Jurnal Elektronik Ilmu Komputer Udayana
Volume 13, No 4. May 2025

p-ISSN: 2301-5373
e-ISSN: 2654-5101

# Popularitas Kursus Online: Analisis Mendalam dengan K-Means di Platform Udemy

Rahmat Hidayat[a1], Ibnu Daqiqil Id[b2*], Rahmah Muthmainah[b3], Muhammad Alfhat Ramadhan[b4], Icha Dwi Yanti[b5], Kornelius Jonathan Manik[b6], Rendi Wahyudi[b7]

[a]Program Studi Manajemen Informatika, Universitas Riau
[b]Program Studi Sistem Informasi, Universitas Riau
Kampus Bina Widya Km 12,5 Simpang Baru Pekanbaru, Indonesia

[1]rahmat.hidayat@lecturer.unri.ac.id
[2]ibnu.daqiqil@lecturer.unri.ac.id
[3]rahmah.muthmainah3281@student.unri.ac.id
[4]muhammad.alfhat6561@student.unri.ac.id
[5]icha.dwi5311@student.unri.ac.id
[6]kornelius.jonathan3893@student.unri.ac.id
[7]rendi.wahyudi0482@student.unri.ac.id

***Abstract***

*Online courses have become increasingly popular in the digital age, offering flexibility, access to materials and enhanced collaboration. However, online course providers face challenges in understanding user preferences to develop effective marketing strategies. This study uses the K-Means clustering algorithm to analyze the popularity of online courses on the Udemy platform. The analysis is based on a Kaggle dataset containing course attributes such as enrollment numbers, reviews, and others. The elbow method was used to determine the optimal number of clusters, with results showing that K=4 is optimal. The clustering quality was assessed using the Davies-Bouldin Index (DBI) and Silhouette Score, yielding a DBI value of 0.65171 and a Silhouette Score of 0.6269, indicating good cluster separation. The study reveals four distinct clusters of courses, each with varying popularity and characteristics. This study can help course providers create more targeted marketing strategies, improving user engagement and satisfaction based on the specific traits of each course group.*

***Keywords:*** *data mining, clustering, k-means, algorithm, online courses, popularity.*

**Abstrak**

Kursus online telah menjadi semakin popular di era digital, menawarkan fleksibilitas, akses ke materi dan kolaborasi yang lebih baik. Namun, penyedia kursus online menghadapi tantangan dalam memahami preferensi pengguna untuk mengembangkan strategi pemasaran yang efektif. Studi ini menggunakan algoritma klastering K-means untuk menganalisis popularitas kursus online di platform Udemy. Analisis ini didasarkan pada dataset Kaggle yang berisi atribut kursus seperti jumlah pendaftara, ulasan, dan lainnya. Metode *Elbow* digunakan untuk menentukan jumlah kluster yang optimal, dengan hasil menunjukkan bahwa K=4 adalah yang optimal. Kualitas pengelompokan dinilai menggunakan *Davies-Bouldin Index* (DBI) dan Skor *Silhouette*, menghasilkan nilai DBI sebesar 0,65171 dan Skor *Silhouette* sebesar 0,6269, yang menunjukkan pemisahan kluster yang baik. Studi ini mengungkapkan empat kluster kursus yang berbeda, masing-masing dengan popularitas dan karakteristik yang bervariasi. Penelitian ini dapat membantu penyedia kursus menciptakan strategi pemasaran yang lebih terarah, meningkatkan keterlibatan dan kepuasan pengguna berdasarkan karakteristik spesifik dari setiap kelompok kursus.

**Kata Kunci:** penambangan data, klastering, k-means, algoritma, kursus online, popularitas.

## 1. Pendahuluan

Dalam era digital saat ini, kegiatan belajar online menjadi suatu hal yang lumrah. Bahkan, di beberapa universitas maupun sekolah tinggi di negara-negara maju seperti Eropa dan Amerika sudah menerapkan sistem ini. Sistem yang sudah diterapkan oleh beberapa negara ialah pengajar dan peserta berinteraksi lewat laptop, melakukan *video call* atau *conference call.* Perlu diketahui, pada dasarnya belajar online dapat melampaui belajar konvensional [1]. Belajar online menuntut siswa untuk memiliki metodologi berpikir yang cukup, khususnya dalam mencerna instruksi yang diberikan secara online [2].

Kelas konvensional sering dianggap kurang efektif karena akses terbatas bagi pelajar di lokasi yang jauh, metode pengajaran yang kaku, dan kurangnya personalisasi [3]. Waktu kelas yang terbatas juga menyulitkan pendalaman materi, sementara interaksi siswa yang minim mengurangi peluang belajar dari satu sama lain. Untuk mengatasi tantangan ini, banyak institusi beralih ke kursus *online* [4]. Kursus *online* menawarkan fleksibilitas waktu dan tempat, akses ke sumber daya pendidikan yang lebih beragam, serta meningkatkan interaksi dan kolaborasi melalui platform digital, menciptakan pengalaman belajar yang lebih efektif dan menarik [5].

Kursus *online* atau "*e-learning*" adalah proses pembelajaran yang memfasilitasi, menyampaikan dan memungkinkan pembelajaran jarak jauh melalui penggunaan teknologi internet [6]. Jenis pembelajaran ini menjadi salah satu bentuk pembelajaran yang paling populer dan berpengaruh, terutama dalam pendidikan pasca-sekolah menengah selama dekade terakhir [7]. Selain fleksibilitas dan kemudahan akses, kursus *online* juga membuka peluang bagi siapa pun untuk mendapatkan sertifikasi resmi dan meningkatkan prospek kerja tanpa batasan usia, menjadikannya alternatif yang ideal bagi banyak pelajar di era digital [8]. E-learning mendorong pembelajaran mandiri dan pengembangan keterampilan belajar. Siswa dapat mengelola waktu mereka sendiri dan mengambil inisiatif dalam pembelajaran mereka [9]. Ini mengajarkan keterampilan penting dalam mengatur diri dan memotivasi diri sendiri untuk mengikuti pendidikan.

Kemajuan kursus *online* yang pesat menjadikannya tidak hanya sebagai alat pembelajaran, tetapi juga sebagai peluang bisnis bagi penyedia layanan pendidikan [10]. Agar kursus online dapat menjangkau lebih banyak pengguna dan menciptakan nilai tambah, diperlukan strategi pemasaran yang efektif. Strategi pemasaran yang disusun secara rinci dan sistematis akan membantu penyedia kursus merancang kegiatan pemasaran yang terarah, sehingga tujuan perusahaan, termasuk meningkatkan keuntungan dan kepuasan pengguna dapat tercapai [11]. Salah satu cara untuk mendukung strategi ini adalah dengan menerapkan algoritma K-Means untuk mengelompokkan jenis kursus online berdasarkan tingkat popularitasnya.

*Clustering* merupakan metode berbasis jarak yang membagi data ke dalam sejumlah klaster berdasarkan atribut numerik [12]. Dalam *clustering data mining*, algoritma K-Means membentuk kelompok baru berdasarkan pembentukan *cluster,* Ini dilakukan berdasarkan kesamaan atau kemiripan karakteristik data.untuk mengidentifikasi karakteristik data berdasarkan nilai jarak terdekat [13]. Proses ini melibatkan pemilihan sejumlah objek sebagai centroid, di mana setiap kursus dikelompokkan berdasarkan jarak *Euclidean* ke *centroid* tersebut [14]. Dengan menggunakan pendekatan ini, kita dapat mengidentifikasi kelompok kursus yang paling diminati dan merumuskan strategi pemasaran yang lebih efektif.

## 2. Metode Penelitian

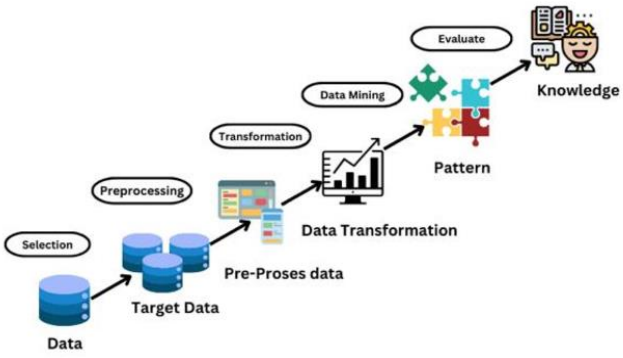
### 2.1 Pengertian Metode Algoritma K-Means *Clustering*

Means *Clustering* suatu metode yang digunakan untuk mencari dan mengelompokkan data yang mirip-mirip agar lebih mudah membedakan dengan data yang lainnya. *Clustering* sendiri merupakan suatu metode data yang tidak memiliki acuan (*unsupervised*) dimana mengelompokkan data yang mirip tanpa perlu contoh atau aturan khusus [15].

p-ISSN: 2301-5373
e-ISSN: 2654-5101

### 2.2 Data Mining

Data mining dikenal dengan istilah *Knowledge Discovery in Database* (KDD) merupakan kegiatan mengumpulkan, mengolah, dan penggunaan data dalam jumlah besar dengan tujuan menemukan pola atau aturan yang dapat disimpan dalam data base atau media penyimpanan lainnya yang dapat membantu dalam membuat keputusan yang lebih baik [16].




**Gambar 1.** Proses Knowledge Discovery in Database (KDD)

1. *Selection*
   *Selection* digunakan untuk menentukan variabel yang diminati sehingga tidak ada persamaan dan tidak terjadi pengulangan yang tidak diperlukan untuk pemrosesan data mining.
2. *Preprocessing*
   Menghilangkan data yang tidak diperlukan seperti menangani *missing value, noise* data serta menangani data-data yang tidak konsisten dan relevan [17].
3. *Transformation*
   Merubah data ke dalam format yang sesuai dalam pengolahan data mining beberapa metode dalam data mining memerlukan format tertentu sebelum dapat diproses pada penambangan data.
4. *Data mining*
   *Data Mining* adalah proses utama pada metode yang diterapkan untuk mendapatkan pengetahuan baru dari data yang telah diproses [18]. Penelitian ini menerapkan teknik *Clustering* yaitu metode K-Means *Clustering*.
5. *Evaluation/Interpretation*
   Pada tahap ini, pola-pola menarik atau model prediksi diidentifikasi dan dimasukkan ke dalam knowledge base. Pola-pola tersebut kemudian dievaluasi untuk memastikan bahwa hasilnya sesuai dengan tujuan yang diinginkan.
   a. Nilai DBI
      DBI menunjukkan seberapa jauh klaster terpisah satu sama lain dan seberapa kompak. Nilai DBI yang lebih kecil menunjukkan hasil *cluster*ing yang lebih baik [19].

$$DBI = \frac{1}{k}\sum_{i\ 1}^{k} \max_{j/i} \left( \frac{S_i + S_j}{d(c_i, c_j)} \right) \quad (1)$$

| | | |
|---|---|---|
| *k* | : | jumlah *cluster* |
| *S* | : | jarak intra-*cluster* untuk *cluster i* |
| *d(ci,cj)* | : | jarak antar centroid dari *cluster i* dan *cluster j* |

   b. *Silhouette Score*
      Nilai *Silhouette* digunakan untuk mengevaluasi seberapa baik objek dalam suatu *cluster* terpisah dari objek lain [20]. Nilai yang mendekati 1 menunjukkan bahwa titik data berada

di dalam klaster yang benar, nilai mendekati 0 menunjukkan data berada di antara dua klister, sedangkan nilai negatif menunjukkan bahwa titik data lebih dekat ke klaster yang salah [21].

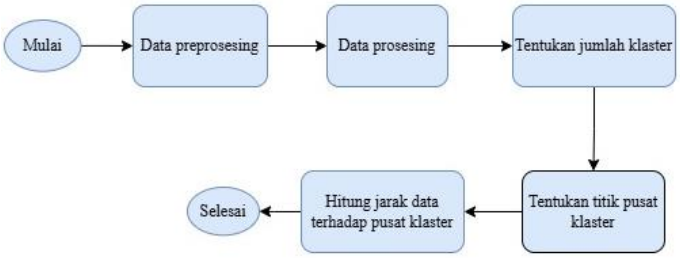
$$s = \frac{b-a}{\max(a,b)} \quad (2)$$

s : *Silhouette Score* untuk titik data
a : Rata-rata jarak antara titik data dan semua titik data lain dalam klaster yang sama
b : Rata-rata jarak antara titik data dan semua titik data dalam klaster terdekat yang berbeda

6. *Knowledge*
Pola-pola yang telah dihasilkan akan ditampilkan kepada pengguna. Pada tahap ini pengetahuan baru yang dihasilkan bisa dimengerti semua orang yang akan dijadikan standar pengambilan kebutuhan [22].

### 2.3 Tahapan perhitungan Algoritma *K-means Clustering Analysis*

Proses perhitungan untuk analisis Algoritma K-Means *Clustering* diuraikan sebagai berikut:




**Gambar 2.** Tahapan perhitungan Algoritma *K-means Clustering*

1. Menentukan jumlah *cluster*
Jumlah *cluster* ditentukan dengan melihat pola dari *elbow method. Elbow method* menunjukkan nilai *within-cluster-sum-of-square* (WCSS) pada sumbu y yang sesuai dengan nilai K yang berbeda.
2. Menentukan titik pusat (*centroid) cluster*
Menentukan titik pusat adalah langkah penting dalam algoritma K-Means. Proses ini dimulai dengan inisialisasi sejumlah centroid secara acak. Setiap titik data kemudian ditugaskan ke klaster dengan centroid terdekat berdasarkan jarak Euclidean. Setelah semua titik data ditugaskan, posisi centroid diperbarui dengan menghitung rata-rata posisi semua titik data dalam klaster tersebut.
3. Menghitung jarak antar titik data objek dan titik pusat *(centroid)*
Rumus *Euclidean Distance* digunakan untuk perhitungan jarak:

$$\mathrm{dist}(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad (3)$$

d (x, y) : Jarak data antara dua titik x dan y
x : Titik pertama dalam ruang dimensi-n (data objek)
y : Titik kedua dalam ruang dimensi-n (data centroid)
j : Jumlah atribut data

### 2.4 Data

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Data yang digunakan dalam penelitian ini berasal dari dataset kursus online platform Udemy, yang berisi informasi mendetail mengenai berbagai kursus yang ditawarkan. Dataset mencakup atribut-atribut yang berkaitan dengan identifikasi kursus.

**Tabel 1.** Deskripsi Data Kursus Online Udemy

| Nama *Field* | Tipe Data | Keterangan |
|---|---|---|
| *course_id* | *integer* | ID unik yang digunakan untuk mengidentifikasi kursus |
| *course_title* | *object* | Judul kursus yang menjelaskan topik atau isi kursus. |
| *url* | *object* | Tautan ke halaman web tempat kursus tersebut dapat diakses. |
| *is_paid* | *boolean* | Menunjukkan apakah kursus berbayar atau gratis. |
| *price* | *integer* | Harga kursus jika berbayar |
| *num_subscribers* | *integer* | Jumlah orang yang telah mendaftar untuk kursus tersebut |
| *num_reviews* | *integer* | Jumlah ulasan yang diberikan oleh peserta kursus |
| *num_lectures* | *integer* | Jumlah kuliah atau modul yang termasuk dalam kursus |
| *level* | *object* | Tingkat kesulitan kursus |
| *content_duration* | *float* | Durasi total konten kursus dalam satuan jam |
| *published_timestamp* | *object* | Tanggal dan waktu ketika kursus diterbitkan |
| *subject* | *object* | Subjek atau kategori utama dari kursus |

### 2.5 *Data Pre-Processing*

Menghapus data *subscribers* dengan nilai 0 dan membersihkan atribut seperti *course_title, url,* dan *published_timestamp* untuk meningkatkan relevansi dan konsistensi dataset. Hal ini dilakukan untuk membuat data yang dianalisis lebih bermakna dan memasukkan hanya informasi yang relevan yang berkontribusi langsung terhadap tujuan analisis. Setelah melalui proses *pre-processing*, data yang sebelumnya berjumlah 3678 mengerucut menjadi 3608.

### 2.6 *Data Transformation*

Pada tahap ini, data mentah diubah menjadi bentuk yang lebih terstruktur dan siap untuk analisis. Transformasi data dilakukan untuk menyederhanakan interpretasi atribut tertentu dengan mengubah data kualitatif menjadi data kuantitatif. Beberapa transformasi dilakukan sebagai berikut:

**Tabel 2.** Data *Transformation is_paid*

| Sebelum | Sesudah |
|---|---|
| *TRUE* | 1 |
| *FALSE* | 0 |

**Tabel 3.** Data *Transformation level*

| Sebelum | Sesudah |
|---|---|
| *All Level* | 1 |
| *Beginner Level* | 2 |
| *Expert Level* | 3 |
| *Intermediate Level* | 4 |

**Tabel 4.** Data *Transformation subject*

| Sebelum | Sesudah |
|---|---|
| *Business Finance* | 1 |
| *Graphic Design* | 2 |
| *Musical Instruments* | 3 |
| *Web Development* | 4 |

Setelah tahap pre-processing dan transformation selesai, data kemudian digabungkan ke dalam dataframe yang tersedia. Proses ini dilakukan dengan tujuan memastikan bahwa semua data berada dalam satu struktur data yang konsisten, yang memungkinkan analisis lebih lanjut dan memungkinkan pengolahan data yang lebih efisien pada tahap berikutnya.

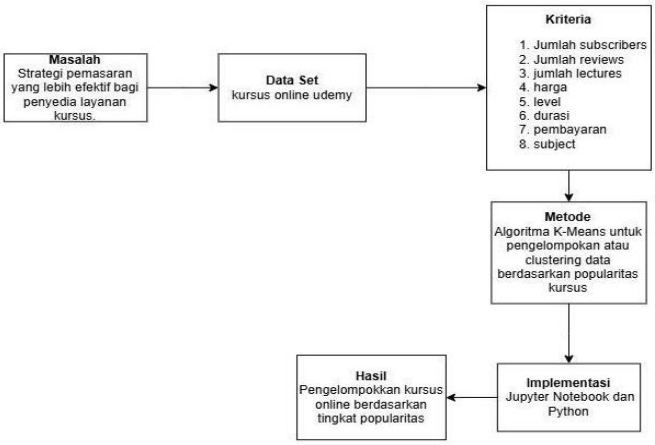
**Tabel 5.** Data yang Sudah Melalui Tahap *Pre-Processing* dan *Transformation*

| No | is_ paid | price | num_subscribers | num_ reviews | num_ lectures | level | content_ duration | sub ject |
|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 200 | 2147 | 23 | 51 | 1 | 1.5 | 1 |
| 2 | 2 | 75 | 2792 | 923 | 274 | 1 | 39 | 1 |
| 3 | 2 | 45 | 2174 | 74 | 51 | 4 | 2.5 | 1 |
| 4 | 2 | 95 | 2451 | 11 | 36 | 1 | 3 | 1 |
| 5 | 2 | 200 | 1276 | 45 | 26 | 4 | 2 | 1 |
| 6 | 2 | 150 | 9221 | 138 | 25 | 1 | 3 | 1 |
| 7 | 2 | 65 | 1540 | 178 | 26 | 2 | 1 | 1 |
| 8 | 2 | 95 | 2917 | 148 | 23 | 1 | 2.5 | 1 |
| 9 | 2 | 195 | 5172 | 34 | 38 | 3 | 2.5 | 1 |
| 10 | 2 | 200 | 827 | 14 | 15 | 1 | 1 | 1 |
| … | … | … | … | … | … | … | … | … |
| 3608 | 2 | 45 | 901 | 36 | 20 | 2 | 2 | 4 |

4. Kerangka Berpikir

   Kerangka berpikir adalah rencana dan diagram yang digunakan untuk menyatukan konsep yang akan dibahas selama penelitian. Ini dibuat untuk membantu peneliti membuat hipotesa yang akan diteliti [23]. Gambar 3 menunjukkan kerangka berfikir yang dirumuskan seperti berikut ini.

p-ISSN: 2301-5373
e-ISSN: 2654-5101



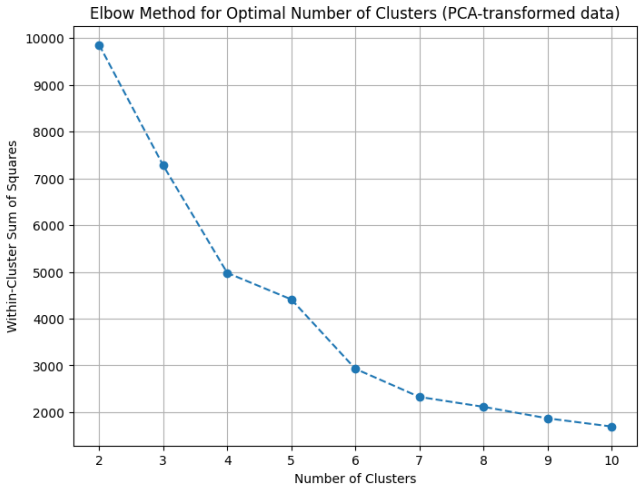

**Gambar 3**. Kerangka Berpikir

## 3. Hasil dan Pembahasan

Pada penelitian ini, tahapan pengujian dilakukan untuk mengevaluasi efektivitas model K-Means dalam mengelompokkan kursus online dari platform Udemy berdasarkan atribut-atribut penting.

### a. Menentukan jumlah *cluster*

Untuk menentukan jumlah klaster yang optimal, peneliti menganalisis nilai WCSS (*Within-Cluster Sum of Squares*) dalam model K-Means. Ditemukan bahwa nilai WCSS mengalami penurunan signifikan saat jumlah klaster bertambah dari 2 menjadi 4, yang menunjukkan peningkatan kekompakan klaster. Namun, setelah mencapai 4 klaster, laju penurunan WCSS mulai melambat, mendanakan bahwa penambahan klaster selanjutnya tidak memberikan peningkatan yang berarti. Dengan demikian, dapat disimpulkan bahwa jumlah klaster yang optimal adalah 4.




**Gambar 4**. *Elbow Method*

### b. Menentukan titik pusat (*centroid) cluster*

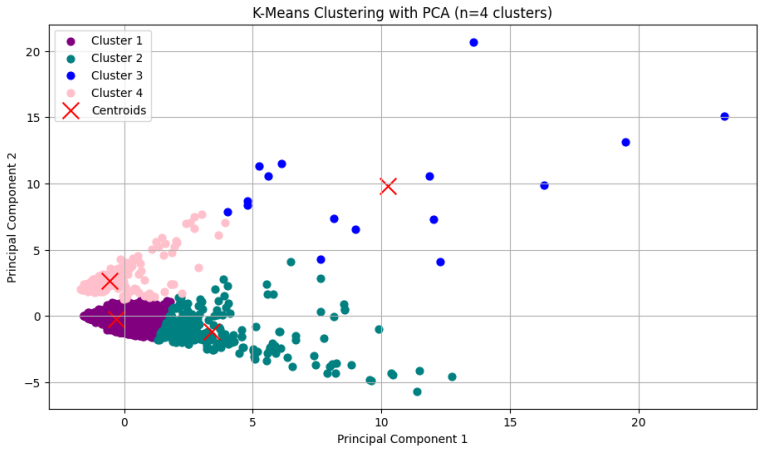
Untuk menentukan centroid, langkah yang dilakukan adalah menghitung rata-rata dari setiap atribut pada anggota *cluster* tersebut. Centroid ini mewakilkan posisi tengah *cluster* berdasarkan nilai atributnya.

**Tabel 6.** Rata-rata atribut setiap *cluster*

| Rata-rata | C1 | C2 | C3 | C4 |
|---|---|---|---|---|
| *num_subscribers* | 1076.48 | 8653.40 | 88291.88 | 15179.95 |
| *num_reviews* | -41.53 | 983.72 | 7831.58 | 986.51 |
| *num_lectures* | 35.21 | 146.04 | 209.36 | -5.36 |

| | | | | |
|---|---|---|---|---|
| *price* | 65.82 | 158.18 | 49.34 | -4.20 |
| *level* | 1.74 | 1.31 | 1.04 | 1.89 |
| *duration* | 3.50 | 16.51 | 24.25 | -1.20 |
| *subject* | 2.41 | 3.15 | 7.10 | 2.93 |
| *Is_paid* | 1.94 | 2.15 | 0.62 | 1.50 |

Atribut *kmeans_pca.cluster_centers* digunakan untuk mengumpulkan klaster, yang menunjukkan posisi data rata-rata dalam setiap klaster. Posisi *centroid* divisualisasikan pada grafik PCA menggunakan *plt.scatter()* dan label "*Centroids*" ditambahkan melalui *plt.legend()* untuk memudahkan identifikasi.




**Gambar 5**. Titik Pusat *(Centroid)* pada tiap *cluster*

**c. Menghitung jarak antar titik data objek dan titik pusat *(centroid)***

Rumus *Euclidean Distance* digunakan untuk menghitung jarak antara setiap titik data objek dan titik pusat (*centroid*) klaster.

Perhitungan Jarak pada seluruh data ke semua pusat *cluster*:

- Jarak data 1 ke *cluster* 1

  √[(2147 - 1076.48)² + (23 - (-41.53))² + (51 - 35.21)² + (200 - 65.82)² + (1 - 1.74)² + (1.5 - 3.50)² + (1 - 2.41)² + (2 - 1.94)²] = 1080.943
- Jarak data 1 ke *cluster* 2

  √[(2147 - 8653.40)² + (23 - 983.72)² + (51 - 146.04)² + (200 - 158.18)² + (1 - 1.31)² + (1.5 - 16.51)² + (1 - 3.15)² + (2 - 2.15)²] = 6577.783
- Jarak data 1 ke *cluster* 3

  √[(2147 - 88291.88)² + (23 - 7831.58)² + (51 - 209.36)² + (200 - 49.34)² + (1 - 1.04)² + (1.5 - 24.25)² + (1 - 7.10)² + (2 - 0.62)²] = 86498.338
- Jarak data 1 ke *cluster* 4

  √[(2147 - 15179.95)² + (23 - 986.51)² + (51 - (-5.36))² + (200 - (-4.20))² + (1 - 1.89)² + (1.5 - (-1.20))² + (1 - 2.93)² + (2 - 1.50)²] = 13070.234

  ……
- Jarak data 3608 ke *cluster* 1

  √[(901 - 1076.48)² + (36 - (-41.53))² + (20 - 35.21)² + (45 - 65.82)² + (2 - 1.74)² + (2 - 3.50)² + (4 - 2.41)² + (2 - 1.94)²] = 193.584
- Jarak data 3608 ke *cluster* 2

p-ISSN: 2301-5373
e-ISSN: 2654-5101

√[(901 - 8653.40)² + (36 - 983.72)² + (20 - 146.04)² + (45 - 158.18)² + (2 - 1.31)² + (2 - 16.51)² + (4 - 3.15)² + (2 - 2.15)²] = 7811.964

- Jarak data 3608 ke *cluster* 3

  √[(901 - 88291.88)² + (36 - 7831.58)² + (20 - 209.36)² + (45 - 49.34)² + (2 - 1.04)² + (2 - 24.25)² + (4 - 7.10)² + (2 - 0.62)²] = 87738.095

- Jarak data 3608 ke *cluster* 4

  √[(901 - 15179.95)² + (36 - 986.51)² + (20 - (-5.36))² + (45 - (-4.20))² + (2 - 1.89)² + (2 - (-1.20))² + (4 - 2.93)² + (2 - 1.50)²] = 14310.658

**Tabel 7.** Jarak data ke setiap titik pusat *(centroid)*

| No | C1 | C2 | C3 | C4 |
|---|---|---|---|---|
| 1 | 1080.943 | 6577.783 | 86498.338 | 13070.234 |
| 2 | 1982.850 | 5863.744 | 85778.569 | 12391.580 |
| 3 | 1103.895 | 6544.636 | 86466.727 | 13038.137 |
| 4 | 1375.834 | 6279.509 | 86196.582 | 12766.728 |
| 5 | 255.7158 | 7437.983 | 87363.900 | 13937.322 |
| 6 | 8146.940 | 1025.826 | 79444.570 | 6021.110 |
| 7 | 512.9712 | 7160.515 | 87089.037 | 13664.102 |
| 8 | 1850.524 | 5798.624 | 85720.154 | 12292.017 |
| 9 | 4098.254 | 3610.449 | 83485.135 | 10055.243 |
| 10 | 289.3896 | 7887.462 | 87813.898 | 14387.323 |
| … | … | … | … | … |
| 3608 | 193.5854 | 7811.964 | 87738.095 | 14310.658 |

**d. Pengelompokkan data berdasarkan jarak minimum *cluster***

Data dikelompokkan berdasarkan jarak paling dekat ke pusat setiap klaster dan ditempatkan pada klaster yang memiliki nilai minimal pada iterasi. Tabel klaster iterasi berikut menunjukkan hasilnya:

**Tabel 8.** Pengelompokkan data

| No | C1 | C2 | C3 | C4 |
|---|---|---|---|---|
| 1 | 1 | | | |
| 2 | 1 | | | |
| 3 | 1 | | | |
| 4 | 1 | | | |
| 5 | 1 | | | |
| 6 | | 1 | | |
| 7 | 1 | | | |
| 8 | 1 | | | |
| 9 | | 1 | | |
| 10 | 1 | | | |
| … | … | … | … | … |
| 3608 | 1 | | | |

**e. Evaluasi**

Hasil evaluasi klaster menunjukkan nilai *Davies-Bouldin Inde*x (DBI) sebesar 0.6517, yang mengindikasikan bahwa klaster yang terbentuk memiliki pemisahan yang baik dan tidak terlalu tumpang tindih.

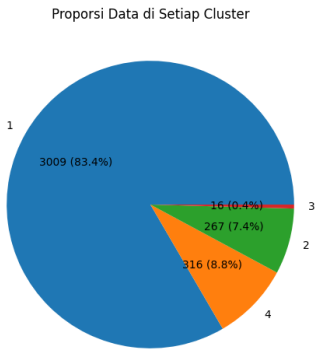
**Tabel 9.** Hasil evaluasi *Davies-Bouldin Inde*x (DBI)

| *Cluster* | Hasil nilai DBI |
|---|---|
| Average within-centroid distance C1 | 0.70 |
| Average within-centroid distance C2 | 1.99 |
| Average within-centroid distance C3 | 6.09 |
| Average within-centroid distance C4 | 0.97 |

Selain itu, nilai *Silhouette Score* untuk 4 klaster adalah 0.6269, yang menunjukkan bahwa struktur klaster cukup kuat, dengan nilai di atas 0.5 menunjukkan bahwa data dalam setiap klaster relatif lebih dekat satu sama lain dibandingkan dengan data di klaster lain. Kedua metrik ini menunjukkan bahwa pembagian klaster yang dihasilkan adalah efektif dan dapat diandalkan.

**f. Knowledge**

Hasil analisis kluster memberikan wawasan mendalam mengenai popularitas data yang diolah. Kluster 3, yang berisi 16 data, menunjukkan bahwa item-item dalam kluster ini memiliki daya tarik yang tinggi dan mungkin menarik perhatian pasar. Kluster 2, dengan 267 data, menunjukkan bahwa item-item di sini juga mendapatkan minat signifikan tetapi tidak sekuat kluster sebelumnya. Sebaliknya, Kluster 1 memiliki jumlah data terbanyak, yaitu 3009, mendanakan bahwa item dalam kluster ini mungkin memerlukan strategi pemasaran yang lebih agresif untuk meningkatkan daya tariknya. Terakhir, kluster 4, yang berisi 316 data, menunjukkan bahwa item-item dalam kluster ini kurang diminati, sehingga bisa menjadi fokus untuk evaluasi lebih lanjut dan perbaikan.




**Gambar** 4. Proporsi Data di Setiap *Cluster*Hasil analisis kluster memberikan

Berdasarkan hasil penelitian yang menunjukkan bahwa metode K-Means berhasil mengelompokkan kursus online Udemy ke dalam empat kluster optimal, maka rekomendasi eksplisit implementasi praktis yang dapat dilakukan oleh pihak manajemen Udemy dalam menerapkan hasil klasterisasi ini dalam strategi pemasaran, pengembangan konten, dan personalisasi layanan. Secara khusus:

- Kluster 1 dan 4, yang merepresentasikan kategori kursus dengan tingkat popularitas yang konsisten dan homogen, dapat dijadikan sebagai prioritas utama dalam promosi, peningkatan kualitas, dan replikasi model kursus sejenis.
- Kluster dengan popularitas rendah sebaiknya dianalisis lebih lanjut untuk mengidentifikasi kekurangan konten, kurangnya visibilitas, atau kebutuhan pembaruan topik agar dapat ditingkatkan efektivitasnya.
- Sistem rekomendasi berbasis klaster dapat dikembangkan, di mana pengguna baru ditawarkan kursus dari klaster yang sesuai dengan minat dominan dan tren popularitas saat ini.
- Dashboard internal berbasis data klasterisasi juga dapat digunakan oleh tim pengembang produk dan konten untuk mengawasi dinamika popularitas kursus secara real-time dan mengambil keputusan berbasis data.

Dengan demikian, hasil klasterisasi ini tidak hanya memberikan wawasan analitis, tetapi juga dapat dijadikan dasar pengambilan keputusan strategis dalam peningkatan kualitas layanan dan kepuasan pengguna platform Udemy.

## 4. Kesimpulan

Kesimpulan yang dihasilkan dari penelitian pengelompokan popularitas kursus online udemy dengan menggunakan metode K-Means, maka terdapat beberapa kesimpulan dari penelitian:

1. Penelitian pengelompokan popularitas kursus online Udemy menggunakan metode K-Means menghasilkan empat kluster optimal berdasarkan analisis *elbow*, yang merepresentasikan berbagai tingkat popularitas. Kluster 1 dan 4 menunjukkan homogenitas yang tinggi dengan rata-rata jarak *centroid* masing-masing 0.70 dan 0.97, sedangkan kluster 3 menunjukkan keragaman terbesar dengan jarak rata-rata 6.09. Hal ini menunjukkan perbedaan signifikan dalam karakteristik data antar-kluster, yang memberikan wawasan mengenai variasi popularitas kursus.
2. Nilai *Davies-Bouldin Indeks* (DBI) sebesar 0.65171 dan *silhouette score* sebesar 0.6269 menunjukkan kualitas klasterisasi yang baik, dengan pemisahan yang jelas antar kluster. Kluster 1, dengan jumlah data terbesar 3009, dapat menjadi fokus strategi pemasaran untuk meningkatkan daya tarik. Kluster 3, meskipun hanya berisi 16 data, menunjukkan kursus dengan daya tarik tinggi, sementara kluster 4 yang kurang diminati dapat dievaluasi lebih lanjut untuk perbaikan.

## Referensi

[1] C. Álvarez-Álvarez, “Video technologies for professor-student interaction in online teaching,” hal. 93–102, 2021, doi: 10.33422/3rd.ictle.2021.02.110.

[2] C. Cortázar *et al.*, “Promoting critical thinking in an online, project-based course,” *Comput. Human Behav.*, vol. 119, no. October 2020, 2021, doi: 10.1016/j.chb.2021.106705.

[3] T. Nguyen *et al.*, “Insights Into Students’ Experiences and Perceptions of Remote Learning Methods: From the COVID-19 Pandemic to Best Practice for the Future,” *Front. Educ.*, vol. 6, no. April, hal. 1–9, 2021, doi: 10.3389/feduc.2021.647986.

[4] J. E. Nieuwoudt, “Investigating synchronous and asynchronous class attendance as predictors of academic success in online education,” *Australas. J. Educ. Technol.*, vol. 36, no. 3, hal. 15–25, 2020, doi: 10.14742/AJET.5137.

[5] J. Mao, “The Impact of Online Courses Towards Students In Autonomous Learning Based On New Media Technologies,” *Rev. Educ. Theory*, vol. 3, no. 3, hal. 27–30, 2020.

[6] M. Bullen dan D. P. Janes, “Learning : Strategies and issues,” *Mak. Transit. to E_Learning Strateg. issues*, vol. 9, no. 3, 2007.

[7] F. A. A. Sebayang, O. Saragih, dan H. Hestina, “Pemanfaatan Media Pembelajaran Online untuk Meningkatkan Pembelajaran Mandiri Di Masa New Normal,” *Pelita Masy.*, vol. 2, no. 1, hal. 64–71, 2020, doi: 10.31289/pelitamasyarakat.v2i1.4222.

[8] G. Anuradha dan G. A. Hema, “Learners Preference Towards Online Courses- a Study,” *Glob. J. Res. Anal.*, no. December, hal. 43–45, 2021, doi: 10.36106/gjra/0206023.

[9] N. B. Segara, A. Suprijono, dan K. G. Setyawan, “The Influence of E-Learning towards Students’ Heutagogy Skills in Higher Education,” *J. Pendidik. dan Pengajaran*, vol. 54, no. 2, hal. 286, 2021, doi: 10.23887/jpp.v54i2.33128.

[10] T. S. Ramli *et al.*, “Pemanfaatan Teknologi Bagi Siswa Dalam Menyokong Peningkatan Ekonomi Digital dan Upaya Menghadapi Era Society 5 .0,” *J. Ilmu Huk. Kenotariatan*, vol. 6, hal. 81–98, 2022.

[11] H. Wijoyo *et al.*, *Strategi Pemasaran Umkm di Masa Pandemi Covid-19*, vol. 5, no. 1. 2022. doi: 10.24014/ekl.v5i1.14713.

[12] Z. Nabila, A. Rahman Isnain, dan Z. Abidin, “Analisis Data Mining Untuk *Cluster*ing Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means,” *J. Teknol. dan Sist. Inf.*, vol. 2, no. 2, hal. 100, 2021, [Daring]. Tersedia pada: http://jim.teknokrat.ac.id/index.php/JTSI

[13] I. Arfyanti, T. Bustomi, dan I. Haristyawan, “Implementasi Data Mining dengan Menerapkan Algoritma K-Means *Cluster*ing untuk Memberikan Rekomendasi Jurusan Kuliah Bagi

Mahasiswa Baru," *J. Media Inform. Budidarma*, vol. 8, no. 2, hal. 988, 2024, doi: 10.30865/mib.v8i2.7429.

[14] T. H. Sardar dan Z. Ansari, "An analysis of MapReduce efficiency in document *cluster*ing using parallel K-means algorithm," *Futur. Comput. Informatics J.*, vol. 3, no. 2, hal. 200–209, 2018, doi: 10.1016/j.fcij.2018.03.003.

[15] M. M. Ghazal dan A. Hammad, "Application of knowledge discovery in database (KDD) techniques in cost overrun of construction projects," *Int. J. Constr. Manag.*, vol. 22, no. 9, hal. 1632–1646, 2022, doi: 10.1080/15623599.2020.1738205.

[16] P. Meilina, "Penerapan Data Mining dengan Metode Klasifikasi Menggunakan Decision Tree dan Regresi," *J. Teknol. Univ. Muhammadiyah Jakarta*, vol. 7, no. 1, hal. 11–20, 2015, [Daring]. Tersedia pada: jurnal.ftumj.ac.id/index.php/jurtek

[17] Sarmini, W. Ma, dan I. Tahyudin, "Sistem Pendukung Keputusan Berbasis K-Means untuk Evaluasi Keberhasilan Bisnis dan Nilai Perusahaan," *J. Sist. Inf. Bisnis*, vol. 4, no. 4, hal. 363–374, 2024, doi: 10.21456/vol14iss4pp363-374.

[18] F. Sembiring, O. Octaviana, dan S. Saepudin, "Implementasi Metode K-Means Dalam Pengklasteran Daerah Pungutan Liar Di Kabupaten Sukabumi (Studi Kasus : Dinas Kependudukan Dan Pencatatan Sipil)," *J. Tekno Insentif*, vol. 14, no. 1, hal. 40–47, 2020, doi: 10.36787/jti.v14i1.165.

[19] F. Zahra, M. A. Ridla, dan N. Azise, "Implementasi Data Mining Menggunakan Algoritma Apriori Dalam Menentukan Persediaan Barang (Studi Kasus : Toko Sinar Harahap)," *JUSTIFY J. Sist. Inf. Ibrahimy*, vol. 3, no. 1, hal. 55–65, 2024, doi: 10.35316/justify.v3i1.5335.

[20] P. J. Rousseeuw, "*Silhouette*s: A graphical aid to the interpretation and validation of *cluster* analysis," *J. Comput. Appl. Math.*, vol. 20, no. C, hal. 53–65, 1987, doi: 10.1016/0377-0427(87)90125-7.

[21] R. Sari, M. I. Arief, dan F. N. Yasir, "Optimasi *Cluster*ing Kualitas Air Menggunakan Principle Component Analysis," vol. 1, no. 3, hal. 24–30, 2024.

[22] C. Zai, "Implementasi Data Mining Sebagai Pengolahan Data," *J. Portal Data*, vol. 2, no. 3, hal. 1–12, 2022, [Daring]. Tersedia pada: http://portaldata.org/index.php/portaldata/article/view/107

[23] E. A. Saputra dan Y. Nataliani, "Analisis Pengelompokan Data Nilai Siswa untuk Menentukan Siswa Berprestasi Menggunakan Metode *Cluster*ing K-Means," *J. Inf. Syst. Informatics*, vol. 3, no. 3, hal. 424–439, 2021, doi: 10.51519/journalisi.v3i3.164.