



## ***Implementation of the K-Means Algorithm for Clustering Student Data***

### **Implementasi Algoritma K-Means untuk Clustering Data Mahasiswa**

**Meila Anriana<sup>1\*</sup>, Allya Dwi Ramadhani<sup>2</sup>, Rita Novita<sup>3</sup>**

<sup>1</sup>Program Studi Sistem Informasi, Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau

<sup>2</sup>Program Studi Teknik Industri, Fakultas Teknologi Industri Universitas Islam Indonesia

<sup>3</sup>Institut Kesehatan Payung Negeri Pekanbaru

E-Mail: [12350321875@students.uin-suska.ac.id](mailto:12350321875@students.uin-suska.ac.id)<sup>1</sup>, [24522050@students.uui.ac.id](mailto:24522050@students.uui.ac.id)<sup>2</sup>,  
[ritanovita@payungnegeri.ac.id](mailto:ritanovita@payungnegeri.ac.id)<sup>3</sup>

*Makalah: Diterima 12 Agustus 2024; Diperbaiki 20 Agustus 2024; Disetujui 2 September 2024*  
*Corresponding Author: Meila Anriana*

#### **Abstrak**

Pengelolaan data mahasiswa aktif dalam jumlah besar memerlukan metode analisis yang mampu mengidentifikasi pola dan karakteristik secara terstruktur. Penelitian ini menerapkan data mining dengan teknik clustering menggunakan algoritma K-Means untuk mengklusterkan data mahasiswa berdasarkan atribut jalur masuk, program studi, fakultas, tahun masuk, dan besaran Uang Kuliah Tunggal (UKT). Dataset yang digunakan berjumlah 34.016 data mahasiswa aktif dari Pusat Teknologi Informasi dan Pangkalan Data (PTIPD) UIN Sultan Syarif Kasim Riau periode 2020–2025, dengan distribusi UKT didominasi nominal Rp3.200.000 dan Rp2.400.000, jumlah mahasiswa tertinggi pada tahun masuk 2020, serta ketidakseimbangan data pada atribut jalur masuk yang didominasi UM-PTKIN dan PMB Ujian Mandiri. Tahapan penelitian meliputi preprocessing data berupa seleksi atribut, penanganan missing values, serta transformasi data melalui one-hot encoding dan normalisasi. Penentuan jumlah cluster optimal dilakukan menggunakan Elbow Method, sedangkan evaluasi menggunakan Silhouette Score. Hasil menunjukkan sebagian besar fakultas terbagi ke dalam dua cluster dan sebagian lainnya tiga cluster, dengan nilai Silhouette Score 0,5441–0,7427. Nilai ini menunjukkan kualitas clustering sangat baik karena cluster dapat dibedakan dengan jelas.

Keyword: clustering, data mahasiswa, elbow method, k-means, silhouette score

#### **Abstract**

*Managing large volumes of active student data requires analytical methods capable of identifying structured patterns and characteristics. This study applies data mining techniques using the K-Means clustering algorithm to cluster student data based on admission pathway, study program, faculty, year of entry, and Tuition Fee (UKT). The dataset consists of 34,016 active student records obtained from the Center for Information Technology and Database (PTIPD) of UIN Sultan Syarif Kasim Riau for the period 2020–2025, with UKT values dominated by Rp3,200,000 and Rp2,400,000, the highest number of students in the 2020 intake year, and an imbalance in admission pathways dominated by UM-PTKIN and PMB Ujian Mandiri. The research stages include data preprocessing through attribute selection, handling missing values, and data transformation using one-hot encoding and normalization. The optimal number of clusters is determined using the Elbow Method, while evaluation is conducted using the Silhouette Score. The results show that most faculties are grouped into two clusters and others into three clusters, with Silhouette Score values ranging from 0.5441 to 0.7427. These values indicate very good clustering quality, as the clusters are clearly distinguishable.*

Keyword: clustering, elbow method, k-means, silhouette score, student data

## 1. PENDAHULUAN

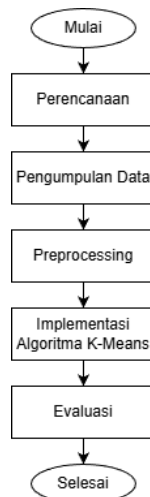
Pengelolaan data mahasiswa di perguruan tinggi penting untuk mendukung pengambilan keputusan berbasis informasi. Setiap tahun, jumlah mahasiswa meningkat dengan karakteristik beragam, seperti jalur penerimaan, program studi, tahun masuk, dan besaran Uang Kuliah Tunggal (UKT). Namun, data mahasiswa selama ini sebagian besar hanya digunakan untuk keperluan administratif, sehingga informasi mengenai sebaran mahasiswa belum terlihat jelas [1], [2]. Universitas Islam Negeri Sultan Syarif Kasim Riau memiliki sembilan fakultas dan lebih dari lima puluh program studi, dengan total mahasiswa 36.235 orang pada periode 2020–2025, rata-rata 6.000–7.000 mahasiswa per tahun. Tingginya volume data ini memerlukan metode analisis yang efisien agar distribusi mahasiswa dapat terlihat secara menyeluruh dan informasi yang diperoleh dapat mendukung pengelolaan akademik. Algoritma K-Means menjadi metode yang sesuai karena mampu mengklasifikasikan data besar dengan cepat dan menghasilkan kluster berdasarkan kesamaan karakteristik mahasiswa [3], [4]. Berbagai kajian relevan sudah memperlihatkan efektivitas algoritma K-Means saat mengklasifikasikan data pendidikan. Sebagai contoh, K-Means berhasil digunakan untuk mengelompokkan calon penerima beasiswa menjadi tiga kategori berdasarkan kriteria tertentu seperti prestasi akademik, ekonomi, dan kebutuhan bantuan [5]. Hasil pengelompokan ini membantu pihak yayasan dalam menentukan penerima beasiswa secara lebih sistematis dan objektif. Selain itu, K-Means juga diterapkan untuk mengelompokkan mahasiswa baru berdasarkan jalur masuk, program studi, dan prestasi akademik mereka [6]. Pendekatan ini memudahkan universitas dalam memahami profil mahasiswa baru serta melakukan perencanaan program akademik yang lebih tepat sasaran.

Beberapa studi juga menggabungkan K-Means dengan metode tambahan untuk meningkatkan akurasi klusterisasi. Misalnya, Lashiyanti mengintegrasikan K-Means dengan Elbow Method serta Silhouette Coefficient guna menetapkan total kluster optimal, menghasilkan nilai Silhouette Score sebesar 0,74 [7]. Pendekatan ini membuktikan bahwa K-Means tidak hanya cepat tetapi juga mampu menghasilkan kelompok yang representatif, sehingga analisis data menjadi lebih valid. Penerapan K-Means pada sekolah oleh Nurahman juga menunjukkan bahwa metode ini dapat membantu pemerintah daerah menentukan prioritas pengembangan fasilitas dan tenaga pendidik, sehingga sumber daya pendidikan dapat dialokasikan secara lebih efektif [8]. Selain di bidang pendidikan, K-Means juga terbukti efektif di sektor sosial dan ekonomi. Sebagai contoh, analisis tingkat kemiskinan di Indonesia menggunakan K-Means berhasil membentuk tiga kelompok utama dengan karakteristik berbeda, dan memperoleh nilai Silhouette Score sebesar 0,7416 [9]. Hal ini memungkinkan pemerintah untuk merancang kebijakan penanggulangan kemiskinan yang lebih tepat sasaran. Di sektor kesehatan, pengelompokan kasus stunting balita berdasarkan usia dan berat badan menggunakan K-Means menunjukkan hasil optimal sehingga intervensi gizi dan kesehatan dapat difokuskan pada kelompok yang paling membutuhkan [10]. Temuan-temuan tersebut memperjelas bahwasanya K-Means merupakan metode yang efisien, adaptif, serta relevan guna menganalisis data besar, baik dalam konteks pendidikan, sosial, maupun ekonomi, karena mampu menghasilkan informasi yang mudah dipahami dan menopang penetapan keputusan berbasis data.

Penelitian ini dilakukan guna mengelompokkan mahasiswa UIN Suska Riau berdasarkan jalur masuk, program studi, tahun masuk, dan besaran UKT. Tujuannya adalah agar PTIPD memperoleh gambaran sebaran mahasiswa secara menyeluruh dan mendapatkan informasi yang bermanfaat untuk pengelolaan akademik serta evaluasi internal universitas. Analisis ini menggunakan algoritma K-Means untuk menampilkan distribusi mahasiswa dengan lebih jelas, membantu identifikasi pola, dan mendukung perencanaan strategis di tingkat universitas.

## 2. METODE DAN BAHAN

Penelitian ini meliputi lima langkah utama, ialah tahap perencanaan dan studi literatur, pengumpulan dataset mahasiswa dari PTIPD UIN Suska Riau, preprocessing dataset untuk memastikan kualitas data dengan memilih kolom relevan, menangani missing values, serta melakukan transformasi dan normalisasi data, kemudian data yang sudah siap diproses menggunakan algoritma K-Means untuk membentuk cluster berdasarkan kesamaan karakteristik mahasiswa, dan kualitas cluster yang terbentuk dievaluasi menggunakan Silhouette Score. Metodologi penelitian bisa dilihat di Gambar 1.



Gambar 1. Alur Penelitian

## 2.1 Data Mining

Data mining ialah proses analisis terhadap kumpulan data besar guna melihat pola, hubungan, dan pengetahuan baru yang berguna dalam pengambilan keputusan. Proses ini mengekstraksi informasi penting melalui tahapan pembersihan, transformasi, analisis, dan interpretasi hasil, serta bersifat interdisipliner karena menggabungkan statistik, kecerdasan buatan, dan machine learning [11]. Dalam kerangka Knowledge Discovery in Database (KDD), data mining merupakan tahap inti yang mencakup proses pembersihan data, integrasi sumber data, seleksi atribut, transformasi, penerapan algoritma analisis, dan evaluasi untuk menghasilkan pola yang valid dan bermanfaat [12]. Secara teknis, proses data mining dimulai dengan pembersihan dan transformasi data agar siap diolah oleh algoritma seperti classification, clustering, dan association rule mining untuk menemukan pola dan hubungan tersembunyi dalam dataset [13]. Hasil akhir berupa pola, aturan, atau model prediktif yang mendukung pengambilan keputusan secara efektif [14].

## 2.2 Clustering

Clustering merupakan bagian dari teknik pada data mining yang dipakai guna mengklasifikasikan data ke dalam banyak kelompok ataupun kluster berdasarkan tingkat kemiripan antar data. Sasaran utama dari clustering ialah menempatkan data yang mempunyai karakteristik serupa ke dalam satu kelompok agar pola dan hubungan tersembunyi di dalam data dapat dikenali dengan lebih mudah. Melalui proses pengelompokan ini, data yang beragam dapat disusun menjadi kelompok yang lebih terstruktur sehingga memudahkan proses analisis selanjutnya [15].

Clustering termasuk ke dalam kategori unsupervised learning karena proses pengelompokan dilakukan tanpa menggunakan label atau kategori awal. Algoritma clustering bekerja dengan mengenali struktur alami data melalui analisis jarak atau tingkat kesamaan antar data, sehingga keterkaitan antar data dapat ditemukan tanpa memerlukan supervisi langsung [16]. Salah satu metode pengukuran jarak yang umum dipakai pada clustering ialah Euclidean Distance, yang perhitungannya dinyatakan dalam bentuk persamaan di bawah.

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

## 2.3 Algoritma K-Means

Algoritma K-Means ialah bagian algoritma clustering yang paling banyak dipakai karena memiliki konsep yang sederhana, efisien, dan mampu mengatasi data pada jumlah banyak. Algoritma ini bekerja dengan menetapkan pusat kluster awal atau centroid, kemudian mengklasifikasikan data berdasarkan jarak paling dekat terhadap pusat kluster itu. Tahap pengelompokan dilakukan dengan cara berulang sampai posisi pusat kluster tidak mengalami perubahan yang signifikan [17].

Penentuan kedekatan antar data dalam algoritma K-Means dilakukan menggunakan Euclidean Distance. Setelah data terbagi ke dalam masing-masing kluster, pusat kluster diperbaharui dengan menghitung nilai rata-rata dari semua data dalam kluster itu. Tahap ini kemudian diulang sampai tercapai kondisi konvergensi atau tidak terjadi perubahan posisi pusat kluster [18]. Perhitungan pembaruan pusat kluster dinyatakan dalam persamaan berikut.

$$\mu_k = \frac{1}{|C_k|} \sum_{x_i \in C_k} x_i \quad (2)$$

Tujuan utama dari algoritma K-Means ialah mengurangi nilai Within-Cluster Sum of Squares, yakni total jarak kuadrat dari setiap data serta pusat klasternya. Fungsi objektif dari algoritma K-Means dinyatakan dalam persamaan berikut.

$$J = \sum_{k=1}^K \sum_{x_i \in C_k} ||x_i - \mu_k||^2 \quad (3)$$

Algoritma K-Means dianggap efektif karena mampu menghasilkan hasil pengelompokan yang mudah dipahami, mempunyai tingkat akurasi yang baik, serta waktu komputasi yang relatif cepat, sehingga banyak digunakan dalam berbagai penerapan analisis data [19].

#### 2.4 Elbow Method

Elbow Method merupakan metode yang digunakan untuk menentukan jumlah kluster optimal dalam algoritma K-Means. Metode ini dilakukan dengan menghitung nilai Within-Cluster Sum of Squares untuk berbagai jumlah kluster, kemudian menampilkannya dalam bentuk grafik. Jumlah kluster optimal ditentukan pada titik di mana penurunan nilai kesalahan mulai melambat secara signifikan dan membentuk sudut menyerupai siku [20].

Nilai Within-Cluster Sum of Squares diperoleh dari total jarak kuadrat pada setiap data serta pusat klasternya. Setelah melewati jumlah kluster tertentu, penurunan nilai kesalahan menjadi tidak signifikan sehingga jumlah kluster tersebut dianggap sebagai jumlah yang paling efisien [21]. Pendekatan Elbow Method bekerja dengan meminimalkan nilai kesalahan pengelompokan dan divisualisasikan dalam grafik hubungan antara jumlah kluster dan nilai error [22]. Selain itu, pendekatan analitik juga dapat digunakan untuk membantu menentukan titik siku secara lebih objektif dan kuantitatif [23].

$$WCSS = \sum_{k=1}^K \sum_{x_i \in C_k} ||x_i - \mu_k||^2 \quad (4)$$

#### 2.5 Silhouette Score

Silhouette Score merupakan metode evaluasi yang dipakai guna menentukan kualitas hasil clustering dengan mengukur tingkat kesesuaian data terhadap kluster yang terbentuk serta jaraknya pada kluster lainnya. Nilai Silhouette ada di rentang minus satu sampai satu, yang mana nilai yang semakin dekat dengan satu memperlihatkan bahwa hasil pengklasifikasian semakin baik [24].

Metode ini mempertimbangkan dua aspek utama, yaitu tingkat kedekatan data dalam satu kluster dan tingkat pemisahan data terhadap kluster terdekat lainnya [25]. Silhouette Score dapat digunakan bersama metode evaluasi lainnya seperti Elbow Method serta Davies–Bouldin Index guna menetapkan struktur kluster yang paling optimal. Pendekatan ini memberikan dasar penilaian secara kuantitatif sehingga hasil evaluasi menjadi lebih objektif dan terukur [26].

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (5)$$

### 3. ANALISA DAN PEMBAHASAN

#### 3.1 Pengumpulan Data

Data yang dipakai di analisis ini merupakan data mahasiswa yang didapatkan dari Pusat Teknologi Informasi dan Pangkalan Data (PTIPD). Dataset berisi data identitas, akademik, dan finansial mahasiswa yang tersimpan dalam sistem internal perguruan tinggi. Data dikumpulkan dalam bentuk dataset terstruktur dan dipakai selaku dasar pada proses analisis clustering.

Guna memberikan gambaran mengenai struktur data yang digunakan, atribut-atribut yang terdapat dalam dataset disajikan di Tabel 1.

**Table 1.** Atribut Dataset

No	Nama Atribut	Tipe Data	Keterangan
1	nim	Numerik	Nomor Induk Mahasiswa
2	nama	Teks	Nama lengkap mahasiswa
3	nama_jurusan	Kategorikal	Program studi mahasiswa
4	nama_jalur_masuk	Kategorikal	Jalur penerimaan mahasiswa
5	nama_fakultas	Kategorikal	Fakultas mahasiswa
6	nik	Numerik	Nomor Induk Kependudukan
7	email	Teks	Alamat email mahasiswa
8	no_selular	Numerik	Nomor telepon seluler
9	no_wa_aktif	Numerik	Nomor WhatsApp aktif

10	jenis_kelamin	Kategorikal	Jenis kelamin mahasiswa
11	ukuran_jaket	Kategorikal	Ukuran jaket almamater
12	asal_sekolah	Teks	Asal sekolah mahasiswa
13	periode	Numerik	Periode akademik
14	jlh_bayar	Numerik	Jumlah pembayaran UKT
15	tahun_masuk	Numerik	Tahun masuk mahasiswa

### 3.2 Preprocessing

Tahap preprocessing dilaksanakan guna menyiapkan data agar dapat diolah menggunakan algoritma K-Means. Tahap ini bertujuan guna memverifikasi bahwa data yang digunakan bersih, relevan, serta berada pada format yang cocok guna proses clustering. Tahapan preprocessing meliputi seleksi atribut, penanganan data tidak lengkap (*missing value*), transformasi data, dan normalisasi.

#### 3.2.1 Seleksi Atribut

Seleksi atribut dilakukan untuk menentukan atribut yang relevan dan digunakan dalam proses clustering. Atribut yang dipilih merupakan atribut akademik dan finansial mahasiswa yang dianggap mampu merepresentasikan karakteristik mahasiswa. Atribut yang digunakan pada penelitian ini ditunjukkan pada Tabel 2.

**Table 2.** Atribut Relevan

No	Nama Atribut	Tipe Data	Keterangan
1	nama_jurusan	Kategorikal	Program studi mahasiswa
2	nama_jalur_masuk	Kategorikal	Jalur penerimaan mahasiswa
3	nama_fakultas	Kategorikal	Fakultas mahasiswa
4	jlh_bayar	Numerik	Jumlah pembayaran UKT
5	tahun_masuk	Numerik	Tahun masuk mahasiswa

#### 3.2.2 Penanganan Data Tidak Lengkap

Pada tahap ini dilakukan pemeriksaan terhadap data yang mempunyai nilai kosong atau tidak lengkap. Ditemukan beberapa data yang tidak memiliki nilai pada atribut jumlah pembayaran (UKT) dan tahun masuk sebagaimana ditunjukkan pada Tabel 3. Untuk mengatasi hal tersebut, nilai kosong digantikan dengan nilai rata-rata dari tiap-tiap atribut. Setelah proses ini dilakukan, seluruh data berada dalam kondisi lengkap dan siap untuk diproses pada tahap selanjutnya.

**Table 3.** Data Tanpa Missing Value

no	nama_fakultas	jlh_bayar	tahun_masuk	nama_jurusan	nama_jalur_masuk
0	Syari'ah dan Hukum	3200000.0	2020.0	D-III Perbankan Syariah	PMB Mandiri
1	Syari'ah dan Hukum	2600000.0	2020.0	D-III Perbankan Syariah	PMB Undangan
2	Syari'ah dan Hukum	3200000.0	2020.0	D-III Perbankan Syariah	PMB Undangan
3	Syari'ah dan Hukum	1925000.0	2020.0	D-III Perbankan Syariah	PMB Undangan
4	Syari'ah dan Hukum	3200000.0	2020.0	D-III Perbankan Syariah	PMB Mandiri
...	...	...	...	...	...
34011	Tarbiyah dan Keguruan	400000.0	2024.0	Pendidikan Profesi Guru	PPG Non PGP
34012	Tarbiyah dan Keguruan	400000.0	2024.0	Pendidikan Profesi Guru	PPG Non PGP
34013	Tarbiyah dan Keguruan	400000.0	2024.0	Pendidikan Profesi Guru	PPG Non PGP
34014	Tarbiyah dan Keguruan	400000.0	2024.0	Pendidikan Profesi Guru	PPG Non PGP
34015	Tarbiyah dan Keguruan	400000.0	2024.0	Pendidikan Profesi Guru	PPG Non PGP

#### 3.2.3 Transformasi Data

Atribut kategorikal yang digunakan, yaitu nama jurusan, nama fakultas, dan nama jalur masuk, ditransformasikan ke dalam bentuk numerik menggunakan metode *one-hot encoding* sebagaimana ditunjukkan pada Tabel 4. Transformasi ini menghasilkan beberapa kolom baru yang merepresentasikan setiap kategori pada atribut tersebut. Proses ini dilakukan agar data kategorikal bisa diproses oleh algoritma K-Means yang berasaskan pada perhitungan jarak.

**Table 4.** Dataset Setelah One Hot Encoding

no	nama_fakultas	jlh_bayar	tahun_masuk	nama_jurusan_administrasi_negara	nama_jurusan_agroteknologi	...	nama_jalur_masuk_spontan	nama_jalur_masuk_umptkin
0	Syari'ah dan Hukum	3200000.0	2020.0	False	False	...	False	False

1	Syari'ah dan Hukum	2600000.0	2020.0	False	False	...	False	False
2	Syari'ah dan Hukum	3200000.0	2020.0	False	False	...	False	False
3	Syari'ah dan Hukum	1925000.0	2020.0	False	False	...	False	False
4	Syari'ah dan Hukum	3200000.0	2020.0	False	False	...	False	False
...	...	...	...	...	...	...	...	...
34011	Tarbiyah dan Keguruan	400000.0	2024.0	False	False	...	False	False
34012	Tarbiyah dan Keguruan	400000.0	2024.0	False	False	...	False	False
34013	Tarbiyah dan Keguruan	400000.0	2024.0	False	False	...	False	False
34014	Tarbiyah dan Keguruan	400000.0	2024.0	False	False	...	False	False
34015	Tarbiyah dan Keguruan	400000.0	2024.0	False	False	...	False	False

### 3.2.4 Normalisasi Data

Normalisasi data dilakukan pada atribut numerik, yaitu jumlah pembayaran (UKT) dan tahun masuk sebagaimana ditunjukkan pada Tabel 5. Metode normalisasi yang digunakan adalah *Min-Max normalization* dengan tujuan menyamakan skala nilai antar atribut. Dengan normalisasi ini, setiap atribut memiliki kontribusi yang seimbang dalam proses clustering sehingga hasil pengelompokan menjadi lebih optimal.

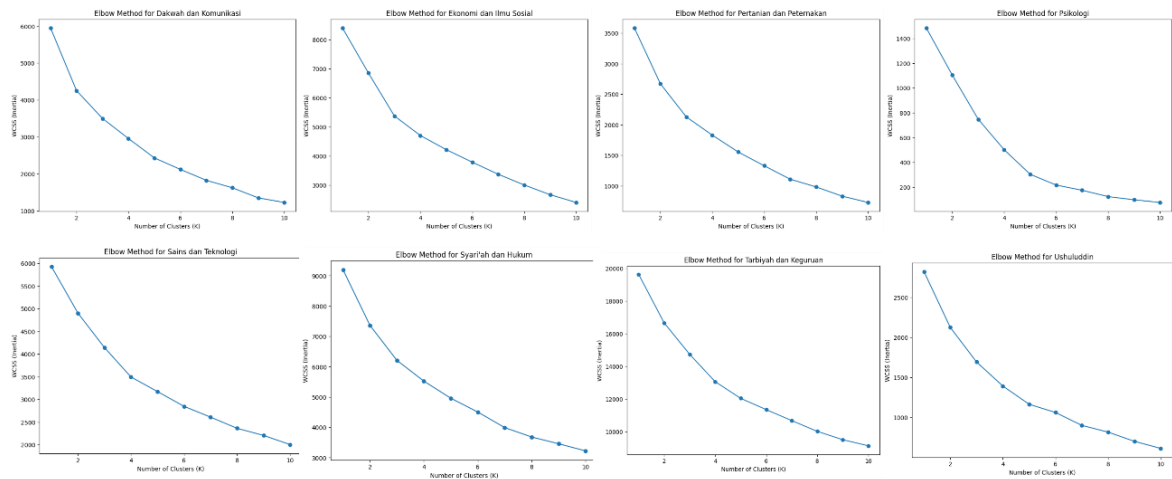
**Table 5.** Dataset Setelah Normalisasi

no	nama_fakultas	jlh_bayar	tahu_nmasuk	nama_jurusan_administrasi_negara	nama_jurusan_teknologi	...	nama_jalur_masuk	nama_jalur_masuk
0	Syari'ah dan Hukum	0.410256	0.0	False	False	...	False	False
1	Syari'ah dan Hukum	0.333333	0.0	False	False	...	False	False
2	Syari'ah dan Hukum	0.410256	0.0	False	False	...	False	False
3	Syari'ah dan Hukum	0.246795	0.0	False	False	...	False	False
4	Syari'ah dan Hukum	0.410256	0.0	False	False	...	False	False
...	...	...	...	...	...	...	...	...
34011	Tarbiyah dan Keguruan	0.051282	0.8	False	False	...	False	False
34012	Tarbiyah dan Keguruan	0.051282	0.8	False	False	...	False	False
34013	Tarbiyah dan Keguruan	0.051282	0.8	False	False	...	False	False
34015	Tarbiyah dan Keguruan	0.051282	0.8	False	False	...	False	False
34015	Tarbiyah dan Keguruan	0.051282	0.8	False	False	...	False	False

## 3.3 Implementasi Algoritma K-Means

### 3.3.1 Penentuan Jumlah Cluster Terbaik (K)

Penetapan total kelompok dilaksanakan menggunakan metode Elbow dengan melihat perubahan nilai WCSS saat jumlah kelompok ditambah. Cara ini bertujuan untuk menemukan titik tempat penurunan nilai mulai melambat, yang memperlihatkan total kelompok yang paling sesuai. Berdasarkan grafik Elbow pada Gambar 2, sebagian besar fakultas menunjukkan bahwa dua kelompok sudah cukup untuk menggambarkan pola data mahasiswa. Namun, Fakultas Ekonomi dan Ilmu Sosial memiliki pola yang berbeda, sehingga diperlukan tiga kelompok agar data dapat terwakili dengan lebih baik. Oleh karena itu, proses pengelompokan data pada setiap fakultas disesuaikan dengan hasil tersebut agar hasil clustering lebih akurat dan sesuai dengan kondisi data.



Gambar 2. Grafik Elbow Method

### 3.3.2 Penentuan Centroid

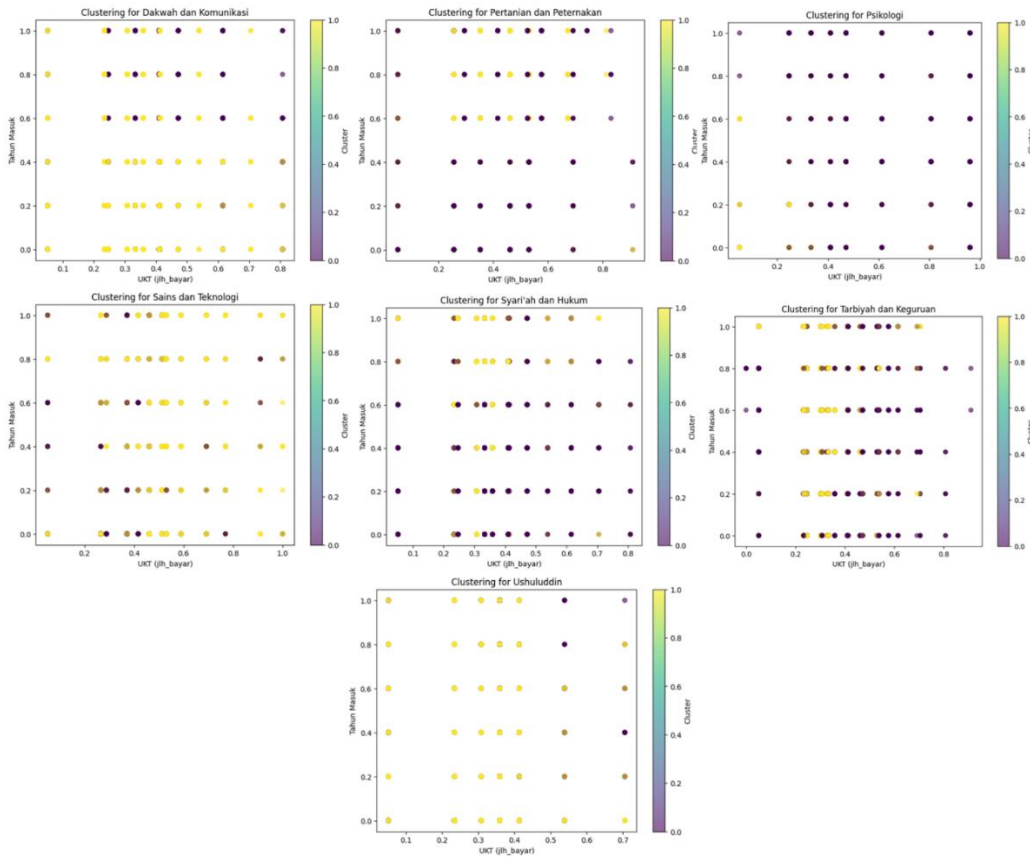
Setelah jumlah kelompok ditentukan, metode K-Means akan mencari titik tengah dari setiap kelompok. Titik tengah ini diperoleh dengan menghitung rata-rata dari data tahun masuk dan jumlah UKT mahasiswa. Proses ini diulang beberapa kali sampai titik tengah tersebut tidak berubah lagi. Hasil akhirnya dipakai guna membagi data mahasiswa kepada kelompok-kelompok yang sesuai. Nilai titik tengah setiap kelompok ditampilkan pada Tabel 6.

Table 6. Centroid Cluster

No	Fakultas	Cluster	UKT	Tahun Masuk
0	Dakwah dan Komunikasi	0	3,646,630.38	2022.25
1	Dakwah dan Komunikasi	1	2,150,220.39	2021.74
2	Ekonomi dan Ilmu Sosial	0	1,975,323.42	2022.20
3	Ekonomi dan Ilmu Sosial	1	5,033,532.93	2022.67
4	Ekonomi dan Ilmu Sosial	2	3,426,582.96	2022.45
5	Pertanian dan Peternakan	0	4,000,511.44	2022.70
6	Pertanian dan Peternakan	1	2,109,234.23	2021.94
7	Psikologi	0	6,107,540.23	2022.52
8	Psikologi	1	2,877,164.10	2022.29
9	Sains dan Teknologi	0	2,364,694.36	2022.22
10	Sains dan Teknologi	1	4,211,830.60	2022.45
11	Syari'ah dan Hukum	0	3,048,347.91	2022.15
12	Syari'ah dan Hukum	1	1,442,901.23	2022.54
13	Tarbiyah dan Keguruan	0	1,084,446.37	2022.80
14	Tarbiyah dan Keguruan	1	3,112,343.22	2022.61
15	Ushuluddin	0	1,489,829.88	2021.73
16	Ushuluddin	1	2,931,123.36	2022.20

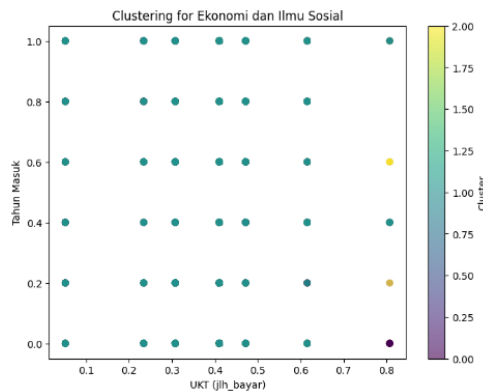
### 3.3.3 Hasil Clustering dengan Algoritma K-Means

Hasil clustering menggunakan algoritma K-Means divisualisasikan untuk mempermudah pemahaman terhadap pola pengelompokan data mahasiswa berdasarkan tahun masuk dan jumlah pembayaran UKT. Visualisasi hasil clustering dibedakan berdasarkan jumlah cluster yang terbentuk, yaitu tiga cluster dan dua cluster.



**Gambar 3.** Visualisasi Cluster Fakultas dengan K=2

Gambar 3 menunjukkan visualisasi hasil clustering pada fakultas-fakultas yang terbagi ke dalam dua cluster. Pada visualisasi tersebut terlihat bahwa satu cluster memiliki sebaran titik data yang lebih padat, sementara cluster lainnya cenderung lebih jarang. Perbedaan kepadatan ini menunjukkan adanya kelompok mahasiswa dengan karakteristik yang lebih dominan jumlahnya dibandingkan kelompok lainnya. Meskipun demikian, kedua cluster masih dapat dibedakan dengan cukup jelas berdasarkan variasi nilai UKT dan tahun masuk, sehingga penggunaan dua cluster dinilai sudah memadai untuk merepresentasikan pola data pada fakultas-fakultas tersebut.



**Gambar 4.** Visualisasi Cluster Fakultas dengan K=3

Gambar 4 menampilkan visualisasi hasil clustering pada Fakultas Ekonomi dan Ilmu Sosial yang terbagi ke dalam tiga cluster. Dari visualisasi tersebut terlihat bahwa sebagian besar data mahasiswa tergolong ke dalam cluster yang didominasi oleh warna toska, sedangkan cluster berwarna kuning hanya mencakup sebagian kecil data, dan cluster berwarna ungu jumlahnya paling sedikit. Kondisi ini menunjukkan bahwa

mayoritas mahasiswa memiliki karakteristik yang relatif mirip, sementara dua cluster lainnya merepresentasikan kelompok mahasiswa dengan karakteristik yang lebih spesifik dan jumlah yang lebih terbatas. Pembagian menjadi tiga cluster ini membantu menggambarkan variasi data yang lebih beragam pada fakultas tersebut.

### 3.4 Evaluasi Hasil Menggunakan Silhouette Score

Tahap evaluasi dilaksanakan guna melihat kualitas hasil clustering yang sudah diperoleh. Evaluasi ini bertujuan untuk mengetahui seberapa baik algoritma K-Means dalam mengelompokkan data mahasiswa pada masing-masing fakultas berdasarkan atribut yang digunakan. Metode evaluasi yang dipakai pada analisis ini ialah Silhouette Score, yang mengukur tingkat kesamaan data pada satu cluster dan tingkat pemisahannya dengan cluster lain.

Nilai Silhouette Score ada di rentang  $-1$  sampai  $1$ , yang mana nilai yang hampir mencapai  $1$  memperlihatkan struktur cluster yang semakin baik dan terpisah dengan jelas. Hasil evaluasi clustering untuk setiap fakultas ditunjukkan pada Tabel 7, yang menyajikan nilai Silhouette Score beserta interpretasinya.

**Table 7.** Nilai Evaluasi Silhouette Score

No	Nama Fakultas	Nilai Silhouette Score	Interpretasi
1	Fakultas Dakwah dan Komunikasi	0.5441	Data sudah terbagi ke dalam kelompok yang cukup rapi dan mudah dibedakan, sehingga hasil clustering termasuk well-structured.
2	Fakultas Ekonomi dan Ilmu Sosial	0.7427	Data terbagi sangat rapi dan perbedaan antar kelompok terlihat sangat jelas, sehingga hasil clustering termasuk well-structured.
3	Fakultas Pertanian dan Peternakan	0.6801	Data tersusun dengan baik dan tiap kelompok dapat dibedakan dengan jelas, sehingga hasil clustering termasuk well-structured.
4	Fakultas Psikologi	0.7042	Data dalam setiap kelompok memiliki kemiripan tinggi dan antar kelompok terlihat jelas perbedaannya, sehingga hasil clustering termasuk well-structured.
5	Fakultas Sains dan Teknologi	0.5635	Data sudah mengelompok dengan rapi dan tidak banyak tercampur antar kelompok, sehingga hasil clustering termasuk well-structured.
6	Fakultas Syari'ah dan Hukum	0.5840	Data telah terbagi dengan baik ke dalam beberapa kelompok yang jelas, sehingga hasil clustering termasuk well-structured.
7	Fakultas Tarbiyah dan Keguruan	0.6443	Pembagian data ke dalam kelompok terlihat rapi dan perbedaannya cukup jelas, sehingga hasil clustering termasuk well-structured.
8	Fakultas Ushuluddin	0.6288	Data sudah terbagi dengan baik ke dalam kelompok yang jelas, sehingga hasil clustering termasuk well-structured.

## 4. KESIMPULAN

Penelitian ini menerapkan algoritma K-Means guna mengklasifikasikan data mahasiswa pada setiap fakultas berdasarkan atribut jalur masuk, program studi, fakultas, tahun masuk, dan besaran jumlah pembayaran Uang Kuliah Tunggal (UKT). Penetapan total cluster optimal dilaksanakan menggunakan Elbow Method. Hasil implementasi memperlihatkan bahwasanya sebagian besar fakultas terbagi ke dalam dua cluster, sementara terdapat fakultas yang terbagi ke dalam tiga cluster, yang menunjukkan adanya perbedaan karakteristik mahasiswa pada masing-masing fakultas. Evaluasi kualitas clustering menggunakan Silhouette Score menghasilkan nilai paling rendah sejumlah  $0,5441$  serta nilai paling tinggi sejumlah  $0,7427$ , yang menandakan bahwa hasil pengelompokan data berada pada kategori sangat baik karena cluster yang terbentuk dapat dibedakan dengan jelas. Dengan demikian, algoritma K-Means bisa menciptakan pengklasifikasian data mahasiswa yang terstruktur dengan baik dan dapat dimanfaatkan sebagai dasar awal analisis pola mahasiswa oleh PTIPD.

## REFERENSI

- [1] A. Putra Pratama *Et Al.*, "Systematic Literature Review: Penerapan K-Means Clustering," 2025.
- [2] F. Khalish, N. M. Piranti, And O. Martadireja, "Implementasi Data Mining Menggunakan Teknik Clustering Dengan Metode K-Means," 2025. [Online]. Available: [Http://Jiip.Stkipyapisdompu.Ac.Id](http://Jiip.Stkipyapisdompu.Ac.Id)

- [3] M. Suyal And S. Sharma, "A Review On Analysis Of K-Means Clustering Machine Learning Algorithm Based On Unsupervised Learning." *Journal Of Artificial Intelligence And Systems*, Vol. 6, No. 1, Pp. 85–95, 2024, Doi: 10.33969/Ais.2024060106.
- [4] E. U. Oti, M. O. Olusola, F. C. Eze, And S. U. Enogwe, "Comprehensive Review Of K-Means Clustering Algorithms," *International Journal Of Advances In Scientific Research And Engineering*, Vol. 07, No. 08, Pp. 64–69, 2021, Doi: 10.31695/Ijasre.2021.34050.
- [5] S. A. Rahmah And J. Antares, "INFORMATIKA Klasterisasi Seleksi Mahasiswa Calon Penerima Beasiswa Yayasan Menggunakan K-Means Clustering," *Jurnal Informatika, Manajemen Dan Komputer*, Vol. 13, No. 2, 2021.
- [6] H. Nopriandi And F. Haswan, "Analisis Klasterisasi Mahasiswa Baru Dalam Memilih Program Studi Dengan Menggunakan Algoritma K-Means," *Journal Of Information System Research (Josh)*, Vol. 3, No. 4, Pp. 666–671, Jul. 2022, Doi: 10.47065/Josh.V3i4.1986.
- [7] A. R. Lashiyanti, I. Rasyid Munthe, F. A. Nasution, And E. P. Korespondensi, "Optimisasi Klasterisasi Nilai Ujian Nasional Dengan Pendekatan Algoritma K-Means, Elbow, Dan Silhouette," *Jurnal Ilmu Komputer Dan Sistem Informasi (Jikoms)*, Vol. 6, No. 1, Pp. 14–20, 2023.
- [8] N. Nurahman, A. Purwanto, And S. Mulyanto, "Klasterisasi Sekolah Menggunakan Algoritma K-Means Berdasarkan Fasilitas, Pendidik, Dan Tenaga Pendidik," *Matrik : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, Vol. 21, No. 2, Pp. 337–350, Mar. 2022, Doi: 10.30812/Matrik.V21i2.1411.
- [9] A. Khalif *Et Al.*, "Klasterisasi Tingkat Kemiskinan Di Indonesia Menggunakan Algoritma K-Means."
- [10] P. Apriyani, A. R. Dikananda, And I. Ali, "Penerapan Algoritma K-Means Dalam Klasterisasi Kasus Stunting Balita Desa Tegalwangi," *Hello World Jurnal Ilmu Komputer*, Vol. 2, No. 1, Pp. 20–33, Mar. 2023, Doi: 10.56211/HelloWorld.V2i1.230.
- [11] S. R. , Durugkar, R. , N. K. K. , Raja, And S. Dan Kumar, "Data Mining And Machine Learning Applications," *Data Mining And Machine Learning Applications*, Pp. 1–19, 2022.
- [12] X. Shu And Y. Ye, "Knowledge Discovery: Methods From Data Mining And Machine Learning," *Soc Sci Res*, Vol. 110, Feb. 2023, Doi: 10.1016/J.Ssresearch.2022.102817.
- [13] E. Saputri, "Teknik Dan Aplikasi Data Mining Di Indonesia: Tinjauan Literatur Satu Dekade (2015-2024)," *It-Explore: Jurnal Penerapan Teknologi Informasi Dan Komunikasi*, Vol. 4, No. 2, Pp. 138–149, Jun. 2025, Doi: 10.24246/Itexplore.V4i2.2025.Pp138-149.
- [14] S. Al Faridzi *Et Al.*, "Pengolahan Data : Pemahaman Gempa Bumi Di Indonesia Melalui Pendekatan Data Mining," 2025.
- [15] M. R. Nahjan, N. Heryana, And A. Voutama, "Implementasi Rapidminer Dengan Metode Clustering K-Means Untuk Analisa Penjualan Pada Toko Oj Cell," 2023.
- [16] A. Nugraha, O. Nurdiawan, And G. Dwilestari, "Penerapan Data Mining Metode K-Means Clustering Untuk Analisa Penjualan Pada Toko Yana Sport," 2022.
- [17] Firmansyah, M Ferrari, And Aziz, "Peningkatan Kinerja Sistem Rekomendasi Wisata Melalui Penerapan Algoritma Collaborative Filtering Dan K-Nearest Neighbors Dengan Metode Klasterisasi K-Means," *Jati (Jurnal Mahasiswa Teknik Informatika)*, Vol. 8, 2024.
- [18] Sekar Setyaningtyas, B. Indarmawan Nugroho, And Z. Arif, "Tinjauan Pustaka Sistematis: Penerapan Data Mining Teknik Clustering Algoritma K-Means," *Jurnal Teknoif Teknik Informatika Institut Teknologi Padang*, Vol. 10, No. 2, Pp. 52–61, Oct. 2022, Doi: 10.21063/Jtif.2022.V10.2.52-61.
- [19] N. Hendrastuty, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa," *Jurnal Ilmiah Informatika Dan Ilmu Komputer (Jima-Ilkom)*, Vol. 3, No. 1, Pp. 46–56, Mar. 2024, Doi: 10.58602/Jima-Ilkom.V3i1.26.
- [20] A. Putri Riani, A. Voutama, And T. Ridwan, "Volume 6 ; Nomor 1," *Januari*, Pp. 164–172, 2023, [Online]. Available: <https://Ojs.Trigunadharma.Ac.Id/Index.Php/Jsk/Index>
- [21] J. Riyono And C. E. Pujiastuti, "Simulation Of The K-Means Clustering Algorithm With The Elbow Method In Making Clusters Of Provincial Poverty Levels In Indonesia," *Jurnal Matematika Mantik*, Vol. 8, No. 2, Pp. 113–123, Dec. 2022, Doi: 10.15642/Mantik.2022.8.2.113-123.
- [22] A. Rajsya And A. Rachman Manga, "Rancang Bangun Penerapan Metode Elbow Pada K-Means Untuk Clustering Data Persediaan Barang," *Literatur Informatika & Komputer*, Vol. 1, No. 4, Pp. 395–403, 2024, Doi: 10.33096/Linier.V1i4.2539.
- [23] I. Herdiana, M. A. Kamal, Triyani, M. N. Estri, And Renny, "A More Precise Elbow Method For Optimum K-Means Clustering," Feb. 2025, [Online]. Available: <http://Arxiv.Org/Abs/2502.00851>
- [24] B. N. Yulisasih, H. Herman, S. Sunardi, And H. Yuliansyah, "Evaluation Of K-Means Clustering Using Silhouette Score Method On Customer Segmentation," *Ilkom Jurnal Ilmiah*, Vol. 16, No. 3, Pp. 330–342, Dec. 2024, Doi: 10.33096/Ilkom.V16i3.2325.330-342.
- [25] T. Rahmawati, Y. Wilandari, And P. Kartikasari, "Analisis Perbandingan Silhouette Coefficient Dan Metode Elbow Pada Pengelompokan Provinsi Di Indonesia Berdasarkan Indikator Ipm Dengan K-Medoids," *Jurnal Gaussian*, Vol. 13, No. 1, Pp. 13–24, Aug. 2024, Doi: 10.14710/J.Gauss.13.1.13-24.

- [26] I. T. Utami, F. Suryaningrum, And D. Ispriyanti, “K-Means Cluster Count Optimization With Silhouette Index Validation And Davies Bouldin Index (Case Study: Coverage Of Pregnant Women, Childbirth, And Postpartum Health Services In Indonesia In 2020),” *Barekeng*, Vol. 17, No. 2, Pp. 707–716, Jun. 2023, Doi: 10.30598/Barekengvol17iss2pp0707-0716.