

## BAB II

### TINJAUAN PUSTAKA

#### 2.1. Data Mining

Data mining merupakan proses pencarian dan analisis terhadap sekumpulan data (database) untuk menemukan pola yang relevan dan bernilai, dengan tujuan mengekstraksi informasi serta pengetahuan yang akurat, potensial, mudah dipahami, dan bermanfaat dalam mendukung pengambilan keputusan. (Petra Refialy et al., 2021a). Data mining merupakan proses mengidentifikasi hubungan pada data yang belum diketahui oleh penggunaan, sebagai dasar pengambilan keputusan penyajian data dibuatkan secara mudah (Petra Refialy et al., 2021b). Tujuan dari data mining adalah untuk mengekstrak informasi dengan metode cerdas dari kumpulan data kemudian mengubah informasi menjadi struktur yang dapat dipahami untuk penggunaan lebih lanjut Menurut (Ramadhan & Kurniawati, 2020) data mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan yaitu:

1. Deskripsi, yaitu mencoba menemukan cara untuk menggambarkan pola dan trend yang kecenderungan terdapat dalam data.
2. Estimasi, yaitu variabel yang target estimasinya lebih kearah numerik dari arah kategori dan memiliki kemiripan dengan klasifikasi. Model yang dibangun juga menggunakan model record lengkap yang menyediakan nilai dari variabel target sebagai nilai prediksi.
3. Prediksi memiliki kemiripan dengan klasifikasi dan estimasi. Beberapa algoritma dan teknik yang digunakan dalam klasifikasi dan estimasi dapat juga digunakan untuk prediksi dalam keadaan yang tepat.
4. Klasifikasi dalam klasifikasi variabel dengan tujuan yang bersifat kategorik. Contohnya : pengklasifikasian persediaan dalam tiga kelas ( persediaan tinggi, persediaan sedang dan persediaan rendah).
5. Pengklusteran, suatu metode untuk mencari dan mengelompokkan data yang memiliki kemiripan karakteristik (*similarity*) antara satu data dengan data

yang lain. Clustering merupakan salah satu metode data mining yang bersifat tanpa arahan (*unsupervised*).

6. Asosiasi, tugas asosiasi dalam data mining adalah menemukan atribut yang muncul dalam suatu waktu.

## 2.2. Clustering

*Clustering* adalah proses kumpulan objek yang dikelompokkan ke dalam objek yang serupa. Dalam analisis data teknik clustering sangat penting dilakukan. Clustering melibatkan pembagian data menjadi beberapa kelompok atau berdasarkan kesamaan atribut antar kumpulan data. Atribut serupa ini direpresentasikan sebagai titik dalam ruang multidimensi (Febriyanti et al., 2022).

Penggunaan *clustering* untuk produk yang laku dan produk yang tidak laku semakin banyak dan data yang dimasukkan maka semakin besar clustering yang terbentuk, pada proses metode *k-means* tujuan dari *clustering* untuk menciptakan kelompok clustering yang terbaik dan optimal sehingga mampu meningkatkan penjualan pengelolaan pada kelebihan persediaan dan memperbaiki ketidakseimbangan dalam penjualan produk (Indriyani & Irfiani, 2019).

### 2.2.1. Analisa Clustering

Analisa *cluster* merupakan analisa statistik yang berguna untuk memisahkan objek ke beberapa kelompok yang memiliki sifat berbeda terhadap kelompok terhadap satu yang lain. Sifat analisis ini berbentuk homogen antar anggota pada kelompok atau dalam kelompok yang mempunyai variasi objek yang berbentuk kecil (Haris Kurniawan et al., 2020). Kegunaan utama analisis *cluster* adalah menghubungkan objek-objek yang memiliki kesamaan terhadap sebuah kelompok atau cluster. Dalam analisis cluster terdapat 6 tahapan dalam pengambilan keputusan antara lain. (Muliono & Sembiring, 2019) :

1. Menentukan desain penelitian analisis *cluster*.
2. Menentukan tujuan asumsi analisis.

3. Memperkirakan dan menurunkan *cluster*.
4. Melakukan interpretasi hasil *cluster*.
5. Mengukur hasil analisis *Cluster* berdasarkan tingkat validasi.

### 2.2.2. Sum Of Square Within Cluster (SSW)

Matriks ketertarikan yang diterapkan pemisah kelompok. Memanfaatkan jumlah kuadrat dalam cluster untuk mengukur jarak pemisah centroid  $C_i$ ,  $C_j$ .  $R_i$ , memberikan penilaian proposi yang menunjukkan kualitas nilai komparatif yang menghubungkan kelompok ke- $i$  dan ke- $j$ . Ukuran selanjutnya menguraikan prosedur untuk menentukan evaluasi.

berikut rumus persamaan menghitung matriks:

$$SSW = \sum_{i=1}^m \sum_{j=1}^m |x_{ij} - c_i|^2$$

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

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

$d$  = Jarak

### 2.2.3. Sum Of Square Between Cluster (SSB)

Tingkat sparasi antar kelompok menggunakan persamaan ini, berikut perhitungan (Santoso & Zulkarnain, 2018) :

$$SSB = \sum_{j=1}^{M-1} \sum_{i=1}^M |c_i - c_j|^2$$

Keterangan:

$d$  = Jarak

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

### 2.2.4. Ratio (Rasio)

Penentuan nilai banding antar cluster dengan rasio apabila kelompok dianggap signifikan sparasi tinggi dan keterikatan rendah, berikut persamaan (Khandava Mulyadien & Enri, n.d.) :

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_{i,j}}$$

Keterangan:

$SSWi$  = Sum of square within-cluster pada centroid  $i$

$SSBi,j$  = Sum of square between cluster data ke  $i$  dengan  $j$  pada cluster yang berbeda

### 2.2.5. Davies Bouldin Index (DBI)

Evaluasi menentukan nilai kualitas berdasarkan sparasi melalui jarak antar kelompok dan afinitas berasal padat dan dekat anggota kelompok (Yuliani, 2021).

Berikut rumus :

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

Berdasarkan hasil perhitungan ada kesimpulan semakin kecil nilai DBI diperoleh (non-negatif  $> 0$ ), kualitas kelompok dihasilkan dianggap semakin baik.

### 2.3. Penelitian Terkait

Penelitian ini tentang “ Implementasi Data Mining Metode *K-Means* terhadap Komisi Customer Penjualan Barang di PT. Dianrijaya Utama Mukti” ini tidak lepas dari acuan penelitian terdahulu yang diharapkan dapat memberikan gambaran dan memperjelas kerangka berfikir dalam pembahasan dan memperkuat landasan teori pada penelitian ini. Beberapa penelitian yang menjadi rujukan dalam penelitian ini yaitu sebagai berikut :

Tabel 2. 1. Hasil Penelitian Terdahulu

No	Nama Dan Tahun	Judul	Tujuan Penelitian	Teknik Analisis Data	Hasil Penelitian
1.	Risa Helilintar dan Intan Nur	Penerapan Algoritma K-Means Clustering Untuk Prediksi	untuk mengetahui prediksi prestasi akademik mahasiswa dengan		Berdasarkan dari hasil cluster kesimpulan yang dapat diambil adalah diperoleh 4

	Farida pada tahun 2018	Prestasi Nilai Akademik Mahasiswa	menggunakan metode k-Means clustering.	kelompok nilai mahasiswa berdasarkan Nilai IPK yaitu, sangat baik, baik, cukup, kurang. Hasil Cluster juga dipengaruhi dari nilai centroid awal yang dipakai dan jumlah data yang dipakai, perbedaan pengambilan data pusat centroid awal yang dipakai juga akan mempengaruhi hasil centroid akhirnya.
2.	Muhamad Khandava Mulyadien dan Ultach Enri pada tahun 2022	Algoritma K-Means Untuk Pengelompokan Bantuan Langsung Tunai (BLT)	membantu mengurangi para warga di Indonesia dari kemiskinan yang tinggi.	Cluster berdasarkan perhitungan SSE yang didapat dengan rincian cluster 6 sebanyak 519 jiwa, cluster 5 sebanyak 130 jiwa, cluster 4 sebanyak 178 jiwa, cluster 3 sebanyak 145 jiwa, cluster 2 sebanyak 466 jiwa, dan cluster 1 sebanyak 178 jiwa. Dengan data uji sebanyak 1632 records. Dari hasil tersebut dapat

					digunakan sebagai rekomendasi untuk penerima bantuan langsung tunai di Desa Dawuan Barat.
3.	Yunita Sari Siregar, dkk. Pada tahun 2023	Implementasi Data Mining Klasifikasi Algoritma Chaid Dalam Menentukan Pola Penerima Mahasiswa Baru	Merancang model pola karakteristik berdasarkan kriteria dalam seleksi penerimaan mahasiswa baru pada program studi teknik informatika menganalisis metode data mining decision tree dengan algoritma CHAID (Chi Squared Automatic Interaction Detection) dalam menentukan siapa	Metode data mining decision tree dengan algoritma CHAID (Chi Squared Automatic Interaction Detection)	Hasil pohon keputusan dalam perancangan pola karakteristik penerimaan calon mahasiswa baru terdapat 43 rule dimana , faktor tertinggi dari ke- 3 faktor yang mempengaruhi pada 75 data testing adalah kriteria Nilai Rapor. Faktor kedua yang juga mempengaruhi hasil keputusan adalah kriteria Ujian CAT (Computer Assisted

			yang lulus berdasarkan kriteria-kriteria yang telah ditentukan	Test). Dan faktor Nilai Wawancara terakhir yang juga mempengaruhi hasil keputusan adalah kriteria.
4.	Talitha Tiara Alifa, dkk. Pada tahun 2024	Implementasi Data Mining Menggunakan Algoritma K-Means Clustering Dalam Analisis Penjualan Produk	Untuk memperdalam pemahaman terhadap pola penjualan produk Fashion Pria, sehingga hasil analisis dapat memberikan wawasan yang berharga dalam mengelola data penjualan.	Penerapan K-Means Clustering diharapkan dapat meningkatkan pemahaman tentang tren penjualan, membantu mengidentifikasi produk yang paling diminati, dan memberikan dasar untuk mengoptimalkan strategi pemasaran. Dengan demikian, penelitian ini berperan penting dalam mengubah informasi menjadi pengetahuan yang dapat meningkatkan kinerja penjualan Fashion Pria. Kesimpulan dari penelitian ini memberikan landasan untuk pengembangan

					strategi bisnis yang lebih cerdas dalam menghadapi dinamika pasar yang terus berubah.
5.	Nur Afiasari, dkk. Pada tahun 2023	Implementasi Data Mining Transaksi Penjualan Menggunakan Algoritma Clustering dengan Metode K-Means	Pemecahan masalah dari informasi tersedia menggunakan data mining.		Mendapatkan 3 kelompok toko bill lights menjadikan adanya produksi dan strategi ulang.
6.	(Triyandana et al., 2022)	kelompok Menu Makanan, Minuman Berdasarkan Tingkat Penjualan.	Untuk mengetahui hasil cluster penjualan makanan dan minuman dengan menggunakan metode Algoritme K-Means.		Penjualan Dpom Coffee menjadi 3 cluster yakni penjualan rendah, sedang, dan tinggi.
7.	(Ramadanti & Muslih, 2022)	Penerapan Data Mining Algoritma K-Means Clustering Pada Populasi	Untuk mengetahui hasil perhitungan cluster pada populasi ayam petelur di Indonesia	Algoritme K-Means Clustering.	Hasil dari ketiga perhitungan yaitu 1 provinsi untuk cluster populasi tinggi yaitu Jawa Timur. Untuk cluster populasi

		Ayam Petelur Di Indonesia.			sedang ada 3 provinsi di antaranya Sumatera Utara, Jawa Barat dan Jawa Tengah. Sedangkan untuk cluster populasi rendah ada 30 provinsi.
8.	(Annisa et al., 2022)	Data Pengguna Air Bersih dari Keluhannya	Untuk mengetahui penggunaan air bersih dengan menggunakan cluster berdasarkan keluhannya	Clustering.	hasil yaitu jenis keluhan pada cluster 1 x 2 yaitu kode 5 jenis keluhan pipa bocor \ peanganan kerusakan menyambung pipa air (gibout join).
9.	(Magdalena & Fahrudin, 2020)	Koperasi Se-Jawa Barat pada Kementerian Koperasi dan UKM.	Untuk mengelompokkan data koperasi di Jawa Barat.		berdasarkan nilai Modal Sendiri, Modal Luar dan Volume Usaha, maka dihasilkan 3 klaster dengan nilai klaster tinggi, sedang dan rendah.
10.	(Ardiansyah et al., 2020)	Menentukan Status Provinsi di Indonesia 2020.	Untuk menentukan jenis wilayah di Indonesia menjadi zona merah, kuning dan hijau COVID-19.	Algoritme K-Means Clustering.	daerah masuk zona merah adalah Jakarta zona kuning yaitu Banten. sebanyak kurang dari 600 pasien yang terinfeksi dikategorikan sebagai zona hijau.

*Sumber : Data diolah pada tahun 2024*

#### 2.4. Data Penjualan

Data penjualan merupakan himpunan informasi yang merekam aktivitas transaksi penjualan suatu produk maupun layanan dalam rentang waktu tertentu. Informasi ini dapat mencakup beragam aspek yang berkaitan dengan proses penjualan tersebut. (Oktaviani & Bahtiar, 2023) :

1. Jumlah produk yang terjual
2. Total pendapatan yang dihasilkan
3. Lokasi penjualan dan jenis produk yang dijual
4. Segmen dan karakteristik pelanggan
5. Perilaku pembelian dan KPI
6. Metrik efisiensi

