

Metode K-Means dalam Mengukur Tingkat Pemahaman Materi Mata Kuliah dan Penilaian Mahasiswa Prodi Informatika

Deti Karmanita^{1✉}, Sarjon Defit², Sumijan²

¹ Fakultas Ilmu Komputer, Universitas Dehasen, Bengkulu, 38228, Indonesia

² Fakultas Ilmu Komputer, Universitas Putra Indonesia "YPTK", Padang, 25221, Indonesia

detikarmanita@unived.ac.id

Abstract

Data mining is a process that uses statistical, mathematical, artificial intelligence, and machine learning techniques to extract and identify useful information and related knowledge from large databases. KDD is often referred to as knowledge discovery in databases. Data mining has five main functions, including clustering, classification, association, sequencing and forecasting. Clustering algorithms attempt to separate existing information sets into homogeneous or similar groups. The degree of similarity of data within a group will result in a greater value, while the difference between groups will result in a smaller value. The K-Means algorithm is part of data mining clustering, where the K-Means algorithm can be used to form new groups of data. The formation of new groups of data in the K-Means algorithm with the process of forming clusters in the process carried out. This study aims to determine the level of understanding of core course material in the Dehasen University Informatics study program received by students. where researchers distributed questionnaires to students to determine the level of understanding of core course material into 4 groups, namely very good, good, quite good and less good. The method used is K-Means with stages namely data selection, pre-processing, data transformation, information extraction and result evaluation. The data consists of 46 core courses taken from the Informatics Study Program curriculum which are assessed by students with an understanding of the course material Very Good as much as 29%, Good as much as 35%, Good Enough as much as 24%, Less Good as much as 12%. This study shows that the implementation of the K-Means method with the support of the RapidMiner application is effective in classifying data on student material understanding and the results can be used for evaluation and improvement of teaching quality.

Keywords: *Data Mining, KDD, Clustering, K-Means, Material Understanding*

Abstrak

Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar. KDD sering disebut juga sebagai penemuan pengetahuan dalam basis data. Data Mining memiliki lima fungsi utama, termasuk pengelompokan (*clustering*), klasifikasi (*classification*), asosiasi (*association*), urutan (*sequencing*) dan peramalan (*forecasting*). Algoritma *clustering* berupaya memisahkan kumpulan informasi yang ada menjadi kelompok-kelompok yang homogen atau sejenis. Tingkat kesamaan data di dalam suatu kelompok akan menghasilkan nilai yang semakin kecil, sementara perbedaan antar kelompok akan menghasilkan nilai yang semakin besar. Algoritma *K-Means* merupakan bagian dari clustering data mining, dimana algoritma *K-Means* dapat dipergunakan untuk pembentukan kelompok baru dari data. Pembentukan kelompok baru dari data pada algoritma *K-Means* dengan proses pembentukan *cluster* pada proses yang dilakukan. Penelitian ini bertujuan untuk mengetahui tingkat pemahaman materi mata kuliah inti pada program studi Informatika Universitas Dehasen yang diterima oleh mahasiswa. dimana peneliti menyebarkan kuisioner kepada mahasiswa untuk menentukan tingkat pemahaman materi mata kuliah inti menjadi 4 kelompok yaitu sangat baik, baik, cukup baik dan kurang baik. Metode yang digunakan adalah K-Means dengan tahapan yaitu pemilihan data, pra-pemrosesan, transformasi data, ekstraksi informasi dan evaluasi hasil. Data terdiri dari 46 mata kuliah inti yang di ambil dari kurikulum Prodi Informatika yang dinilai oleh mahasiswa dengan pemahaman materi mata kuliah Sangat Baik sebanyak 29%, Baik sebanyak 35%, Cukup Baik sebanyak 24%, Kurang Baik sebanyak 12%. Penelitian ini menunjukkan bahwa implementasi metode K-Means dengan dukungan aplikasi RapidMiner efektif dalam mengelompokkan data pemahaman materi mahasiswa dan hasilnya dapat digunakan untuk evaluasi dan peningkatan kualitas pengajaran.

Kata kunci: *data mining, KDD, clustering, k-means, pemahaman materi*

KomtekInfo is licensed under a Creative Commons Attribution-Share Alike 4.0 International License.



1. Pendahuluan

Perkembangan teknologi Indonesia yang semakin pesat telah menandakan adanya perubahan transisi pada era kehidupan manusia yang baru. Hal ini tidak dapat

dipungkiri bahwa kehidupan manusia telah memasuki di era digital 5.0 yang memberikan dampak dan efek dengan bertujuan untuk mempermudah segala pekerjaan manusia. Dalam bidang teknologi memberikan pengaruh terhadap sektor bidang di

kehidupan manusia baik dari segi ekonomi, sosial, kesehatan hingga di pemerintahan. Terdapat banyak perubahan yang terjadi akibat dari perkembangan teknologi salah satunya terjadinya pemanfaatan sistem digital yang awalnya setiap data di kolektifkan secara konvensional saat ini data yang dikumpulkan dari instansi sudah berbasis komputer. Melalui pemanfaatan teknologi dan informasi tiap harinya setiap instansi dapat menghasilkan hingga triliunan byte data. Data-data tersebut berupa tulisan, gambar, suara, angka, dan lain-lain. Dari sekian banyaknya data yang di kumpulkan setiap waktunya hal ini mengarah kepada pemanfaatan Big data yang dapat mengolah serta menganalisis data-data tersebut menjadi lebih mudah dibaca dan dicari apabila data tersebut di butuhkan [1].

Data mining atau yang sering disebut KDD (*Knowledge Discovery in Database*) adalah proses yang mencakup pengumpulan dan pemanfaatan riwayat data untuk menemukan pola dan hubungan dalam kumpulan data yang besar. Penambangan data menghasilkan informasi yang dapat digunakan untuk membuat keputusan di masa mendatang. Salah satu teknik yang dikenal dalam penambangan data adalah pengelompokan atau pengelompokan data (*clustering*) [2].

Data adalah informasi dan angka yang dikumpulkan untuk digunakan dalam berbagai jenis penelitian dan analisis sebagai dukungan untuk pandangan atau pendapat [3]. Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar [4].

Data Mining memfasilitasi pengguna untuk mengungkap informasi yang mungkin tidak akan terlihat oleh mereka saat melakukan eksplorasi data dalam basis data. Data Mining adalah salah satu komponen dalam proses *Knowledge Discovery in Databases* (KDD), yang melibatkan serangkaian langkah, termasuk pemilihan data, pra-pemrosesan, transformasi data, ekstraksi informasi, dan evaluasi hasil. KDD sering disebut juga sebagai penemuan pengetahuan dalam basis data. Data Mining memiliki lima fungsi utama, termasuk pengelompokan (*clustering*), klasifikasi (*classification*), asosiasi (*association*), urutan (*sequencing*) dan peramalan (*forecasting*) [5].

Clustering adalah salah satu teknik yang termasuk dalam fungsi data mining. Algoritma clustering digunakan untuk mengelompokkan sejumlah data ke dalam kelompok-kelompok tertentu, yang disebut cluster. Data mining, yang sering disebut sebagai *Knowledge Discovery in Database* (KDD), adalah proses yang mencakup pengumpulan dan penggunaan data historis untuk mengidentifikasi pola, hubungan, atau keteraturan dalam data berukuran besar [6].

Algoritma *clustering* berupaya memisahkan kumpulan informasi yang ada menjadi kelompok-kelompok yang homogen atau sejenis. Tingkat kesamaan data di dalam suatu kelompok akan menghasilkan nilai yang semakin besar, sementara perbedaan antar kelompok akan menghasilkan nilai yang semakin kecil [7].

Algoritma *K-Means* merupakan bagian dari clustering data mining, dimana algoritma *K-Means* dapat dipergunakan untuk pembentukan kelompok baru dari data. Pembentukan kelompok baru dari data pada algoritma *K-Means* dengan proses pembentukan *cluster* pada proses yang dilakukan [8].

Berdasarkan penelitian sebelumnya oleh Afiasari, dkk mengelompokkan 160 produk di Toko Bill Lights berdasarkan penjualan dengan metode K-Means. Hasilnya: Cluster 0 (Produk Laku): 88 produk dengan penjualan tinggi, Cluster 1 (Produk Tidak Laku): 26 produk dengan penjualan nol, Cluster 2 (Produk Kurang Laku): 46 produk dengan penjualan rendah. Hasil ini membantu toko dalam mengelola stok dan strategi penjualan lebih efektif [9].

Penelitian juga dilakukan oleh Fadila dan Rinaldi mengelompokkan jumlah UMKM di kabupaten/kota di Jawa Barat berdasarkan data dari tahun 2016 hingga 2023 dengan metode algoritma K-Means dan di implementasikan menggunakan aplikasi RapidMiner, ditemukan bahwa pengelompokan terbaik adalah pada $K=2$ dengan nilai Davies Bouldin Index (DBI) sebesar 0.619. Cluster pertama (Cluster 0) terdiri dari 112 data dengan jumlah UMKM tinggi, sedangkan cluster kedua (Cluster 1) terdiri dari 104 data dengan jumlah UMKM sedang. Total data yang dianalisis adalah 216 [10].

Selanjutnya penelitian yang dilakukan oleh Ferdy Pangestu, dkk tentang pengelompokan data obat-obatan dengan metode Algoritma K-Means berdasarkan karakteristiknya menjadi empat kategori: obat bebas, obat bebas terbatas, obat keras, dan psikotropika & narkotika. Pengelompokan ini memudahkan dalam pemantauan stok obat dan pengelolaan persediaan di Puskesmas Kecamatan Sawah Besar. Akurasi pengelompokan data obat telah dihitung, dengan nilai akurasi untuk cluster 1 sebesar 66,23%, cluster 2 sebesar 7,69%, dan cluster 3 sebesar 23,07% [11].

Inas Ajeng Nur Afifah dan Heri Nurdianto melakukan penelitian tentang meningkatkan efisiensi pengelolaan buku di Perpustakaan Daerah Kota Metro, dengan metode algoritma K-Means. Data: 200 buku, kategori filsafat, psikologi, agama, ilmu sosial, bahasa. Data peminjaman Januari-Desember. Metode: K-Means Clustering, 3 cluster (prioritas tinggi, dipertimbangkan, bukan prioritas). Penghitungan jarak dengan Euclidean Distance. Hasil: Cluster 0: Buku peminjaman tinggi (prioritas tinggi). Cluster 1: Buku peminjaman sedang (dipertimbangkan). Cluster 2: Buku peminjaman rendah (bukan prioritas) [12].

Penelitian oleh Muhammad Rafi Nahjan, dkk untuk menyusun strategi pemasaran, seperti menyesuaikan harga dan profit pada transaksi serta mempertimbangkan promosi untuk meningkatkan penjualan dengan metode clustering K-Means didapat transaksi yang masuk dalam Cluster 2 paling diminati oleh pelanggan, sedangkan transaksi di Cluster 1 kurang diminati meskipun memiliki keuntungan lebih besar. Cluster 0 menunjukkan hasil yang cukup baik [13].

Menurut Dodi Nofri Yoliadi dengan menggunakan algoritma clustering untuk mengelompokkan data penjualan barang elektronik dari CV. Berkah Elektronik menjadi dua cluster: produk laku dan tidak laku. Data yang dianalisis meliputi merk, jenis, nama, harga dan jumlah barang terjual. Hasilnya menunjukkan bahwa 12 produk tidak laku dan 28 produk laku, konsisten baik dalam pengolahan manual maupun menggunakan software Tanagra 1.4.48 [14].

Penelitian menggunakan algoritma K-Means clustering dilakukan oleh Slamet Pujiono dkk untuk analisis penjualan produk fashion. Cluster: Data dibagi menjadi 2 cluster: Cluster_0: 616 item produk terlaris. Cluster_1: 8 item produk kurang laris. DBI: Nilai terbaik $k=2$ dengan DBI -0.310 , menunjukkan pembagian cluster optimal. Rekomendasi: Cluster_0: Tingkatkan stok untuk memenuhi permintaan tinggi. Cluster_1: Lakukan strategi promosi dan perbaikan produk [15].

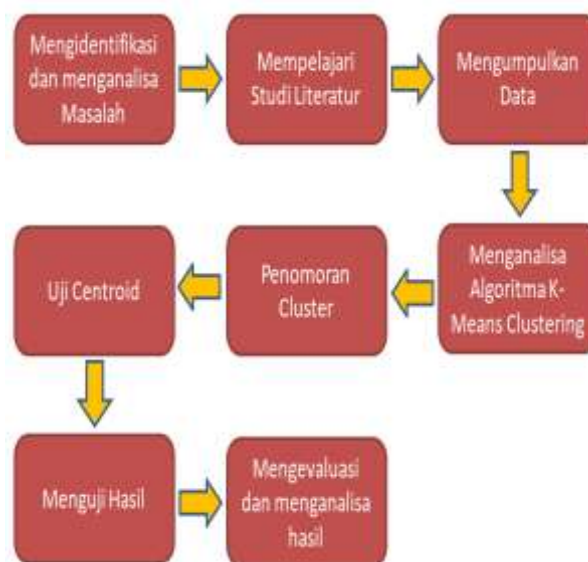
Penerapan metode K-Means sebagai alat analisis untuk mengukur tingkat pemahaman mahasiswa terhadap materi mata kuliah inti dalam Program Studi Informatika. Dengan menggunakan teknik clustering ini, diperoleh gambaran yang lebih objektif dan berbasis data mengenai variasi pemahaman di antara mahasiswa. Hasil analisis ini tidak hanya memungkinkan penilaian yang lebih mendalam dan adil, tetapi juga dapat digunakan untuk meningkatkan kualitas pengajaran melalui strategi pembelajaran yang lebih adaptif dan tepat sasaran.

Penelitian ini juga berkontribusi pada literatur akademik dengan memperluas penggunaan teknologi dan data science di bidang pendidikan, khususnya dalam evaluasi akademik. Secara keseluruhan, penerapan K-Means dapat mendukung pengambilan keputusan yang lebih efektif dalam perencanaan kegiatan belajar mengajar dan peningkatan sistem penilaian, sehingga menciptakan lingkungan belajar yang lebih inklusif dan mendorong keberhasilan akademik mahasiswa.

2. Metodologi Penelitian

Pada tahapan ini akan diuraikan mengenai metodologi penelitian dan kerangka kerja. Kerangka kerja merupakan Langkah-langkah yang digunakan untuk menyelesaikan permasalahan yang di bahas. Lebih

jelasan mengenai kerangka kerja ini dapat dilihat pada Gambar 1.



Gambar 1. Kerangka Kerja Penelitian

Berdasarkan alur kerja yang ditampilkan pada Gambar 1, setiap tahapan dalam alur tersebut dapat dijelaskan secara rinci. Setiap langkah dalam alur kerja ini menunjukkan urutan proses yang harus diikuti dari awal hingga akhir. Penjelasan detail dari tiap langkah akan membantu memahami bagaimana setiap tahap saling berhubungan dalam keseluruhan proses.

1. Menganalisa dan mengidentifikasi masalah

Pada tahap ini dilakukan analisis terhadap masalah dengan memahami masalah yang ada yaitu masih terbatasnya pemanfaatan *clustering* data. Pada tahap ini kemudian akan dirumuskan masalah yang akan menjadi objek penelitian. Perumusan masalah dilakukan untuk menentukan masalah apa saja yang terdapat pada objek penelitian serta memberikan batasan dari permasalahan yang akan diteliti

2. Melakukan studi literatur

Tahap ini merupakan tahap dimana literatur-literatur yang membahas mengenai data mining, clustering dan literatur-literatur lain yang berkaitan dengan penelitian dikumpulkan dan dijadikan referensi untuk membantu proses penelitian ini.

3. Mengumpulkan data

Dalam pengumpulan data dilakukan observasi yaitu pengamatan secara langsung ditempat penelitian sehingga masalah yang ada dapat diketahui dengan jelas. Kemudian dilakukan wawancara dan kuesioner yang bertujuan untuk mendapatkan informasi data yang dibutuhkan. Selain itu juga dilakukan studi kepustakaan yaitu dengan membaca buku-buku yang menunjang dalam

melakukan penganalisaan terhadap data dan informasi yang didapat

4. Menganalisa Algoritma K-Means *Clustering*

Pada tahap ini dilakukan Analisa algoritma yang digunakan pada pengklasteran data yang menjadi topik penelitian ini. Algoritma yang di Analisa pada penelitian ini adalah K-Means *Clustering*

5. Penomoran Klaster

Proses untuk mengelompokan data ke dalam beberapa *cluster* atau kelompok sehingga data dalam satu *cluster* memiliki tingkat kemiripan yang maksimum dan data antar *cluster* memiliki kemiripan yang minimum.

6. Uji Centroids

Menguji hasil data berdasarkan proses cluster untuk menemukan nilai yang terbaik.

7. Mengimplementasi menggunakan aplikasi RapidMiner

Secara umum pengelompokan (*clustering*) data dimulai dengan menentukan jumlah cluster K. kemudian menentukan K sebagai pusat cluster secara acak (random). Pada percobaan ini akan ditentukan data secara acak sebagai titik pusat awal untuk memperhitungkan jarak dari seluruh kelompok cluster yang akan dibentuk. Cluster dari setiap mata kuliah dengan 4 kriteria yang telah ditentukan yaitu cluster sangat baik, cluster baik, cluster cukup dan cluster kurang baik.

8. Mengevaluasi dan menganalisa hasil

Berdasarkan mekanisme pengujian yang telah dilakukan terhadap mahasiswa dengan menggunakan K-Means *Clustering* di dapatkan hasil analisisnya dan dirumuskan kesimpulan dari hasil penelitian tersebut

Setelah jumlah cluster diketahui, baru proses cluster dilakukan tanpa mengikuti proses hierarki. Metode ini biasa disebut dengan K-Means Clustering. langkah-langkah melakukan clustering dengan metode K-Means adalah sebagai berikut [16]:

- a. Pilih jumlah cluster k.
- b. Inisialisasi k pusat cluster ini bisa dilakukan dengan berbagai cara. Namun yang paling sering dilakukan adalah dengan cara random. Pusat-pusat cluster diberi nilai awal dengan angka-angka random,

- c. Alokasikan semua data/ objek ke cluster terdekat.

Kedekatan dua objek ditentukan berdasarkan jarak kedua objek tersebut. Demikian juga kedekatan suatu data ke cluster tertentu ditentukan jarak antara data dengan pusat cluster. Dalam tahap ini perlu dihitung jarak tiap data ke tiap pusat cluster. Jarak paling antara satu data dengan satu cluster tertentu akan menentukan suatu data masuk dalam cluster mana. Untuk menghitung jarak semua data ke setiap titik pusat cluster dapat menggunakan teori jarak Euclidean yang dirumuskan sebagai berikut:

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

Dimana:

- D (i,j) = Jarak data ke i ke pusat cluster j
- X_{ki} = Data ke i pada atribut data ke k
- X_{kj} = Titik pusat ke j pada atribut ke k

Persamaan menghitung jarak Euclidean antara data ke-i dan pusat cluster ke-j, berdasarkan perbedaan nilai atribut masing-masing. Jarak ini diperoleh dengan menjumlahkan kuadrat perbedaan untuk setiap atribut, kemudian mengambil akar kuadrat dari jumlah tersebut. Hasilnya menunjukkan seberapa dekat data ke-i dengan pusat cluster ke-j.

- d. Hitung kembali pusat cluster dengan keanggotaan cluster yang sekarang. Pusat cluster adalah rata-rata dari semua data/objek dalam cluster tertentu. Jika dikehendaki bisa juga menggunakan median dari cluster tersebut. Jadi rata-rata (mean) bukan satu-satunya ukuran yang bisa dipakai
- e. Tugaskan lagi setiap objek memakai pusat cluster yang baru. Jika pusat cluster tidak berubah lagi maka proses clustering selesai. Atau, kembali ke langkah nomor 3 sampai pusat cluster tidak berubah lagi

3. Hasil dan Pembahasan

Data yang digunakan untuk proses pengelompokan mata kuliah ini diperoleh untuk mengetahui tingkat pemahaman mahasiswa terhadap materi mata kuliah inti pada Prodi Informatika Fakultas Ilmu Komputer Universitas Dehasen Bengkulu. Data-data yang diambil berupa kuisioner yang disebarkan kepada mahasiswa yang telah mengambil mata kuliah tersebut, berikut daftar pernyataan kuisioner seperti Tabel 1

Tabel 1. Daftar Pertanyaan Kuisioner

No	Pernyataan
1	Seberapa baik pemahaman konsep dasar yang diajarkan dalam mata kuliah ini
2	Apakah penjelasan dosen selama kuliah membantu Anda dalam memahami materi
3	Apakah diskusi kelompok yang dilakukan dalam kelas baik dalam memahami materi
4	Dosen menjelaskan materi dengan jelas dan mudah dimengerti

- 5 Bagaimana Anda menilai metode pengajaran yang digunakan oleh dosen
- 6 Bagaimana penggunaan waktu di luar kelas untuk mempelajari materi mata kuliah
- 7 Apakah dosen memberikan instruksi dengan jelas untuk tugas-tugas yang diberikan
- 8 Apakah umpan balik yang diberikan oleh dosen pada tugas dan ujian membantu Anda untuk memahami kesalahan dan memperbaikinya?
- 9 Buku teks yang direkomendasikan membantu saya dalam memahami materi.
- 10 Seberapa sering Anda menggunakan buku teks yang disarankan oleh dosen
- 11 Materi tambahan (seperti artikel, video, dll.) yang disediakan oleh dosen bermanfaat
- 12 Seberapa mampu Anda menerapkan konsep yang dipelajari dalam mata kuliah ini
- 13 Bagaimana penilaian Anda dengan materi yang diajarkan dengan tujuan karir Anda
- 14 Bagaimana Anda menilai kepuasan keseluruhan terhadap mata kuliah ini

Berdasarkan kuisioner yang telah dilakukan, diperoleh tabel berikut. Tabel ini menggambarkan tingkat rekapitulasi hasil pemahaman mahasiswa terhadap pemahaman mahasiswa terhadap berbagai aspek materi mata kuliah. Hasil tersebut disajikan dalam bentuk yang telah diajarkan.

Tabel 2. Data Rekapitulasi Hasil Pemahaman Mahasiswa Terhadap Mata Kuliah

NO	MATA KULIAH	EVALUASI MATA KULIAH (BERDASARKAN KUISIONER)													
		1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	Organisasi & Arsitektur Komputer	4,00	1,11	2,89	1,95	1,32	1,68	2,84	2,74	2,84	1,11	3,79	3,42	2,42	3,05
2	Rekayasa Perangkat Lunak	4,00	3,00	2,00	1,00	2,00	3,00	1,08	3,88	2,08	1,04	4,00	2,00	2,00	3,00
3	Manajemen Proyek TI	4,00	3,00	2,70	1,80	2,05	2,60	2,40	2,70	2,10	2,55	3,50	2,40	2,40	1,45
4	Analisa & Perancangan Sistem	4,00	3,70	4,00	2,70	2,00	2,50	2,40	2,60	2,50	2,35	3,00	2,70	2,90	3,00
5	Analisis Algorithma	4,00	3,91	3,40	3,60	3,20	3,00	3,20	3,60	3,00	3,10	3,00	3,20	3,82	3,55
6	Mobile Computing	4,00	2,84	2,53	2,42	2,58	2,63	2,68	2,63	2,63	2,32	2,58	2,74	2,79	2,74
7	Sistem Pakar	4,00	2,13	2,30	2,04	1,87	2,39	2,43	2,13	2,35	2,39	3,65	3,52	2,74	3,26
8	Data Warehouse & Data Mining	4,00	3,00	3,70	3,96	3,20	3,00	3,00	3,23	3,00	3,08	3,40	3,92	3,92	3,08
9	Sistem Cerdas	4,00	3,10	3,90	2,60	2,80	3,70	2,90	3,60	3,40	3,30	3,60	2,40	3,40	3,90
10	Komunikasi Data	4,00	2,62	3,61	2,57	2,43	2,76	2,14	2,29	2,19	2,62	2,76	2,95	2,86	2,90
11	Teori Bahasa Automata	4,00	2,68	2,86	2,91	2,91	2,95	2,86	2,55	2,68	2,68	2,95	2,68	2,73	3,05
12	Sistem Operasi	4,00	1,70	2,96	2,00	2,17	2,39	2,13	2,52	2,35	2,17	3,83	3,39	2,57	3,48
13	Basis Data	4,00	3,00	3,14	3,66	3,48	3,86	3,72	3,86	3,90	3,38	3,86	2,93	3,90	3,97
14	Riset Operasional	4,00	3,05	3,00	3,05	2,80	3,00	3,05	2,55	2,80	2,85	3,05	2,95	2,95	2,70
15	Pengembangan Berbasis Platform	4,00	1,32	2,89	1,95	1,47	1,79	2,84	2,74	2,84	1,26	3,79	3,42	2,42	3,05
16	Jaringan Komputer	4,00	3,93	3,93	2,67	3,41	3,52	4,00	3,96	2,85	3,59	2,30	3,00	2,19	3,00
17	Logika Fuzzy	4,00	3,00	3,80	2,40	2,90	3,20	3,80	4,00	2,30	3,00	4,00	2,00	2,00	2,00

Data pada Tabel 2 digunakan untuk mengelompokkan mata kuliah ke dalam kategori "Sangat Baik (C1)", "Baik (C2)", "Cukup Baik (C3)", dan "Kurang Baik (C4)". Pengelompokan ini didasarkan pada poin-poin pernyataan kuisioner yang tercantum pada Tabel 1. Langkah pertama adalah melakukan pemrosesan data,

termasuk normalisasi, agar atribut berada pada skala yang seragam. Selanjutnya, dipilih metode klasterisasi yang sesuai dan diterapkan untuk membentuk klaster mata kuliah berdasarkan atribut tersebut. Akhirnya, hasil klasterisasi dievaluasi untuk memastikan kesesuaiannya dengan kategori yang diinginkan.

Tabel 3. Titik Pusat Awal Dari Setiap Cluster

NO	MATA KULIAH	EVALUASI MATA KULIAH (BERDASARKAN KUISIONER)													
		1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	Organisasi & Arsitektur Komputer	4,00	1,11	2,89	1,95	1,32	1,68	2,84	2,74	2,84	1,11	3,79	3,42	2,42	3,05
2	Teori Bahasa Automata	4,00	2,68	2,86	2,91	2,91	2,95	2,86	2,55	2,68	2,68	2,95	2,68	2,73	3,05
3	Sistem Cerdas	4,00	3,10	3,90	2,60	2,80	3,70	2,90	3,60	3,40	3,30	3,60	2,40	3,40	3,90
4	Logika Fuzzy	4,00	3,00	3,80	2,40	2,90	3,20	3,80	4,00	2,30	3,00	4,00	2,00	2,00	2,00

Tabel 3 menunjukkan titik pusat awal dari setiap cluster berdasarkan evaluasi mata kuliah. Setiap baris dalam tabel ini mewakili nilai evaluasi untuk mata kuliah tertentu pada berbagai indikator (1 hingga 14). Titik pusat ini digunakan sebagai centroid awal dalam

proses clustering K-Means untuk menentukan posisi awal cluster dan membagi data ke dalam kelompok yang sesuai.

1. Perhitungan Jarak Pusat Cluster

Perhitungan jarak antara data dengan pusat kluster awal dilakukan menggunakan persamaan Euclidean Distance. Persamaan ini memungkinkan pengukuran jarak secara geometris antara dua titik dalam ruang multidimensi. Berikut adalah hasil perhitungan jarak tersebut tersaji pada Tabel 4.

Tabel 4. Hasil Perhitungan Jarak Iterasi 1

DATA	C ₁	C ₂	C ₃	C ₄
1	0,000	3,377	4,504	4,373
2	3,833	3,925	4,543	4,284
3	3,393	2,421	3,767	2,724
4	3,589	1,950	2,633	3,099
5	6,408	4,179	3,065	4,350
6	3,089	0,998	2,930	3,083
7	2,190	2,136	3,579	3,895
8	5,643	3,738	2,919	3,950
9	4,504	2,368	0,000	2,912
10	3,171	1,203	3,089	3,453
11	3,377	0,000	2,368	2,782
12	1,926	2,248	3,370	3,727
13	5,250	3,072	1,949	3,728
14	3,633	0,724	2,332	2,693
15	0,318	3,101	4,243	4,130
16	5,361	2,882	2,674	2,634
17	4,373	2,782	2,912	0,000

Selanjutnya, bandingkan jarak setiap mata kuliah ke centroid dan pilih jarak yang paling dekat. Jarak ini

Tabel 6. Pusat Cluster Baru

Centroid Baru	Data 1	Data 2	Data 3	Data 4	Data 5	Data 6	Data 7	Data 8	Data 9	Data 10	Data 11	Data 12	Data 13	Data 14
C ₁	4,00	1,78	2,69	1,73	1,74	2,22	2,22	2,97	2,53	1,40	3,85	3,06	2,35	3,15
C ₂	4,00	2,86	2,85	2,50	2,38	2,69	2,57	2,49	2,46	2,54	3,07	2,85	2,77	2,73
C ₃	4,00	3,25	3,76	3,56	3,57	3,88	3,66	3,87	3,82	3,44	3,87	3,31	3,76	3,63
C ₄	4,00	3,47	3,87	2,54	3,16	3,36	3,90	3,98	2,58	3,30	3,15	2,50	2,10	2,50

Langkah selanjutnya adalah menghitung nilai Euclidean Distance dari semua data ke titik pusat kluster baru (C₁, C₂, C₃, C₄), seperti yang telah dilakukan pada langkah sebelumnya. Perhitungan ini bertujuan untuk menentukan seberapa dekat setiap data dengan pusat kluster yang baru, sehingga pengelompokan dapat diperbarui dan disempurnakan. Hasil perhitungan ini akan digunakan untuk menilai apakah data perlu dikelompokkan ulang ke kluster yang berbeda berdasarkan jarak terpendek ke pusat kluster yang baru.

Tabel 7. Hasil Pengelompokan Cluster 2

Kelompok Cluster	Anggota Kelompok	Jumlah
C ₁	1, 2, 7, 12, 15	5
C ₂	3, 4, 6, 10, 11, 14	6
C ₃	5, 8, 9, 13	4
C ₄	16, 17	2

menunjukkan bahwa mata kuliah tersebut termasuk dalam cluster yang memiliki pusat centroid terdekat. Dengan cara ini, data dikelompokkan sesuai dengan cluster yang tepat, sebagaimana ditunjukkan dalam Tabel 5.

Tabel 5. Hasil Pengelompokan Cluster 1

Kelompok Cluster	Anggota Kelompok	Jumlah
C ₁	1, 2, 12, 15	4
C ₂	3, 4, 6, 7, 10, 11, 14	7
C ₃	5, 8, 9, 13	4
C ₄	16, 17	2

Hasil clustering menunjukkan bahwa Cluster C₂ adalah yang terbesar dengan 7 anggota, menandakan kepadatan data tertinggi. Cluster C₁ dan Cluster C₃ masing-masing memiliki 4 anggota, menunjukkan ukuran cluster yang moderat. Cluster C₄ adalah yang terkecil dengan hanya 2 anggota, mencerminkan ukuran cluster yang paling kecil.

2. Penentuan pusat cluster baru

Setelah semua nilai centroid baru dihitung, hasilnya dapat dilihat pada tabel berikut. Tabel ini menampilkan posisi terbaru dari pusat-pusat cluster setelah iterasi perhitungan dilakukan. Perubahan posisi centroid menunjukkan hasil akhir dari proses clustering.

Hasil clustering menunjukkan bahwa Cluster C₂ adalah yang terbesar dengan 6 anggota, menunjukkan kepadatan data yang tinggi. Cluster C₁ mengikuti dengan 5 anggota, sementara Cluster C₃ memiliki 4 anggota, menunjukkan ukuran yang lebih moderat. Cluster C₄ adalah yang terkecil dengan hanya 2 anggota, menunjukkan bahwa cluster ini memiliki jumlah anggota yang paling sedikit.

Tabel 8. Hasil Pengelompokan Cluster 3

Kelompok Cluster	Anggota Kelompok	Jumlah
C ₁	1, 2, 7, 12, 15	5
C ₂	3, 4, 6, 10, 11, 14	6
C ₃	5, 8, 9, 13	4
C ₄	16, 17	2

Pada pengujian iterasi ke-2 dan ke-3, tidak terjadi perubahan nilai centroid, yang menunjukkan bahwa hasilnya sama dengan centroid pada iterasi sebelumnya. Oleh karena itu, proses iterasi dinyatakan selesai dengan diperolehnya 4 klaster dalam 3 iterasi. Berikut ini adalah tabel hasil akhir analisis perhitungan pemahaman mahasiswa terhadap mata kuliah menggunakan algoritma K-Means.

Tabel 9. Hasil Analisa K-Means Clustering

Cluster	Hasil Akhir		
	Centroid Akhir	Jumlah Anggota	Nama Mata Kuliah
1	1, 2, 7, 12, 15	5	Organisasi & Arsitektur Komputer Rekayasa Perangkat Lunak Sistem Pakar Sistem Operasi Pengembangan Berbasis Platform Manajemen Proyek TI Analisa & Perancangan Sistem
2	3, 4, 6, 10, 11, 14	6	Mobile Computing Komunikasi Data Teori Bahasa Automata Riset Operasional Analisis Algoritma
3	5, 8, 9, 13	4	Data Warehouse & Data Mining Sistem Cerdas Basis Data
4	16, 17	2	Jaringan Komputer Logika Fuzzy

Dari tabel di atas, maka didapat bahwa pemahaman materi mata kuliah “Sangat Baik” sebanyak 29% terdiri dari 5 mata kuliah yaitu Organisasi & Arsitektur Komputer, Rekayasa Perangkat Lunak, Sistem Pakar, Sistem Operasi dan Pengembangan Berbasis Platform. Pemahaman materi mata kuliah “Baik” sebanyak 35% terdiri dari 6 mata kuliah yaitu Manajemen Proyek TI, Analisa & Perancangan Sistem, Mobile Computing, Komunikasi Data, Teori Bahasa Automata, Riset Operasional, pemahaman materi mata kuliah “Cukup Baik” sebanyak 24% terdiri dari 4 mata kuliah, yaitu Analisis Algoritma, Data Warehouse & Data Mining, Sistem Cerdas dan Basis Data. Sedangkan pemahaman materi mata kuliah “Kurang Baik” sebanyak 12% terdiri dari 2 mata kuliah, yaitu Jaringan Komputer dan Logika Fuzzy.

4. Kesimpulan

Kesimpulan dari penelitian yang dilakukan adalah Metode K-Means dengan efektif untuk mengelompokkan data tingkat pemahaman mahasiswa terhadap mata kuliah inti di Program Studi Informatika, Universitas Dehasen Bengkulu, dalam penelitian ini menunjukkan bahwa pendekatan ini dapat mendukung pengambilan keputusan yang lebih efektif dalam perencanaan kegiatan belajar mengajar

serta peningkatan sistem penilaian di lingkungan akademik.

Daftar Rujukan

- [1] Mustari, K. A., Assiroj, P., Hartati, B., & Samuel, F. (2024). *Implementasi Data Mining Pada Instansi Pemerintahan (Systematic Literature Review)*. 8(3), 3137–3142
- [2] Lestari, P., Suarna, N., & Prihartono, W. (2024). *Implementasi Data Mining Clustering Dalam Mengelompokkan Penduduk Penyandang Disabilitas Menggunakan Algoritma K-Means*. 8(1), 958–966.
- [3] Setiawan, K., & Ramdhani, F. (2024). *Application of Data Mining for Customer Prediction at PT . XYZ Using Linear Regression Algorithm Penerapan Data Mining untuk Prediksi Pelanggan di PT . XYZ Menggunakan Algoritma Linear Regression*. 4(April), 490–497.
- [4] Sasmita, G. W., & Rahmadayanti, F. (2024). *Implementasi Data Mining Menggunakan Algoritma Apriori dalam Menentukan Persediaan Obat*. 8, 13893–13897.
- [5] Liantoni, F., Sukmaningrum, A. H., Qurin, M. T., & Putri, N. V. (2024). *Analisis Klasifikasi 7 Jenis Dry Beans dengan Penerapan Metode Naïve Bayes Classification Analysis of 7 Types of Dry Beans Using the Naïve Bayes Method*. 12(2), 248–252. <https://doi.org/10.26418/justin.v12i2.73034>
- [6] Pujiono, S., Astuti, R., & Basysyar, F. M. (2024). *Implementasi Data Mining Untuk Menentukan Pola Penjualan Produk Menggunakan Algoritma K-Means Clustering*. 8(1).
- [7] Lestari, P., Suarna, N., & Prihartono, W. (2024). *Implementasi Data Mining Clustering Dalam Mengelompokkan Penduduk Penyandang Disabilitas Menggunakan Algoritma K-Means*. 8(1), 958–966.
- [8] Faran, J., & Triayudi, A. (2024). *Penerapan Algoritma K-Means Data Mining untuk Clustering Kinerja Karyawan Koperasi*. 4(4), 2096–2108. <https://doi.org/10.30865/klik.v4i4.1728>
- [9] Afiasari, N., Suarna, N., & Rahaningsih, N. (2023). *Implementasi Data Mining Transaksi Penjualan Menggunakan Algoritma Clustering dengan Metode K-Means E-commerce K-Means melakukan analisis penerapan Data Mining dalam mengelompokkan jumlah*. 9, 100–110.
- [10] Fadila, R., & Rinaldi, A. R. (2024). *Penerapan Data Mining Dalam Mengelompokkan Jumlah Umkm Berdasarkan Kabupaten Kota Menggunakan K-Means Clustering*. 8(2), 1446–1450.
- [11] Ferdy Pangestu, F. P., Nur Yasin, N. Y., Ronald Chistover Hasugian, R. C., & Yunita, Y. (2023). *Penerapan Algoritma K-Means Untuk Mengklasifikasi Data Obat*. *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, 12(1), 53–62. <https://doi.org/10.32736/sisfokom.v12i1.1461>
- [12] Afifah, I. A. N., & Nurdiyanto, H. (2023). *Data mining clustering dalam pengelompokan buku perpustakaan menggunakan algoritma k-means*. 8(3), 802–814.
- [13] Nahjan, M. R., Heryana, N., Voutama, A., Komputer, F. I., Karawang, U. S., & Miner, R. (2023). *Implementasi Rapidminer Dengan Metode Clustering K-Means Untuk Analisa Penjualan Pada Toko Oj Cell*. 7(1), 101–104.
- [14] Yoliadi, D. N. (2023). *Data mining Dalam Analisis Tingkat Penjualan Barang Elektronik*. 3.
- [15] Pujiono, S., Astuti, R., & Basysyar, F. M. (2024). *Implementasi Data Mining Untuk Menentukan Pola Penjualan Produk Menggunakan Algoritma K-Means Clustering*. 8(1).
- [16] Kurniawan, R. A., Hasibuan, M. S., Piramida, P., & Ramadhan, R. S. (2022). *Penerapan Algoritma K-Means Untuk Clustering Tempat Makan Di Batubara*. *Journal of Computer Science and Informatics Engineering (CoSIE)*, 01(1), 10–18. <https://doi.org/10.55537/cosie.v1i1.27>