BAB II

LANDASAN TEORI

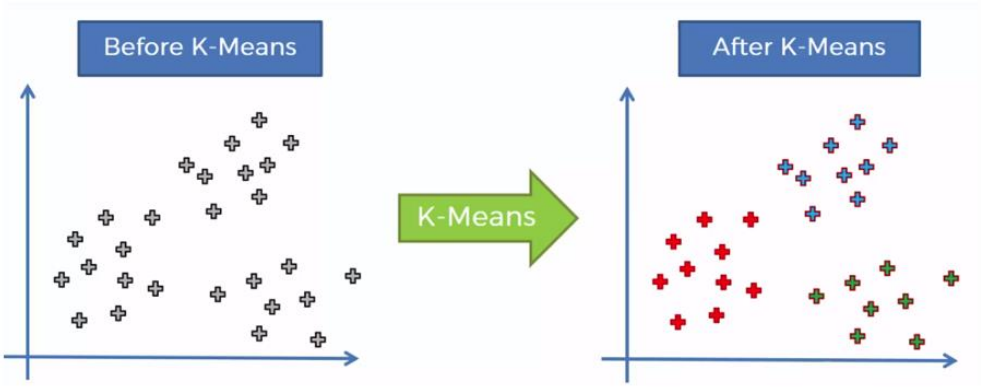
2.1 K-Means Clustering

K-Means adalah metode klasterisasi yang umum digunakan dalam analisis data untuk mengelompokkan data ke dalam sejumlah cluster berdasarkan kedekatan data dengan centroid. Menurut Khakim et al., metode ini efektif dalam mengelompokkan data berdasarkan parameter tertentu, seperti dalam analisis jaringan dokumentasi hukum. Menurut (Noviyanto, 2020) Clustering merujuk pada pengelompokan dokumen, observasi atau kasus pada kelas yang objeknya mirip. Klaster adalah kumpulan dokumen yang mirip satu sama lain dan berbeda dengan dokumen pada klaster lain. Clustering berbeda dengan Clasification, pada Clustering tidak ada target variabel untuk dikelompokkan. Algoritma Clustering mencoba untuk membagi kumpulan data menjadi klaster yang anggotanya relatif sama, dimana kemiripan dokumen di klaster yang sama tinggi, dan kemiripan dokumen di klaster lain kecil.

Clustering dianggap sebagai metode pembelajaran tanpa pengawasan yang paling penting, di mana masalah seperti itu adalah tentang menemukan pola dalam kumpulan data yang tidak berlabel. Cluster Clustering membagi kumpulan data menjadi beberapa kelompok dimana kesamaan pada suatu kelompok tertentu lebih besar daripada pada kelompok lainnya (Kamila, Khairunnisa and Mustakim, 2019).

Penggunaan algoritma pengelompokan tergantung pada jenis data yang tersedia untuk tujuan dan aplikasi tertentu. Jika analisis klaster digunakan sebagai

alat deskriptif atau eksplorasi, beberapa algoritma dapat dicoba pada data yang sama untuk mendapatkan apa yang diungkapkan oleh data tersebut. Secara umum metode Clustering dapat diklasifikasikan menjadi beberapa kategori, salah satunya adalah kategori metode partisi. Metode partisi ini didasarkan pada awalnya menentukan jumlah grup dan kemudian menetapkan kembali objek secara iteratif untuk menemukan grup yang terletak di suatu titik. Salah satu algoritma yang populer dalam penerapan metode partisi ini adalah algoritma K-Means dan algoritma K-Medoids (Kamila, Khairunnisa and Mustakim, 2019).



Gambar 2.1 *Algoritma K-Means Clustering*

2.2 Knowledge Discovery in Database (KDD)

Metode Knowledge Discovery in Databases (KDD) adalah proses yang sistematis untuk menemukan informasi yang berguna dari data yang besar dan kompleks. Salah satu teknik yang sering digunakan dalam KDD adalah algoritma K-Means Clustering, yang berfungsi untuk mengelompokkan data ke dalam cluster berdasarkan karakteristik yang sama. Menurut Qirom, K-Means Clustering dapat diterapkan dalam konteks medis, seperti dalam pengelompokan pasien hipertensi berdasarkan karakteristik tertentu, di mana evaluasi jumlah cluster yang optimal dilakukan menggunakan Davies Bouldin Index (DBI) Qirom (K. Handoko, 2020). Hal ini menunjukkan bahwa K-Means tidak hanya efektif dalam pengelompokan, tetapi juga dalam memberikan wawasan yang lebih dalam tentang data yang dianalisis.

(Sapitri, 2024) menambahkan bahwa K-Means adalah metode pembelajaran tanpa pengawasan yang digunakan untuk mengelompokkan data yang belum dilabel ke dalam cluster yang berbeda. Penelitiannya berfokus pada pengelompokan jumlah penduduk miskin di Jawa Barat, yang bertujuan untuk memberikan informasi yang berguna dalam penyaluran bantuan kepada masyarakat. Proses ini juga melibatkan tahapan KDD, yang mencakup seleksi data, pra-pemrosesan, dan transformasi data, sebelum akhirnya melakukan pengelompokan.

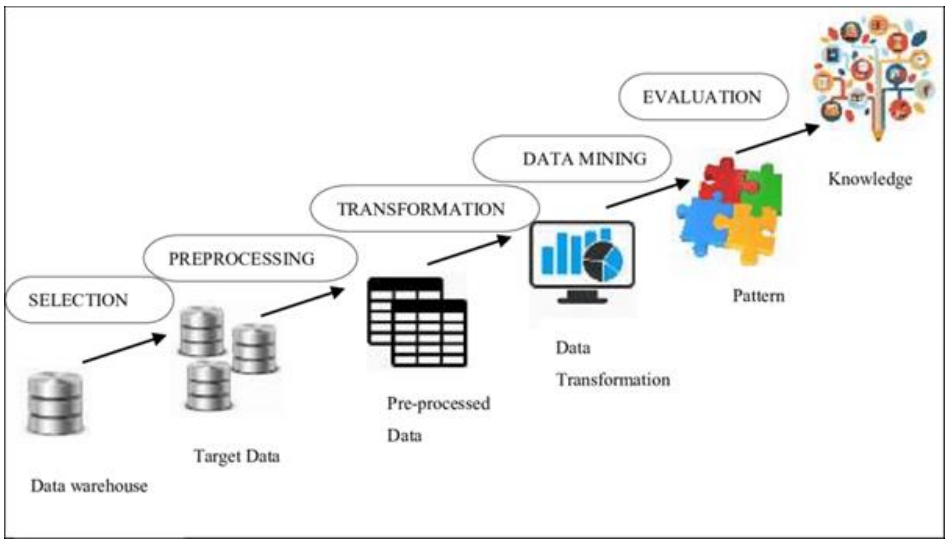
(Gustipartsani, 2023) menerapkan K-Means dalam analisis data kunjungan wisatawan di Kabupaten Karawang, dengan tujuan untuk meningkatkan pengelolaan objek wisata. Dalam penelitian ini, KDD digunakan untuk

mengidentifikasi pola pengunjung dan membantu dalam pengambilan keputusan terkait pengembangan wisata. Hal ini menunjukkan fleksibilitas K-Means dalam berbagai konteks aplikasi, dari kesehatan hingga pariwisata.

(N Azwanti, 2023) juga menggunakan K-Means dalam analisis data pemilu untuk memetakan pola persaingan antar partai politik di setiap daerah pemilihan. Dengan menerapkan KDD, peneliti dapat melakukan seleksi, pra-pemrosesan, dan evaluasi kluster untuk mendapatkan wawasan yang lebih baik tentang dinamika politik. Penelitian ini menyoroti pentingnya KDD dalam membantu pengambilan keputusan strategis berdasarkan data yang kompleks.

Knowledge discovery memiliki kelemahan pada data yang memiliki integrasi atau navigasi yang berbeda.. Pendekatan baru diperlukan jika suatu dimensi yang ada di dalam data juga meningkat. Maka didapatkan kesimpulan pengertian Knowledge Discovery Data (KDD) merupakan sejumlah data yang besar yang diekstrak melalui tahapan penggalian dan analisis untuk mendapatkan informasi dan pengetahuan yang berguna (Marisa et al, 2021).

K-Means Clustering dalam kerangka KDD merupakan alat yang sangat berguna untuk menganalisis dan mengelompokkan data dalam berbagai bidang. Dengan pendekatan yang sistematis, peneliti dapat mengidentifikasi pola dan hubungan dalam data yang kompleks, yang pada gilirannya dapat mendukung pengambilan keputusan yang lebih baik.



Gambar 2.2 *Proses Knowledge Discovery in Database (KDD)*

2.3 Keunggulan dan Keterbatasan K-Means

Metode K-Means memiliki keunggulan yang signifikan dalam pengelompokan data, termasuk kemudahan implementasi dan kecepatan dalam proses klasterisasi. K-Means efektif dalam mengelompokkan data yang memiliki struktur yang jelas dan dapat digunakan dalam berbagai aplikasi, seperti pemasaran dan analisis data sosial (C Wahana,2023). Namun, metode ini juga memiliki keterbatasan, terutama terkait dengan pemilihan titik awal centroid yang dapat mempengaruhi hasil klasterisasi dan berpotensi menyebabkan konvergensi pada solusi lokal. Selain itu, K-Means kurang efektif dalam menangani data dengan distribusi yang tidak teratur atau adanya outlier, yang dapat menyebabkan hasil kluster yang tidak representative. Oleh karena itu, meskipun K-Means merupakan alat yang berguna dalam analisis data, penting untuk mempertimbangkan konteks dan karakteristik data sebelum penerapannya.

Kelebihan K-means adalah kemampuannya untuk menangani dataset besar dan kompleks dengan efisiensi yang tinggi. Namun, metode ini juga memiliki kelemahan, seperti sensitivitas terhadap pemilihan titik awal centroid dan ketidakmampuan untuk menangani cluster dengan bentuk yang tidak bulat (MR Nahjan, 2023) Oleh karena itu, pemahaman yang mendalam tentang karakteristik data sangat penting untuk penerapan K-means yang efektif.

2.4 Sembako

Sembako, atau sembilan bahan pokok, merujuk pada kebutuhan dasar yang esensial bagi masyarakat, termasuk beras, minyak, gula, dan bahan makanan lainnya. Menurut Jazuli, pembagian paket sembako dalam kegiatan sosial merupakan bentuk kepedulian terhadap masyarakat yang membutuhkan, menunjukkan solidaritas sosial yang kuat. Selain itu, Salamah meneliti efektivitas program sembako selama pandemi COVID-19, yang menunjukkan bahwa pemahaman masyarakat terhadap program ini masih perlu ditingkatkan agar lebih tepat sasaran dan efektif (Dwy Ayu Istiqomah, 2022). Dalam konteks modern, aplikasi teknologi untuk perdagangan sembako juga mulai berkembang, seperti yang diungkapkan oleh Julianti et al., yang menunjukkan perlunya sistem penjualan berbasis teknologi untuk meningkatkan efisiensi dan jangkauan pasar. Dengan demikian, sembako tidak hanya berfungsi sebagai kebutuhan dasar, tetapi juga sebagai indikator solidaritas sosial dan kemajuan teknologi dalam distribusi kebutuhan pokok masyarakat.

Sembako, singkatan dari "sembilan bahan pokok," merujuk pada kebutuhan dasar yang penting bagi masyarakat Indonesia, termasuk beras, minyak goreng, gula, dan lainnya. Sembako memiliki peran krusial dalam perekonomian, terutama bagi usaha mikro, kecil, dan menengah (UMKM) yang menjual produk ini di pasar lokal. Dalam konteks ini, toko sembako sering kali berfungsi sebagai penyedia utama kebutuhan sehari-hari bagi masyarakat, dan persaingan di antara mereka sangat ketat. Selain itu, aplikasi teknologi seperti sistem informasi berbasis web untuk memantau harga dan penjualan sembako semakin berkembang, membantu pedagang dalam mengelola bisnis mereka. Program pemerintah juga berupaya meningkatkan aksesibilitas sembako, terutama selama masa krisis seperti pandemi COVID-19, meskipun efektivitasnya masih perlu ditingkatkan.

2.5 Data Mining

Data Mining adalah serangkaian proses untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu basis data. Informasi yang dihasilkan diperoleh dengan cara mengekstraksi dan mengenali pola yang penting atau menarik dari data yang terdapat pada basis data. Data mining terutama digunakan untuk mencari pengetahuan yang terdapat dalam basis data yang besar sehingga sering disebut Knowledge Dicovery Database (KDD) (Vulandari, 2019).. Girsang et al. menekankan bahwa data mining dapat membantu dalam memprediksi hasil, seperti dalam penentuan penerima program bantuan pemerintah. Selain itu, Takdirillah menjelaskan bahwa teknik data mining, seperti algoritma Apriori, dapat digunakan untuk menganalisis data transaksi guna mendukung strategi penjualan.

Data Mining adalah proses penggalian informasi dan pola yang bermanfaat dari suatu data yang sangat besar. Proses data mining terdiri dari pengumpulan data, ekstraksi data, analisa data, dan statistik data. Ia juga umum dikenal sebagai knowledge discovery, knowledge extraction, data/pattern analysis, information harvesting, dan lainnya (Arhami dan Nasir, 2020).

2.5.1 Langkah-langkah utama dalam proses data mining

Proses data mining melibatkan beberapa langkah utama yang saling terkait, dimulai dari pemrosesan data hingga analisis hasil. Langkah – langkah nya adalah sebagai berikut :

1. Langkah pertama adalah pengumpulan data, di mana data yang relevan dikumpulkan dari berbagai sumber.
2. Selanjutnya, tahap pemrosesan data (preprocessing) dilakukan untuk membersihkan dan menyiapkan data, termasuk penghapusan noise dan pengisian nilai yang hilang
3. Setelah data siap, teknik data mining seperti klasifikasi, asosiasi, dan klustering diterapkan untuk mengekstrak pola dan informasi yang berguna
4. Setelah penerapan teknik mining, hasil yang diperoleh perlu dievaluasi dan dianalisis untuk memastikan keakuratan dan relevansinya
5. Langkah terakhir adalah penyajian hasil, di mana informasi yang diperoleh disajikan dalam format yang mudah dipahami untuk pengambilan keputusan lebih lanjut

Proses ini mencerminkan kompleksitas dan interkoneksi antara berbagai langkah dalam data mining, yang merupakan bagian integral dari penemuan pengetahuan dalam basis data

2.5.2 Tahapan data mining mendukung implementasi metode K-Means

Tahapan data mining yang mendukung implementasi metode K-Means meliputi beberapa langkah penting, yaitu pengumpulan data, pra-pemrosesan, pemilihan fitur, pengelompokan, dan evaluasi hasil. Pengumpulan data dilakukan untuk memperoleh informasi yang relevan dari sumber yang beragam, seperti data penjualan atau data Kesehatan. Selanjutnya, pra-pemrosesan data penting untuk membersihkan dan menyiapkan data agar bebas dari noise dan konsisten, yang merupakan langkah krusial sebelum pengelompokan. Setelah data siap, pemilihan fitur dilakukan untuk menentukan atribut yang paling relevan untuk analisis. Metode K-Means kemudian diterapkan untuk mengelompokkan data berdasarkan kesamaan karakteristik, seperti dalam pengelompokan pelanggan atau produk. Terakhir, evaluasi hasil clustering penting untuk menilai efektivitas pengelompokan yang dilakukan, seringkali menggunakan metrik seperti silhouette score atau elbow method untuk menentukan jumlah cluster yang optimal. Dengan demikian, tahapan ini sangat mendukung implementasi metode K-Means dalam berbagai konteks analisis data.

2.6 Rapid Miner

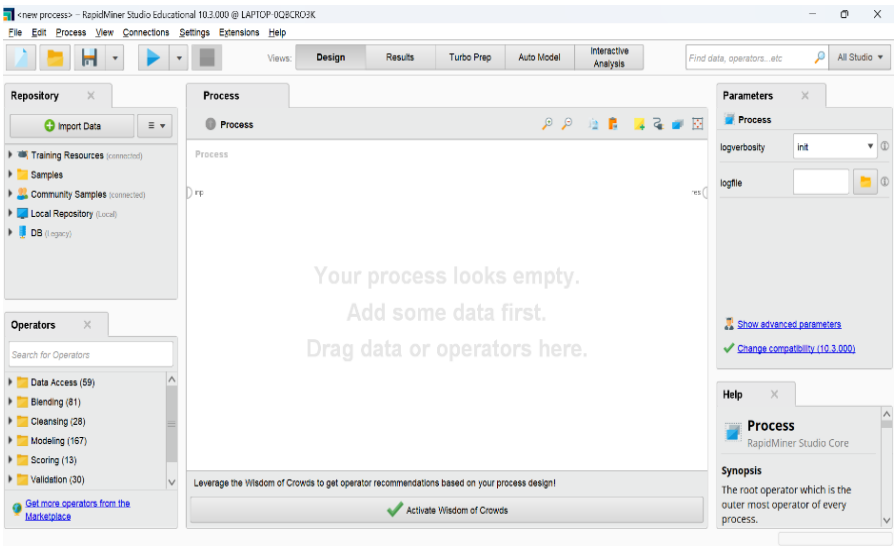
Rapid Miner adalah sebuah software untuk pengolahan data mining. Rapid Miner adalah sebuah solusi untuk melakukan analisis terhadap data mining, text

mining dan analisis prediksi. “RapidMiner menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik”(Setiawan dalam jurnal Purwanto & Darmadi, 2019:45). Rapid Miner memiliki kurang lebih 500 operator data mining, termasuk input, output, data preprocessing dan visualisasi. RapidMiner merupakan software yang berdiri sendiri untuk analisis data dan sebagai mesin data mining yang diintegrasikan pada produknya sendiri. RapidMiner ditulis menggunakan bahasa java sehingga dapat bekerja disemua sistem operasi

RapidMiner adalah platform perangkat lunak yang digunakan untuk analisis data dan pembelajaran mesin. Menurut para ahli, RapidMiner memungkinkan pengguna untuk melakukan eksplorasi data, pemodelan, dan evaluasi hasil dengan antarmuka yang intuitif, sehingga memudahkan pengguna yang tidak memiliki latar belakang teknis untuk menerapkan teknik analisis data yang komplek. RapidMiner juga mendukung berbagai metode analisis, termasuk analisis prediktif dan analisis statistik, yang dapat diintegrasikan dengan berbagai sumber data. Lebih lanjut, RapidMiner dikenal karena kemampuannya dalam mengotomatisasi proses analisis data, yang membantu dalam pengambilan keputusan berbasis data. Dengan demikian, RapidMiner menjadi alat yang sangat berharga dalam berbagai bidang, termasuk bisnis, kesehatan, dan penelitian akademis, karena dapat meningkatkan efisiensi dan efektivitas analisis data.

RapidMiner adalah perangkat lunak yang digunakan untuk analisis data dan pengolahan data mining. Perangkat ini memungkinkan pengguna untuk melakukan analisis teks, ekstraksi pola dari dataset, dan menggabungkan metode statistik serta

kecerdasan buatan untuk menghasilkan informasi yang berkualitas tinggi dari data yang diolah. RapidMiner juga mendukung berbagai algoritma machine learning, termasuk algoritma klasifikasi seperti Naïve Bayes dan Decision Tree, yang sering digunakan dalam penelitian untuk memprediksi dan mengklasifikasikan data. Keunggulan RapidMiner terletak pada antarmuka grafisnya yang memudahkan pengguna dalam merancang alur kerja analisis. Selain itu, RapidMiner dapat digunakan dalam berbagai aplikasi, mulai dari analisis sentimen di media sosial hingga pengembangan refrigeran ramah lingkungan.



Gambar 2.3 *Tampilan RapidMiner Versi 10.3*

2.7 Peneliti Terdahulu

Berikut adalah beberapa penelitian terdahulu yang relevan dengan penggunaan metode *K-Means*

Tabel 3.1 *Penelitian Terdahulu*

Referensi Penelitian	1
Judul	Implementasi Data Mining Untuk Menentukan Persediaan Stok Obat Di Apotek K-24 Menggunakan Metode K-Means Clustering
Nama Penulis	Desy Ayu Ramadhanty, Renita Syafitri, Errissya Rasywir, Despita Meisak
Tahun	2022
Hasil	Pengolahan data mining telah berkembang sangat pesat, beradaptasi dengan segala bentuk analisis data. Pada dasarnya, data mining dapat menganalisis data untuk menggunakan teknik perangkat lunak untuk menemukan pola dalam kumpulan data tersembunyi. Manajemen persediaan yang tinggi dan tidak ekonomis karena beberapa produk mungkin memiliki ruang dan kelebihan. Hal ini tentu sangat merugikan pelaku usaha seperti tempat kesehatan Apotek K-24. Metode K-Means sudah menjadi salah satu teknik data mining yang digunakan untuk merancang strategi persediaan atau buku pesanan

	yang efektif menggunakan data transaksi penjualan bisnis. Tujuan penelitian dari penelitian ini adalah untuk menerapkan algoritma KMeans, dan data transaksi obat dari Apotek K-24 di berikan sebagai contoh tipikal. Hasil analisis untuk penelitian ini menggunakan 20 buah data.

Referensi Penelitian	2
Judul	Data Mining untuk Pengelompokan Saham pada Sektor Energi dengan Metode K-Means
Nama Penulis	Anggi Srimurdianti Sukamto, Wawan Setiawan, Enda Esyudha Pratama
Tahun	2023
Hasil	Saham adalah kepemilikan hak oleh perorangan (pemegang saham) pada suatu perusahaan berdasarkan pemberian modal sehingga dianggap memiliki kepemilikan dan pengawasan perusahaan tersebut berdasarkan bagian tertentu. Menurut data dari Indonesia Stock Exchange (IDX) pada tahun 2020, jumlah investor di Pasar Modal Indonesia yang terdiri dari investor saham, reksadana dan obligasi, mengalami kenaikan 56 persen yaitu 3,87 juta Single Investor Identification (SID) sampai pada tanggal 29 Desember 2020. Kenaikan ini

	menjadi 4 kali lipat lebih tinggi sejak 4 tahun terakhir. Investor saham juga mengalami kenaikan sebanyak 53 persen menjadi 1,68 juta SID. Hal tersebut menunjukkan besarnya minat masyarakat terhadap keikutsertaan pada kepemilikan saham. Namun dalam berinvestasi terdapat risiko. Risiko dalam berinvestasi di pasar modal sebenarnya dapat diminimalisir dengan pemilihan saham yang benar terutama dalam hal fundamendal perusahaan. Penelitian ini menggunakan metode K-Means untuk mengelompokan saham sesuai dengan karakteristiknya. Berdasarkan perhitungan, didapatkan sebanyak 5 kali Iterasi untuk 4 Kelas/Cluster yang telah didefinisikan diawal. Selain itu, didapatkan hasil bahwa Kelas/Cluster 1 dan 4 diisi oleh emiten-emiten yang memiliki fundamental buruk serta Kelas/Cluster 2 berisi emiten pemberi dividen yang tinggi. Secara keseluruhan, didapatkan kesimpulan bahwa algoritma K-Mean dapat digunakan untuk membantu para Investor dalam melakukan pencarian emiten yang sesuai dengan karakteristik yang diinginkan.
Referensi Penelitian	3
Judul	Tinjauan Pustaka Sistematis Pada Data Mining : Studi Kasus Algoritma K-Means Clustering

Nama Penulis	Sekar Setyaningtyas, Bangkit Indarmawan Nugroho, Zaenul Arif
Tahun	2022
Hasil	Data Mining adalah metode untuk menganalisis pola dan karakteristik di masa depan serta untukmengumpulkan informasi tak terduga yang belum pernah terlihat sebelumnya dari database yang besar.Dalam data mining, clustering adalah salah satu teknik yang berguna untuk analisis data. Salah satualgoritma data mining adalah algoritma K-Means yang merupakan teknik clustering berdasarkanpembagian jarak. Tujuan yang ingin dicapai dalam paper ini yakni menganalisis teknik clusteringmenggunakan algoritma K-Means dalam data mining dengan melakukan review secara mendalam danmengevaluasi penelusuran melalui literatur terpilih berdasarkan kriteria tertentu dan studi yang dipilihakan diproses untuk menjawab pertanyaan penelitian. Tinjauan Pustaka Sistematis (Systematic LiteratureReview/SLR) merupakan sebuah metode penelitian yang bertujuan untuk mengidentifikasi danmengevaluasi hasil penelitian dengan teknik terbaik berdasarkan prosedur yang spesifik dari hasilperbandingan. Berdasarkan pemilihan literatur

	publikasi jurnal, Pattern Recognition, Knowledge-BasedSystem, Applied Soft Computing dan IEEE Access dapat menjadi rujukan utama terkait algoritma KMeans. Hasil perbandingan metode menunjukkan bahwa Euclidean Distance memiliki keunggulanperhitungan jarak yang lebih baik, sehingga metode tersebut dapat dijadikan sebagai pilihan utamaterkait teori perhitungan jarak pada algoritma K-Means.

Referensi Penelitian	4
Judul	Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan
Nama Penulis	Gustientiedina Gustientiedina, M. Hasmil Adiya, Yenny Desnelita
Tahun	2019
Hasil	Perencanaan dari kebutuhan obat-obatan yang tepat dapat membuat pengadaan obat-obatan menjadi efektif dan efisien sehinggaobat-obatan dapat tersedia dengan cukup sesuai dengan kebutuhan serta dapat diperoleh pada saat yang diperlukan. Menganalisa pemakaian obat, perencanaan dan pengendalian obat-obatan dapat dilakukan pada data miningyaitu dengan

	clusterisasi.Metode yang akan di pakai untuk clustering data obat-obatan adalah algoritma K-Means yang mana merupakan metode clustering dengan non hirarki yang mempartisi data â€“ dataÂ kedalam cluster dimana data â€“ datadengan karakteristik sama akan dikelompokkan padasatu cluster dan data â€“ data dengan karakteristik yang berbeda akan dikelompokkan padacluster lainnya.Tujuan penelitian ini yaitumengelompokkan data obat-obatan pada rumah sakitsehingga dapat digunakan dalam acuan pengambilan keputusan perencanaan dan pengendaliaan persediaan obat-obatan di rumah sakit.

Referensi Penelitian	5
Judul	Implementasi Data Mining Untuk Menentukan Persediaan Stok Obat Di Enok Menggunakan Metode K-Means Clustering
Nama Penulis	Ferlanda, Septi Andryana, Eri Mardiani
Tahun	2021
Hasil	Pengolahan data mining telah berkembang sangat pesat, beradaptasi dengan segala bentuk analisis data. Pada dasarnya, data mining dapat menganalisis data untuk menggunakan teknik perangkat lunak untuk

	menemukan pola dalam kumpulan data tersembunyi. Manajemen persediaan yang tinggi dan tidak ekonomis karena beberapa produk mungkin memiliki ruang dan kelebihan. Hal ini tentu sangat merugikan pelaku usaha seperti tempat kesehatan Apotek Enok. Metode K-Means sudah menjadi salah satu teknik data mining yang digunakan untuk merancang strategi persediaan atau buku pesanan yang efektif menggunakan data transaksi penjualan bisnis. Tujuan penelitian dari penelitian ini adalah untuk menerapkan algoritma K-Means, dan data transaksi obat dari Apotek Enok di berikan sebagai contoh tipikal. Hasil analisis untuk penelitian ini menggunakan 20 buah data. Pengumpulan data obat yang di lakukan dengan algortma K-Means diulang sebanyak tiga kali, kemudian didapatkan hasil kelompok yaitu kelompok 1 berisi 6 obat kerja lambat dan kelompok 2 terdapat 14 obat kerja cepat. Pencarian cluster ini menggunakan fitur web untuk menemukan produk lambat dan obat cepat.

Referensi Penelitian	6

Judul	Pengelompokan Data Nilai Siswa Menggunakan Metode K-Means Clustering
Nama Penulis	Aditia Yudistira, Rio Andika
Tahun	2023
Hasil	Dalam proses pembelajaran terdapat 3 kategori penilaian yaitu nilai siswa, disiplin, serta sikap. Hasil pengelompokan data nilai siswa menggunakan metode *K-Means clustering* menunjukan bahwa berdasarkan hasil cluster data siswa menggunakan dataset siswa dalam satu semester, maka didapatkan *cluster* 0 berjumlah 59 siswa, *cluster* 1 berjumlah 94 siswa, dan *cluster* 2 berjumlah 1 siswa. Hasil pengujian menggunakan *elbow method* maka jumlah cluster yang baik yang digunakan adalah 3 *cluster*, sehingga dalam penelitian ini menggunaka 3 *cluster* yaitu *cluster* 0, *cluster* 1, dan *cluster* 2. Hasil pengujian menggunakan *silhouette coefficient* maka jumlah *cluster* yang baik yang digunakan adalah 3 *cluster* dengan nilai *silhouette coefficient* yaitu 0.489, dan lebih baik dari nilai *silhouette coefficient cluster* lainnya.

Referensi Penelitian	7
Judul	Penerapan Data Mining untuk Memprediksi Siswa Berprestasi dengan Menggunakan Algoritma K Nearest Neighbor
Nama Penulis	Sri Widaningsih
Tahun	2022
Hasil	Proses penentuan siswa berprestasi di SMK Tunas Sinar Mandiri Cianjur tidak hanya ditentukan dari nilai akademik saja, tetapi dipertimbangkan beberapa aspek non akademik seperti jenis kelamin, keaktifan ekstrakulikuler, presensi, dan kepribadian. Untuk nilai akademik, diambil dari nilai mata pelajaran produktif kelas X dan XI. Untuk memprediksi siswa mana yang berprestasi di kelas XII dapat diterapkan teknik data mining dengan menggunakan algoritma k nearest neighbor (kNN) dengan k = 1. Tahapan proses data mining mengikuti tahapan dalam Knowledge Discovery di Databases (KDD). Tahapan ini dimulai dari selection, preprocessing, transformation, data mining, dan evaluation/interpretation. Proses transformasi data menggunakan metoda min- max dan rumus jarak yang digunakan yaitu jarak Euclidean. Pembuatan aplikasi data mining ini mengikuti paradigma

	waterfall. Dari hasil analisis dan desain dengan menggunakan unified modeling langunge (UML) dihasilkan beberapa halaman utama yaitu kelola data siswa, penentuan variabel dan bobot variabel, Dengan penerapan teknik data mining ini dapat mempermudah pihak sekolah mempersiapkan siswa yang akan mendapatkan beasiswa