

Implementation of K-Means Clustering to Analyze the Factors of Study Completion Delay of Informatics Engineering Students at STMIK Widya Cipta Dharma

*Nur Fadilah Putri Sukma¹⁾, Wahyuni²⁾, Eka Arriyanti³⁾

^{1,2,3)}Teknik Informatika, STMIK Widya Cipta Dharma

Jl. M. Yamin, Gn. Kelua, Kec. Samarinda Ulu, Kota Samarinda, Kalimantan Timur 75123

E-mail: nurfadilah19055@gmail.com

ABSTRACT

Delays in study completion have become a significant issue in higher education as they affect the accreditation and quality of academic programs. At STMIK Widya Cipta Dharma, more than half of the Informatics Engineering students from the 2017–2021 cohorts failed to graduate on time. Therefore, an in-depth analysis is needed to identify the factors influencing this condition. This study aims to analyze the dominant factors and patterns related to delayed graduation using the K-Means Clustering method. The data used include academic factors such as GPA, semester GPA, total credits taken, and repeated courses; as well as non-academic factors such as financial constraints, employment status, organizational involvement, and distance from home to campus. The entire analysis process follows the CRISP-DM methodology and is implemented using the Python programming language on the Jupyter Notebook platform. The results of this study show that students with delayed study completion can be grouped into several clusters with distinct characteristics. On the academic side, the main differences between clusters are determined by a combination of total credits completed and unfinished credits. On the non-academic side, the most significant variations are influenced by employment status, organizational activity, and working hours. These findings confirm that delayed graduation is not solely related to academic progress but is also influenced by extracurricular activities. Therefore, monitoring both academic and non-academic aspects is essential for early detection and effective student guidance.

Keywords: Clustering, Student Graduation Delay, Academic Factors, Non-Academic Factors, CRISP-DM

Implementasi K-Means Clustering untuk Menganalisis Faktor Keterlambatan Penyelesaian Studi Mahasiswa Prodi Teknik Informatika di STMIK Widya Cipta Dharma

ABSTRAK

Keterlambatan penyelesaian studi menjadi salah satu permasalahan signifikan dalam dunia pendidikan tinggi karena berdampak terhadap akreditasi dan mutu program studi. Di STMIK Widya Cipta Dharma, lebih dari separuh mahasiswa Program Studi Teknik Informatika angkatan 2017–2021 tidak berhasil menyelesaikan studi tepat waktu. Oleh karena itu, diperlukan analisis yang mendalam terhadap faktor-faktor yang memengaruhi kondisi tersebut. Penelitian ini bertujuan untuk menganalisis faktor dominan dan pola keterlambatan penyelesaian studi mahasiswa dengan menerapkan metode K-Means Clustering. Data yang digunakan terdiri dari faktor akademik seperti IPK, IPS, jumlah SKS yang telah diambil, serta SKS yang diulang; dan faktor non-akademik seperti kendala keuangan, status pekerjaan, keaktifan organisasi, dan jarak tempat tinggal ke kampus. Seluruh proses analisis mengikuti tahapan CRISP-DM dan diimplementasikan menggunakan bahasa pemrograman Python melalui platform Jupyter Notebook. Hasil dari penelitian ini menunjukkan bahwa mahasiswa yang mengalami keterlambatan studi dapat dikelompokkan ke dalam beberapa cluster dengan karakteristik yang berbeda. Pada sisi akademik, perbedaan antar cluster ditentukan oleh kombinasi total SKS yang telah ditempuh dan SKS yang belum diselesaikan. Sedangkan pada sisi non-akademik, variasi yang paling menonjol dipengaruhi oleh status pekerjaan, keaktifan organisasi, dan jumlah jam kerja. Temuan ini menegaskan bahwa keterlambatan studi tidak hanya disebabkan oleh faktor akademik, tetapi juga aktivitas di luar perkuliahan. Oleh karena itu, pemantauan kedua aspek tersebut penting untuk mendukung upaya deteksi dini dan pembinaan mahasiswa secara berkelanjutan.

Kata Kunci: Clustering, Keterlambatan Kelulusan Mahasiswa, Faktor Akademik, Faktor Non-Akademik, CRISP-DM

1. PENDAHULUAN

Keterlambatan dalam menyelesaikan studi mahasiswa merupakan masalah yang penting dalam pendidikan tinggi, khususnya di program studi Teknik Informatika. Perbedaan kemampuan mahasiswa untuk menyelesaikan studi tepat waktu menjadi salah satu indikator utama dalam menilai keberhasilan dan kelayakan suatu program studi. Hal ini sejalan dengan penilaian akreditasi yang dilakukan oleh Badan Akreditasi Nasional Perguruan Tinggi (BAN-PT), yang menekankan pentingnya kelulusan tepat waktu sebagai ukuran kualitas pendidikan (Badan Akreditasi Nasional Perguruan Tinggi Jakarta, 2019). Kelulusan mahasiswa tepat waktu sangat penting dalam mengukur kualitas program studi dalam sebuah penilaian akreditasi. Berdasarkan data yang diperoleh dari Sistem Informasi Akademik (SIAK) STMIK Widya Cipta Dharma, dilakukan analisis terhadap jumlah mahasiswa baru dan jumlah mahasiswa yang berhasil menyelesaikan studi tepat waktu dalam periode tahun ajaran 2017/2018 hingga 2020/2021. Rangkuman data tersebut disajikan pada tabel 1 berikut:

Tabel 1. Data Kelulusan Mahasiswa Tepat Waktu

Tahun Ajaran Masuk	Jumlah Mahasiswa Baru	Tahun Lulus (Tepat Waktu)	Jumlah Lulus Tepat Waktu	Persentase Kelulusan Tepat Waktu
2017/2018	166	2021/2022	94	56,63%
2018/2019	130	2022/2023	70	53,85%
2019/2020	128	2023/2024	36	28,13%
2020/2021	151	2024/2025	60	39,74%
Total	575 Mahasiswa		260	45,22%

Sumber: Sistem Informasi Akademik (SIAK) STMIK Widya Cipta Dharma

Berdasarkan tabel di atas, terlihat bahwa persentase kelulusan tepat waktu cenderung naik turun antar angkatan. Secara keseluruhan sekitar 45,22% mahasiswa berhasil lulus tepat waktu, sementara 54,78% lainnya belum menyelesaikan studi sesuai masa studi normal.

Kondisi ini menunjukkan perlunya evaluasi dan perbaikan dalam sistem pembelajaran yang ada. Dengan menerapkan teknik *K-Means Clustering*, diharapkan dapat diidentifikasi kelompok-kelompok mahasiswa berdasarkan faktor-faktor yang mempengaruhi keterlambatan mereka. Analisis ini tidak hanya akan memberikan informasi yang bermanfaat bagi program studi Teknik Informatika, tetapi juga dapat membantu merumuskan strategi untuk meningkatkan kualitas pendidikan dan mendorong mahasiswa agar dapat menyelesaikan studi tepat waktu. Oleh karena itu, penelitian ini bertujuan untuk mengimplementasikan *K-Means Clustering* dalam menganalisis faktor keterlambatan penyelesaian studi mahasiswa program studi Teknik Informatika di STMIK Widya Cipta Dharma. Dengan demikian, diharapkan penelitian ini dapat memberikan kontribusi yang signifikan dalam upaya meningkatkan tingkat kelulusan tepat waktu mahasiswa.

Berdasarkan permasalahan yang ada maka dibuat penelitian sebagai skripsi, yaitu “Implementasi *K-Means Clustering* untuk Menganalisis Faktor Keterlambatan Penyelesaian Studi Mahasiswa Prodi Teknik Informatika di STMIK Widya Cipta Dharma.”

2. RUANG LINGKUP

Agar penelitian ini tidak melebar lebih luas dari pembahasan yang dimaksud, maka dibuatlah batasan masalah ini untuk membatasi pada ruang lingkup penelitian sebagai berikut:

1. Penelitian hanya dilakukan pada mahasiswa Program Studi Teknik Informatika STMIK Widya Cipta Dharma.
2. Data yang digunakan mencakup faktor akademik (IPS, IPK, jumlah SKS yang ditempuh, jumlah SKS yang belum diambil, dan jumlah SKS yang diulang) serta faktor non-akademik (keuangan, status pekerjaan, status pernikahan, keaktifan dalam organisasi, dan jarak tempat tinggal ke kampus).
3. Data yang digunakan berasal dari mahasiswa tahun ajaran 2017/2028 sampai dengan 2020/2021 yang telah melewati batas studi normal (lebih dari 8 semester untuk Program S1) dan mahasiswa semester 8 tahun ajaran 2021/2022 yang mengalami kendala akademik.
4. Implementasi sistem hanya dilakukan pada platform *website* Interaktif, tanpa mempertimbangkan *deploy* ke platform desktop atau mobile.

3. BAHAN DAN METODE

3.1 Implementasi

Menurut Syafitri (2021), Implementasi merujuk pada tindakan yang dilakukan untuk mencapai tujuan-tujuan yang telah ditetapkan dalam suatu keputusan. Tindakan ini bertujuan untuk mengubah keputusan-keputusan tersebut menjadi pola-pola operasional dan berusaha mencapai perubahan, baik yang besar maupun kecil, sesuai dengan yang telah diputuskan sebelumnya. Pada dasarnya, implementasi juga merupakan upaya untuk memahami apa yang seharusnya terjadi setelah program dilaksanakan. Secara etimologis, implementasi berasal dari bahasa Inggris yaitu *to implement*. Dalam kamus besar *Webster*, *to implement* (mengimplementasikan) berarti *to provide the means for carrying out* (menyediakan sarana untuk melaksanakan sesuatu) dan *to give practical effect to* (untuk menimbulkan dampak/akibat terhadap sesuatu).

3.2 Clustering K-Means

Menurut Fairuzabadi, dkk (2024), Clustering merupakan metode dalam *unsupervised learning* yang digunakan untuk membagi data ke dalam kelompok atau klaster berdasarkan kemiripan karakteristik. Setiap klaster berisi data yang memiliki sifat atau ciri yang lebih serupa satu sama lain dibandingkan dengan data yang berada di klaster berbeda. Teknik ini sering

dimanfaatkan untuk mengeksplorasi struktur data dan mengungkap pola tersembunyi di dalamnya.

Menurut Fairuzabadi, dkk (2024), *K-Means* merupakan salah satu algoritma unsupervised learning yang paling umum digunakan dalam proses clustering. Teknik clustering sendiri bertujuan untuk mengelompokkan data ke dalam beberapa grup berdasarkan kemiripan karakteristik, sehingga data dalam satu kluster memiliki tingkat kesamaan yang lebih tinggi dibandingkan dengan data di kluster lain. Dalam algoritma *K-Means*, data dibagi ke dalam sejumlah kluster sebanyak k , di mana setiap data akan dipetakan ke kluster dengan centroid terdekat yang merepresentasikan pusat dari kelompok tersebut.

3.3 Analisis

Menurut Darmawati (2023), analisis merupakan suatu proses yang mencakup berbagai aktivitas seperti memecah, membedakan, serta mengklasifikasikan suatu hal agar dapat dikelompokkan kembali berdasarkan kriteria tertentu. Setelah itu, dilakukan pencarian hubungan antar komponen tersebut untuk kemudian ditafsirkan maknanya.

3.4 Jupyter Notebook

Menurut Rantau (2022), Jupyter Notebook adalah aplikasi pengembangan interaktif berbasis web yang dijalankan di browser default. Setiap blok kode dapat dieksekusi secara independen, memberikan kemudahan dan fleksibilitas dalam penggunaannya. Aplikasi ini memungkinkan pengguna untuk menggabungkan berbagai jenis konten dalam satu notebook, sehingga hasil dari kode, visualisasi, persamaan, dan teks biasa dapat diakses dalam satu tempat.

3.5 CRISP-DM

Menurut Rivanthio, dkk (2020), Cross Industry Standard Process for Data Mining (CRISP-DM) digunakan sebagai standard proses data mining sekaligus sebagai metode penelitian. Menurut Daniel T. Larose, Cross-Industry Standard Process for Data Mining (CRISP-DM) yang dikembangkan tahun 1996 oleh analisis dari beberapa industri seperti Daimler Chrysler, SPSS dan NCR. CRISP-DM menyediakan standar proses data mining sebagai strategi pemecahan masalah secara umum dari bisnis atau unit penelitian.

Menurut Alwi (2024), CRISP-DM terdiri dari enam tahap utama, yaitu pemahaman bisnis (*business understanding*), pemahaman data (*data understanding*), persiapan data (*data preparation*), pemodelan (*modeling*), evaluasi (*evaluation*), dan penerapan (*deployment*). Berikut adalah penjelasan di setiap tahapannya.

1. Business Understanding

Tahapan ini bertujuan untuk memperoleh pemahaman yang menyeluruh terhadap kebutuhan pengguna atau klien. Kegiatan yang dilakukan mencakup

penetapan tujuan bisnis, analisis situasi berdasarkan ketersediaan sumber daya, perumusan tujuan dalam pengumpulan data, serta penyusunan rencana kerja proyek secara sistematis.

2. Data Understanding

Tahap selanjutnya adalah pemahaman data, yang mencakup proses identifikasi, pengumpulan, dan analisis dataset yang relevan untuk mendukung tercapainya tujuan proyek. Aktivitas utama dalam fase ini meliputi pengumpulan data awal, deskripsi data, eksplorasi terhadap karakteristik data, serta pemeriksaan terhadap kualitas data yang tersedia.

3. Data Preparation

Tahapan ini sering disebut sebagai *data mining*, yang berfokus pada penyusunan dataset akhir yang akan digunakan dalam proses pemodelan. Kegiatan dalam tahap ini mencakup peningkatan kualitas data agar memenuhi standar dan kebutuhan dari metode pemodelan yang akan diterapkan.

4. Modeling

Pada tahap ini dilakukan pembangunan dan penilaian berbagai model dengan menerapkan beragam metode pemodelan. Empat tugas utama dalam fase ini meliputi pemilihan teknik pemodelan yang tepat, perancangan skenario pengujian, pembuatan model, serta evaluasi terhadap performa model yang dihasilkan.

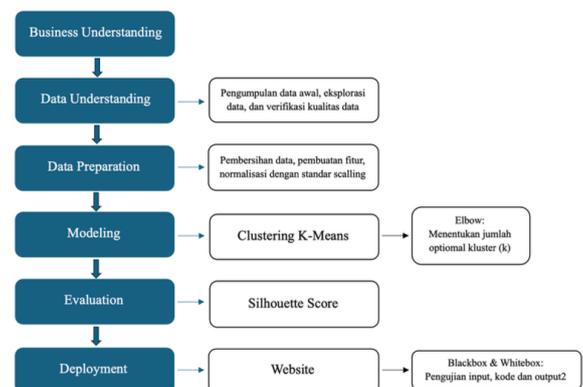
5. Evaluation

Tahap evaluasi bertujuan untuk menentukan apakah model yang dibangun telah sesuai dengan kebutuhan bisnis serta merumuskan tindakan lanjutan yang harus dilakukan. Tiga aktivitas utama dalam fase ini mencakup penilaian terhadap hasil yang diperoleh, pelaksanaan tinjauan terhadap proses keseluruhan, dan penetapan langkah selanjutnya yang perlu diambil.

6. Deployment

Ini merupakan langkah terakhir dan paling penting dalam proses CRISP-DM. Perencanaan untuk implementasi dimulai dari fase pemahaman bisnis dan harus mencakup tidak hanya cara untuk menciptakan nilai dari model, tetapi juga cara untuk mengubah skor keputusan serta mengintegrasikan keputusan tersebut.

4. HASIL DAN PEMBAHASAN



Gambar 1 Alur Penelitian

4.1 Business Understanding

Keterlambatan penyelesaian studi mahasiswa menjadi isu penting yang memengaruhi kualitas pendidikan tinggi, termasuk di Prodi Teknik Informatika STMIK Widya Cipta Dharma. Berdasarkan data SIAK tahun ajaran 2017/2018–2020/2021, tercatat 54,78% mahasiswa tidak lulus tepat waktu. Faktor-faktor penyebab keterlambatan terbagi menjadi dua, yaitu akademik (IPS, IPK, jumlah SKS, SKS belum diambil, dan SKS diulang) serta non-akademik (kondisi keuangan, pekerjaan, status pernikahan, keaktifan organisasi, dan jarak tempat tinggal).

Penelitian ini menerapkan metode *K-Means Clustering* untuk mengelompokkan mahasiswa berdasarkan kemiripan karakteristik akademik dan non-akademik. Tujuannya adalah mengidentifikasi pola umum dan faktor dominan penyebab keterlambatan. Visualisasi hasil clustering disajikan melalui website sederhana yang menampilkan data dalam bentuk tabel, grafik PCA, dan barplot secara interaktif. Proses ini mengacu pada model CRISP-DM, meliputi pemahaman masalah, analisis data, perancangan atribut, hingga visualisasi berbasis web.

4.2 Business Understanding

Data yang digunakan dalam penelitian ini terdiri dari dua jenis, yaitu data akademik dan non-akademik, yang dianalisis secara terpisah karena perbedaan sumber dan struktur. Data akademik diperoleh dari Pusat Komputer (PUSKOM) STMIK Widya Cipta Dharma, berupa rekam nilai mahasiswa Program Studi Teknik Informatika dari tahun ajaran 2014/2015 hingga 2020/2021, dengan total 121.920 baris data dari 2.116 mahasiswa. Data ini mencakup informasi akademik seperti NIM, angkatan, mata kuliah, nilai, dan tanggal kelulusan, yang digunakan untuk menghitung atribut seperti IPK, IPS, total SKS, SKS belum diambil, dan SKS diulang.

Sementara itu, data non-akademik dikumpulkan melalui wawancara, kuesioner, dan observasi terhadap mahasiswa semester akhir dan yang melewati masa studi normal. Sebanyak 77 responden memberikan informasi terkait kondisi non-akademik, seperti kendala keuangan, status pekerjaan, status pernikahan, keaktifan dalam organisasi, jarak tempat tinggal ke kampus, serta moda transportasi. Data ini merepresentasikan faktor-faktor eksternal yang berpotensi memengaruhi keterlambatan studi.

4.3 Data Preparation

Tahap ini bertujuan untuk mempersiapkan data yang akan digunakan dalam proses clustering. Seluruh proses dilakukan secara terpisah antara data akademik dan data non-akademik karena kedua jenis data ini memiliki sumber, struktur, dan format yang berbeda. Pemisahan ini juga diperlukan agar proses pengolahan berjalan lebih terfokus dan sesuai dengan karakteristik masing-masing jenis data.

4.3.1 Data Akademik

Data akademik diolah menggunakan Jupyter Notebook dengan bahasa Python melalui beberapa tahapan, yakni pembersihan data, seleksi atribut, dan normalisasi. Proses pembersihan mengidentifikasi bahwa sebagian besar atribut bersih dari *missing values*, kecuali pada kolom judul skripsi dan tanggal lulus yang wajar karena belum semua mahasiswa menyelesaikan studi.

Data kemudian disaring agar hanya mencakup mahasiswa Program Studi Teknik Informatika angkatan 2017–2020. Selanjutnya dihitung total SKS, IPK, rata-rata IPS, SKS diulang, serta SKS yang belum diambil. Informasi kelulusan ditentukan berdasarkan durasi studi, dan mahasiswa dikategorikan sebagai “Tepat Waktu” atau “Tidak Tepat Waktu”. Data mahasiswa yang tergolong “Tidak Tepat Waktu” menjadi fokus analisis.

```
[18]: # Gabungkan berdasarkan kolom 'nim'
df_gabungan = df_akademik_proses.merge(angkatan_lulus[['nim', 'status_lulus']], on='nim', how='left')

# Filter hanya mahasiswa yang tidak lulus tepat waktu
akademik_prosesnon = df_gabungan[df_gabungan['status_lulus'] == 'Tidak Tepat Waktu'].copy()
akademik_prosesnon

[18]:
```

	nim	sks_total	sks_diulang	ipk	ips	sks_belum_diambil	status_lulus
0	0051b5ff83	136.0	0.0	3.735294	3.738528	8.0	Tidak Tepat Waktu
1	007bfc33a2	149.0	5.0	1.926174	1.932343	0.0	Tidak Tepat Waktu
2	008f2da56f	43.0	0.0	1.651163	1.670552	101.0	Tidak Tepat Waktu
3	00d9efc2a2	146.0	2.0	3.780822	3.783271	0.0	Tidak Tepat Waktu
4	013d0a8b91	152.0	8.0	3.467105	3.466909	0.0	Tidak Tepat Waktu
...
640	fbe37d4015	142.0	0.0	3.295775	3.295238	2.0	Tidak Tepat Waktu
641	fc1dc74900	145.0	1.0	3.048276	3.051186	0.0	Tidak Tepat Waktu
642	fcf9c769a7	146.0	2.0	2.883562	2.895681	0.0	Tidak Tepat Waktu
643	ff5e684c4c	151.0	7.0	3.337748	3.338783	0.0	Tidak Tepat Waktu
644	ffbad04a58	127.0	0.0	3.212598	3.189646	17.0	Tidak Tepat Waktu

446 rows x 7 columns

Gambar 2 Fitur Akademik

Tahap akhir adalah normalisasi fitur akademik menggunakan metode Min-Max Scaling untuk menyamakan skala antar atribut sebelum diterapkan pada model *K-Means Clustering*.

```
[19]: from sklearn.preprocessing import MinMaxScaler

fitur_akademik = akademik_prosesnon[['ipk', 'ips', 'sks_total', 'sks_belum_diambil', 'sks_diulang']]
scaler = MinMaxScaler()
# normalisasi
fitur_akademik_scaled = scaler.fit_transform(fitur_akademik)
# Opsi frame baru: hasil normalisasi
akademik_scaled_df = pd.DataFrame(fitur_akademik_scaled, columns=[f'ccol' for col in fitur_akademik.columns])
# 3. Tambahkan kembali kolom 'nim' agar bisa diambil
akademik_scaled_df[['nim']] = akademik_prosesnon[['nim']].values
akademik_scaled_df

[19]:
```

	ipk	ips	sks_total	sks_belum_diambil	sks_diulang	nim
0	0.953144	0.954100	0.776148	0.057954	0.000000	0051b5ff83
1	0.491507	0.483378	0.852071	0.000000	0.166667	007bfc33a2
2	0.421331	0.420363	0.224892	0.736619	0.000000	008f2da56f
3	0.946781	0.946519	0.833202	0.000000	0.266667	00d9efc2a2
4	0.884719	0.884781	0.868822	0.000000	0.266667	013d0a8b91
...
441	0.840991	0.840989	0.810601	0.014388	0.000000	fbe37d4015
442	0.777838	0.778685	0.828402	0.000000	0.033333	fc1dc74900
443	0.758805	0.738999	0.834320	0.000000	0.066667	fcf9c769a7
444	0.851701	0.852082	0.863905	0.000000	0.233333	ff5e684c4c
445	0.819768	0.814021	0.721893	0.123302	0.000000	ffbad04a58

446 rows x 6 columns

Gambar 3 Normalisasi Akademik

4.3.2 Data Non-Akademik

Sebelum dianalisis menggunakan Jupyter Notebook, data mahasiswa dipra-proses terlebih dahulu di Microsoft Excel, mencakup penghapusan data duplikat, penanganan *missing values*, serta penyaringan hanya pada mahasiswa yang tidak lulus tepat waktu. Selanjutnya, data non-akademik dimapping untuk menyederhanakan kategori jawaban menjadi bentuk yang

lebih seimbang, seperti klasifikasi status kerja, organisasi, transportasi, dan pembiayaan.

Beberapa atribut yang tidak relevan atau tidak memiliki variasi, seperti status pernikahan, dihapus dari dataset. Setelah mapping, data kategori diubah menjadi nilai numerik proporsional (0–1) sesuai tingkat intensitasnya terhadap keterlambatan studi. Tahap akhir berupa normalisasi dengan metode Min-Max Scaling, untuk memastikan kesetaraan skala antar fitur sebelum proses *K-Means Clustering*.

```
[95]:
```

	KKU	Biaya	SKerja	JKerja	Organisasi	JOrganisasi	Jarak	Transportasi
0	0.0	0.0	0.0	0.0	1.0	1.0	0.5	0.0
1	0.5	0.0	0.0	0.0	1.0	1.0	0.5	0.0
2	1.0	0.0	0.0	0.0	0.0	0.0	0.5	0.0
3	1.0	0.0	1.0	0.5	0.0	0.0	1.0	0.0
4	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0
...
67	0.0	1.0	0.0	0.0	1.0	1.0	1.0	0.0
68	0.0	0.0	1.0	0.5	0.0	0.0	0.5	0.0
69	0.5	1.0	1.0	0.5	0.0	0.0	1.0	0.0
70	0.5	0.0	1.0	0.5	0.0	0.0	0.5	0.0
71	1.0	0.0	1.0	0.5	0.0	0.0	1.0	0.0

72 rows x 8 columns

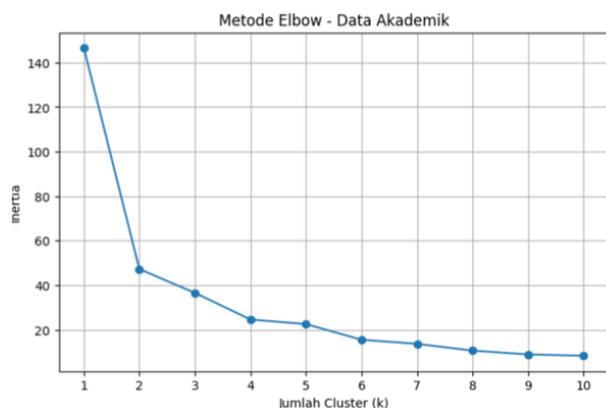
Gambar 4 Normalisasi Non-Akademik

4.4 Modeling

Tahap modeling merupakan inti dari penelitian ini, di mana algoritma *K-Means Clustering* digunakan untuk mengelompokkan mahasiswa berdasarkan karakteristik yang serupa. Mengingat data terbagi menjadi akademik dan non-akademik, pemodelan dilakukan secara terpisah untuk masing-masing jenis data. Proses dimulai dengan penentuan jumlah cluster optimal menggunakan metode Elbow, diikuti penerapan algoritma *K-Means*, serta interpretasi hasil melalui visualisasi dan analisis centroid.

4.4.1 Modeling Data Akademik

Pada tahap ini, data akademik mahasiswa yang telah dinormalisasi digunakan untuk proses clustering dengan algoritma *K-Means*. Tujuannya adalah mengelompokkan mahasiswa berdasarkan karakteristik akademik yang berkaitan dengan keterlambatan studi.



Gambar 5 Elbow Data Akademik

Gambar 5 menentukan jumlah cluster optimal dilakukan menggunakan metode Elbow, yang menunjukkan nilai optimal pada $k = 4$. *K-Means Clustering* kemudian diterapkan pada data akademik menggunakan lima atribut: IPK, IPS, total SKS, SKS belum diambil, dan SKS diulang. Setiap mahasiswa dikelompokkan ke dalam salah satu dari empat cluster berdasarkan kemiripan karakteristik.

```
[281]: from sklearn.cluster import KMeans
# Buang kolom 'nim' untuk proses clustering
data_for_clustering = akademik_scaled_df.drop('nim', axis=1)
# Terapkan K-Means dengan k = 4
kmeans = KMeans(n_clusters=4, random_state=42)
akademik_scaled_df['cluster'] = kmeans.fit_predict(data_for_clustering)
akademik_scaled_df

[281]:
```

	ipk	ips	sks_total	sks_belum_diambil	sks_diulang	nim	cluster
0	0.953144	0.954100	0.775148	0.057554	0.000000	0051b5ff83	2
1	0.491507	0.493378	0.852071	0.000000	0.166667	007b1c33a2	2
2	0.421331	0.426363	0.224852	0.726619	0.000000	008f20a56f	3
3	0.964761	0.965519	0.834320	0.000000	0.066667	00d9efc2a2	2
4	0.884710	0.884781	0.869822	0.000000	0.266667	013d0a8b91	2
...
441	0.840991	0.840969	0.810651	0.014388	0.000000	fbe37dd015	2
442	0.777836	0.778685	0.828402	0.000000	0.033333	fc1dc74600	2
443	0.735805	0.738999	0.834320	0.000000	0.066667	fc9c769a7	2
444	0.851701	0.852082	0.863905	0.000000	0.233333	ff5e684c4c	2
445	0.819766	0.814021	0.721893	0.122302	0.000000	ffb4d04a58	2

446 rows x 7 columns

Gambar 6 K-Means Data Akademik

Gambar 6 menunjukkan hasil penerapan *K-Means* dengan 4 cluster pada data akademik mahasiswa. Pengelompokan dilakukan berdasarkan IPK, IPS, jumlah SKS yang telah diambil, SKS belum diambil, dan SKS yang diulang. Setiap mahasiswa diberi label cluster sesuai kemiripan karakteristik akademiknya.

```
[25]: # Hitung rata-rata tiap kolom untuk masing-masing cluster
average_per_cluster = akademik_scaled_df.groupby('cluster').mean(numeric_only=True)
average_per_cluster

[25]:
```

cluster	ipk	ips	sks_total	sks_belum_diambil	sks_diulang
0	0.743883	0.746408	0.912856	0.000000	0.509091
1	0.636513	0.642539	0.329487	0.599400	0.000000
2	0.831025	0.832235	0.818627	0.016339	0.053968
3	0.215727	0.211805	0.159049	0.806624	0.000000

Gambar 7 Avarage Cluster Data Akademik

Gambar 7 memperlihatkan rata-rata nilai tiap atribut akademik pada masing-masing cluster. Cluster 2 memiliki nilai IPK, IPS, dan SKS tertinggi, sedangkan Cluster 3 menunjukkan karakteristik mahasiswa dengan capaian akademik terendah dan jumlah SKS belum diambil tertinggi.

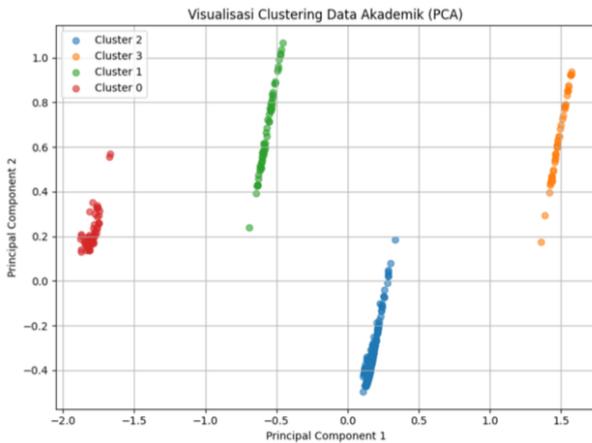
```
[25]: # Hitung jumlah anggota tiap cluster
cluster_counts = akademik_scaled_df['cluster'].value_counts().sort_index()

# Tampilkan hasilnya
for i, count in cluster_counts.items():
    print(f"Cluster {i}: {count} mahasiswa")

Cluster 0: 55 mahasiswa
Cluster 1: 60 mahasiswa
Cluster 2: 273 mahasiswa
Cluster 3: 58 mahasiswa
```

Gambar 8 Jumlah Mahasiswa per Cluster Akademik

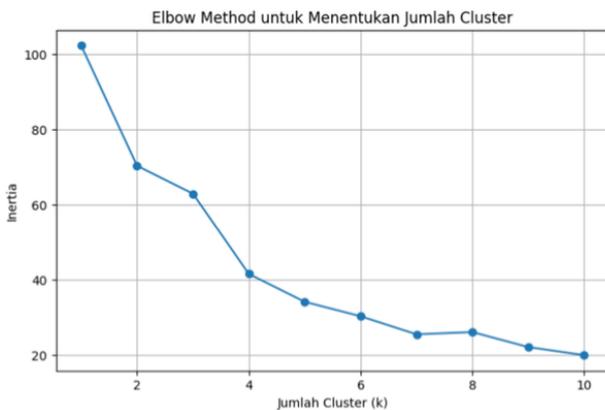
Gambar 8 menunjukkan cluster 2 merupakan kelompok terbesar dengan 273 mahasiswa, sementara cluster lainnya berisi antara 50 hingga 60 mahasiswa.



Gambar 9 Visualisasi PCA Data Akademik

4.4.2 Modeling Data Non-Akademik

Pada tahap ini, data non-akademik mahasiswa yang telah melalui proses transformasi dan normalisasi digunakan untuk melakukan clustering dengan algoritma K-Means.



Gambar 10 Elbow Data Non-Akademik

Gambar 10 memperlihatkan grafik Elbow Method pada data non-akademik untuk menentukan jumlah cluster optimal. Secara visual, titik siku pertama memang tampak berada di $k = 2$, menunjukkan penurunan inertia yang signifikan. Namun, pemilihan jumlah cluster $k = 4$ tetap digunakan untuk memperoleh hasil pengelompokan yang lebih terperinci dan representatif terhadap variasi karakteristik mahasiswa. Dengan $k = 4$, hasil clustering dapat menangkap lebih banyak pola keterlambatan dari sisi non-akademik yang beragam, sehingga lebih bermanfaat dalam analisis lanjutan.

```
[21]: from sklearn.cluster import KMeans
# Inisialisasi dan fit model KMeans
kmeans = KMeans(n_clusters=4, random_state=42, n_init='auto')
kmeans.fit(nonakademik_scaled_df)
# Tambahkan hasil cluster ke DataFrame
nonakademik_scaled_df['cluster'] = kmeans.labels_
nonakademik_scaled_df.head(72)
```

```
[21]:
```

	KKU	Biaya	SKerja	JKerja	Organisasi	JOrganisasi	Jarak	Transportasi	cluster
0	0.0	0.0	0.0	0.0	1.0	1.0	0.5	0.0	3
1	0.5	0.0	0.0	0.0	1.0	1.0	0.5	0.0	3
2	1.0	0.0	0.0	0.0	0.0	0.0	0.5	0.0	2
3	1.0	0.0	1.0	0.5	0.0	0.0	1.0	0.0	1
4	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	2
...
67	0.0	1.0	0.0	0.0	1.0	1.0	1.0	0.0	3
68	0.0	0.0	1.0	0.5	0.0	0.0	0.5	0.0	1
69	0.5	1.0	1.0	0.5	0.0	0.0	1.0	0.0	1
70	0.5	0.0	1.0	0.5	0.0	0.0	0.5	0.0	1
71	1.0	0.0	1.0	0.5	0.0	0.0	1.0	0.0	1

Gambar 11 K-Means Data Non-Akademik

Gambar 11 menunjukkan hasil penerapan algoritma K-Means pada data non-akademik mahasiswa. Pengelompokan dilakukan berdasarkan variabel-variabel seperti kendala keuangan, status kerja, keaktifan organisasi, jarak tempat tinggal, dan moda transportasi. Setiap mahasiswa diklasifikasikan ke dalam salah satu dari empat cluster.

```
[14]: # Hitung rata-rata tiap kolom untuk masing-masing cluster
average_per_cluster = nonakademik_scaled_df.groupby('cluster').mean(numeric_only=True)
average_per_cluster
```

```
[14]:
```

cluster	KKU	Biaya	SKerja	JKerja	Organisasi	JOrganisasi	Jarak	Transportasi
0	0.583333	0.333333	1.0	0.750000	1.0	1.0	0.416667	0.333333
1	0.409091	0.227273	0.0	0.000000	0.0	0.0	0.477273	0.136364
2	0.300000	0.400000	0.0	0.000000	1.0	1.0	0.533333	0.066667
3	0.586207	0.586207	1.0	0.689655	0.0	0.0	0.741379	0.068966

Gambar 12 Average Cluster Data Non-Akademik

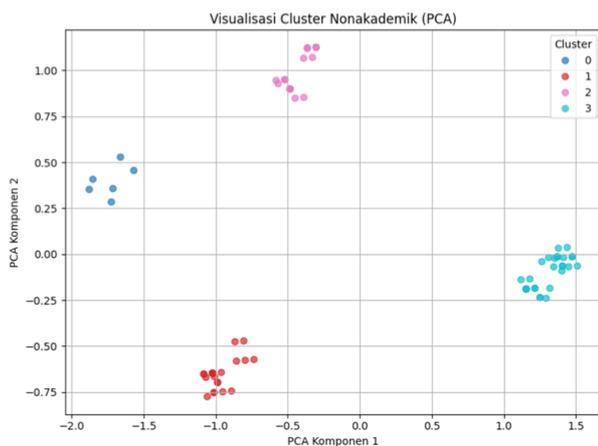
Gambar 12 menyajikan nilai rata-rata setiap atribut non-akademik pada masing-masing cluster. Nilai ini digunakan sebagai representasi pusat (centroid) dari tiap cluster, yang merefleksikan karakteristik umum mahasiswa dalam kelompok tersebut berdasarkan faktor-faktor non-akademik.

```
[24]: # Hitung jumlah anggota tiap cluster
cluster_counts = nonakademik_scaled_df['cluster_custom'].value_counts().sort_index()
# Tampilkan hasilnya
for i, count in cluster_counts.items():
    print(f"Cluster {i}: {count} mahasiswa")
```

```
Cluster 0: 6 mahasiswa
Cluster 1: 22 mahasiswa
Cluster 2: 15 mahasiswa
Cluster 3: 29 mahasiswa
```

Gambar 13 Jumlah Mahasiswa per Cluster NonAkademik

Gambar 13 menunjukkan jumlah anggota pada masing-masing cluster hasil pengelompokan data non-akademik berdasarkan centroid yang telah ditentukan. Cluster 3 memiliki jumlah mahasiswa terbanyak, yaitu 29 orang, sedangkan cluster 0 merupakan yang paling sedikit dengan hanya 6 mahasiswa.



Gambar 14 Visualisasi PCA Non-Akademik

4.4.3 Hasil Clustering Data Akademik

Pemodelan data akademik dengan algoritma *K-Means Clustering* menghasilkan empat cluster yang merepresentasikan karakteristik mahasiswa berdasarkan IPK, IPS, total SKS, SKS belum diambil, dan SKS diulang. Analisis nilai centroid tiap cluster mengungkapkan pola dominan dalam setiap kelompok, yang memberikan gambaran perbedaan capaian akademik dan potensi keterlambatan studi di antara mahasiswa.

Tabel 2 Centroid Data Akademik

Cluster	IPK	IPS	SKS Total	SKS Belum Diambil	SKS Diulang
0	0.743883	0.746408	0.912856	0.000000	0.509099
1	0.636513	0.642539	0.329487	0.599405	0.000000
2	0.831025	0.832235	0.818627	0.016338	0.053968
3	0.215727	0.211805	0.159049	0.806623	0.000000

- Cluster 0 (55 Mahasiswa) Mahasiswa dalam cluster ini memiliki IPK dan IPS tinggi (sekitar 0,74), serta total SKS yang telah diambil juga tinggi (sekitar 0,91). SKS yang belum diambil hampir tidak ada, namun jumlah SKS yang diulang cukup tinggi (sekitar 0,51). Pola ini menunjukkan bahwa mahasiswa telah menempuh hampir seluruh beban studi, namun terhambat oleh banyaknya pengulangan mata kuliah. Keterlambatan dalam penyelesaian studi pada kelompok ini cenderung disebabkan oleh hambatan akademik spesifik, yaitu pengulangan yang memperpanjang durasi studi.
- Cluster 1 (60 Mahasiswa) Mahasiswa dalam cluster ini memiliki capaian IPK dan IPS sedang (sekitar 0,63), namun total SKS yang telah diambil masih rendah (sekitar 0,33), dan SKS yang belum diambil cukup tinggi (sekitar 0,60). Tidak terdapat pengulangan SKS yang berarti. Kondisi ini menggambarkan mahasiswa yang berada pada tahap awal studi, namun sudah mengalami keterlambatan akibat ritme studi yang lambat, seperti pengambilan

SKS yang minim per semester atau kendala non-akademik sejak awal.

- Cluster 2 (273 Mahasiswa) Cluster ini menunjukkan mahasiswa dengan performa akademik paling tinggi. IPK dan IPS berada di kisaran 0,83, SKS yang telah ditempuh tinggi (sekitar 0,82), serta SKS yang belum diambil dan diulang sangat rendah (sekitar 0,016 dan 0,054). Meskipun secara akademik sangat baik, mahasiswa dalam cluster ini tetap termasuk kategori tidak lulus tepat waktu. Hal ini mengindikasikan bahwa keterlambatan pada kelompok ini kemungkinan besar dipengaruhi oleh faktor non- akademik atau kendala lainnya.
- Cluster 3 (58 Mahasiswa) Mahasiswa dalam cluster ini memiliki IPK dan IPS sangat rendah (sekitar 0,21), total SKS rendah (sekitar 0,15), dan SKS belum diambil sangat tinggi (sekitar 0,81). Tidak ada pengulangan SKS yang signifikan. Pola ini menunjukkan tingkat keterlambatan yang parah, bahkan dapat merepresentasikan mahasiswa yang tidak aktif kuliah, mengalami drop-out informal, atau menghadapi kendala berat dalam melanjutkan studi.

Sebagian besar mahasiswa berada di Cluster 2, menunjukkan performa akademik tinggi namun tetap mengalami keterlambatan studi. Hal ini mengindikasikan bahwa keterlambatan tidak selalu disebabkan oleh faktor akademik, melainkan kemungkinan besar oleh faktor non-akademik seperti cuti, pekerjaan, organisasi, atau hambatan administratif. Cluster 0 dan 3 mencerminkan hambatan akademik, seperti pengulangan SKS tinggi (Cluster 0) dan SKS yang masih minim (Cluster 3). Sementara Cluster 1 menggambarkan mahasiswa yang terhambat sejak awal studi, meskipun berada di tahap awal perkuliahan.

Tabel 3 Selisih Centroid Data Akademik

Fitur	Max	Min	Selisih (Dominan)
SKS Belum Diambil	0.806623	0.000000	0.806623
SKS Total	0.912856	0.159049	0.753807
IPS	0.832235	0.211805	0.620430
IPK	0.831025	0.215727	0.615298
SKS Diulang	0.509099	0.000000	0.509099

Berdasarkan hasil analisis centroid pada masing-masing cluster, ditemukan bahwa fitur yang paling membedakan antar kelompok mahasiswa adalah jumlah SKS yang belum diambil dan total SKS yang telah ditempuh. Semakin besar jumlah SKS yang belum diambil dan semakin rendah total SKS yang telah diselesaikan, maka semakin besar kemungkinan mahasiswa tergolong ke dalam kelompok keterlambatan studi. Hal ini mengindikasikan bahwa keterlambatan penyelesaian studi lebih dominan dipengaruhi oleh progres akademik yang belum optimal, dibandingkan sekadar capaian nilai IPK atau IPS. Dengan demikian,

monitoring terhadap beban SKS yang telah dan belum diselesaikan dapat menjadi indikator penting dalam mendeteksi potensi keterlambatan sejak dini.

4.4.4 Hasil Clustering Data Non-Akademik

Selain faktor akademik, keterlambatan studi juga dipengaruhi oleh aspek non-akademik yang tidak selalu tercermin dari nilai akademik. Untuk itu, dilakukan pemodelan *K-Means Clustering* terhadap data non-akademik guna mengidentifikasi pola keterlambatan berdasarkan kondisi personal dan sosial mahasiswa. Variabel yang dianalisis meliputi kendala keuangan, status dan durasi kerja, keterlibatan dalam organisasi, jarak tempat tinggal, serta moda transportasi. Hasil analisis centroid tiap cluster memberikan gambaran pola dominan dan karakteristik umum mahasiswa dalam setiap kelompok.

Tabel 4 Centroid Data Non-Akademik

Cluster	KKU	Biaya	SKerja	JKerja
0	0.583333	0.333333	1.000000	0.750000
1	0.409091	0.227273	0.000000	0.000000
2	0.300000	0.400000	0.000000	0.000000
3	0.586207	0.586207	1.000000	0.689655

Tabel 5 Centroid Data Non-Akademik (Lanjutan)

Cluster	Organisasi	JOrganisasi	Jarak	Transportasi
0	1.000000	1.000000	0.416667	0.333333
1	0.000000	0.000000	0.477273	0.136364
2	1.000000	1.000000	0.533333	0.066667
3	0.000000	0.000000	0.741379	0.068966

- Cluster 0 (6 mahasiswa) Mahasiswa dalam cluster ini menghadapi beban non-akademik yang sangat tinggi. Mereka mengalami kendala keuangan sedang (0,58), sebagian masih dibiayai orang tua (Biaya = 0,33), dan seluruhnya bekerja penuh waktu (SKerja = 1,00) dengan rata-rata jam kerja lebih dari 20 jam per minggu (0,75). Selain itu, mereka juga aktif dalam dua organisasi atau lebih (Organisasi = 1,00) dan menghabiskan lebih dari 10 jam per minggu untuk organisasi (JOrganisasi = 1,00). Rata-rata jarak ke kampus 5 -10 km (Jarak = 0,42), dan transportasi yang digunakan adalah kendaraan pribadi (0,33). Kesimpulan: Mahasiswa ini terbebani oleh pekerjaan, organisasi, dan kondisi ekonomi, yang memengaruhi waktu dan energi untuk menyelesaikan studi tepat waktu.
- Cluster 1 (22 mahasiswa) Kelompok ini memiliki beban non-akademik yang ringan. Kendala keuangan tergolong ringan (0,41) dan mayoritas dibiayai oleh orang tua (Biaya = 0,23). Tidak ada yang bekerja (SKerja = 0,00) maupun aktif organisasi (Organisasi = 0,00). Rata-rata jarak ke kampus 5–10 km (Jarak = 0,48) dan transportasi yang digunakan merupakan kendaraan pribadi

(0,14). Kesimpulan: Mahasiswa dalam cluster ini tidak memiliki hambatan non-akademik yang besar, namun tetap terlambat, kemungkinan karena kurangnya motivasi atau manajemen waktu yang kurang baik. 3. Cluster 2 (15 mahasiswa) Mahasiswa dalam cluster ini memiliki kendala keuangan ringan (0,30), sebagian membiayai sendiri dan sebagian dibiayai orang tua (Biaya = 0,40). Tidak ada yang bekerja (SKerja = 0,00), tapi mereka aktif dalam lebih dari satu organisasi (1,00) dan menghabiskan >10 jam per minggu untuk organisasi (1,00). Jarak ke kampus 5-10 km (0,53) dan transportasi yang digunakan merupakan kendaraan pribadi (0,07). Kesimpulan: Keterlambatan studi di cluster ini kemungkinan besar dipengaruhi oleh tingginya aktivitas organisasi, bukan oleh faktor ekonomi atau pekerjaan.

- Cluster 3 (29 mahasiswa) Kelompok ini memiliki kendala keuangan tinggi (0,59) dan mayoritas membiayai kuliah sendiri (Biaya = 0,59). Seluruh mahasiswa bekerja penuh waktu (1,00) dengan jam kerja antara 10–20 jam per minggu (0,69). Tidak ada yang aktif organisasi. Jarak ke kampus lebih dari 10 km (0,74) dan mereka menggunakan kendaraan pribadi (0,07). Kesimpulan: Keterlambatan studi dalam cluster ini sangat dipengaruhi oleh faktor ekonomi, keharusan bekerja, serta lokasi tempat tinggal yang jauh dari kampus.

Hasil clustering data non-akademik menunjukkan bahwa keterlambatan studi tidak selalu disebabkan oleh beban berat seperti pekerjaan atau masalah keuangan. Mahasiswa dengan beban ringan pun dapat mengalami keterlambatan karena faktor internal, seperti kurang disiplin dan manajemen waktu yang buruk. Sebaliknya, mahasiswa dengan beban tinggi cenderung terlambat karena waktu dan energi terbagi. Temuan ini menegaskan bahwa manajemen waktu merupakan faktor kunci dalam kelulusan tepat waktu, sehingga pembinaan perlu diberikan tidak hanya pada mahasiswa dengan beban berat, tetapi juga pada mereka yang tampak tidak memiliki hambatan nyata.

Tabel 6 Selisih Centroid Data Akademik

Fitur	Max	Min	Selisih (Dominan)
SKerja	1.000000	0.000000	1.000000
Organisasi	1.000000	0.000000	1.000000
JOrganisasi	1.000000	0.000000	1.000000
JKerja	0.750000	0.000000	0.750000
Biaya	0.586207	0.227273	0.358934
Jarak	0.741379	0.416667	0.324713
KKU	0.586207	0.300000	0.286207
Transportasi	0.333333	0.066667	0.266667

Hasil analisis centroid menunjukkan bahwa faktor non-akademik yang paling dominan membedakan antar cluster keterlambatan studi adalah status kerja (SKerja),

keaktifan dalam organisasi (Organisasi dan JOrganisasi), serta jumlah jam kerja (JKerja). Ketiga faktor tersebut memiliki selisih nilai maksimum antar cluster sebesar 1.0 untuk SKerja dan Organisasi, serta 0.75 untuk JKerja, yang mengindikasikan variasi signifikan antar kelompok mahasiswa. Hal ini menegaskan bahwa aktivitas non-akademik seperti bekerja dan aktif di organisasi berperan besar dalam memengaruhi keterlambatan penyelesaian studi. Di sisi lain, faktor seperti transportasi, KJU, dan jarak tempat tinggal memiliki kontribusi lebih rendah, ditunjukkan oleh selisih yang relatif kecil. Dengan demikian, keterlambatan studi mahasiswa tidak hanya disebabkan oleh faktor akademik, tetapi juga sangat dipengaruhi oleh keterlibatan dalam aktivitas di luar perkuliahan.

4.5 Evaluation

Evaluasi dilakukan setelah proses clustering untuk memastikan bahwa pembentukan cluster bersifat bermakna dan tidak hanya secara matematis. Metrik yang digunakan adalah Silhouette Score, yang mengukur seberapa baik data dikelompokkan:

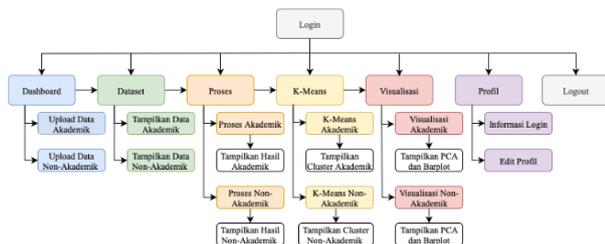
1. Nilai mendekati 1: pemisahan antar cluster sangat baik.
2. Nilai mendekati 0: data berada di batas antar cluster.
3. Nilai negatif: data mungkin berada di cluster yang salah.

Hasil evaluasi:

1. Data akademik: Silhouette Score sebesar 0.8096, menunjukkan pembentukan cluster yang sangat baik.
2. Data non-akademik: Silhouette Score sebesar 0.5318, menunjukkan pemisahan sedang namun tetap bermakna.

Hasil ini mengonfirmasi bahwa pemilihan 4 cluster merupakan jumlah paling optimal untuk kedua jenis data.

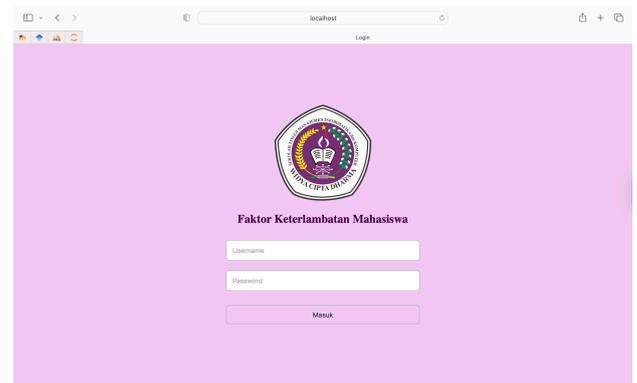
4.6 Deployment



Gambar 15 Sitemap Website

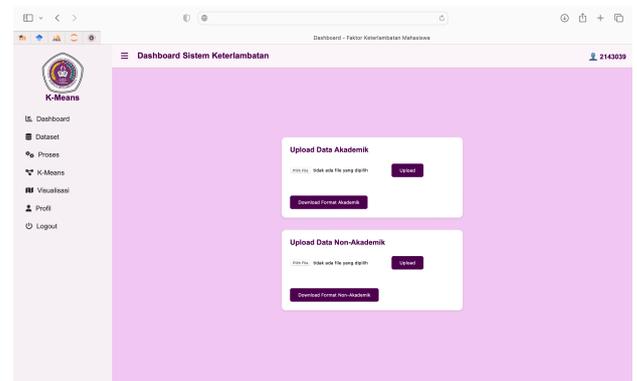
Gambar 15 menampilkan sitemap dari website yang dikembangkan untuk menganalisis keterlambatan studi mahasiswa. Navigasi dimulai dari halaman login sebagai pintu masuk utama. Setelah login, pengguna dapat mengakses enam menu utama: Dashboard, Dataset, Proses, K-Means, Visualisasi, dan Profil, serta menu Logout. Menu Dashboard digunakan untuk mengunggah data akademik dan non-akademik, sementara Dataset

menampilkan data tersebut dalam bentuk tabel. Menu Proses menyediakan fitur preprocessing secara terpisah untuk tiap jenis data. Menu K-Means digunakan untuk melakukan clustering, dan hasilnya divisualisasikan melalui menu Visualisasi dengan pendekatan PCA dan barplot. Menu Profil menampilkan data pengguna dan opsi perubahan profil. Struktur sitemap ini mendukung alur kerja sistem berbasis model CRISP-DM dan memudahkan analisis secara bertahap dan sistematis.



Gambar 16 Halaman Login

Gambar 16 menunjukkan tampilan form login yang digunakan untuk mengakses sistem. Pengguna diminta memasukkan username dan password yang sesuai. Jika data valid, pengguna diarahkan ke dashboard; jika tidak, sistem menampilkan pesan kesalahan. Form ini bertujuan menjaga keamanan akses terhadap website dan data.



Gambar 17 Halaman Dashboard

Gambar 17 menampilkan halaman dashboard utama pada sistem. Halaman ini menyediakan dua fitur utama, yaitu upload data akademik dan upload data non-akademik, yang akan digunakan dalam proses analisis keterlambatan studi. Tampilan dirancang dengan antarmuka sederhana dan warna konsisten untuk memudahkan navigasi pengguna.



nim	angkatan	kidprodi	kdskn	namamatakuliah	skmaks	nilaimutu	nilairahur	bobot	ta	sest	ja
18600309a	2014	57401	3C238	Praktikum Komputer Akuntansi	2.0	80.0	A	4.0	20142015	1	
18600309a	2014	57401	3P410	Praktikum Jaringan Komputer 2	2.0	90.8	A	4.0	20152016	2	
18600309a	2014	57401	3P5211	Praktikum Multi Media Lanjut	2.0	83.0	C	2.0	20162017	1	
18600309a	2014	57401	3W1201	Pendidikan Agama Islam	2.0	73.8	B	3.0	20142015	1	
18600309a	2014	57401	3W1205	Pendidikan Pancasila	2.0	78.0	B	3.0	20142015	1	
18600309a	2014	57401	3W1206	Bahasa Inggris 1	2.0	90.0	A	4.0	20142015	1	
18600309a	2014	57401	3W1207	Matematika Dasar	2.0	82.75	A	4.0	20142015	1	
18600309a	2014	57401	3W1208	Algoritma dan Pemrograman Dasar	2.0	92.45	A	4.0	20142015	1	
18600309a	2014	57401	3W1209	Praktikum Algoritma dan Pemrograman Lanjut	2.0	78.75	B	3.0	20142015	1	

Gambar 18 Halaman Dataset Akademik

Nim	TA	KKKU	Biaya	SKerja	JKerja	SNikah	Organisasi	JOrganisasi	Jarak	Transportasi
2043038	2020	Tidak	Orang Tua	Tidak	Tidak	Belum menikah	Aktif dalam 1 organisasi	Kurang dari 5 jam	5-10 km	Sepeda motor
2043036	2020	Ya, orang tua	Ya, orang tua	Tidak	Tidak	Belum menikah	Belum menikah	Belum menikah	5-10 km	Sepeda motor
2043113	2020	Ya, sedang	Orang tua	Tidak	Tidak	Belum menikah	Tidak aktif	Tidak aktif	5-10 km	Sepeda motor
2043044	2020	Ya, sedang	Orang tua dan Beasiswa	Ya, sedang	Kurang dari 10 jam	Belum menikah	Tidak aktif	Tidak aktif	Lebih dari 10 km	Sepeda motor
2043007	2020	Ya, sedang	Orang tua dan Hasil Kerja sendiri	Tidak	Tidak	Belum menikah	Tidak aktif	Tidak aktif	Kurang dari 5 km	Sepeda motor
2043037	2020	Ya, sedang	Orang tua	Tidak	Tidak	Belum menikah	Tidak aktif	Tidak aktif	Kurang dari 5 km	Sepeda motor
2043024	2020	Ya, sedang	Beasiswa	Tidak	Tidak	Belum menikah	Aktif dalam 1 organisasi	5-10 jam	5-10 km	Sepeda motor
2043041	2020	Ya, sedang	Orang tua	Tidak	Tidak	Belum menikah	Aktif dalam 1 organisasi	5-10 jam	5-10 km	Sepeda motor
2043032	2020	Tidak	Orang tua	Tidak	Tidak	Belum menikah	Tidak aktif	Tidak aktif	5-10 km	Sepeda motor

Gambar 19 Halaman Dataset Non-Akademik

Gambar 18 dan 19 menampilkan tampilan dataset akademik dan non-akademik yang telah diunggah ke dalam sistem. Tampilan ini memudahkan pengguna untuk memverifikasi dan memantau data akademik sebelum dilakukan proses lebih lanjut.

ipk	ips	skstotal	sksbelumdiambil	sksdilang	nm
0.85314402271896	0.95409881478797	0.77514703984628	0.657939986453238	0.0	005104863
0.4915065695540386	0.4533777191334895	0.8520710259171587	0.18666666666666666	0.07011033a2	007011033a2
0.42313184678245	0.426362724061258	0.3248520710059714	0.72616750059713	0.0	00802a6f
0.964761454888939	0.96518505427946	0.834319526272189	0.00666666666666667	0.00666666666666667	00964c2a2
0.864708618874732	0.8847807346587006	0.8682824852071005	0.0	0.26666666666666666	0130a0b91
0.8791213892050071	0.8803408399655596	0.834319526272189	0.0	0.06666666666666667	014ac8aad
0.80450157088122	0.8057903896021075	0.8224852071005971	0.0	0.0	015aa173c
0.851343012794176	0.851796545505458	0.8682824852071005	0.0	0.26666666666666666	0178a5660
0.839738406687395	0.84193596073733	0.824023668639052	0.0	0.03333333333333333	03378c2be
0.8027298850574713	0.8043805411629456	0.8224852071005971	0.0	0.0	03462afef
0.55981688237548	0.5807679383506255	0.6019474566213017	0.258992805753957	0.0	03a35755b
0.127862698655173	0.12786387798836555	0.6887973964970414	0.892096300525218	0.0	040d3a6d0
0.7197275021028073	0.7336130784937825	0.6952248520710058	0.151079136686475	0.0	043cd253e

Gambar 20 Halaman Processed Data Akademik

Nim	TA	KKKU	Biaya	SKerja	JKerja	Organisasi	JOrganisasi	Jarak	Transportasi
2043038	2020	0.0	0.0	0.0	0.0	1.0	1.0	0.5	0.0
2043036	2020	0.5	0.0	0.0	0.0	1.0	1.0	0.5	0.0
2043113	2020	1.0	0.0	0.0	0.0	0.0	0.0	0.5	0.0
2043044	2020	1.0	0.0	1.0	0.5	0.0	0.0	1.0	0.0
2043007	2020	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0
2043037	2020	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2043024	2020	0.5	1.0	0.0	0.0	1.0	1.0	0.5	0.0
2043041	2020	1.0	0.0	0.0	0.0	1.0	1.0	0.5	0.0
2043032	2020	0.0	0.0	0.0	0.0	0.0	0.0	0.5	0.0
2043055	2020	0.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0
1943098	2019	0.5	1.0	1.0	1.0	0.0	0.0	1.0	0.0
1943055	2019	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2043016	2019	1.0	0.0	1.0	1.0	0.0	0.0	0.5	0.0

Gambar 21 Halaman Processed Data Non-Akademik

Gambar 20 dan 21 memperlihatkan hasil preprocessing data dalam bentuk tabel yang telah dinormalisasi ke dalam nilai numerik berskala 0 hingga 1. Proses ini penting untuk mempersiapkan data agar dapat digunakan dalam algoritma clustering secara seimbang.

cluster	ipk	ips	skstotal	sksbelumdiambil	sksdilang
0	0.7439	0.7464	0.9129	0.0000	0.5091
1	0.8365	0.6425	0.3295	0.5994	0.0000
2	0.8310	0.8322	0.8186	0.0163	0.0540
3	0.2157	0.2118	0.1590	0.8068	0.0000

nm	ipk	ips	skstotal	sksbelumdiambil	sksdilang	cluster
005104863	0.8531	0.9541	0.7751	0.6579	0.0000	2
007011033a2	0.4915	0.4534	0.8521	0.1867	0.0000	2
00802a6f	0.4213	0.4264	0.3249	0.7266	0.0000	3
00964c2a2	0.9648	0.9655	0.8343	0.0000	0.0667	2
0130a0b91	0.8847	0.8848	0.8688	0.0000	0.2667	2
014ac8aad	0.8791	0.8803	0.8343	0.0000	0.0667	2
015aa173c	0.8045	0.8058	0.8225	0.0000	0.0000	2
0178a5660	0.8511	0.8518	0.8688	0.0000	0.2667	2
03378c2be	0.8397	0.8419	0.8284	0.0000	0.0333	2

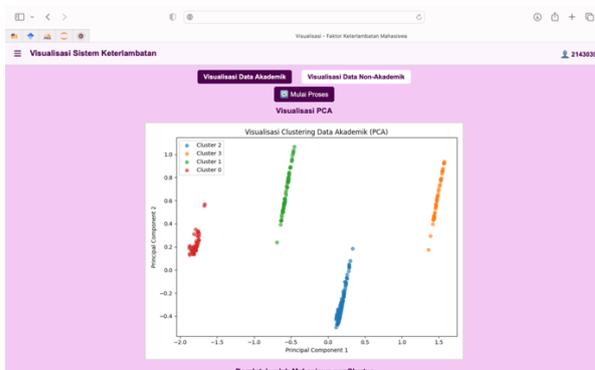
Gambar 22 Halaman K-Means Data Akademik

cluster	KKKU	Biaya	SKerja	JKerja	Organisasi	JOrganisasi	Jarak	Transportasi
0	0.5833	0.3333	1.0000	0.7500	1.0000	1.0000	0.4167	0.3333
1	0.4091	0.2275	0.0000	0.0000	0.0000	0.0000	0.4773	0.1564
2	0.3000	0.4000	0.0000	0.0000	1.0000	1.0000	0.5333	0.0667
3	0.5882	0.5882	1.0000	0.8887	0.0000	0.0000	0.7414	0.0660

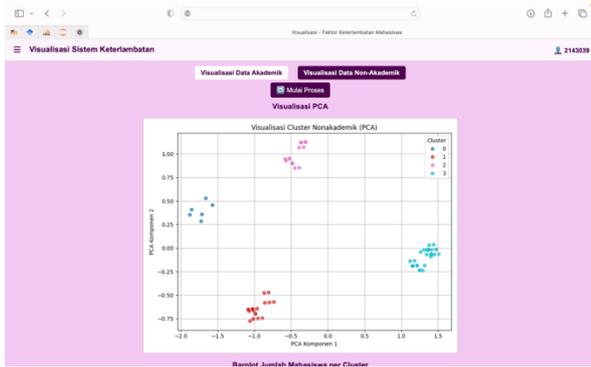
Nim	KKKU	Biaya	SKerja	JKerja	Organisasi	JOrganisasi	Jarak	Transportasi	cluster
2043038	0.0	0.0	0.0	0.0	1.0	1.0	0.5	0.0	2
2043036	0.5	0.0	0.0	0.0	1.0	1.0	0.5	0.0	2
2043113	1.0	0.0	0.0	0.0	0.0	0.0	0.5	0.0	1
2043044	1.0	0.0	1.0	0.5	0.0	0.0	1.0	0.0	3
2043007	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	2
2043037	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1
2043024	0.5	1.0	0.0	0.0	1.0	1.0	0.5	0.0	2
2043041	1.0	0.0	0.0	0.0	1.0	1.0	0.5	0.0	2
2043032	0.0	0.0	0.0	0.0	0.0	0.0	0.5	0.0	1

Gambar 23 Halaman K-Means Data Non-Akademik

Gambar 22 dan 23 menampilkan hasil clustering K-Means terhadap data akademik yang telah dinormalisasi serta ciri khas tiap cluster berdasarkan nilai centroid tiap fitur. Tabel ini berisi data akademik dalam bentuk skala (scaled), seperti IPK, IPS, total SKS, SKS belum diambil, dan SKS diulang, beserta informasi cluster yang ditetapkan untuk setiap mahasiswa.



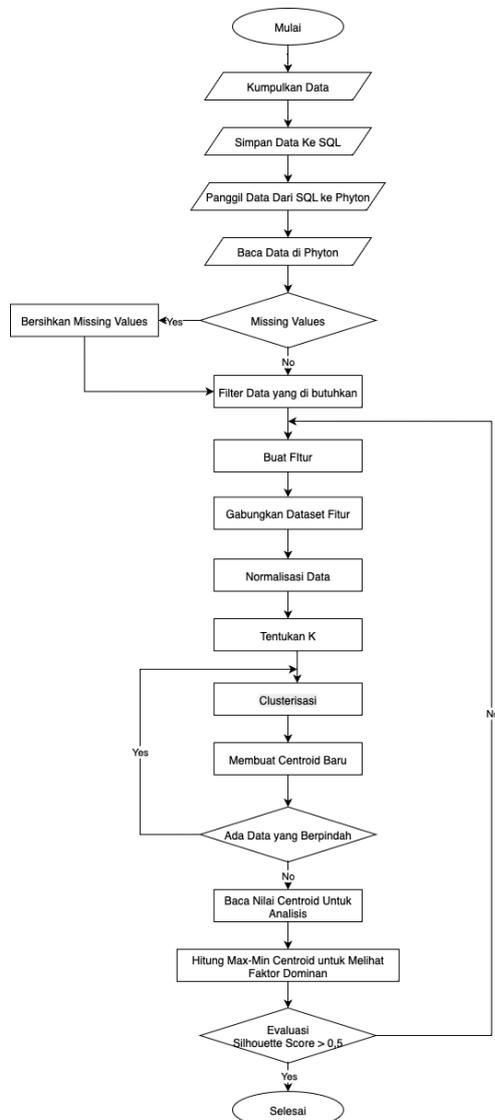
Gambar 24 Halaman Visualisasi Data Akademik



Gambar 25 Halaman Visualisasi Data Non-Akademik

Gambar 24 dan 25 menampilkan visualisasi hasil clustering data non-akademik menggunakan PCA memperlihatkan jumlah mahasiswa per cluster.

4.7 Flowchart Proses Analisis Data



Gambar 26 Flowchart Proses Analisis Data

5. KESIMPULAN

Melalui penerapan metode K-Means Clustering pada data mahasiswa yang mengalami keterlambatan studi, penelitian ini berhasil memetakan pola karakteristik mahasiswa ke dalam beberapa kluster dengan perbedaan yang cukup mencolok antar kelompok. Proses pengelompokan tersebut turut membantu dalam mengidentifikasi faktor dominan yang berperan dalam keterlambatan penyelesaian studi. Adapun kesimpulan dari penelitian ini dapat dirangkum sebagai berikut:

1. Implementasi *K-Means Clustering* dilakukan melalui tahapan sistematis, dimulai dari pengumpulan dan penyimpanan data ke SQL, pemanggilan serta pembacaan data di Python, hingga pembersihan missing values. Data yang telah bersih kemudian difilter, dipilih fitur-fitur yang relevan, dinormalisasi, dan ditentukan nilai K menggunakan metode elbow. Proses clustering dilakukan secara iteratif hingga tidak ada lagi perpindahan data antar kluster. Setelah itu, centroid dianalisis untuk mengidentifikasi faktor dominan berdasarkan nilai maksimum dan minimum. Evaluasi dilakukan menggunakan Silhouette Score, dan apabila hasilnya belum optimal, proses akan kembali pada tahap pengolahan fitur untuk perbaikan. Tahapan ini memastikan akurasi dan efektivitas dalam mengidentifikasi faktor keterlambatan studi mahasiswa.
2. Hasil pengelompokan data akademik menunjukkan bahwa mahasiswa yang mengalami keterlambatan studi memiliki karakteristik yang beragam. Pola-pola tersebut mencakup pengulangan mata kuliah, ritme studi yang cenderung lambat sejak awal, performa akademik yang tinggi namun tetap mengalami keterlambatan karena faktor di luar akademik, serta progres akademik yang masih rendah.
3. Hasil pengelompokan data non-akademik menunjukkan bahwa keterlambatan studi dapat dipengaruhi oleh beragam kondisi, mulai dari beban pekerjaan yang tinggi, keterlibatan aktif dalam organisasi, hingga tanggungan ekonomi pribadi. Di sisi lain, terdapat pula mahasiswa dengan beban non-akademik yang relatif ringan namun tetap mengalami keterlambatan, yang mengindikasikan bahwa faktor seperti pengelolaan waktu dan motivasi pribadi juga berperan penting.
4. Faktor dominan yang memengaruhi keterlambatan studi berasal dari dua sisi utama, yaitu progres akademik yang belum optimal dan tekanan aktivitas non-akademik. Dari sisi akademik, keterlambatan umumnya terjadi ketika mahasiswa belum menyelesaikan sebagian besar beban studinya. Sementara itu, dari sisi non-akademik, keterlambatan sering kali dipengaruhi oleh tanggung jawab di luar kampus, seperti pekerjaan dan keterlibatan organisasi, terutama jika mahasiswa belum sepenuhnya berhasil mengelola waktu secara

seimbang antara tuntutan akademik dan berbagai aktivitas di luar perkuliahan.

5. Hasil Silhouette Score pada data akademik mencapai 0.8096 yang mencerminkan kualitas pengelompokan yang optimal dengan pemisahan antar klaster yang cukup jelas dan kekompakan dalam masing-masing klaster yang tinggi. Sementara itu, nilai Silhouette Score tertinggi yang diperoleh dari data non-akademik adalah sebesar 0,5318. Nilai ini dipengaruhi oleh karakteristik data yang bersifat kategorikal serta jumlah responden yang terbatas, sehingga menghasilkan distribusi klaster yang kurang terpisah secara optimal.

6. SARAN

Berdasarkan kesimpulan dari hasil analisis data dan evaluasi clustering, terdapat beberapa saran yang dapat menjadi pertimbangan untuk pengembangan penelitian selanjutnya, peningkatan sistem pemantauan akademik di perguruan tinggi, serta sebagai acuan bagi mahasiswa dalam memahami dan mengatasi faktor keterlambatan studi. Adapun saran dari penelitian ini dapat dirangkum sebagai berikut:

1. Bagi perguruan tinggi: Memperkuat Bimbingan dan Konseling Mahasiswa karena keterlambatan studi tidak hanya dipengaruhi faktor akademik, tetapi juga tekanan non-akademik seperti pekerjaan dan aktivitas organisasi. Perguruan tinggi perlu meningkatkan layanan bimbingan akademik dan konseling psikologis untuk membantu mahasiswa mengelola waktu, stres, dan motivasi belajar secara lebih efektif.
2. mahasiswa: Mahasiswa diharapkan dapat meningkatkan kemampuan dalam mengelola waktu dan menyeimbangkan antara aktivitas akademik dan non-akademik, karena keterlambatan studi sering kali dipengaruhi oleh tekanan di luar perkuliahan, seperti pekerjaan atau kegiatan organisasi. Selain itu, mahasiswa disarankan untuk lebih aktif berkomunikasi dengan dosen wali sebagai pihak pembimbing akademik, agar dapat memperoleh arahan, dukungan, serta solusi yang tepat terhadap permasalahan studi yang dihadapi. Peran dosen wali sangat penting dalam membantu mahasiswa menjaga motivasi belajar dan menyelesaikan studi tepat waktu.
3. Bagi penelitian selanjutnya: Menambahkan variabel seperti tingkat kehadiran, semangat belajar, kondisi psikologis, serta memperluas jumlah data yang digunakan pada data non-akademik, dapat membantu menghasilkan pengelompokan yang lebih akurat dan merepresentasikan kondisi mahasiswa secara lebih menyeluruh.
4. Sistem yang dikembangkan sebaiknya mampu menerima input data baru dan tersimpan dalam database secara berkelanjutan setiap tahun dan dilengkapi dengan fitur penyebaran serta pengolahan kuesioner secara otomatis untuk data non-akademik,

sehingga proses analisis dapat dilakukan secara efisien dan terintegrasi dalam satu platform.

5. Indeksasi Silhouette Coefficient dari cluster pada penerapan algoritma yang lain mungkin perlu digunakan.
6. Penelitian selanjutnya dengan alur analisis dan metode yang sama wajib menggunakan dataset berbeda dengan tahun berurutan guna meminimalisir bias. contoh: penelitian ini datasetnya 2017-2020, maka penelitian selanjutnya menggunakan dataset tahun 2019-2022.

7. REFERENSI

- Ahmad, S., & Rahma, N. (2022). Analisis faktor penyebab keterlambatan penyelesaian studi pada mahasiswa program studi Informatika. *Jurnal Manajemen Pendidikan*, 8(1), 12–25. <https://journals.ums.ac.id/jmp/article/download/12134/7114>
- Alwi, M. N. (2024, June 19). CRISP-DM: Tahapan, studi kasus, kelebihan, dan kekurangan. Dicoding. <https://www.dicoding.com/blog/crisp-dm-tahapan-studi-kasus-kelebihan-dan-kekurangan>.
- Amrullah, A., Purnamasari, I., Sari, B. N., Garno, & Voutama, A. (2022). Analisis cluster faktor penunjang pendidikan menggunakan algoritma K-Means (Studi kasus: Kabupaten Karawang). *JIRE (Jurnal Informatika & Rekayasa Elektronika)*, 5(2).
- Arriyanti, E., Lailiyah, S., & Kurniawan, A. PENERAPAN METODE K-MEANS CLUSTERING UNTUK PENGELOMPOKAN JUDUL SKRIPSI MAHASISWA (STUDI KASUS: STMIK WIDYA CIPTA DHARMA).
- Badan Akreditasi Nasional Perguruan Tinggi Jakarta. (2019). *Pedoman Penilaian*. Jakarta: Badan Akreditasi Nasional Perguruan Tinggi.
- Brown, M. S. (2014). *Data mining for dummies*.
- Dacwanda, D. O., & Nataliani, Y. (2021). Implementasi k-Means clustering untuk analisis nilai akademik siswa berdasarkan nilai pengetahuan dan keterampilan. *AITI: Jurnal Teknologi Informasi*, 18(2), 125-138.
- Darmawati. (2023). Analisis manajemen pembelajaran pendidikan Pancasila dalam meningkatkan pemahaman nilai-nilai Pancasila pada mahasiswa semester I Prodi Pendidikan Jasmani Unimerz tahun 2022. *Journal of Innovation Research and Knowledge*, 2(10), 3937-3946.

- Dhewayani, F. N., Amelia, D., Alifah, D. N., Sari, B. N., & Jajuli, M. (2022). Implementasi K-Means clustering untuk pengelompokan daerah rawan bencana kebakaran menggunakan model CRISP-DM. *Jurnal Teknologi dan Informasi (JATI)*, 12(1), 64-76.
- Fairuzabadi, M., Adytia, P., Wahyuni, Prastyo, P. H., Resha, M., Anwar, N., Intan, I., Amiruddin, M. R. K., Wulan, N., Sekti, B. A., & Erna, A. (2024). *Machine learning: Konsep, algoritma dan implementasi*. Yayasan Kita Menulis.
- Febriyanti, N. M. D., Sudana, A. A. K. O., & Piarsa, I. N. (2021) Implementasi Black Box Testing pada Sistem Informasi Manajemen Dosen. 2, 6.
- Galela, M. R. (2023). *KMS:: Koefisien Silhouette untuk Menentukan Jumlah Kluster Ideal*. diakses dari Kementerian Keuangan Republik Indonesia. <https://klc2.kemenkeu.go.id/kms/knowledge/koefisien-silhouette-untuk-menentukan-jumlah-kluster-ideal-d0e05a09/detail/>
- Helmud, E. (2021). OPTIMASI BASIS DATA ORACLE MENGGUNAKAN COMPLEX VIEWSTUDI KASUS : PT. BERKAT OPTIMIS SEJAHTERA (PT.BOS) PANGKALPINANG. *Jurnal Informatika*, 80-81.
- Lahema, J. F., Amaliah, T. H., & Usman. (2024). Analisis faktor penyebab keterlambatan penyelesaian studi mahasiswa akuntansi ditinjau dari teori atribusi (Studi kasus mahasiswa Universitas Negeri Gorontalo). *Tangible Jurnal*, 9(2), 275-290.
- Londjo, M. F. (2021). Implementasi white box testing dengan teknik basis path pada pengujian form login. *Jurnal Siliwangi Seri Sains dan Teknologi*, 7(2), 35-40. <https://jurnal.unsil.ac.id/index.php/jssainstek/article/view/4086>
- Maesaroh, S., Afiyati, A., Hakim, L., Sari, Y. S., Yusuf, M., Perkasa, E. B., Utami, W. S., Saptadi, N. T. S., Mutmainah, S., Khairunnas, K., Harahap, E. P., Alamin, Z., Karima, I. S., Saputra, A., & Mubarak, R. (2024). *Bahasa pemrograman Python*. PT Sada Kurnia Pustaka.
- Pamungkas, C. A. (2017). *Pengantar dan Implementasi Basis Data*. Yogyakarta: Deepublish.
- Purba W. Siawin W, Hardih, Nababan M, Dharshinni NP dan Aisyah S. 2018. Implementasi Data Mining untuk Pengelompokan dan Prediksi Karyawan yang Berpotensi PHK dengan Algoritma K-Means Clustering, *Jurnal Sistem Informasi Ilmu Komputer Prima*. Vol: 2, No. 1.
- Putri, N. A., & Hidayat, R. (2022). Analisis faktor penyebab keterlambatan penyelesaian studi mahasiswa Akuntansi Universitas Negeri Gorontalo. *Jurnal Ilmiah Akuntansi*, 2(1), 77-88. <https://ojs.stie-tdn.ac.id/index.php/TB/article/view/547>
- Rahayu, P. T., Daryanto, & A'yun, Q. (2022). Perbandingan Algoritma K-Nearest Neighbor Dan Gaussian Naive Bayes Pada Klasifikasi Penyakit Diabetes Melitus. *Jurnal Smart Teknologi*, 366-373.
- Rahmawati, I., & Suharto, E. (2021). Analisis faktor-faktor keterlambatan mahasiswa tingkat akhir. *Jurnal Pendidikan dan Kebudayaan*, 17(2), 45-56. <https://ejournal.unesa.ac.id/index.php/jupe/article/download/58151/46262>
- Rangan, A. Y., Ukkas, M. I., & Qomariah, S. (2018). Implementasi Metode Support Vector Mechine dan K-Means Clustering dengan Support Vector Mechine Tema Karya Ilmiah pada Stmik Widya Cipta Dharma. *PROSIDING SNITT POLTEKBA*, 3(1), 274-279.
- Rantau, H. (2022). Pembuatan Kode Program Dan Simulasi Skema Masakan ACD Menggunakan Jupyter Notebook Pada Pemograman Python Anaconda. Yogyakarta: Politeknik LPP Yogyakarta.
- Rohman, Y. A. (2020). *Pengenalan NumPy, Pandas, Matplotlib*. Medium.com. <https://medium.com/@yasirabd/pengenalan-numpy-pandas-matplotlib-b90bafd36c0>
- Sahi, Ahmad. 2020. Aplikasi Test Potensi Akademik Seleksi Saringan Masuk LP3I Berbasis Web Online menggunakan Framework Codeigniter, *Jurnal Teknologi Informasi dan Komunikasi*. Vol : 7, No. 1.