

## **SYSTEMATIC LITERATURE REVIEW: PENERAPAN K-MEANS CLUSTERING**

Albert Putra Pratama<sup>1</sup>, Vita Nurul Fathya<sup>2</sup>, Muhammad Fahrury Romdendine<sup>3</sup>

<sup>1</sup>Program Studi Manajemen Teknologi Keimigrasian, Politeknik Imigrasi

<sup>2</sup>Program Studi Manajemen Teknologi Keimigrasian, Politeknik Imigrasi

<sup>3</sup>Program Studi Manajemen Teknologi Keimigrasian, Politeknik Imigrasi

<sup>1</sup>putraalbert07@gmail.com, <sup>2</sup>vitafathya@gmail.com, <sup>3</sup>romdendine@poltekim.ac.id

### **ABSTRACT**

*This study is a systematic literature review on the application of K-Means Clustering in various sectors using the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) approach. The clustering technique, particularly K-Means, is widely used in big data analysis due to its simplicity and efficiency. Although this method is popular, the main challenges include determining the optimal number of clusters, handling outliers, and computational limitations when applied to large-scale data. This research analyzes sectors that apply K-Means Clustering, such as industry, education, healthcare, and finance. The findings of this study are expected to provide insights into the trends in the use of clustering methods and offer recommendations on the most suitable tools for applying clustering based on data characteristics .*

*Keywords: data mining, k-means, clustering, PRISMA, clustering algorithms*

### **ABSTRAK**

Penelitian ini merupakan kajian literatur sistematis tentang penerapan *K-Means Clustering* dalam berbagai sektor dengan menggunakan pendekatan PRISMA (*Preferred Reporting Items for Systematic Reviews and Meta-Analyses*). Teknik clustering, khususnya K-Means, banyak digunakan dalam analisis data besar karena kesederhanaan dan efisiensinya. Meskipun metode ini populer, tantangan utama yang dihadapi adalah penentuan jumlah *cluster* yang optimal, penanganan *outlier*, serta keterbatasan komputasi pada data berskala besar. Dalam penelitian ini, dilakukan analisis terhadap sektor-sektor yang menerapkan K-Means Clustering, seperti industri, pendidikan, kesehatan, dan keuangan. Hasil kajian ini diharapkan dapat memberikan wawasan tentang tren penggunaan metode clustering serta memberikan rekomendasi mengenai alat yang paling sesuai untuk penerapan clustering berdasarkan karakteristik data.

*Kata Kunci: data mining, k-means, clustering, PRISMA, algoritma clustering*

#### **A. Pendahuluan**

Di era digital saat ini, data menjadi salah satu aset paling

berharga dalam pengambilan keputusan. Dengan semakin berkembangnya teknologi informasi,

volume data yang dihasilkan setiap hari semakin besar dan kompleks. Oleh karena itu, penting bagi organisasi dan individu untuk memanfaatkan teknik yang efektif dalam mengelola dan menganalisis data tersebut. Salah satu teknik yang banyak digunakan untuk menganalisis data adalah *data mining*, yang mencakup berbagai metode seperti klasifikasi, prediksi, dan *clustering*. Teknik *clustering* memungkinkan pengelompokan data berdasarkan kemiripan karakteristik, sehingga memudahkan analisis lebih lanjut serta membantu dalam pengambilan keputusan berbasis data. Dalam konteks ini, *clustering* berperan penting dalam mengidentifikasi pola atau grup yang tidak terlihat sebelumnya, yang dapat meningkatkan pengambilan keputusan strategis. [1]

Dengan pertumbuhan volume data yang pesat, terutama dalam bidang kesehatan, ekonomi, dan teknologi, teknik *clustering* menjadi semakin relevan. Organisasi kini menghadapi tantangan besar dalam memproses data besar dan kompleks, yang jika tidak dikelola dengan baik, dapat menghambat pengambilan keputusan yang tepat. Oleh karena

itu, penting bagi organisasi untuk memanfaatkan teknik *clustering* yang efektif, baik dalam hal efisiensi maupun akurasi, agar dapat memperoleh wawasan berharga dari data yang ada. Salah satu metode *clustering* yang paling banyak digunakan adalah K-Means Clustering, yang telah terbukti efektif dalam mengelompokkan data dalam berbagai sektor.

K-Means Clustering adalah metode *clustering* yang paling populer dan banyak digunakan dalam berbagai aplikasi. Metode ini terkenal karena kesederhanaannya, kemudahan implementasi, dan efisiensinya dalam menangani data dalam jumlah besar [2]. Proses kerja K-Means sangat sederhana, yaitu dengan menentukan jumlah cluster yang diinginkan, kemudian algoritma akan mengelompokkan data ke dalam cluster berdasarkan kedekatannya dengan pusat cluster (centroid). Setiap iterasi algoritma akan memperbaiki posisi pusat cluster hingga hasilnya stabil. Penerapannya pun sangat luas, mulai dari sektor kesehatan untuk mengelompokkan pasien berdasarkan kondisi medis, sektor bisnis dalam analisis penjualan dan segmentasi pelanggan, hingga

pendidikan untuk memahami pola prestasi akademik mahasiswa [3].

Dalam dunia bisnis dan perdagangan, K-Means Clustering terbukti efektif dalam mengelompokkan pelanggan berdasarkan pola pembelian mereka. Dengan cara ini, strategi pemasaran bisa lebih tepat sasaran dan sesuai dengan kebutuhan pelanggan [4]. Misalnya, perusahaan dapat mengidentifikasi pelanggan dengan kebiasaan membeli produk tertentu, sehingga dapat menawarkan produk yang lebih relevan kepada mereka. Selain itu, dalam industri properti, metode ini juga digunakan untuk menentukan harga rumah berdasarkan lokasi dan luas tanah, yang membantu pembeli dan penjual dalam menentukan harga pasar yang lebih kompetitif. Dalam konteks kesehatan, K-Means digunakan untuk mengelompokkan pasien berdasarkan gejala atau kondisi medis mereka, yang memudahkan dokter dalam memberikan diagnosis dan perawatan yang lebih tepat sasaran.

Namun, meskipun K-Means terkenal cepat dan efisien, metode ini memiliki beberapa kelemahan. Salah satunya adalah sulitnya menentukan jumlah cluster yang optimal.

Menentukan jumlah cluster yang tepat sangat penting agar hasil *clustering* dapat digunakan secara maksimal dalam analisis data. Untuk mengatasi masalah ini, teknik tambahan seperti Elbow Method dan Silhouette Score sering digunakan dalam penelitian guna menemukan jumlah cluster yang paling sesuai [5]. Teknik-teknik ini membantu mengidentifikasi jumlah cluster yang memberikan hasil terbaik berdasarkan varians data. Selain itu, K-Means juga menghadapi tantangan dalam penanganan outlier, karena sensitif terhadap data yang tidak representatif. Outlier dapat memengaruhi posisi pusat cluster, yang pada gilirannya dapat memengaruhi hasil analisis.

Selain itu, K-Means memiliki keterbatasan dalam hal komputasi ketika diterapkan pada data berskala besar. Dengan semakin besar dataset, waktu komputasi yang dibutuhkan untuk menjalankan algoritma juga semakin meningkat, sehingga mempengaruhi efisiensi dan kecepatan proses. Oleh karena itu, meskipun K-Means adalah metode yang sangat berguna, dalam beberapa kasus, penelitian telah menunjukkan bahwa teknik alternatif seperti DBSCAN dan K-Medoids bisa

lebih baik, terutama ketika data memiliki kepadatan yang tidak merata atau mengandung banyak *outliers* [2].

Meskipun banyak penelitian yang telah mengkaji penggunaan K-Means dalam berbagai sektor, masih terdapat gap dalam literatur yang membahas penerapan metode ini secara komprehensif di berbagai bidang dan perbandingannya dengan metode *clustering* lainnya. Sebagian besar penelitian sebelumnya lebih fokus pada penerapan satu metode dalam satu sektor atau aplikasi, tanpa memberikan perbandingan yang mendalam tentang kelebihan dan kekurangan tiap metode clustering dalam konteks yang berbeda. Penelitian ini bertujuan untuk mengisi gap tersebut dengan melakukan kajian literatur sistematis yang mencakup penerapan K-Means Clustering di berbagai sektor, seperti kesehatan, pendidikan, bisnis, dan ekonomi. Selain itu, penelitian ini juga membandingkan K-Means dengan metode *clustering* alternatif seperti DBSCAN dan K-Medoids, guna memberikan wawasan yang lebih luas mengenai metode mana yang paling sesuai untuk jenis data dan tujuan analisis tertentu.

Tujuan penelitian ini adalah untuk:

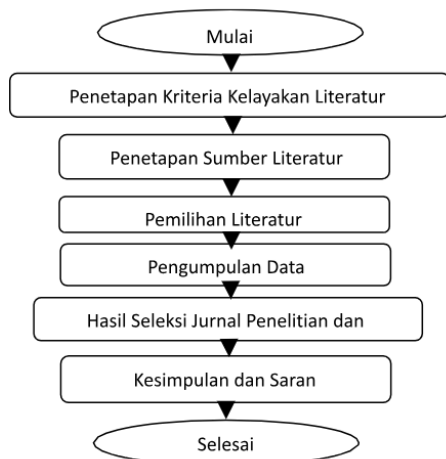
1. Menyajikan kajian literatur tentang penerapan K-Means Clustering di berbagai sektor.
2. Menganalisis dan membandingkan kelebihan dan kekurangan K-Means Clustering dengan metode *clustering* lainnya, seperti DBSCAN dan K-Medoids.
3. Memberikan rekomendasi metode *clustering* yang paling sesuai untuk berbagai jenis data berdasarkan karakteristik dan tujuan analisis.

Dengan tujuan tersebut, penelitian ini diharapkan dapat memberikan kontribusi yang lebih luas bagi peneliti, akademisi, dan praktisi dalam memilih metode *clustering* yang tepat berdasarkan jenis data yang dihadapi dan tujuan analisis yang ingin dicapai. Selain itu, hasil dari penelitian ini juga akan membantu memperkaya pemahaman kita tentang bagaimana K-Means Clustering diterapkan dalam berbagai sektor serta bagaimana tantangan dalam penerapan metode ini dapat diatasi dengan teknik-teknik tambahan yang ada.

## **B. Metode Penelitian**

Dalam kajian literatur ini, metode penelitian yang digunakan adalah

PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses). PRISMA merupakan suatu kerangka kerja yang dirancang untuk meningkatkan transparansi dan kualitas dalam sistematik review dengan menyediakan pedoman yang jelas dalam proses seleksi dan analisis literatur [6]. Metode ini banyak digunakan dalam penelitian berbasis *systematic review* karena memberikan struktur yang sistematis dalam mengidentifikasi, memilih, dan menilai kualitas literatur yang relevan dengan topik penelitian. Dengan adanya pedoman ini, diharapkan proses penelitian menjadi lebih terstandarisasi dan hasil yang diperoleh lebih dapat dipercaya.



Gambar 1 *Flowchart* Penelitian

PRISMA terdiri dari empat tahap utama yang dijalankan secara

bertahap. Keempat tahap ini, yaitu kelayakan, sumber literatur, pemilihan literatur, dan pengumpulan data, merupakan proses penting yang harus dilalui agar literatur yang digunakan dalam kajian ini dapat dipertanggungjawabkan secara ilmiah dan valid. Setiap tahapan memiliki peranan yang sangat penting dalam memastikan bahwa penelitian ini menggunakan literatur yang relevan, berkualitas, dan sesuai dengan kriteria yang telah ditetapkan. Berikut adalah penjelasan rinci mengenai setiap tahapan yang diikuti dalam penelitian ini:

#### 1. Kelayakan

Tahap pertama dalam metode PRISMA adalah kelayakan, yang bertujuan untuk mengevaluasi literatur berdasarkan teks lengkapnya. Pada tahap ini, literatur yang relevan dan sesuai dengan topik penelitian akan dipilih, sementara literatur yang tidak memenuhi kriteria akan dikeluarkan. Kriteria eksklusi diterapkan untuk menghindari bias dalam penelitian. Misalnya, studi yang tidak memiliki data eksperimen atau yang tidak menggunakan metode *clustering* yang dibahas dalam kajian ini akan

disaring. Beberapa kriteria inklusi yang digunakan adalah sebagai berikut:

- IC1: Artikel dan skripsi yang dipertimbangkan harus merupakan penelitian asli yang sudah dipelajari dan ditulis dalam bahasa Indonesia. Ini penting karena penelitian ini berfokus pada literatur yang relevan dengan konteks lokal dan penelitian yang dapat diakses oleh peneliti di Indonesia.
- IC2: Artikel dan skripsi yang ditinjau harus berasal dari jurnal yang bereputasi, dan harus dipublikasikan antara tahun 2021 hingga 2025. Pembatasan tahun ini bertujuan untuk memastikan bahwa literatur yang digunakan adalah literatur terbaru, yang mencerminkan tren terkini dalam penerapan metode *clustering* dalam *data mining*. Dengan demikian, kajian ini akan lebih relevan dengan perkembangan terbaru dalam bidang tersebut.
- IC3: Artikel dan skripsi yang ditinjau harus bertujuan untuk menganalisis metode *Clustering* dalam *Data Mining*. Hal ini

memastikan bahwa semua literatur yang dipilih benar-benar berfokus pada teknik *clustering*, yang menjadi inti dari penelitian ini.

## 2. Sumber Literatur

Pada tahap ini, dilakukan pencarian literatur dari berbagai database akademik, seperti Scopus, IEEE Xplore, Springer, dan Google Scholar, dengan menggunakan kata kunci “clustering” dan “data mining”. Database ini dipilih karena mereka menyediakan jurnal ilmiah yang memiliki standar kualitas yang tinggi dan sering digunakan dalam penelitian akademik. Pencarian dilakukan dengan memilih artikel yang dipublikasikan dalam jurnal terkemuka, serta artikel yang membahas secara spesifik tentang teknik *clustering* dan aplikasinya dalam *data mining*. Untuk mendapatkan sumber literatur yang lebih akurat, kata kunci yang digunakan telah disesuaikan dengan topik penelitian. Pencarian literatur dilakukan dengan menggunakan kombinasi kata kunci yang relevan, sehingga artikel yang ditemukan benar-benar berfokus pada metode K-Means

dan *clustering* secara umum dalam analisis *data mining*. Penelitian ini bertujuan untuk memaksimalkan jangkauan literatur yang relevan, dengan fokus pada kualitas dan relevansi topik penelitian.

### 3. Pemilihan Literatur

Setelah literatur dikumpulkan, dilakukan penyaringan awal berdasarkan judul dan abstrak untuk mengeliminasi studi yang tidak relevan dengan topik penelitian. Penyaringan ini dilakukan untuk memastikan bahwa literatur yang dipilih benar-benar berkaitan dengan metode *clustering* yang digunakan dalam penelitian ini. Artikel yang memiliki kesamaan topik tetapi tidak memenuhi kriteria inklusi akan dikeluarkan dari analisis lebih lanjut. Penyaringan ini bertujuan untuk memastikan bahwa hanya literatur yang berkualitas tinggi dan relevan dengan tujuan penelitian yang dimasukkan dalam kajian ini. Proses ini juga mengurangi potensi bias dan meningkatkan kualitas hasil yang diperoleh.

### 4. Pengumpulan Data

Pengumpulan data dilakukan dengan cara manual melalui pembuatan formulir ekstraksi data.

Formulir ini digunakan untuk mencatat informasi penting dari setiap literatur yang relevan, seperti tahun publikasi, sektor implementasi, metode *clustering* yang digunakan, serta tools atau perangkat lunak yang diterapkan dalam penelitian tersebut. Penelitian ini memanfaatkan jurnal dengan kata kunci "Data Mining Clustering" yang telah dipilih berdasarkan kriteria yang telah disebutkan sebelumnya. Berdasarkan sumber-sumber literatur yang terkumpul dan kriteria inklusi, terdapat jurnal dan skripsi yang layak menjadi kandidat terpilih untuk memenuhi syarat penelitian ini. Proses pemilihan dilakukan terhadap 100.000 artikel berdasarkan kata kunci yang digunakan, dan akhirnya hanya terdapat 30 artikel terpilih yang memenuhi semua kriteria dan syarat penelitian ini. Tabel 1 menampilkan hasil dari seleksi jurnal penelitian yang dikumpulkan, yang menunjukkan jumlah literatur yang terpilih dan relevansi setiap artikel dengan topik yang dibahas.

Dengan mengikuti tahapan-tahapan yang terstruktur dan

sistematis dalam pendekatan PRISMA, penelitian ini dapat memastikan bahwa literatur yang digunakan adalah yang paling relevan dan berkualitas. Proses ini membantu mengidentifikasi dan memilih literatur yang tepat yang akan memberikan wawasan berharga untuk penelitian lebih lanjut mengenai penerapan metode *clustering* dalam *data mining*. Metode ini juga memastikan bahwa hanya literatur yang memiliki kualitas tinggi dan relevansi dengan topik yang dipilih yang akan dimasukkan ke dalam kajian, sehingga hasil penelitian ini akan memberikan kontribusi yang signifikan terhadap pemahaman tentang *clustering* dan aplikasinya dalam berbagai sektor. Dengan langkah-langkah yang jelas dan transparan, diharapkan penelitian ini dapat memberikan hasil yang lebih mendalam dan valid, serta menjadi referensi yang berguna bagi peneliti, akademisi, dan praktisi yang tertarik pada topik ini.

Tabel 1 Sumber Publikasi

Sumber	Penemuan Studi (Berdasarkan judul dan Kata Kunci)	Kandidat	Terpilih
	Data Mining,		

	Clustering		
Crossref	1000	400	12
Google Scholar	21.800	200	10
Sematic Scholar	248.000	1000	8
<b>Total</b>	<b>270.800</b>	1600	30

### C. Hasil Penelitian dan Pembahasan

Penelitian ini bertujuan menyelidiki penerapan dari Metode Clustering dalam data mining. Metode yang dimaksud adalah K-Means. Penelitian ini akan mengidentifikasi sektor apa saja yang sering menggunakan K-Means dalam penelitiannya. Selain ini akan diidentifikasi juga tools yang digunakan untuk clustering. Pada tabel ada disajikan data seperti sumber publikasi, tahun publikasi, klasifikasi tabel, sektor implementasi dan tools yang digunakan dalam Clustering. Tabel 2 menunjukkan sumber publikasi yang digunakan dalam penelitian ini:

Tabel 2 Sumber Publikasi

No	Judul	Tahun	Tipe
1	Clustering Data Nilai...	2021	Artikel
2	Data Mining Menggunakan...	2022	Jurnal
3	Implementasi Data Mining...	2023	Jurnal
4	Implementasi Data Mining...	2021	Jurnal
5	Implementasi Data Mining...	2024	Jurnal
6	Penerapan Data Mining...	2023	Artikel
7	Penerapan Data Mining...	2022	Jurnal
8	Penerapan Data Mining...	2024	Jurnal
9	Penerapan Data Mining...	2023	Jurnal
10	Penerapan Data Mining ...	2021	Jurnal
11	Penerapan Data Mining...	2021	Jurnal



12	Penerapan Data Mining...	2023	Jurnal
13	Penerapan Data Mining...	2023	Skripsi
14	Analisis Data Hasil...	2022	Jurnal
15	Data Mining Strategi...	2022	Jurnal
16	Implementasi Data Mining...	2021	Jurnal
17	Klasterisasi Data Penjualan...	2023	Jurnal
18	Penerapan Algoritma K-Means...	2023	Jurnal
19	Penerapan Data Mining...	2021	Jurnal
20	Penerapan Metode K-Means...	2023	Artikel
21	Pengelompokan Barang...	2022	Jurnal
22	Sistem Predksi Ketepatan...	2021	Jurnal
23	Implementasi Data Mining...	2021	Jurnal
24	Klasterisasi Pola Penjualan...	2021	Jurnal
25	Optimasi Cluster K-Means...	2021	Jurnal
26	Penerapan Algoritma K-Means...	2021	Jurnal
27	Penerapan Algoritma K-Means	2021	Jurnal
28	Penerapan Data Mining...	2021	Jurnal
29	Penerapan Data Mining...	2022	Jurnal
30	Penerapan Metode K-Means...	2021	Jurnal

Tabel 3 menunjukkan sektor-sektor yang menerapkan teknik clustering dalam penelitian yang telah dianalisis, memberikan wawasan mendalam mengenai aplikasi teknik ini di berbagai bidang. Sektor-sektor yang paling banyak menggunakan teknik clustering termasuk Industri dan Perdagangan, Pendidikan, Kesehatan dan Lingkungan, serta Keuangan dan Ekonomi. Setiap sektor ini memanfaatkan clustering untuk berbagai tujuan, seperti segmentasi pelanggan, analisis data akademik, pengelompokan pasien berdasarkan kondisi medis, dan analisis keuangan. Selain itu, tabel tersebut juga merinci jenis perangkat lunak atau tools yang digunakan oleh masing-masing sektor

dalam implementasi metode clustering. Ini termasuk alat pemrograman seperti Python (Scikit-Learn) yang paling banyak digunakan karena fleksibilitasnya dalam menangani data besar dan kompleks, serta kemudahan dalam mengimplementasikan berbagai algoritma *clustering*. Rapid Miner, yang berbasis GUI, lebih banyak dipilih di sektor Pemerintahan dan Pendidikan karena antarmukanya yang mudah digunakan, memungkinkan analisis *clustering* dilakukan tanpa membutuhkan keahlian pemrograman yang mendalam. Selain itu, alat lain seperti R, Matlab, dan Weka juga digunakan, dengan R lebih dikenal dalam analisis statistik yang mendalam, Matlab unggul dalam pemrosesan matematis dan Weka memberi kemudahan bagi pemula dalam mengaplikasikan metode *clustering*. Informasi ini memberikan gambaran yang jelas mengenai preferensi sektor terhadap tools tertentu dalam proses analisis data menggunakan metode clustering, dan menunjukkan bagaimana alat yang dipilih disesuaikan dengan jenis data, kompleksitas analisis, dan keterampilan pengguna di masing-masing sektor. Keberagaman

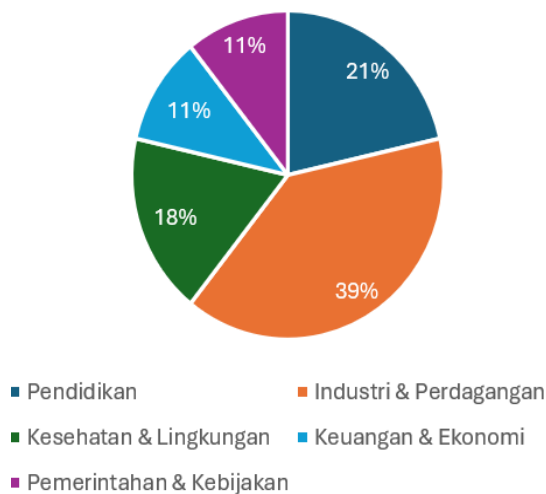
pemilihan tools ini juga mencerminkan berbagai pendekatan yang diambil dalam penerapan teknik clustering untuk menghasilkan analisis yang lebih akurat dan efektif sesuai dengan kebutuhan sektor masing-masing.

Tabel 3 Sektor Implementasi dan Tools

Penulis	Tahun	Sektor Implementasi	Tools
Indra Maulana dkk.	2021	Pendidikan	Python (Scikit-Learn)
Rafki Mauliadi	2022	Industri & Perdagangan	Python (Scikit-Learn)
Affani Putri Riyandoro dkk.	2023	Industri & Perdagangan	Python (Scikit-Learn)
Rini Astuti dkk.	2021	Kesehatan & Lingkungan	Python (Scikit-Learn)
Wismoyo Bagas Laksono dkk.	2024	Kesehatan & Lingkungan	Python (Scikit-Learn)
Ramadhan a dkk.	2023	Industri & Perdagangan	R
Yudi Wibowo	2022	Industri & Perdagangan	Matlab
Teneshya Lidia Putri dkk.	2024	Keuangan & Ekonomi	Python (Scikit-Learn)
Ahmad Rifqi dkk.	2023	Kesehatan & Lingkungan	Python (Scikit-Learn)
L. Magdalena dkk.	2021	Pemerintahan & Kebijakan	Rapid Miner
Saut Parsaoran Tamba dkk.	2021	Industri & Perdagangan	Rapid Miner
Muhammad Sholeh dkk.	2023	Pendidikan	Python (Scikit-Learn)
M. Afriko Yunaldo dkk.	2023	Pendidikan	Python (Scikit-Learn)
Siska Febriani	2022	Kesehatan & Lingkungan	Rapid Miner

Mangapul Siahaan dkk.	2022	Pemerintahan dan Kebijakan	Rapid Miner
Cici Armayani dkk.	2021	Industri dan Perdagangan	Python (Scikit-Learn)
Elin Mayoana Fitri dkk.	2023	Industri dan Perdagangan	Python (Scikit-Learn)
Ilwan Syafrinal	2023	Industri dan Perdagangan	Rapid Miner
Septian Isnanto dkk.	2021	Pendidikan	Python (Scikit-Learn)
Ana Neva dkk.	2023	Transportasi dan Pariwisata	Python (Scikit-Learn)
Achmad Fikri Sallaby dkk.	2022	Industri dan Perdagangan	Rapid Miner
Mochamad Rizky Alvin Fernanda dkk.	2021	Pendidikan	Matlab
Suhandio Handoko dkk.	2021	Teknologi dan Telekomunikasi	Python (Scikit-Learn)
Sabrina Aulia Rahma	2021	Industri dan Perdagangan	Rapid Miner
Andre Winarta dkk.	2021	Pemerintahan dan Kebijakan	Python (Scikit-Learn)
Nita Mirantika dkk	2021	Kesehatan dan Lingkungan	R
Yunita Ratna Sari dkk.	2021	Keuangan dan Ekonomi	Rapid Miner
Januardi Nasir	2021	Pendidikan	Weka
Agung Nugraha dkk.	2021	Industri dan Perdagangan	Python (Scikit-Learn)
Juniar Hutagalung dkk	2021	Keuangan dan Ekonomi	Python (Scikit-Learn)

### Sektor Implementasi Clustering



Gambar 2 Diagram Sektor Implementasi

Diagram ini menunjukkan distribusi sektor implementasi untuk 30 jurnal yang diunggah, dengan Industri dan Perdagangan mencatatkan jumlah tertinggi (11 jurnal), yang mencerminkan pentingnya metode *clustering* dalam sektor ini. Di sektor industri, K-Means Clustering digunakan secara luas untuk mengelompokkan data pelanggan berdasarkan pola pembelian dan transaksi, yang memungkinkan perusahaan untuk merancang strategi pemasaran yang lebih tepat sasaran dan personal. Selain itu, analisis clustering juga digunakan untuk segmentasi pasar, yang dapat meningkatkan efisiensi dalam merancang kampanye

pemasaran yang lebih relevan dan efisien. Sebagai contoh, Handoko et al. (2020) menunjukkan bagaimana teknik clustering digunakan untuk meningkatkan pengalaman pelanggan melalui pemahaman yang lebih mendalam tentang perilaku konsumsi mereka.

Sektor Pendidikan menyumbangkan 6 jurnal yang berfokus pada penggunaan *clustering* untuk menganalisis data mahasiswa dan merancang strategi promosi pendidikan yang lebih efektif. Teknik clustering digunakan untuk mengelompokkan mahasiswa berdasarkan prestasi akademik atau pola belajar mereka, yang dapat membantu dalam perencanaan pembelajaran yang lebih *personalized*. Sebagai contoh, penelitian oleh Maulana & Rosalina (2021) menunjukkan bagaimana *K-Means Clustering* membantu institusi pendidikan memahami dinamika prestasi akademik mahasiswa dan meningkatkan kurikulum serta metode pengajaran yang disesuaikan dengan kebutuhan masing-masing kelompok mahasiswa.

Di sektor Kesehatan dan Lingkungan, yang mencatatkan 5 jurnal, metode *clustering* digunakan

untuk menganalisis data pasien, pola penyakit, dan kualitas udara. Teknik ini sangat berguna dalam memahami tren kesehatan di berbagai wilayah dan memetakan potensi masalah kesehatan yang perlu mendapatkan perhatian lebih. Mirantika et al. (2021) menjelaskan bagaimana data clustering dapat digunakan untuk memetakan distribusi penyakit atau menentukan wilayah yang membutuhkan intervensi medis cepat, serta mengelompokkan daerah berdasarkan tingkat keparahan masalah kesehatan. Selain itu, analisis kualitas udara juga menjadi fokus sektor ini, dengan teknik clustering membantu dalam mengidentifikasi pola polusi udara yang dapat mempengaruhi kebijakan lingkungan.

Sektor Keuangan dan Ekonomi, dengan 3 jurnal, lebih menekankan pada analisis data finansial dan segmentasi pasar untuk merancang strategi investasi dan pengelolaan risiko yang lebih efektif. Clustering digunakan untuk mengelompokkan nasabah atau produk finansial berdasarkan karakteristik tertentu, seperti kebiasaan transaksi atau preferensi investasi. Dengan analisis ini, lembaga keuangan dapat

merancang produk dan layanan yang lebih sesuai dengan kebutuhan berbagai kelompok nasabah. Di sektor ini, misalnya, penelitian oleh Teneshya Lidia Putri et al. (2024) memperlihatkan bagaimana *K-Means* digunakan untuk mengelompokkan data nasabah bank berdasarkan riwayat transaksi mereka, sehingga bank dapat menyesuaikan strategi pemasaran dan penawaran produk secara lebih tepat.

Pemerintahan dan Kebijakan serta Transportasi & Pariwisata masing-masing menyumbangkan 3 dan 1 jurnal, yang menggambarkan pemanfaatan clustering dalam analisis kebijakan publik dan mobilitas masyarakat. Di sektor pemerintahan, clustering digunakan untuk menganalisis data administratif dan sosial, yang membantu pemerintah dalam membuat kebijakan yang lebih efektif dan tepat sasaran. L. Magdalena et al. (2021) mengilustrasikan bagaimana *clustering* digunakan untuk mengidentifikasi daerah-daerah yang membutuhkan alokasi sumber daya yang lebih besar berdasarkan data statistik sosial dan ekonomi. Di sektor transportasi, penelitian oleh Ana Neva dkk. (2023) menunjukkan bahwa *K-*

*Means* digunakan untuk memahami pola perjalanan wisatawan dan perilaku mobilitas mereka, yang membantu dalam merancang kebijakan transportasi publik yang lebih efisien dan ramah pengguna.

Sektor Teknologi dan Telekomunikasi, meskipun hanya menyumbangkan 1 jurnal, menunjukkan potensi besar dalam penerapan *clustering*, terutama dalam analisis data e-commerce dan energi. Suhandio Handoko et al. (2021) menunjukkan penggunaan clustering untuk menganalisis data pelanggan dalam industri telekomunikasi, seperti analisis penggunaan paket data pelanggan untuk merancang paket yang lebih sesuai dengan kebutuhan mereka. Penerapan clustering dalam sektor teknologi dan telekomunikasi semakin penting seiring dengan berkembangnya layanan berbasis data dan kebutuhan untuk mempersonalisasi pengalaman pengguna.

Secara keseluruhan, K-Means Clustering telah terbukti menjadi metode yang sangat fleksibel dan berguna dalam berbagai sektor. Distribusi sektor yang menunjukkan dominasi di Industri dan Perdagangan, diikuti oleh Pendidikan

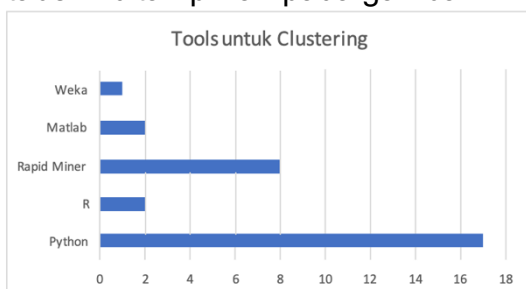
dan Kesehatan, menegaskan pentingnya teknik ini dalam mengelompokkan data dan menganalisis pola untuk pengambilan keputusan yang lebih baik. Pemilihan sektor yang beragam juga menunjukkan bahwa clustering tidak hanya terbatas pada satu sektor tertentu, tetapi dapat diterapkan secara luas untuk meningkatkan efisiensi dan efektivitas analisis data dalam berbagai bidang.

Tabel 4 Tools untuk Clustering

Penulis	Metode				
	A	B	C	D	E
Indra Maulana dkk.	✓				
Rafki Mauliadi	✓				
Affani Putri Riyandoro dkk.	✓				
Rini Astuti dkk.	✓				
Wismoyo Bagas Laksono dkk.	✓				
Ramadhana dkk.		✓			
Yudi Wibowo				✓	
Teneshya Lidia Putri dkk.	✓				
Ahmad Rifqi dkk.	✓				
L. Magdalena dkk.			✓		
Saut Parsaoran Tamba dkk.			✓		
Muhammad Sholeh dkk.	✓				
M. Afriko Yunaldo dkk.	✓				
Siska Febriani				✓	
Mangapul Siahaan					✓

dkk.		
Cici Armayani dkk.	✓	
Elin Mayoana Fitri dkk.	✓	
Ilwan Syafrinal		✓
Septian Isnanto dkk.	✓	
Ana Neva dkk.	✓	
Achmad Fikri Sallaby dkk.		✓
Mochamad Rizky Alvin		✓
Fernanda dkk.		
Suhandio Handoko dkk.	✓	
Sabrina Aulia Rahma		✓
Andre Winarta dkk.	✓	
Yunita Ratna Sari dkk.		✓
Nita Mirantika dkk		✓
Januardi Nasir		✓
Agung Nugraha dkk.	✓	
Juniar Hutagalung dkk.	✓	

Berdasarkan Tabel 4 di atas, anotasi tools untuk clustering yang digunakan adalah sebagai berikut: A: Python, B: R, C: Rapid Miner, D: Matlab, E: Weka. Kesimpulan dari tabel 4 ditampilkan pada gambar 2.



**Gambar 3 Diagram Tools untuk Clustering**

Kolom Tools menunjukkan beragam perangkat lunak yang digunakan dalam penerapan *data mining* dan *clustering* pada penelitian yang telah dianalisis. Setiap alat memiliki kelebihan dan kekurangannya sendiri, serta digunakan sesuai dengan kebutuhan sektor dan kompleksitas data yang dihadapi. Python dengan Scikit-Learn menjadi alat yang paling banyak digunakan karena fleksibilitas, skalabilitas, dan kemudahan dalam implementasi algoritma *clustering*. Python memungkinkan pengguna untuk melakukan analisis pada data besar dan kompleks, terutama dengan adanya berbagai library seperti Pandas dan NumPy, yang memungkinkan manipulasi dan pembersihan data secara efisien. Scikit-Learn menawarkan berbagai algoritma *clustering* yang mudah diterapkan, termasuk K-Means, DBSCAN, dan Agglomerative Clustering, serta menyediakan berbagai metode evaluasi hasil clustering. Keunggulan Python terletak pada kemampuan untuk mengolah data dalam jumlah besar, serta kemudahan untuk menyesuaikan algoritma dengan kebutuhan analisis yang lebih spesifik,

menjadikannya pilihan utama di sektor-sektor seperti industri, teknologi, dan penelitian ilmiah (Handoko et al., 2020).

Di sisi lain, Rapid Miner, yang berbasis GUI, lebih populer di sektor Pemerintahan dan Pendidikan karena antarmukanya yang intuitif dan kemudahan dalam penggunaannya tanpa memerlukan pemrograman yang mendalam. Dengan Rapid Miner, pengguna dapat menerapkan metode *clustering* tanpa memerlukan keterampilan pemrograman yang mendalam, sehingga lebih banyak dipilih oleh praktisi atau peneliti yang tidak memiliki latar belakang teknis yang kuat. Rapid Miner menawarkan berbagai algoritma dan alat untuk analisis data, namun memiliki keterbatasan dalam fleksibilitas dan skalabilitas bila dibandingkan dengan Python. Misalnya, ketika menangani dataset yang sangat besar atau ketika peneliti membutuhkan penyesuaian algoritma yang lebih mendalam, Python cenderung lebih unggul (Mirantika et al., 2021). Meskipun demikian, Rapid Miner tetap menjadi pilihan utama di sektor-sektor yang mengutamakan kemudahan penggunaan, seperti dalam pendidikan dan pemerintahan, di

mana pemrosesan data lebih terfokus pada analisis deskriptif dan perencanaan strategis.

Sementara itu, R dan Matlab sering digunakan untuk analisis statistik dan numerik yang lebih kompleks. R dikenal karena kekuatannya dalam analisis statistik, dan sering digunakan dalam penelitian-penelitian yang memerlukan perhitungan statistik yang mendalam dan pengolahan data berbasis model matematis. R memiliki berbagai paket seperti *cluster* dan *factoextra* yang memungkinkan implementasi algoritma *clustering* dengan kemudahan dalam visualisasi hasil. Keunggulan utama R terletak pada kemampuannya untuk mengelola analisis statistik yang lebih lanjut, tetapi kurang fleksibel dibandingkan dengan Python dalam hal kemampuan pemrograman untuk tujuan yang lebih kompleks (Febriani, 2022). Di sisi lain, Matlab lebih unggul dalam pemrosesan matematis dan numerik yang rumit, serta sangat berguna untuk penelitian yang berfokus pada simulasi atau analisis berbasis matriks dan komputasi numerik yang memerlukan pemahaman mendalam tentang aljabar linier dan teori kontrol.

Meskipun Matlab kurang digunakan dalam *data mining* dibandingkan Python, alat ini tetap populer di kalangan ilmuwan teknik dan matematika.

Selain itu, Weka adalah perangkat lunak yang banyak digunakan oleh pemula dalam dunia *data mining* karena antarmukanya yang sederhana dan mudah dipahami. Weka menawarkan berbagai algoritma *clustering* seperti K-Means, Simple K-Means, dan EM clustering, yang dapat digunakan tanpa memerlukan keterampilan pemrograman. Keunggulan utama Weka adalah kemudahan penggunaan bagi pengguna yang tidak memiliki latar belakang teknis. Namun, Weka memiliki keterbatasan dalam hal fleksibilitas dan skalabilitas, terutama ketika menghadapi dataset yang lebih besar dan lebih kompleks. Dibandingkan dengan Python, yang memungkinkan pengguna untuk melakukan penyesuaian mendalam dan menangani volume data yang lebih besar, Weka lebih terbatas dalam kemampuannya untuk menangani data besar dan untuk pengolahan data yang memerlukan pemrograman tingkat lanjut (Putri & Dana, 2024). Meski demikian, Weka

tetap menjadi pilihan bagi pengguna yang baru memulai di bidang *data mining* atau yang membutuhkan alat untuk melakukan analisis dasar dengan cepat.

Pemilihan alat atau tools yang digunakan dalam penerapan *clustering* sangat bergantung pada kebutuhan analisis, kemampuan pemrograman, dan jenis data yang digunakan. Python dengan Scikit-Learn menjadi pilihan utama bagi peneliti yang bekerja dengan data besar dan kompleks, karena fleksibilitas dan skalabilitasnya. Bagi sektor yang mengutamakan kemudahan penggunaan dan antarmuka grafis, Rapid Miner dan Weka menjadi pilihan yang lebih sesuai, meskipun dengan keterbatasan dalam menangani dataset besar dan kustomisasi algoritma yang lebih dalam. Sementara itu, R dan Matlab tetap menjadi pilihan bagi mereka yang membutuhkan analisis statistik yang lebih mendalam dan pemrosesan numerik yang rumit. Pemilihan alat ini menunjukkan beragam pendekatan yang digunakan dalam analisis data, mulai dari penggunaan alat berbasis pemrograman lanjutan hingga penggunaan perangkat lunak



berbasis GUI yang ramah bagi pengguna non-teknis.

#### **D. Kesimpulan**

Berdasarkan kajian literatur dengan pendekatan PRISMA, penerapan K-Means Clustering terbukti efektif di berbagai sektor, seperti Industri dan Perdagangan, Pendidikan, Kesehatan dan Lingkungan, serta Keuangan dan Ekonomi. K-Means sangat populer karena kesederhanaan, efisiensi, dan kemampuannya dalam menangani data besar, meskipun metode ini menghadapi beberapa tantangan, seperti penentuan jumlah cluster yang optimal. Tantangan ini sering diatasi dengan menggunakan teknik seperti Elbow Method dan Silhouette Score. Selain itu, K-Means juga menghadapi masalah dalam penanganan outlier dan keterbatasan komputasi saat diterapkan pada data besar. Python dengan Scikit-Learn adalah alat yang paling banyak digunakan dalam penelitian ini, berkat fleksibilitasnya dalam menangani berbagai jenis data dan kemudahan implementasi berbagai algoritma clustering.

Penelitian ini memberikan kontribusi yang signifikan bagi peneliti, akademisi, dan praktisi di bidang

clustering. Bagi peneliti, kajian ini memperkaya pemahaman mengenai berbagai metode clustering yang digunakan di sektor-sektor berbeda serta tantangan yang dihadapi dalam penerapannya. Bagi akademisi, penelitian ini menjadi referensi untuk penelitian lebih lanjut, seperti pengembangan metode clustering hybrid atau peningkatan akurasi dalam pengelompokan data. Bagi praktisi, penelitian ini memberikan panduan praktis dalam memilih alat yang sesuai, seperti Python (Scikit-Learn) atau Rapid Miner, untuk analisis data, serta memberikan wawasan dalam pengambilan keputusan berbasis data.

Penelitian lanjutan sebaiknya mengeksplorasi integrasi K-Means Clustering dengan algoritma lain seperti DBSCAN atau K-Medoids untuk mengatasi masalah outliers dan data dengan kepadatan tidak merata. Selain itu, penggunaan Fuzzy Clustering bisa dieksplorasi untuk data yang tidak bisa dikelompokkan secara kaku, seperti data yang memiliki nilai ambiguitas. Penelitian juga dapat fokus pada penerapan clustering pada big data dan data tidak terstruktur, seperti teks atau gambar, yang memerlukan teknik pemrosesan

yang lebih canggih. Penggunaan teknik pemrosesan seperti MapReduce atau cloud computing juga dapat menjadi fokus utama untuk menangani data besar, mempercepat analisis, dan mengoptimalkan performa pemrosesan. Pemilihan metode dan tools yang tepat, disesuaikan dengan karakteristik data yang ada, akan meningkatkan akurasi dan relevansi hasil analisis.

#### **DAFTAR PUSTAKA**

- Fariq, A. (2011). Perkembangan dunia konseling memasuki era globalisasi. *Pedagogi*, II Nov 2011(Universitas Negeri Padang), 255-262
- S. Handoko, F. Fauziah, and E. T. E. Handayani, "Implementasi Data Mining Untuk Menentukan Tingkat Penjualan Paket Data Telkomsel Menggunakan Metode K-Means Clustering," *J. Ilm. Teknol. dan Rekayasa*, vol. 25, no. 1, pp. 76–88, 2020, doi: 10.35760/tr.2020.v25i1.2677.
- A. Winarta and W. J. Kurniawan, "Optimasi Cluster K-Means Menggunakan Metode Elbow Pada Data Pengguna Narkoba Dengan Pemrograman Python," *JTIK (Jurnal Tek. Inform. Kaputama)*, vol. 5, no. 1, pp. 113–119, 2021, doi: 10.59697/jtik.v5i1.593.
- N. Mirantika, "Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Penyebaran Covid-19 di Provinsi Jawa Barat," *Nuansa Inform.*, vol. 15, no. 2, pp. 92–98, 2021, doi: 10.25134/nuansa.v15i2.4321.
- Y. Wibowo, "Penerapan Data Mining Dalam Pengelompokan Data Member Card Mitra10 Untuk Meningkatkan Rewards Terhadap Konsumen dengan Metode Fuzzy Subtractive Clustering," *TIN Terap. Inform. Nusant.*, vol. 2, no. 8, pp. 471–475, 2022, doi: 10.47065/tin.v2i8.993.
- S. Febriani, "Analisis Data Hasil Diagnosa Untuk Klasifikasi Gangguan Kepribadian Menggunakan Algoritma C4.5 Siska Febriani Sistem Informasi \* ) Rohmansyah@gmail.com," vol. 2, no. 9, pp. 1–9, 2022.
- A. Rifqi and R. T. Aldisa, "Penerapan Data Mining Untuk Clustering Kualitas Udara," *J. Sist. Komput. dan Inform.*, vol. 5, no. 2, p. 289, 2023, doi: 10.30865/json.v5i2.7145
- I. Syafrinal and E. L. Febrianti, "Penerapan Algoritma K-Means Pada Aplikasi Data Mining Untuk Menentukan Pola Penjualan (Studi Kasus: Zahra Mart)," *J. Digit*, vol. 13, no. 1, p. 31, 2023, doi: 10.51920/jd.v13i1.320.
- F., F. T. Kesuma, and S. P. Tamba, "Penerapan Data Mining Untuk Menentukan Penjualan Sparepart Toyota Dengan Metode K-Means Clustering," *J. Sist. Inf. dan Ilmu Komput. Prima(JUSIKOM PRIMA)*, vol. 2,

- no. 2, pp. 67–72, 2020, doi: 10.34012/jusikom.v2i2.376.
- R. Mauliadi, “Data Mining Menggunakan Algoritma K-Means Clustering dalam Analisis Tingkat Potongan Harga Terhadap Harga Jual Sepeda Motor Honda,” *J. Inform. Ekon. Bisnis*, vol. 4, pp. 7–9, 2022, doi: 10.37034/infkeb.v4i4.156.
- S. Aulia, “Klasterisasi Pola Penjualan Pestisida Menggunakan Metode K-Means Clustering (Studi Kasus Di Toko Juanda Tani Kecamatan Hutabayu Raja),” *Djtechno J. Teknol. Inf.*, vol. 1, no. 1, pp. 1–5, 2021, doi: 10.46576/djtechno.v1i1.964.
- Y. R. Sari, A. Sudewa, D. A. Lestari, and T. I. Jaya, “Penerapan Algoritma K-Means Untuk Clustering Data Kemiskinan Provinsi Banten Menggunakan Rapidminer,” *CESS (Journal Comput. Eng. Syst. Sci.)*, vol. 5, no. 2, p. 192, 2020, doi: 10.24114/cess.v5i2.18519.
- F. D. Ana Neva, “Penerapan Metode K-Means Clustering dalam Mengelompokkan Jumlah Wisatawan Asing di Jawa Barat,” *Algor*, vol. 2, pp. 141–149, 2023, [Online]. Available: <https://jurnal.buddhidharma.ac.id/index.php/algor/article/view/1872%0Ahttps://jurnal.buddhidharma.ac.id/index.php/algor/article/download/1872/1495>
- E. M. Fitri, R. R. Suryono, and A. Wantoro, “Klasterisasi Data Penjualan Berdasarkan Wilayah Menggunakan Metode K-Means Pada Pt Xyz,” *J. Komputasi*, vol. 11, no. 2, pp. 157–168, 2023, doi: 10.23960/komputasi.v11i2.12582.
- A. Fikri Sallaby, R. Tri Alinse, V. Novita Sari, and T. Ramadani, “Pengelompokan Barang Menggunakan Metode K-Means Clustering Berdasarkan Hasil Penjualan Di Toko Widya Bengkulu PENGELOMPOKAN BARANG MENGGUNAKAN METODE K-MEANS CLUSTERING BERDASARKAN HASIL PENJUALAN DI TOKO WIDYA BENGKULU,” *J. Media Infotama*, vol. 18, no. 1, p. 2022, 2022.
- M. Siahaan, “Data Mining Strategi Pembangunan Infrastruktur Menggunakan Algoritma K-Means,” *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 11, no. 3, pp. 316–324, 2022, doi: 10.32736/sisfokom.v11i3.1453.
- Ramadhana, Islamiyah, and A. P. A. Masa, “Penerapan Data Mining Menggunakan Metode K-Means Clustering Pada Data Ekspor Batubara,” *Adopsi Teknol. dan Sist. Inf.*, vol. 2, no. 1, pp. 35–42, 2023, doi: 10.30872/atasi.v2i1.595.
- L. Magdalena and R. Fahrudin, “Penerapan Data Mining Untuk Koperasi Se-Jawa Barat Menggunakan Metode Clustering pada Kementerian Koperasi dan UKM,” *J. Digit*, vol. 9, no. 2, p. 190, 2020, doi: 10.51920/jd.v9i2.120.
- A. Putri Riyandoro, A. Voutama, and Y. Umaidah, “Implementasi Data

- Mining Clustering K- Means Dalam Menggolongkan Beragam Merek Laptop,” JATI (Jurnal Mhs. Tek. Inform., vol. 7, no. 2, pp. 1372–1377, 2023, doi: 10.36040/jati.v7i2.6816.
- J. Nasir, “Penerapan Data Mining Clustering Dalam Mengelompokan Buku Dengan Metode K-Means,” *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 11, no. 2, pp. 690–703, 2021, doi: 10.24176/simet.v11i2.5482.
- S. Isnanto and S. Widodo, “Penerapan Data Mining Pada Penerimaan Mahasiswa Baru Dengan Algoritma K-Means Clustering,” *J. Tek. Inf. dan Komput.*, vol. 4, no. 2, p. 158, 2021, doi: 10.37600/tekinkom.v4i2.367.
- M. R. A. Fernanda, P. Sokibi, and R. Fahrudin, “Sistem Prediksi Ketepatan Kelulusan Mahasiswa Berdasarkan Data Akademik Dan Non Akademik Menggunakan Metode K-Means (Studi Kasus : Universitas Catur Insan Cendekia),” *J. Digit*, vol. 11, no. 1, p. 89, 2021, doi: 10.51920/jd.v11i1.182.
- R. Astuti and K. Ukar, “Implementasi Data Mining dengan Metode Clustering Algoritma K- Means untuk Pengelompokan Data Tilang di Instansi Pemerintah,” *Media Inform.*, vol. 20, no. 2, pp. 109–121, 2021, doi: 10.37595/mediainfo.v20i2.77.
- C. Armayani, A. Fauzi, and H. Sembiring, “Implementasi Data Mining Pengelompokan Jumlah Data Produktivitas Ubinan Tanaman Pangan Berdasarkan Jenis Ubinan Dengan Metode Clustering Dikab Langkat (Studi Kasus : Badan Pusat Statistik Langkat),” *J. Inform. Kaputama*, vol. 5, no. 1, pp. 185–196, 2021, doi: 10.59697/jik.v5i1.318.
- B. Laksono, Y. Syahidin, and Y. Yunengsih, “Implementasi Data Mining Klasterisasi Data Pasien Rawat Inap dengan Algoritma K-Means Clustering,” *J. Teknol. Sist. Inf. dan Apl.*, vol. 7, no. 2, pp. 621–627, 2024, doi: 10.32493/jtsi.v7i2.39354.
- A. Nugraha, O. Nurdiawan, and G. Dwilestari, “Penerapan Data Mining Metode K-Means Clustering Untuk Analisa Penjualan Pada Toko Yana Sport,” JATI (Jurnal Mhs. Tek. Inform., vol. 6, no. 2, pp. 849–855, 2022, doi: 10.36040/jati.v6i2.5755.
- T. Lidia Putri and R. Danar Dana, “Penerapan Data Mining Pada Clustering Data Harga Rumah Dki Jakarta Menggunakan Algoritma K-Means,” JATI (Jurnal Mhs. Tek. Inform., vol. 8, no. 1, pp. 1174–1179, 2024, doi: 10.36040/jati.v8i1.8957.
- I. Maulana and U. Rosalina, “Clustering Data Nilai Ujian Akhir Semester Menggunakan Algoritma Data Mining K-Means,” *PERISKOP J. Sains dan Ilmu Pendidik.*, vol. 1, no. 2, pp. 76–85, 2020, doi: 10.58660/periskop.v1i2.10.
- M. Sholeh, S. Suraya, and D. Andayati, “Penerapan Data Mining pada Model Clustering

- Data Kuesioner Mahasiswa terhadap Kinerja Dosen,” J. Eksplora Inform., vol. 13, no. 2, pp. 208– 217, 2024, doi: 10.30864/eksplora.v13i2.751.
- J. Hutagalung and F. Sonata, “Penerapan Metode K-Means Untuk Menganalisis Minat Nasabah,” J. Media Inform. Budidarma, vol. 5, no. 3, p. 1187, 2021, doi: 10.30865/mib.v5i3.3113.
- I. Teknologi, D. A. N. Bisnis, M. A. Had, and I. Putri, “Kementerian pendidikan, kebudayaan, riset, dan teknologi institut teknologi dan bisnis palcomtech skripsi sistem pendukung keputusan siswa berprestasi berbasis,” 2023.