SYSTEMATIC LITERATURE REVIEW: PEMETAAN KARAKTERISTIK IMPLEMENTASI DATA MINING PADA BERBAGAI BIDANG

Amhar Nabil Makarim[1], Gunawan Ari Nursanto[2], Priati Assiroj[3]
[1,2,3]Manajemen Teknologi Keimigrasian, Politeknik Imigrasi dan Pemasyarakatan
[1]amharnabilmakarim@gmail.com, [2]gunawan@poltekim.ac.id,
[3]priati.assiroj@poltekim.ac.id

ABSTRACT

This study aims to map the characteristic of data mining implementation based on 37 journals reviewed systematically. The method employed was a Systematic Literature Review (SLR) with PRISMA approach, consisting of identification, screening, eligibility, and inclusion stages. The data were analyzed descriptively to identify methods, algorithms, data types, and implementation result of data mining across various domains. The findings reveal that the most dominant task was clustering, with 27 articles, while K-Means was the most frequently used algorithm, appearing in 26 studies. The most widely studied implementation domain was economics and business with 12 articles, followed by health with 7 articles, social and government with 6 articles, also tourism and education with 5 articles each. In addition to K-Means, other algorithms identified in the reviewed studies included Apriori, FP-Growth, K-NN, Decision Tree/C4.5, Naïve Bayes, K-Medoids, and U-KMeans. In terms of evaluation, classification studies commonly used accuracy, association rule studies relied on support and confidence, while clustering studies tended to use measures such as the Davies-Bouldin Index and silhouette score. This study concludes that K-Means remains the most dominant algorithm in data mining implementation. The result of this review are expected to serve as a reference for understanding trends in data mining implementation and for supporting more focused future research.

***Keywords**: Data Mining, K-Means, Clustering*

ABSTRAK

Penelitian ini bertujuan memetakan karakteristik implementasi data mining berdasarkan 37 jurnal yang dikaji secara sistematis. Metode dilakukan dengan Systematic Literature Review (SLR) dengan pendekatan PRISMA, melalui tahapan identifikasi, penyaringan, kelayakan, dan inklusi artikel. Data dianalisis secara deskriptif untuk mengidentifikasi kecenderungan metode, algoritma, jenis data, serta hasil implementasi data mining pada berbagai bidang. Hasil artikel ini menunjukkan bahwa metode yang paling dominan adalah clustering dengan 27 artikel, sedangkan algoritma yang paling banyak digunakan Adalah K-Means dengan 26 artikel. Bidang implementasi yang paling banyak diteliti adalah ekonomi dan bisnis sebanyak 12 artikel, diikuti kesehatan 7 artikel, sosial dan pemerintahan 6 artikel, serta pariwisata dan pendidikan masing-masing 5 artikel. Selain K-Means,

algoritma lain yang digunakan meliputi Apriori, FP-Growth, K-NN, Decision Tree/C4.5, Naïve Bayes, K-Medoids, dan U-KMeans. Dari sisi evaluasi, penelitian klasifikasi umumnya menggunakan akurasi, penelitian association rule menggunakan support dan confidence, sedangkan penelitian clustering menggunakan ukuran seperti Davies-Bouldin Index dan silhouette score. Penelitian ini menyimpulkan bahwa K-Means masih menjadi algoritma yang paling dominan dalam implementasi data mining. Hasil kajian ini diharapkan dapat menjadi rujukan dalam memahami tren implementasi data mining serta mendukung penelitian lanjutan yang lebih terarah.

Kata Kunci: Data Mining, K-Means, Clustering

A. Pendahuluan

Perkembangan teknologi informasi telah mendorong pertumbuhan data dalam berbagai sektor sehingga organisasi tidak lagi hanya membutuhkan penyimpanan data juga pengolahan data untuk menghasilkan informasi dalam pengambilan keputusan. Dalam konteks ini, data mining berperan penting karena mampu mengekstraksi pola, hubungan, dan pengetahuan tersembunyi dari data berukuran besar. Salah satu teknik yang dominan digunakan adalah clustering, dimana mengelompokkan data berdasarkan kemiripan karakteristik, dengan algoritma K-Means sebagai metode yang paling populer karena sederhana, cepat, dan mudah diimplementasikan. Namun, K-Means juga memiliki keterbatasan, terutama pada penentuan jumlah cluster dan inisialisasi centroid awal (Sinaga & Yang, 2020).

Fenomena tersebut tercermin pada 37 jurnal yang dianalisis dalam penelitian ini. 37 jurnal tersebut didapat dari empat basis data yaitu, Google Scholar, Semantic Scholar, Crossref, dan OpenAlex dengan total 11.733 artikel diidentifikasi dan kemudian diseleksi menjadi 37 jurnal terpilih. Implementasi data mining ditemukan pada berbagai bidang, seperti penjualan menu food and beverage, produk fashion, frozen food, dan pengelompokan customer; kunjungan wistawan di Karawang, Jawa Timur, Kota Cirebon, Yogyakarta, dan Riau; kesehatan seperti penyebaran COVID-19, ISPA, gangguan tidur, HIV/AIDS, penyakit rumah sakit, dan pasien rawat inap; hingga bidang pendidikan dan sosial seperti penilaian kinerja dosen, pola belajar mahasiswa, perilaku penggunaan media sosial siswa,

penyandang disabilitas, serta lowongan pekerjaan. Sebagian besar penelitian tersebut menunjukkan kecenderungan penggunaan metode clustering, khususnya penggunaan algoritma K-Means, karena dinilai efektif dalam memetakan data ke dalam kelompok tinggi, sedang, dan rendah sesuai kebutuhan analisis (Syahputra et al., 2022; Gustipartsani et al., 2023; Gunawan & Purwayoga, 2022; Dewi et al., 2022; Lestari et al., 2024; Mubarok et al., 2025).

Penelitian ini dilakukan untuk menyusun *systematic literature review* terhadap 37 jurnal terkait implementasi data mining. Penelitian ini bertujuan mengidentifikasi karakteristik penelitian yang meliputi metode yang digunakan, algoritma yang diterapkan, jenis data yang diolah, dan hasil utama yang diperoleh dari masing-masing studi.

Hasil penelitian ini diharapkan bisa memberikan gambaran yang lebih terstruktur mengenai tren implementasi *data mining*, menjadi rujukan akademik dalam memahami dominasi penggunaan algoritma tertentu, serta membantu peneliti menentukan atribut analisis yang relevan untuk penelitian lanjutan, khususnya pada kajian *clustering* menggunakan K-Means.

B. Metode Penelitian

Systematic Literature Review (SLR) dengan pendekatan PRISMA (*Preferred Reporting Items for Systematic Reviews and Meta-Analyses*) digunakan dalam penelitian ini. Metode ini digunakan untuk memastikan proses pencarian, seleksi, dan pelaporan artikel dilakukan secara sistematis, transparan, dan dapat ditelusuri kembali. Dalam penelitian ini, PRISMA digunakan untuk menelaah artikel-artikel yang membahas implementasi *data mining*, khususnya yang berkaitan dengan metode, algoritma, data, dan hasil penelitian.

Tahapan utama PRISMA dalam penelitian ini meliputi:

1. Identifikasi (*Identification*)
   Pada tahap ini ditentukan kata kunci yang relevan dengan topik penelitian, kemudian dilakukan penelusuran artikel dari sumber literatur yang digunakan. Tahap ini bertujuan memperoleh kumpulan awal artikel yang berkaitan dengan implementasi *data mining*.
2. Penyaringan (*Screening*)
   Artikel yang telah ditemukan diseleksi dengan menghapus data duplikat, kemudian disaring berdasarkan judul,

abstrak, dan kata kunci menyesuaikan dengan fokus penelitian.

3. Kelayakan (*Eligibility*)
   Artikel yang lolos tahap penyaringan selanjutnya dibaca secara lebih rinci untuk menilai kesesuaiannya dengan kriteria penelitian, baik dari segi topik, metode, maupun hasil yang dilaporkan. Artikel yang tidak relevan tidak dilanjutkan ke tahap berikutnya.
4. Inklusi (*Included Studies*)

Artikel yang memenuhi semua kriteria ditetapkan sebagai studi akhir dan dianalisis lebih lanjut. Data yang didapatkan dari artikel terpilih kemudian diekstraksi ke dalam tabel sintesis untuk memetakan karakteristik penelitian.

Data yang telah dihimpun dari 4 sumber dengan kandidat 57 jurnal. Selanjutnya diseleksi kembali menjadi 37 jurnal. Hasil jurnal yang terpilih disajikan dalam Tabel 1 sebagai berikut:

Tabel 1 Pengumpulan Data

Source	**Studi Menemukan (berdasarkan judul dan kata kunci) "Implementasi *Data Mining*"**	**Kandidat**	**Artikel Terpilih**
Google Scholar	8006	25	15
Semantic Scholar	1763	7	6
OpenAlex	997	9	5
Crossref	967	16	11
Total	**11.733**	**57**	**37**

C. Hasil Penelitian dan Pembahasan

Tabel 2 Daftar Jurnal Terpilih

No	Judul	Tahun
1	Implementasi Data Mining untuk Pengelompokan…(Santoso & Zulkarnain, 2018)	2018
2	Implementasi Data Mining dalam Klasifikasi…(Purnomo & Prasetyaningrum, 2020)	2020
3	Unsupervised K-Means Clustering Algorithm (Sinaga & Yang, 2020)	2020
4	Clustering Tingkat Penjualan Menu…(Syahputra et al., 2022)	2022
5	Data Mining Menggunakan Algoritma K-Means…(Gunawan & Purwayoga, 2022)	2022
6	Data Mining Menggunakan Algoritma K-Means…(Ariyanto, 2022)	2022
7	Implementasi Data Mining Algoritma K-Means…(Hidayah et al., 2022)	2022
8	Implementasi Data Mining dalam Menganalisa…(Sihombing et al., 2022)	2022
9	Implementasi Data Mining Menggunakan Metode…(Simanjuntak et al., 2022)	2022
10	Detecting Learning Patterns in Tertiary…(Tuyishimire et al., 2022)	2022
11	Penerapan Data Mining untuk Clustering…(Dewi et al., 2022)	2022
12	Penerapan Data Mining Algoritma K-Means…(Ramadanti & Muslih, 2022)	2022
13	Penerapan Data Mining dalam Pemilihan…(Gustrianda & Mulyana, 2022)	2022
14	Penerapan Data Mining pada Tata...(Andini et al., 2022)	2022
15	The Application of K-Means Clustering…(Abdullah et al., 2022)	2022
16	Data Mining Clustering Menggunakan Algoritma...(Gustipartsani et al., 2023)	2023
17	Implementasi Algoritma K-Means dalam Analisis...(Sari et al., 2023)	2023

18	Implementasi Data Mining dalam Mengelompokkan…(Waruwu et al., 2023)	2023
19	Implementasi Data Mining dalam Pengelompokan…(Saragih et al., 2023)	2023
20	Implementasi Data Mining Menggunakan Metode…(Amran et al., 2023)	2023
21	Implementasi Data Mining pada Hasil…(Syahputri et al., 2023)	2023
22	Implementasi Data Mining untuk Pengelompokan…(Sartika & Jumadi, 2023)	2023
23	Implementasi Data Mining FP-Growth untuk…(Komariyah et al., 2023)	2023
24	Implementasi RapidMiner dengan Metode Clustering…(Nahjan et al., 2023)	2023
25	Penerapan Data Mining Menggunakan Algoritma…(Pratama & Jaman, 2023)	2023
26	Implementasi Algoritma K-Means Clustering dalam…(Maulida et al., 2024)	2024
27	Implementasi Data Mining dalam Mengelompokkan...(Lestari et al., 2024)	2024
28	Implementasi Data Mining pada Instansi...(Mustari et al., 2024)	2024
29	Implementasi Data Mining untuk Menentukan...(Pujiono et al., 2024)	2024
30	Implementasi K-Means Clustering Menggunakan RapidMiner…(Liyanto et al., 2024)	2024
31	Implementasi Teknik Data Mining pada...(Wulandari et al., 2024)	2024
32	Implementasi Data Mining Klasterisasi Data…(Laksono et al., 2024)	2024
33	Implementasi Data Mining dalam Penerapan…(Hamsiah, 2025)	2025
34	Implementasi Data Mining untuk Clustering…(Mubarok et al., 2025)	2025
35	Implementasi Data Mining untuk Klasifikasi…(Fhardana & Siddik, 2025)	2025
36	Implementasi Data Mining untuk Menentukan…(Ambroini & Sari, 2025)	2025
37	Implementasi Data Mining dalam Pengelompokan…(Utari & Iqbal, 2026)	2026

Berdasarkan proses seleksi yang telah dilakukan, penelitian ini menghasilkan 37 jurnal terpilih yang menjadi fokus kajian. Daftar jurnal tersebut disajikan pada Tabel 2. Dilihat dari tahun terbit, publikasi tersebar pada rentang 2018-2026, dengan konsentrasi terbesar pada periode 2022-2024. Tahun 2022 menjadi periode paling dominan dengan 12 artikel, diikuti tahun 2023 sebanyak 10 artikel, dan tahun 2024 sebanyak 7 artikel. Pola ini menunjukkan bahwa penelitian mengenai implementasi *data mining* mengalami peningkatan pada beberapa tahun terakhir, terutama ketika kebutuhan pengolahan data untuk mendukung keputusan praktis di berbagai sektor semakin tinggi.

Tabel 3 Analisis Jurnal

No	**Judul**	**Clustering Tingkat Penjualan Menu (2022)**
1	**Metode**	KDD
	Teknik	*Clustering*
	Algoritma	K-Means
	Data	Data transaksi penjualan
	Jumlah Data	Histori transaksi selama 1 tahun
	Evaluasi	Tidak dilaporkan
	Hasil	K-Means membagi histori transaksi menu F&B selama 1 tahun ke 4 cluster: penjualan tinggi 5 menu, sedang 11, rendah 25, dan sangat rendah 131. Paper tidak melaporkan akurasi atau validasi kuantitatif.
2	**Judul**	**Data Wisatawan di Kabupaten Karawang (2023)**

	Metode	KDD
	Teknik	Clustering
	Algoritma	K-Means
	Data	Data kunjungan wisatawan dari opendata.jabarprov.go.id
	Jumlah Data	214.794 wisatawan
	Evaluasi	K optimal = 3; metrik evaluasi numerik tidak dilaporkan jelas
	Hasil	Data kunjungan wisatawan Karawang dikelompokkan ke 3 cluster (rendah, sedang, tinggi). Nilai DBI 0,422 pada k=3 digunakan sebagai konfigurasi terbaik.
3	**Judul**	**Potensi Penyebaran Virus Corona di Kota Cirebon (2022)**
	Metode	Tidak memakai KDD/CRISP-DM sebagai framework utama. Metodenya berupa alur penelitian 3 tahap: pra-penelitian, persiapan penelitian, proses penelitian; di dalamnya ada studi pustaka, pengumpulan data, cleaning, clustering, perhitungan K-Means
	Teknik	Clustering
	Algoritma	K-Means
	Data	Variabel kontak erat, suspek, terkonfirmasi dari beberapa kelurahan
	Jumlah Data	22 Kelurahan
	Evaluasi	Tidak dilaporkan
	Hasil	Sebanyak 22 kelurahan Kota Cirebon dibagi menjadi 3 cluster potensi penyebaran COVID-19: tinggi 4, sedang 5, dan rendah 13 kelurahan. Paper tidak melaporkan metrik akurasi kuantitatif.
4	**Judul**	**Klasifikasi Penyakit Infeksi Saluran Pernafasan Akut (ISPA) (2022)**
	Metode	Metodenya berupa kerangka kerja penelitian: penentuan masalah/tujuan, mempelajari literatur, pengumpulan data, analisa data, dengan teknik observasi, wawancara, studi pustaka. KDD hanya dijadikan landasan teori
	Teknik	Clustering
	Algoritma	K-Means
	Data	Pasien Klinik Rahmatan Lil Alamin
	Jumlah Data	250 data pasien
	Evaluasi	Tidak dilaporkan
	Hasil	Dari 250 data pasien, K-Means menghasilkan 3 cluster ISPA: biasa 141, sedang 17, dan berat 7. Paper menekankan hasil pengelompokan, tetapi tidak menyajikan akurasi numerik.
5	**Judul**	**Penentuan Gangguan Tidur Seseorang berdasarkan Gaya Hidup (2024)**
	Metode	Tidak memakai framework baku. Metodenya adalah tahapan penelitian custom: pengumpulan data dari Kaggle, preprocessing, proses K-Means clustering, evaluasi silhouette score
	Teknik	Clustering
	Algoritma	K-Means
	Data	Dataset Sleep Health and Lifestyle (Kaggle)
	Jumlah Data	Tidak dilaporkan jelas
	Evaluasi	*Silhouette Score* pada data *Sleep Duration* dan *Quality of Sleep* untuk jumlah cluster sebanyak 3 memiliki skor 0,6319914771456847 menunjukkan *medium structure*. *Silhouette Score* pada data *Heart Rate* dan *Daily Steps* dengan

		jumlah cluster sebanyak 3 memiliki skor 0,7233444078004578 menunjukkan *strong structure*; penelitian ini juga memakai Elbow Method
	Hasil	Elbow method memilih k=3. Model Sleep Duration-Quality of Sleep memperoleh silhouette 0,632, sedangkan model Heart Rate-Daily Steps memperoleh silhouette 0,723; hasil cluster memetakan insomnia, sleep apnea, dan tidak ada gangguan tidur.
6	**Judul**	**Klasterisasi Penyebaran Penyakit HIV/AIDS (2023)**
	Metode	Metode penelitian yang dinyatakan adalah kuantitatif dan eksperimental. Bukan KDD/CRISP-DM
	Teknik	Clustering
	Algoritma	K-Means
	Data	Kasus HIV/AIDS per kecamatan di Karawang; 1992-2022
	Jumlah Data	2.052 penduduk
	Evaluasi	Tidak dilaporkan
	Hasil	Kasus HIV/AIDS per kecamatan di Karawang dipetakan ke 4 cluster: rendah 73%, sangat tinggi 3%, sedang 7%, dan tinggi 17%. Paper juga menampilkan uji DBI dengan nilai terendah 0,073 pada k=2, sedangkan hasil akhir dibahas pada 4 cluster.
7	**Judul**	**Clustering Penyakit di RS Panti Waluyo Surakarta (2022)**
	Metode	Metode pengembangan sistem yang dipakai eksplisit Prototyping
	Teknik	Clustering
	Algoritma	K-Means
	Data	Data rekam medis Februari 2019; umur, kode penyakit, kecamatan
	Jumlah Data	200 data
	Evaluasi	Tidak dilaporkan
	Hasil	Sebanyak 200 data rekam medis dikelompokkan ke 3 cluster berdasarkan umur, kode penyakit, dan kecamatan. Cluster dominan menunjukkan penyakit Chronic ischaemic heart disease, tetapi metrik akurasi tidak dilaporkan.
8	**Judul**	**Mengelompokkan Penduduk Penyandang Disabilitas (2024)**
	Metode	Metodenya berupa penelitian kuantitatif dengan tahapan Preliminary Study, Literature Review, Data Collecting, Data Preparation, Analisis Data, Evaluasi. Jurnal ini juga menunjukkan tahapan Data Selection, Preprocessing, Data Transformation, Data Mining, Evaluation, Knowledge yang sangat mirip KDD, tetapi ditulis sebagai alur analisis penelitian.
	Teknik	Clustering
	Algoritma	K-Means
	Data	Kabupaten/kota di Jawa Barat
	Jumlah Data	27 kabupaten/kota
	Evaluasi	DBI = 0,121
	Hasil	Data penyandang disabilitas kabupaten/kota di Jawa Barat menghasilkan 2 cluster optimal dengan DBI 0,121. Cluster 1 berisi 27 kabupaten/kota dan Cluster 2 berisi 2 kabupaten/kota.
9	**Judul**	**Klasifikasi Kunjungan Wisatawan Kota Yogyakarta (2020)**
	Metode	Metodenya ditulis sebagai workflow data mining: pengumpulan data, pre-processing, clustering, analisis. Jurnal ini juga menjelaskan KDD secara cukup kuat pada bagian preprocessing, jadi bisa disebut KDD-oriented workflow, walau tidak setegas paper KDD murni.

	Teknik	Clustering
	Algoritma	K-Means
	Data	Data wisatawan
	Jumlah Data	48 data; 3 atribut: objek wisata, bulan, tahun
	Evaluasi	Tidak dilaporkan
	Hasil	Data kunjungan wisatawan Kota Yogyakarta diproses dalam 5 iterasi dan dibagi ke 3 cluster, masing-masing 24 data (50%), 11 data (23%), dan 13 data (27%). Tidak ada akurasi kuantitatif yang dilaporkan.
10	**Judul**	**Pola Penjualan Roti (2022)**
	Metode	Metodenya adalah kerangka kerja algoritma FP-Growth, bukan CRISP-DM/KDD eksplisit.
	Teknik	Association Rule
	Algoritma	FP-Growth
	Data	Data transaksi
	Jumlah Data	30 transaksi September 2021 pada French Bakery
	Evaluasi	Support dan confidence
	Hasil	FP-Growth pada 30 transaksi French Bakery menghasilkan pola pembelian untuk strategi marketing. Rule terkuat adalah Batter Keju -> Donat Ceres dengan support sebesar 23,33% dan confidence sebesar 77,78%.
11	**Judul**	**Mengelompokkan Data Penduduk Kurang Mampu (2023)**
	Metode	Metode penelitian yang ditulis adalah tahapan penelitian/data collecting, terdiri dari observasi, wawancara, studi literatur; lalu diterapkan K-Means clustering.
	Teknik	Clustering
	Algoritma	K-Means
	Data	Data penduduk Desa Hiliwaele II
	Jumlah Data	34 penduduk
	Evaluasi	Tidak dilaporkan jelas
	Hasil	Data dari penduduk kurang mampu Desa Hiliwaele II dibagi menjadi 2 cluster setelah 3 iterasi, yaitu cluster normal 15 orang dan cluster rendah 19 orang. Paper tidak melaporkan evaluasi kuantitatif seperti DBI atau silhouette.
12	**Judul**	**Sebaran Mahasiswa Baru pada STIE-SAK (2025)**
	Metode	Metodenya berupa kerangka penelitian custom: pengumpulan data, pengolahan data (seleksi, cleaning, transformasi, data mining, evaluasi), clustering, analisis. Tidak eksplisit KDD/CRISP-DM.
	Teknik	Clustering
	Algoritma	K-Medoid
	Data	Data mahasiswa baru 2024 (nama, gender, asal sekolah, alamat)
	Jumlah Data	322 mahasiswa
	Evaluasi	Tidak dilaporkan jelas
	Hasil	K-Medoid membentuk 3 cluster sebaran mahasiswa baru STIE-SAK: Kerinci tertinggi, Sungai Penuh sedang, dan luar daerah rendah. Paper menyebut silhouette coefficient, tetapi nilainya tidak dirinci.
13	**Judul**	**Perilaku Penggunaan Media Sosial Siswa SMA (2026)**
	Metode	Metodenya adalah tahapan penelitian custom: identifikasi masalah, pengumpulan data, studi literatur, penerapan algoritma K-Means, kesimpulan.
	Teknik	Clustering

	Algoritma	K-Means
	Data	Kuesioner tertutup; frekuensi, durasi, jumlah platform
	Jumlah Data	20 data
	Evaluasi	Tidak dilaporkan
	Hasil	Dengan normalisasi Min-Max dan K=3, perilaku penggunaan media sosial siswa SMA dikelompokkan menjadi rendah, sedang, dan tinggi. Studi tidak menyajikan akurasi atau validasi internal cluster secara numerik.
14	**Judul**	**Potensi Kejahatan pada Kapolsek Beringin (2023)**
	Metode	Metode penelitian yang dinyatakan adalah deskriptif kualitatif dan field research, disertai pengumpulan data dan studi pustaka.
	Teknik	Clustering
	Algoritma	K-Means
	Data	Data kejahatan pada wilayah Polsek Beringin
	Jumlah Data	27 wilayah
	Evaluasi	Tidak dilaporkan
	Hasil	Data potensi kejahatan Kapolsek Beringin dikelompokkan ke 2 cluster, yaitu tingkat kejahatan tinggi dan rendah. Wilayah tinggi meliputi Emplasmen Kwala Namu, Beringin, Denai Kuala, Denai Lama, Kubah Sentang, dan Pematang Biara; metrik akurasi tidak dilaporkan.
15	**Judul**	**Mempolakan Penjualan pada Showroom Motor Bekas (2023)**
	Metode	Metode yang digunakan adalah data mining klasifikasi menggunakan Decision Tree C4.5, lalu evaluasi confusion matrix/akurasi.
	Teknik	Classification
	Algoritma	Decision Tree/C4.5
	Data	Data motor
	Jumlah Data	27 data
	Evaluasi	Tidak dilaporkan jelas
	Hasil	Decision Tree/C4.5 pada 27 data showroom motor bekas mencapai accuracy 100%. Hasil ini menunjukkan performa klasifikasi sangat tinggi, meskipun ukuran data relatif kecil.
16	**Judul**	**Kenaikan Pangkat Dinas Ketenagakerjaan Kota Medan (2022)**
	Metode	Tidak ada framework baku yang dinyatakan eksplisit. Metodenya adalah data mining klasifikasi dengan Naïve Bayes Classifier, berbasis observasi dan wawancara, diimplementasikan dengan Python.
	Teknik	Classification
	Algoritma	Naïve Bayes
	Data	Data primer Dinas Ketenagakerjaan; atribut pendidikan, golongan, usia, prestasi, masa kerja
	Jumlah Data	38 data
	Evaluasi	Tidak dilaporkan
	Hasil	Naïve Bayes digunakan untuk mengklasifikasikan kelayakan kenaikan pangkat PNS berdasarkan atribut pendidikan, golongan, usia, prestasi, dan masa kerja. Studi menyebut hasil klasifikasi akurat dan efektif, tetapi tidak mencantumkan angka akurasi numerik.
17	**Judul**	**Hasil Penjualan Makanan Beku (2023)**

	Metode	Metodenya adalah workflow 5 langkah: studi literatur, analisis permasalahan, pengumpulan data, implementasi, evaluasi dan analisis hasil.
	Teknik	Clustering
	Algoritma	K-Means
	Data	Data penjualan *frozen food* Agustus 2021
	Jumlah Data	22 data
	Evaluasi	Tidak dilaporkan
	Hasil	K-Means pada 22 data penjualan frozen food membagi produk menjadi 2 cluster, yaitu sangat laku dan laku. Hasil ini membantu penentuan stok, tetapi paper tidak menampilkan evaluasi kuantitatif seperti DBI atau silhouette.
18	**Judul**	**Implementasi Data Mining pada Instansi Pemerintahan (SLR) (2024)**
	Metode	PRISMA
	Teknik	SLR / Literature Review
	Algoritma	Multi-algorithm data mining
	Data	Artikel-artikel implementasi data mining di instansi pemerintahan
	Jumlah Data	Bukan studi dataset empiris
	Evaluasi	Tidak ada
	Hasil	Paper ini merupakan systematic literature review tentang implementasi data mining pada instansi pemerintahan, sehingga tidak melaporkan akurasi model tunggal. Hasil utamanya berupa sintesis bahwa data mining mendukung efisiensi layanan publik dan pengambilan keputusan.
19	**Judul**	**Clustering Lowongan Pekerjaan (2025)**
	Metode	Metode yang dipakai eksplisit CRISP-DM, dengan tahap business understanding, data understanding, data preparation, modeling, evaluation, deployment.
	Teknik	Clustering
	Algoritma	K-Means
	Data	Data lowongan pekerjaan
	Jumlah Data	33.000 *record*
	Evaluasi	Silhouette Score = 0,78; DBI = 0,55
	Hasil	Pada 33.000 data LinkedIn Job Postings 2023, K-Means membentuk 4 cluster lowongan kerja. Nilai silhouette 0,78048 dan DBI 0,55457 menunjukkan kualitas cluster yang baik serta mendukung dashboard segmentasi pekerjaan.
20	**Judul**	**Klasifikasi Kunjungan Wisatawan Riau (2025)**
	Metode	Tidak ada framework baku seperti KDD/CRISP-DM yang ditegaskan. Metodenya adalah pipeline klasifikasi: pengumpulan data, preprocessing, transformasi data, implementasi K-NN, pembagian data latih-uji, dan evaluasi confusion matrix.
	Teknik	Classification
	Algoritma	K-NN
	Data	Data kunjungan 2021-2023
	Jumlah Data	11 data tempat wisata
	Evaluasi	Evaluasi hasil menggunakan confusion matrix untuk menguji akurasi metode K-Nearest Neighbor juga menunjukkan hasil yang baik.

	Hasil	K-NN mengklasifikasikan kunjungan wisatawan Riau dengan accuracy 93,75% pada split 80/20. Akurasi per label adalah 100% untuk ramai, 75% untuk cukup ramai, dan 96% untuk tidak ramai.
21	**Judul**	**Pola Penjualan pada Coffee Shop (Menrabic) (2025)**
	Metode	Metode penelitian yang saya temukan adalah pendekatan kuantitatif dan dikategorikan sebagai penelitian eksperimental. Jurnal ini tidak menegaskan KDD/CRISP-DM sebagai framework utama.
	Teknik	Association Rule
	Algoritma	Apriori
	Data	Data transaksi penjualan makanan/minuman Menrabic Coffee Shop
	Jumlah Data	293 data transaksi
	Evaluasi	Minimum support = 0,5%; minimum confidence = 15%; contoh rule juga melaporkan lift, misalnya 4,27
	Hasil	Apriori pada data transaksi Menrabic Coffee Shop menemukan rule terkuat Coffee Abay Sanger Chocolate dengan support 0,7%, confidence 17,95%, dan lift 4,27.
22	**Judul**	**Pola Penjualan Produk Fashion (Merek X) (2024)**
	Metode	Metode yang dipakai eksplisit KDD (Knowledge Discovery in Database), dengan tahapan data selection, preprocessing, transformation, data mining, interpretation/evaluation.
	Teknik	Clustering
	Algoritma	K-Means
	Data	Data penjualan April-Juli
	Jumlah Data	624 *record*
	Evaluasi	DBI = -0,310
	Hasil	Pada 624 data penjualan produk fashion April-Juli, K-Means terbaik dipilih pada k=2 dengan DBI -0,310 menurut paper. Hasil akhir menempatkan 616 item pada cluster paling laris dan 8 item pada cluster kurang laris.
23	**Judul**	**Pengelompokan Customer PT. Kasana Teknindo Gemilang (2018)**
	Metode	Analisa sistem berjalan dan flowchart/tahapan proses algoritma K-Means sebagai metode kerja penelitian.
	Teknik	Clustering
	Algoritma	K-Means
	Data	Data transaksi customer; atribut quantity, jumlah transaksi, total harga
	Jumlah Data	Data dari 13 *customer*
	Evaluasi	Tidak dilaporkan jelas
	Hasil	K-Means mengelompokkan customer PT. Kasana Teknindo Gemilang berdasarkan quantity, jumlah transaksi, dan total harga ke 3 cluster potensial. Hasil utamanya berupa penentuan centroid customer potensial C1-C3 tanpa metrik evaluasi kuantitatif.
24	**Judul**	**Wilayah Pelanggaran Lalu Lintas Polres Bengkulu (2023)**
	Metode	Metode penelitian yang disebut eksplisit adalah SDLC (System Development Life Cycle).
	Teknik	Clustering
	Algoritma	K-Means
	Data	Data pelanggaran lalu lintas wilayah Polres Bengkulu

	Jumlah Data	Tidak dilaporkan jelas
	Evaluasi	Tidak dilaporkan
	Hasil	Wilayah pelanggaran lalu lintas Polres Bengkulu dikelompokkan menjadi 3 cluster. Tanah Patah dan Betungan masuk kelompok tinggi, beberapa wilayah lain masuk sedang, sedangkan Rawa Makmur dan Sungai Hitam masuk rendah; akurasi numerik tidak dilaporkan.
25	**Judul**	**Analisis Pola Pembelian Online Shop Ayu Collection (2023)**
	Metode	Metode penelitian yang dinyatakan adalah penelitian perpustakaan dan penelitian lapangan (observasi dan wawancara). Bukan CRISP-DM; KDD hanya dibahas sebagai konsep teoretis.
	Teknik	Association Rule
	Algoritma	FP-Growth
	Data	Data transaksi
	Jumlah Data	250 transaksi; 17 item produk fashion
	Evaluasi	Minimum support = 40%; confidence = 50%
	Hasil	FP-Growth pada 250 transaksi Ayu Collection menghasilkan aturan asosiasi untuk rekomendasi produk. Dengan minimum support 40% dan confidence 50%, rule tertinggi mencapai confidence 79,1%.
26	**Judul**	**Kunjungan Wisatawan Asing Jawa Timur (2024)**
	Metode	KDD
	Teknik	Clustering
	Algoritma	K-Means
	Data	Data kunjungan wisatawan mancanegara (ASIA) 2016-2023 via Bandara Juanda dari BPS Jatim
	Jumlah Data	50 negara
	Evaluasi	DBI = 0,293
	Hasil	Data wisatawan mancanegara ASIA melalui Bandara Juanda 2016-2023 dikelompokkan menjadi 2 cluster optimal dengan DBI 0,293. Cluster 0 berisi 2 data dan cluster 1 berisi 48 data.
27	**Judul**	**Analisa Penjualan pada Toko OJ Cell (2023)**
	Metode	Tidak ada framework baku yang ditegaskan. Yang digunakan adalah alur metode penelitian sistematis: tahap pengumpulan data lalu pemrosesan di RapidMiner dengan K-Means.
	Teknik	Clustering
	Algoritma	K-Means
	Data	Data transaksi
	Jumlah Data	139 *record*
	Evaluasi	Tidak dilaporkan; hanya hasil cluster 50, 24, dan 65 item
	Hasil	RapidMiner dan K-Means membagi 139 transaksi OJ Cell ke 3 cluster penjualan dengan komposisi 50, 24, dan 65 item. Cluster profit tertinggi tercatat pada cluster 1 sebesar Rp5.549,95, tetapi paper tidak menampilkan DBI atau silhouette.
28	**Judul**	**Data Kunjungan Wisatawan Hotel Bintang Kota Cirebon (2024)**
	Metode	KDD
	Teknik	Clustering
	Algoritma	K-Means

	Data	Data kunjungan Hotel Bintang tahun 2021
	Jumlah Data	240 *record*
	Evaluasi	Uji DBI beberapa K: K2 = 0,865; K3 = 0,590; K4 = 0,456; K5 = 0,430; K6 = 0,458; K7 = 0,413; K8 = 0,421; K9 = 0,477; K10 = 0,499. Nilai terbaik: DBI = 0,413 pada K = 7
	Hasil	Sebanyak 240 data kunjungan hotel bintang Kota Cirebon tahun 2021 dibagi menjadi 7 cluster, dari tingkat kunjungan paling rendah hingga sangat tinggi. Nilai DBI terbaik 0,413 pada k=7, dengan cluster tertinggi mencapai 6.702.111 pengunjung.
29	**Judul**	**Klasterisasi Data Pasien Rawat Inap (2024)**
	Metode	Metode yang digunakan adalah kerangka kerja penelitian: analisis masalah, kajian teori, pengumpulan data, analisis dan pengolahan data, analisa hasil, pengambilan keputusan; sumbernya lewat library search dan field research.
	Teknik	Clustering
	Algoritma	K-Means
	Data	Data pasien
	Jumlah Data	10.000 data pasien
	Evaluasi	Tidak ada metrik evaluasi formal yang jelas; hanya hasil pembagian cluster
	Hasil	Klasterisasi 10.000 pasien rawat inap menggunakan variabel kategori usia dan diagnosis menghasilkan 4 cluster. Elbow menunjukkan titik optimal pada k=4 dan silhouette score 0,60 menandakan struktur cluster cukup baik.
30	**Judul**	**Detecting Learning Patterns (2022)**
	Metode	Dynamic K-Means clustering dengan cluster migration analysis untuk evaluasi perkembangan belajar mahasiswa.
	Teknik	Clustering
	Algoritma	K-Means
	Data	Data mahasiswa dari South African tertiary institution / Univ. of Johannesburg
	Jumlah Data	703 mahasiswa
	Evaluasi	Silhouette = 0,39 untuk 5 cluster; Silhouette = 0,58 untuk 3 cluster
	Hasil	Paper internasional ini menggunakan K-Means untuk mendeteksi pola belajar mahasiswa dan menganalisis perpindahan antarklaster performa. Silhouette 0,39 saat 5 cluster dan 0,58 saat 3 cluster menunjukkan bahwa konfigurasi 3 cluster lebih kompak.
31	**Judul**	**Kebiasaan Konsumen dan Prediksi Stok (2023)**
	Metode	Metode berupa workflow Apriori terstruktur: pengumpulan data, preprocessing, implementasi Apriori di Google Colab/Python, dan evaluasi aturan dengan lift ratio.
	Teknik	Association Rule
	Algoritma	Apriori
	Data	Data transaksi Toko Elektronik WK
	Jumlah Data	Data transaksi selama 30 hari
	Evaluasi	Minimum confidence = 0,6; ditemukan 16 aturan asosiasi; beberapa rule memiliki confidence = 1,0; lift ratio = 1,0 atau lebih; support tertinggi juga dilaporkan 100%
	Hasil	Apriori pada data transaksi 30 hari Toko Elektronik WK menghasilkan 16 aturan asosiasi untuk kebiasaan konsumen dan prediksi stok. Beberapa rule memiliki confidence 1,0, misalnya Amplifier Mini -> Amplifier, dengan lift ratio 1,0.

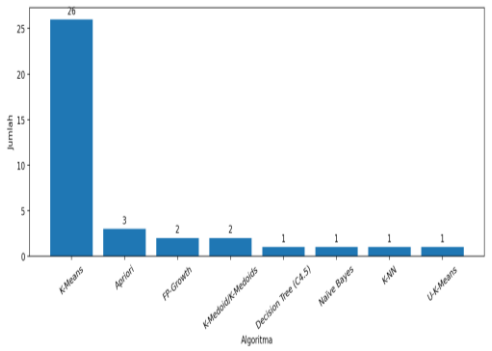
32	**Judul**	**Clustering Penilaian Kinerja Dosen STMIK Primakara (2022)**
	Metode	Metode yang dipakai eksplisit KDD, dengan tahapan data selection, preprocessing/cleaning, transformation, data mining, interpretation/evaluation.
	Teknik	Clustering
	Algoritma	K-Means
	Data	Data mahasiswa
	Jumlah Data	983 data mahasiswa
	Evaluasi	DBI = 0,270
	Hasil	K-Means pada 983 data mahasiswa menghasilkan beberapa cluster yaitu, 4 cluster penilaian kinerja dosen: sangat baik 312, baik 401, cukup baik 189, dan kurang baik 81. Nilai DBI 0,270 menunjukkan hasil cluster relatif baik karena mendekati nol.
33	**Judul**	**Populasi Ayam Petelur di Indonesia (2022)**
	Metode	Metode berupa studi literatur, pengumpulan data sekunder dari BPS, lalu pengolahan manual di Excel dan bantuan RapidMiner/Orange.
	Teknik	Clustering
	Algoritma	K-Means
	Data	Data ayam petelur
	Jumlah Data	Data ayam petelur pada 34 provinsi di Indonesia tahun 2016-2020
	Evaluasi	Tidak dilaporkan
	Hasil	Populasi ayam petelur 34 provinsi Indonesia periode 2016-2020 dikelompokkan menjadi 3 cluster. Hasil konsisten antara perhitungan manual, RapidMiner, dan Orange, yaitu 1 provinsi tinggi, 3 sedang, dan 30 rendah; paper tidak melaporkan metrik akurasi formal.
34	**Judul**	**Pemilihan Produk Unggulan PT XYZ (2022)**
	Metode	Metode berupa kerangka pemikiran penelitian: analisa & identifikasi masalah, pengumpulan data, analisis sistem, implementasi metode, pengujian, hasil pengujian.
	Teknik	Comparative Clustering
	Algoritma	K-Means dan K-Medoids
	Data	Sampel produk, laporan penjualan, dan stok gudang
	Jumlah Data	91 sampel produk
	Evaluasi	DBI K-Means = -0,430; DBI K-Medoids = -1,392.
	Hasil	Studi komparatif ini mengelompokkan 91 produk PT XYZ ke 3 cluster produk unggulan, sedang, dan rendah menggunakan K-Means dan K-Medoids. Paper melaporkan DBI K-Means -0,430 dan K-Medoids -1,392, tetapi interpretasi hasil antarbagian paper tampak tidak sepenuhnya konsisten.
35	**Judul**	**Tata Letak Buku di Perpustakaan Sintong Bingei (2022)**
	Metode	Metode yang digunakan adalah rancangan penelitian/flowchart penelitian: pengumpulan data, pengolahan ke Excel, implementasi algoritma Apriori, pengolahan RapidMiner, evaluasi rancangan.
	Teknik	Association Rule
	Algoritma	Apriori
	Data	Database peminjaman buku perpustakaan
	Jumlah Data	Tidak dilaporkan

	Evaluasi	Rule utama: support = 50%, confidence = 86%; rule lain juga muncul, misalnya 41% dan 83%
	Hasil	Apriori pada data peminjaman buku perpustakaan menghasilkan pola tata letak buku terkuat Ilmu Murni -> Ilmu Sosial dengan support 50% dan confidence 86%.
36	**Judul**	**Province Clustering of COVID-19 Risk in Indonesia (2022)**
	Metode	Metode yang dinyatakan adalah Methods and Statistical Analysis, yaitu analisis K-Means clustering berbasis data COVID-19 menggunakan R.
	Teknik	Clustering
	Algoritma	K-Means
	Data	Data SATGAS COVID-19 per 19 April 2020
	Jumlah Data	3032 kasus COVID
	Evaluasi	Elbow, Silhouette, dan Gap Statistic disebut untuk pemilihan cluster, tetapi nilai akhir masing-masing metrik tidak dilaporkan jelas pada ringkasan hasil; hasil akhirnya 3 cluster
	Hasil	Paper internasional ini mengelompokkan 34 provinsi Indonesia berdasarkan kasus confirmed, recovered, dan death COVID-19. Meskipun elbow, silhouette, dan gap statistic menunjukkan k optimal 2, hasil akhir dipresentasikan menjadi 3 kelompok provinsi.
37	**Judul**	**Unsupervised K-Means Clustering Algorithm (2020)**
	Metode	Metode yang digunakan adalah merancang U-k-means sebagai unsupervised learning schema, lalu diuji secara eksperimen dan dibandingkan dengan metode lain.
	Teknik	Algorithm Development
	Algoritma	U-K-Means
	Data	Eksperimen simulasi dan benchmark clustering
	Jumlah Data	Studi teoretis/simulasi
	Evaluasi	Menggunakan banyak indeks: DB, Silhouette Width, CH, Gap Statistic, dll., tetapi bukan satu nilai evaluasi tunggal karena diuji pada beberapa skenario eksperimen
	Hasil	Paper metode ini mengusulkan U-K-Means yang dapat menentukan jumlah cluster optimal secara otomatis tanpa inisialisasi awal. Kontribusi utamanya bukan angka akurasi tunggal, melainkan peningkatan skema unsupervised K-Means.

Berdasarkan Tabel 3, karakteristik penelitian didominasi oleh metode clustering, yaitu sebanyak 27 artikel atau 72,97% dari total kajian. Sementara itu, *association rule* muncul pada 5 artikel (13,51%), *classification* pada 3 artikel (8,11%), serta masing-masing 1 artikel untuk *systematic literature review* dan *algorithm development* (2,7%). Dominasi *clustering* menunjukkan bahwa sebagian besar penelitian implementasi *data mining* masih berfokus pada kebutuhan pemetaan kelompok, segmentasi objek, dan penentuan prioritas berdasarkan

karakteristik data. Pola ini tampak pada penelitian di bidang penjualan, kesehatan, pariwisata, pendidikan, dan pemerintah yang umumnya menggunakan *clustering* untuk memisahkan objek ke dalam kategori tinggi, sedang, dan rendah sesuai kebutuhan analisis (Syahputra et al., 2022; Gunawan & Purwayoga, 2022; Dewi et al., 2022; Mustari et al., 2024).

Gambar 1 Penggunaan Algoritma



Dari sisi algoritma, Gambar 1 menunjukkan bahwa K-Means merupakan algoritma yang paling dominan dengan 26 kemunculan atau sekitar 70,27%. Setelah itu, algoritma Apriori muncul sebanyak 3 kali (8,11%), FP-Growth dan K-Medoids masing-masing 2 kali (5,41%), sedangkan K-NN, Decision Tree/C4.5, Naïve Bayes, dan U-KMeans masing-masing 1 kali (2,7%). Dominasi K-Means mengindikasikan bahwa penelitian implementasi *data mining* masih cenderung memilih algoritma yang sederhana, cepat, dan mudah diterapkan pada data numerik maupun data tabular. Hal ini juga sejalan dengan karakter K-Means yang masih dianggap efektif untuk proses segmentasi awal, meskipun memiliki keterbatasan pada penentuan jumlah cluster dan sensitivitas terhadap centroid awal (Sinaga & Yang, 2020). Di sisi lain, keberadaan Apriori dan FP-Growth menunjukkan bahwa *data mining* pada data transaksi lebih banyak diarahkan untuk menemukan pola pembelian, sedangkan K-NN, Naïve Bayes, dan Decision Tree digunakan ketika data telah memiliki label kelas dan tujuan analisisnya bersifat prediktif (Arman et al., 2023; Fhardana & Siddik, 2025).

Tabel 3 juga menunjukkan bahwa data yang digunakan sangat beragam, baik dari jenis maupun ukuran. Pada skala kecil, terdapat penelitian yang hanya menggunakan 20 sampel perilaku media sosial siswa atau 22 data penjualan *frozen food*. Pada skala menengah, terdapat penelitian dengan 200 data rekam medis, 240 data kunjungan hotel, 624 data penjualan produk fashion, dan 983 data mahasiswa. Pada skala besar, terdapat penelitian dengan 33.000 data lowongan pekerjaan. Variasi ini

menunjukkan bahwa implementasi *data mining* pada literatur yang ditinjau tidak terbatas pada satu ukuran data tertentu, tetapi telah diterapkan mulai dari data sederhana berbasis kuesioner sampai data besar berbasis platform digital. Namun demikian, perbedaan ukuran data tersebut juga memengaruhi kualitas pelaporan hasil, terutama pada bagian evaluasi model.

Tabel 4 Implementasi Data Mining pada Berbagai Bidang

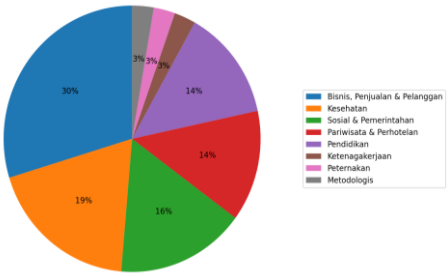
Bidang	Implementasi	Jumlah
Ekonomi & Bisnis	Clustering Tingkat Penjualan Menu (2022)	**12**
	Implementasi Data Mining Dalam Menganalisa Pola Penjualan Roti (2022)	
	Implementasi Data Mempolakan Penjualan pada Showroom Motor Bekas (2023)	
	Implementasi Data Mining pada Hasil Penjualan Makanan Beku (2023)	
	Implementasi Data Mining untuk Clustering Lowongan Pekerjaan (2025)	
	Implementasi Data Mining untuk Menentukan Pola Penjualan pada Coffee Shop (2025)	
	Implementasi Data Mining untuk Menentukan Pola Penjualan Produk (2024)	
	Implementasi Data Mining untuk Pengelompokan Customer PT. Kasana Teknindo Gemilang (2018)	
	Implementasi Data Mining FP-Growth untuk Analisis Pola Pembelian (2023)	
	Implementasi RapidMiner untuk Analisa Penjualan pada Toko OJ Cell (2023)	
	Penerapan Data Mining Menggunakan untuk Mengetahui Kebiasaan Konsumen dan Prediksi Stok Produk (2023)	
	Penerapan Data Mining Dalam Pemilihan Produk Unggulan (2022)	
Kesehatan	Data Mining untuk Mengetahui Potensi Penyebaran Virus Corona di Kota Cirebon (2022)	**7**
	Data Mining untuk Klasifikasi Penyakit Infeksi Saluran Pernafasan Akut (2022)	
	Implementasi Algoritma K-Means Clustering dalam Penentuan Gangguan Tidur (2024)	
	Implementasi Algoritma K-Means dalam Analisis Klasterisasi Penyebaran Penyakit HIV/AIDS (2023)	
	Implementasi Data Mining untuk Clustering Penyakit di RS Panti Waluyo Surakarta (2022)	
	Implementasi Data Mining Klasterisasi Data Pasien Rawat Inap (2024)	
	The Application of K-Means Clustering for Province Clustering in Indonesia of the Risk of the COVID-19 (2022)	
Sosial & Pemerintahan	Implementasi Data Mining Clustering dalam Mengelompokkan Penduduk Penyandang Disabilitas (2024)	**6**
	Implementasi Data Mining Dalam Mengelompokkan Data Penduduk Kurang Mampu (2023)	

	Implementasi Data Mining Dalam Pengelompokkan Data Potensi Kejahatan pada Kapolsek Beringin (2023)	
	Implementasi Data Mining untuk Data Kenaikan Pangkat Dinas Ketenagakerjaan Kota Medan (2022)	
	Implementasi Data Mining pada Pemerintahan (2024)	
	Implementasi Data Mining untuk Pengelompokkan Wilayah Pelanggaran Lalu Lintas pada Polres Bengkulu (2023)	
Pariwisata	Data Mining pada Data Kunjungan Wisatawan di Kabupaten Karawang (2023)	**5**
	Implementasi Data Mining dalam Klasifikasi Kunjungan Wisatawan di Kota Yogyakarta (2020)	
	Implementasi Data Mining Untuk Klasifikasi Kunjungan Wisatawan Riau (2025)	
	Implementasi K-Means Clustering pada Kunjungan Wisatawan Asing di Provinsi Jawa Timur (2024)	
	Implementasi Teknik Data Mining pada Data Kunjungan Wisatawan di Hotel Bintang Kota Cirebon (2024)	
Pendidikan	Implementasi Data Mining dalam Sebaran Mahasiswa Baru pada STIE-SAK (2025)	**5**
	Implementasi Data Mining dalam Pengelompokan Perilaku Penggunaan Media Sosial Siswa SMA (2026)	
	Detecting Learning Patterns in Tertiary Education (2022)	
	Penerapan Data Mining untuk Clustering Penilaian Kinerja Dosen (2022)	
	Penerapan Data Mining pada Tata Letak Buku di Perpustakaan Sintong Bingei (2022)	
Pertanian & Peternakan	Penerapan Data Mining pada Populasi Ayam Petelur di Indonesia (2022)	**1**
Metodologi & Teoretis	Unsupervised K-Means Clustering Algorithm (2020)	**1**

Berdasarkan Tabel 4 dan divisualisasikan kembali pada Gambar 2, bidang implementasi *data mining* yang paling dominan adalah Ekonomi & Bisnis dengan 12 artikel (32,43%). Posisi berikutnya adalah Kesehatan dengan 7 artikel (18,92%), Sosial & Pemerintah dengan 6 artikel (16,22%), serta Pariwisata dan Pendidikan masing-masing 5 artikel (13,51%). Sementara itu, Pertanian & Peternakan serta Metodologi & Teoretis masing-masing hanya diwakili oleh 1 artikel (2,7%). Dominasi bidang ekonomi dan bisnis menunjukkan bahwa *data mining* paling banyak dimanfaatkan untuk keperluan operasional yang langsung berkaitan dengan penjualan, persediaan, segmentasi pelanggan, dan rekomendasi promosi. Bidang kesehatan berada pada urutan kedua karena data pasien, penyakit, dan risiko wilayah sangat sesuai untuk dianalisis melalui pengelompokan. Bidang sosial,

pemerintahan, dan pariwisata juga cukup menonjol karena kebutuhan pemetaan wilayah, pelayanan publik, dan pengelolaan destinasi wisata semakin membutuhkan pendekatan berbasis data (Syahputra et al., 2022; Gunawan & Purwayoga, 2022; Mustari et al., 2024; Dewi et al., 2022).

Gambar 2 Distribusi Bidang



Secara umum, hasil kajian ini memperlihatkan dua pola utama. Pertama, implementasi *data mining* dalam 37 jurnal sangat didominasi oleh pendekatan clustering, khususnya K-Means. Kedua, pemanfaatan *data mining* paling banyak diarahkan pada bidang ekonomi dan bisnis, diikuti oleh kesehatan, yang menunjukkan bahwa kebutuhan segmentasi dan pengelompokkan data paling besar terdapat pada sektor-sektor yang berorientasi pada efisiensi dan dukungan keputusan. Temuan ini menunjukkan bahwa penelitian implementasi *data mining* masih memiliki ruang pengembangan, terutama pada aspek standardisasi evaluasi, konsistensi istilah metodologis, dan penguatan studi komparatif antaralgoritma. Dengan demikian penelitian ini tidak hanya memetakan kecenderungan algoritma dan bidang implementasi, tetapi juga menegaskan perlunya penelitian lanjutan yang lebih kuat secara metodologi.

D. Kesimpulan dan Saran

Berdasarkan hasil penelitian dengan metode *systematic literature review* terhadap 37 jurnal, bisa disimpulkan bahwa implementasi *data mining* paling banyak diterapkan pada bidang ekonomi dan bisnis, diikuti kesehatan, sosial dan pemerintahan, pariwisata, serta pendidikan. Dari sisi metode, penelitian didominasi oleh clustering, sedangkan dari sisi algoritma, K-Means adalah algoritma yang dominan digunakan. Dominasi K-Means menunjukkan bahwa algoritma ini masih menjadi pilihan utama dalam penelitian implementasi *data mining* karena sederhana, mudah diterapkan, dan efektif untuk pengelompokan data pada berbagai konteks, seperti penjualan, wisata, kesehatan, pendidikan, dan pemerintahan.

Penelitian ini memberikan kontribusi berupa pemetaan yang lebih terstruktur mengenai metode, algoritma, jenis data, dan hasil implementasi *data mining* berdasarkan 37 jurnal yang ditelaah. Hasil penelitian ini juga dapat menjadi dasar dalam menentukan atribut analisis untuk penelitian lanjutan, khususnya pada kajian implementasi *data mining* berbasis *clustering* dan K-Means.

Berdasarkan hasil penelitian, ada beberapa saran yang dapat dilaksanakan. Pertama, penelitian selanjutnya dapat mengembangkan kajian komparatif antaralgoritma, misalnya antara K-Means dengan K-Medoids, DBSCAN, atau algoritma optimasi lainnya, untuk pemahaman yang lebih mendalam tentang kelebihan dan keterbatasan masing-masing metode. Kedua, penelitian lanjutan juga disarankan untuk menggunakan dataset yang lebih besar, lebih beragam, dan berasal dari berbagai sumber agar hasil implementasi *data mining* memiliki tingkat generalisasi yang lebih baik.

DAFTAR PUSTAKA

Abdullah, D., Susilo, S., Ahmar, A. S., Rusli, R., & Hidayat, R. (2022). The application of K-means clustering for province clustering in Indonesia of the risk of the COVID-19 pandemic based on COVID-19 data. Quality & Quantity, 56, 1283–1291. doi:10.1007/s11135-021-01176-w

Ambroini, A., & Sari, I. P. (2025). Implementasi data mining untuk menentukan pola penjualan pada coffee shop menggunakan algoritma Apriori. Jurnal Sistem Informasi dan Ilmu Komputer, 3(3), 88–98. doi:10.59581/jusiik-widyakarya.v3i3.5624

Andini, Y., Hardinata, J. T., & Purba, Y. P. (2022). Penerapan data mining pada tata letak buku di perpustakaan Sintong Bingei Pematangsiantar dengan metode Apriori. Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURASIK), 7(1), 13–18.

Ariyanto, D. (2022). Data mining menggunakan algoritma K-means untuk klasifikasi penyakit infeksi saluran pernafasan akut. Jurnal Sistim Informasi dan Teknologi, 4(1), 13–18.

Arman, S. A., Untari, R. T., & Erion. (2023). Implementasi data mining menggunakan metode decision tree dalam mempolakan penjualan pada showroom motor bekas. Journal of Science and Social Research, 6(2), 398–403.

Dewi, N. L. P. P., Purnama, I. N., & Utami, N. W. (2022). Penerapan data mining untuk clustering penilaian kinerja dosen menggunakan algoritma K-means (Studi kasus: STMIK Primakara). Jurnal Ilmiah Teknologi Informasi Asia, 16(2).

Fhardana, R. M., & Siddik, M. (2025). Implementasi data mining untuk

klasifikasi kunjungan wisatawan Riau menggunakan K-NN. Jurnal SANTI, 5(2).

Gustrianda, R., & Mulyana, D. I. (2022). Penerapan data mining dalam pemilihan produk unggulan dengan metode algoritma K-means dan K-medoids. Jurnal Media Informatika Budidarma, 6(1), 27–34. doi:10.30865/mib.v6i1.3294

Gunawan, H., & Purwayoga, V. (2022). Data mining menggunakan algoritma K-means clustering untuk mengetahui potensi penyebaran virus Corona di Kota Cirebon. Jurnal SISFOKOM, 11(1), 1–8. doi:10.32736/sisfokom.v11i1.1316

Gustipartsani, K., Rahaningsih, N., Dana, R. D., & Mustafa, I. Y. (2023). Data mining clustering menggunakan algoritma K-means pada data kunjungan wisatawan di Kabupaten Karawang. JATI (Jurnal Mahasiswa Teknik Informatika), 7(6).

Hamsiah. (2025). Implementasi data mining dalam penerapan clustering algoritma K-medoid sebaran mahasiswa baru pada STIE-SAK. Jurnal Ilmiah ILKOMINFO, 8.

Hidayah, A., Lestari, W., & Purwanto, E. (2022). Implementasi data mining algoritma K-means untuk clustering penyakit di RS Panti Waluyo Surakarta.

Komariyah, S., Anwar, S., & Nurhakim, B. (2023). Implementasi data mining FP-growth untuk analisis pola pembelian pada transaksi penjualan. Jurnal Manajemen dan Bisnis Ekonomi, 1(2), 62–75.

Laksono, W. B., Syahidin, Y., & Yunengsih, Y. (2024). Implementasi data mining klasterisasi data pasien rawat inap dengan algoritma K-means clustering. Jurnal Teknologi Sistem Informasi dan Aplikasi, 7(2), 621–627. doi:10.32493/jtsi.v7i2.39354

Lestari, P., Suarna, N., & Prihartono, W. (2024). Implementasi data mining clustering dalam mengelompokkan penduduk penyandang disabilitas menggunakan algoritma K-means. JATI (Jurnal Mahasiswa Teknik Informatika), 8(1).

Liyanto, E. W., Homaidi, A., & Lutfi, A. (2024). Implementasi K-means clustering menggunakan RapidMiner dalam pengelompokan data kunjungan wisatawan asing di Provinsi Jawa Timur. E-Link.

Maulida, E., Hidayati, A. F., Silalahi, F. Y., & Prasetyo, Y. D. (2024). Implementasi algoritma K-means clustering dalam penentuan gangguan tidur seseorang berdasarkan gaya hidup. Journal Information Engineering and Educational Technology, 8(1).

Mubarok, R., Syahputra, A. A., Permana, A. T., Sholiah, L., & Tarwoto. (2025). Implementasi data mining untuk clustering lowongan pekerjaan menggunakan metode algoritma K-means. Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi), 9(2), 703–712. doi:10.35870/jtik.v9i2.3438

Mustari, K. A., Assiroj, P., Hartati, B., & Samuel, F. (2024). Implementasi data mining pada instansi pemerintahan (Systematic literature review). JATI (Jurnal Mahasiswa Teknik Informatika), 8(3).

Nahjan, M. R., Heryana, N., & Voutama, A. (2023). Implementasi

RapidMiner dengan metode clustering K-means untuk analisa penjualan pada toko OJ Cell. JATI (Jurnal Mahasiswa Teknik Informatika), 7(1).

Pratama, O., & Jaman, J. H. (2023). Penerapan data mining menggunakan algoritma Apriori untuk mengetahui kebiasaan konsumen dan prediksi stok produk (Studi kasus toko elektronik WK). JATI (Jurnal Mahasiswa Teknik Informatika), 7(3).

Purnomo, B. S., & Prasetyaningrum, P. T. (2020). Implementasi data mining dalam klasifikasi kunjungan wisatawan di Kota Yogyakarta menggunakan algoritma K-means. JIFTI: Jurnal Ilmiah Teknologi Informasi dan Robotika, 2(1).

Pujiono, S., Astuti, R., & Basysyar, F. M. (2024). Implementasi data mining untuk menentukan pola penjualan produk menggunakan algoritma K-means clustering. JATI (Jurnal Mahasiswa Teknik Informatika), 8(1).

Ramadanti, E., & Muslih, M. (2022). Penerapan data mining algoritma K-means clustering pada populasi ayam petelur di Indonesia. RABIT: Jurnal Teknologi dan Sistem Informasi Univrab, 7(1), 1–7. doi:10.36341/rabit.v7i1.2155

Saragih, R. P., Zunaidi, M., & Hafizah. (2023). Implementasi data mining dalam pengelompokkan data potensi kejahatan menggunakan algoritma K-means clustering pada Kapolsek Beringin. Jurnal Sistem Informasi TGD, 2(4), 665–675.

Sari, T. P., Hananto, A. L., Novalia, E., Tukino, & Hilabi, S. S. (2023). Implementasi algoritma K-means dalam analisis klasterisasi penyebaran penyakit HIV/AIDS. Infotek: Jurnal Informatika dan Teknologi, 6(1), 104–114. doi:10.29408/jit.v6i1.7423

Santoso, T. B., & Zulkarnain, R. (2018). Implementasi data mining untuk pengelompokan customer PT. Kasana Teknindo Gemilang menggunakan algoritma K-means. Jurnal Satya Informatika, 3(1), 60–70.

Sihombing, L. K., Tugiono, & Pane, U. F. S. S. (2022). Implementasi data mining dalam menganalisa pola penjualan roti menggunakan algoritma FP-growth. Jurnal Sistem Informasi TGD, 1(3), 228–238.

Simanjuntak, A. Y., Simatupang, I. S. S., & Anita. (2022). Implementasi data mining menggunakan metode Naïve Bayes classifier untuk data kenaikan pangkat dinas ketenagakerjaan Kota Medan. Journal of Science and Social Research, 5(1), 85–91.

Sinaga, K. P., & Yang, M.-S. (2020). Unsupervised K-means clustering algorithm. IEEE Access, 8. doi:10.1109/ACCESS.2020.2988796

Syahputra, H., Mayola, L., & Guswandi, D. (2022). Clustering tingkat penjualan menu (food and beverage) menggunakan algoritma K-means. Jurnal KomtekInfo, 9(1), 29–33. doi:10.35134/komtekinfo.v9i1.274

Syahputri, N. I., Rismayanti, & Sundari, S. (2023). Implementasi data mining pada hasil penjualan makanan beku menggunakan K means clustering. Jurnal Ilmu Komputer dan Sistem Informasi (JIRSI), 2(2), 174–182.

Tuyishimire, E., Mabuto, W., Gatabazi, P., & Bayisingize, S. (2022). Detecting learning patterns in tertiary education using K-means clustering. Information, 13(2), 94. doi:10.3390/info13020094

Utari, & Iqbal, M. (2026). Implementasi data mining dalam pengelompokan perilaku penggunaan media sosial siswa SMA menerapkan algoritma K-means. Journal of Informatics Management and Information Technology, 6(1), 44–51. doi:10.47065/jimat.v6i1.959

Waruwu, A., Yetri, M., & Setiawan, F. (2023). Implementasi data mining dalam mengelompokkan data penduduk kurang mampu menggunakan metode K-means clustering. Jurnal Sistem Informasi TGD, 2(6), 945–955.

Wulandari, S., Astuti, R., & Basysyar, F. M. (2024). Implementasi teknik data mining pada data kunjungan wisatawan menggunakan K-means clustering di hotel bintang Kota Cirebon. JATI (Jurnal Mahasiswa Teknik Informatika), 8(1).