

Metode *Elbow* adalah pendekatan yang digunakan untuk mendapatkan informasi dalam menentukan jumlah *cluster* optimal dengan cara menganalisis persentase perbandingan antara jumlah *cluster* ( $k$ ) yang menyebabkan pembentukan sudut tajam pada suatu titik. Nilai  $k$  dalam konteks perpaduan dengan *K-Means*

adalah hasil dari grafik yang menggambarkan hubungan antara jumlah *cluster* dan penurunan kesalahan. Jumlah *cluster* ( $k$ ) yang muncul setelah pengujian dengan *K-Means* dievaluasi menggunakan teknik SSE[37].

### 2.2.8 Davies Bouddin Index (DBI)

*Davies Bouddin Index* merupakan salah satu metode evaluasi internal yang mengukur evaluasi *cluster* pada suatu metode pengelompokan yang didasarkan pada nilai kohesi dan separasi. Dalam suatu pengelompokan, kohesi didefinisikan sebagai jumlah dari kedekatan data terhadap *centroid* dari *cluster* yang diikuti. Sedangkan separasi didasarkan pada jarak antar *centroid* dari *clusternya*. Semakin kecil nilai DBI yang diperoleh (non-negatif  $\geq 0$ ), maka semakin baik *cluster* yang diperoleh dari pengelompokan K-means yang digunakan [38]. Indeks *Davies-Bouldin* berfungsi sebagai metrik untuk menilai kualitas pemisahan *cluster* sekaligus mengukur kekompakan *cluster* tersebut.

*Sum of square within cluster* (SSW) merupakan persamaan yang digunakan untuk mengetahui matrik kohesi dalam sebuah *cluster* ke- $i$  yang dirumuskan sebagai berikut[38]:

$$SSW_i = \frac{1}{m_i} \sum_{j=i}^{m_i} d(x_i, c_i) \quad (2.4)$$

**Keterangan:**

$m$  = Jumlah data dalam *cluster* ke- $i$

$c$  = Centroid *cluster* ke- $i$

$d$  = Jarak

*Sum of square between cluster* (SSB) merupakan persamaan yang digunakan untuk mengetahui separasi antar *cluster* yang dihitung menggunakan persamaan[38]:

$$SSB_{ij} = d(C_i, C_j) \quad (2.5)$$

**Keterangan:**

$d$  = Jarak

$c$  = Centroid *cluster* ke- $i$

Setelah nilai kohesi dan separasi diperoleh, kemudian dilakukan pengukuran rasio untuk mengetahui nilai perbandingan antara *cluster* ke-*i* dan *cluster* ke-*j*. *Cluster* yang baik adalah *cluster* yang memiliki nilai kohesi sekecil mungkin dan separasi yang sebesar mungkin. Nilai rasio dihitung menggunakan persamaan sebagai berikut[38]:

$$R_{ij} = \frac{SSW_i + SSW_j}{SSB_{ij}} \quad (2.6)$$

Nilai rasio yang diperoleh tersebut digunakan untuk mencari nilai daviesbouldin index (DBI). Persamaan yang diterapkan untuk menghitung indeks *Davies-Bouldin* sebagai berikut [14]:

$$DB = \frac{1}{n} \sum_{i=1}^n \max_{j \neq i} \left( \frac{S_i + S_j}{d(c_i, c_j)} \right) \quad (2.7)$$

**Keterangan:**

$n$  = Jumlah *cluster*

$S_i$  = Jarak rata-rata (kemiripan) antara setiap titik data dalam *cluster* “*i*”

$d(c_i, c_j)$  = Jarak antara pusat-pusat *cluster* “*i*” dan “*j*”

### 2.2.9 *Rapid Miner*

*Rapid Miner* adalah perangkat lunak yang memiliki antarmuka pengguna berbasis GUI (*Graphical User Interface*) untuk mempermudah penggunaannya. Perangkat lunak ini bersifat open source dan dibangun menggunakan bahasa pemrograman Java dengan lisensi GNU *Public Licence*, sehingga dapat dijalankan pada berbagai sistem operasi. Dengan *Rapid Miner*, tidak diperlukan keahlian koding khusus karena semua fasilitas telah disediakan. Perangkat ini dirancang khusus untuk keperluan *Data Mining*[39].