Computing and Education Technology Journal (CETJ)
https://ppjp.ulm.ac.id/journals/index.php/cetj/index
Volume 4 Nomor 1 Tahun 2024
Hal 28-41

e-ISSN : 2830-1706
p-ISSN : 2830-1692

Implementasi Algoritma *K-Means* untuk Prediksi Potensi Waktu Kelulusan Mahasiswa Berdasarkan Hasil Akademik Di Jurusan Pendidikan MIPA FKIP ULM

Muhammad Ibnu Alkautsar[1*], Harja Santana Purba[2], Novan Alkaf Bahraini Saputra[3], Nuruddin Wiranda[4], Rizky Pamuji[5]

[1,2,3,4,5] Pendidikan Komputer, Fakultas Keguruan dan Ilmu Pendidikan, Universitas Lambung Mangkurat, Indonesia
E-mail: [1*]ibnualkautsar56@gmail.com, [2]harja.sp@ulm.ac.id, [3]novan.saputra@ulm.ac.id, [4]nuruddin.wd@ulm.ac.id, [5]rizky.pamuji @ulm.ac.id
Email Korespondensi: ibnualkautsar56@gmail.com

***Submitted*: 29-04-2024; *Accepted*: 31-05-2024; *Published*: 31-05-2024**

DOI: https://doi.org/10.20527/cetj.v4i1.12328

Abstrak

Manajemen pendidikan tinggi yang kompleks memerlukan inovasi untuk meningkatkan efisiensi waktu kelulusan mahasiswa. Permasalahan yang ada pada Jurusan Pendidikan MIPA FKIP ULM dalam memahami keterlambatan waktu kelulusan mahasiswa adalah belum adanya sistem dan metode yang efisien dalam memecahkan masalah keterlambatan ini. Sehingga dalam penelitian ini metode machine learning dengan algoritma klastering K-Means diusulkan untuk memecahkan masalah keterlambatan. Jenis data yang digunakan dalam penelitian adalah IPK dan SKS mahasiswa angkatan 2016 sampai 2019 dari semester 1 sampai 5. Dalam penggunaannya data tersebut harus di preprocessing terlebih dahulu agar, bisa diinputkan ke dalam algoritma K-Means. Output yang dihasilkan adalah klaster mahasiswa berdasarkan hasil akademik, dengan label klaster Tepat Waktu Cumlaude, Tepat Waktu Sangat Memuaskan, Tidak Tepat Waktu Sangat Memuaskan, dan Tidak Tepat Waktu Memuaskan. Dalam hasil evaluasi algoritmanya, K-means dinyatakan tidak dapat memecahkan permasalahan keterlambatan waktu kelulusan mahasiswa. Evaluasi dilakukan dengan menggunakan metode Davies-Bouldien Index yang menghasilkan skor sebesar 0,67 (tidak memuaskan). Lalu Adjusted Rand Index sebesar 0,17 (buruk). Namun dengan menggunakan metode Calinski-Harabasz Index, K-means memperoleh skor sebesar 338,28 yang berarti bahwa K-means cocok digunakan untuk klastering data hasil akademik. Hasil penelitian memberikan informasi tentang distibusi jumlah data mahasiswa, tren Indeks Prestasi Kumulatif Mahasiswa, daftar mahasiswa setiap klaster potensi waktu kelulusannya.

Kata Kunci: Evaluasi Performa Klastering, Klastering Potensi Waktu Kelulusan, K-Means, Label Klaster.

Abstract

Complex higher education management requires innovation to improve the efficiency of student graduation time. The existing problem in the Department of Mathematics and Science Education at FKIP ULM in understanding the delay in student graduation time is the lack of an efficient system and method to solve this delay problem. Therefore, in this research, the machine learning method with the K-Means clustering algorithm is proposed to solve the delay problem. The data used in this study are the GPA and credits of students from the 2016 to 2019 cohorts from semester 1 to 5. In its use, the data must be preprocessed first so that it can be input into the K-Means algorithm. The output produced is clusters of students based on academic results, with cluster labels On-Time Cum Laude, On-Time Very Satisfactory, Not On-Time Very Satisfactory, and Not On-Time Satisfactory. In the evaluation of the algorithm, K-means was found unable to solve the problem of student graduation time delay. The evaluation was conducted using the Davies-Bouldin Index method which resulted in a score of 0.67 (unsatisfactory). Then the Adjusted Rand Index was 0.17 (poor). However, using the Calinski-Harabasz Index method, K-means obtained a score of 338.28, which means that K-means is suitable for clustering academic result data. The research results provide information about the distribution of the number of student data, the trend of Student Cumulative Grade Point Average, and the list of students in each cluster with the potential time for graduation.

Keywords: K-means, Label Clusters, Performance Evaluation Clusters, Potential Graduation Time Clusters.

How to cite: Alkautsar, M. I., Purba, H. S., Saputra, N. A. B., Wiranda, N., Pamuji, R. (2024). Implementasi Algoritma K-Means Untuk Prediksi Potensi Waktu Kelulusan Mahasiswa Berdasarkan Hasil Akademik Di Jurusan Pendidikan MIPA FKIP ULM. *Computing and Education Technology Journal (CETJ)*, 4(1), 28-41. doi: https://doi.org/10.20527/cetj.v4i1.12328

© 2024 Computing and Education Tehcnology Journal (CETJ)

1. PENDAHULUAN

Dalam dunia pendidikan, waktu kelulusan mahasiswa menjadi isu penting yang memerlukan perhatian serius karena merupakan indikator keberhasilan sistem pendidikan (Agwil et al., 2020). Tingkat kelulusan yang tidak tepat waktu menandakan adanya masalah dalam proses pembelajaran dan dapat berdampak pada mahasiswa secara individu, lembaga pendidikan, dan masyarakat secara keseluruhan. Masalah ini mencakup biaya tambahan, kesempatan karir yang hilang, dan dampak psikologis seperti kurangnya percaya diri atau frustrasi. Oleh karena itu, perhatian terhadap mahasiswa yang tidak lulus tepat waktu sangatlah penting (Al Husaini et al., 2022).

Di Universitas Lambung Mangkurat, jurusan Pendidikan MIPA FKIP mengalami peningkatan tingkat kelulusan, meskipun masih terdapat mahasiswa yang mengalami kesulitan untuk lulus tepat waktu. Faktor-faktor seperti masalah akademik dan kurangnya motivasi menjadi penyebab utama. Oleh karena itu, diperlukan suatu sistem untuk membantu mengevaluasi dan memprediksi waktu kelulusan mahasiswa berdasarkan hasil akademik (Agwil et al., 2020). Dalam hal ini, teknik clustering dapat digunakan untuk mengelompokkan mahasiswa berdasarkan tingkat kelulusan mereka. Salah satu teknik clustering yang dapat digunakan adalah algoritma K-Means.

Indeks Prestasi Kumulatif (IPK) dan jumlah SKS yang diperoleh mahasiswa dalam penelitian ini merupakan atribut yang digunakan untuk mengelompokkan data ke dalam kelompok atau cluster berdasarkan kesamaan ciri atau atribut. Algoritma K-Means digunakan dalam penelitian ini untuk mengelompokkan siswa ke dalam kelompok berdasarkan potensi waktu kelulusannya. (Primartha, 2018).

Penelitian ini bertujuan untuk mengelompokkan mahasiswa Pendidikan MIPA FKIP ULM berdasarkan potensi waktu kelulusan mereka dengan mempertimbangkan hasil akademik. Data yang digunakan mencakup periode 2016-2019, dari semester 1 hingga 5, di Fakultas Keguruan dan Ilmu Pendidikan ULM.

Beberapa penelitian terkait implementasi algoritma K-Means dalam clustering potensi waktu kelulusan mahasiswa telah dilakukan. Rusdiana & Sam'ani (2016) menggunakan algoritma K-Means untuk mengklasifikasikan data mahasiswa STMIK Palangkaraya berdasarkan jumlah beban studi, IPK, dan tesis kelulusan, dengan tingkat akurasi sebesar 70% (Rusdiana & Sam'ani, 2016). Nur Addini (2021) mengembangkan aplikasi prediksi waktu kelulusan mahasiswa berdasarkan cluster dari parameter IPK dan kehadiran mahasiswa, dengan tujuan membantu pihak kampus dan mahasiswa dalam memprediksi tingkat kelulusan tepat waktu (Nur Addini & Rosian Adhy, 2021). Rahmawati (2019) menggunakan algoritma clustering K-Means untuk mengidentifikasi kemampuan akademik mahasiswa, memberikan wawasan bagi perguruan tinggi dalam mengoptimalkan penjaminan mutu (Rahmawati et al., 2019). Sementara itu, Helilintar & Farida (2018) menggunakan K-Means Clustering di Universitas Negeri Kediri untuk mengelompokkan data berdasarkan predikat kelulusan, memberikan pemahaman lebih baik tentang prestasi akademik mahasiswa untuk pengambilan keputusan terkait kelulusan (Helilintar & Farida, 2018). Lalu ada penelitian yang diangkat oleh Santos (2009) untuk mengevaluasi algoritma K-means secara eksternal, agar mengetahui apakah hasil prediksi sudah tepat dengan data fakta (Santos & Embrechts, 2009). Untuk evaluasi algoritma dirujuk pada penelitian Mughnyanti (2020), dinyatakan bahwa salah satu metode evaluasi internal yang dapat diterapkan adalah Davies-Bouldien Index (Mughnyanti et al., 2020). Sementara itu, penelitian yang diangkat oleh Firman Ashari (2023) menyatakan bahwa ada beberapa metode evaluasi internal yang dapat diterapkan ke algoritma K-means, seperti analisis Elbow, Calinski-Harabasz Index, Davies-Bouldien Index (Firman Ashari et al., 2023).

Berdasakan latar belakang yang telah dipaparkan maka dilakukan penelitian tentang "implemetasi algoritma k-means untuk clustering potensi waktu kelulusan mahasiswa berdasarkan hasil akademik di Jurusan Pendidikan MIPA FKIP ULM".

2. METODE

2.1. Jenis Penelitian

Penelitian ini menerapkan pendekatan campuran yaitu kualitatif dan kuantitatif, dengan jenis penelitian *explanatory sequential design*. Pendekatan campuran (*mixed methods*), atau kombinasi, atau hibrid, dan sejenisnya adalah penelitian yang menggabungkan dua bentuk data secara bersamaan dengan mencampurkan data tersebut, dengan memberikan prioritas pada salah satu atau kedua bentuk data.

Explanatory sequential design menekankan fase kuantitatif, diikuti oleh fase kualitatif (Othman et al., 2020). Tujuan dari fase kualitatif kedua seringkali untuk menjelaskan hasil ditemukan pada fase kuantitatif pertama, dan terkadang untuk menjelaskan *outlier* yang tidak sepenuhnya konsisten dengan data yang dikumpulkan.

2.2. Teknik Pengumpulan Data

Dalam penelitian ini menerapkan teknik pengumpulan data dengan metode studi dokumentasi. Studi dokumentasi diartikan sebagai upaya untuk memperoleh data dan informasi berupa catatan tertulis/gambar yang tersimpan berkaitan dengan masalah yang diteliti. Dokumen merupakan fakta dan data tersimpan dalam berbagai bahan yang berbentuk dokumentasi. Dokumentasi didapatkan dari dokumen yang dikelola oleh operator program studi Jurusan Pendidikan Matematika dan Ilmu Pengetahuan Alam, Fakultas Keguruan dan Ilmu Pendidikan ULM. Dokumentasi yang diperlukan oleh peneliti adalah data identitas dan hasil akademis dari mahasiswa. Dokumentasi tersebut berupa daftar NIM, nama, Angkatan, perolehan IPK tiap semester, perolehan SKS tiap semester, dan masa studi mahasiswa.

2.3. Teknik Analisis

Penelitian ini menggunakan tiga teknik analisis, antara lain teknik penamaan label, analisis statistik deskriptif dan teknik analisis kuantitatif. Berikut penjelasan mengenai teknik analisis penelitian:

1) Teknik Penamaan Label

Dalam menentukan penamaan label klaster potensi waktu kelulusan, kami merujuk pada pedoman akademik ULM tahun 2018. Berdasarkan Pasal 36, mahasiswa diatur untuk menyelesaikan studi dalam 8 semester, dengan rentang minimal 7 semester dan maksimal 12 semester. Pasal 55 memberikan arahan mengenai predikat kelulusan, seperti Memuaskan (IPK 2,76-3,00), Sangat Memuaskan (IPK 3,01-3,50), Pujian (IPK 3,51-4,00) dengan syarat menyelesaikan studi tepat waktu dalam 8 semester atau lebih cepat, serta menyelesaikan semua mata kuliah tanpa perlu mengulang. Mahasiswa yang memperoleh IPK 3,51-4,00, namun tidak menyelesaikan studi tepat waktu, akan mendapat predikat Sangat Memuaskan (TIM, 2018).

2) Analisis Statistik Deskriptif

Statistik deskriptif pada dasarnya merupakan proses transformasi data penelitian dalam bentuk yang lebih mudah dipahami dan diinterpretasikan. Tabulasi menyajikan ringkasan, pengaturan, penyusunan data dalam bentuk numerik dan grafik. Statistik deskriptif umumnya digunakan oleh peneliti untuk memberikan informasi mengenai karakteristik variabel penelitian sekaligus mensupport variabel yang diteliti. Kegiatan yang berhubungan dengan statistik deskriptif seperti menghitung mean (rata-rata), median, modus, mencari standar deviasi dan melihat kemencengan distribusi data dan sebagainya.

3) Teknik Analisis Kuantitatif

Teknik analisis kuantitatif pada penelitian ini dirancang untuk mendalami pemahaman terhadap pola akademik mahasiswa melalui penerapan algoritma clustering K-Means. Berikut adalah penjelasan lebih panjang mengenai setiap langkah dalam analisis kuantitatif ini:

a. Penerapan Algoritma Kmeans

Dalam penerapan data akademik mahasiswa yang terkumpul diproses menggunakan algoritma clustering K-Means. Algoritma ini digunakan untuk mengelompokkan mahasiswa berdasarkan pola akademik yang mungkin tidak terlihat secara langsung. Pemilihan jumlah optimal dari cluster menjadi langkah kritis untuk memastikan hasil clustering yang signifikan.

b. Analisis Cluster

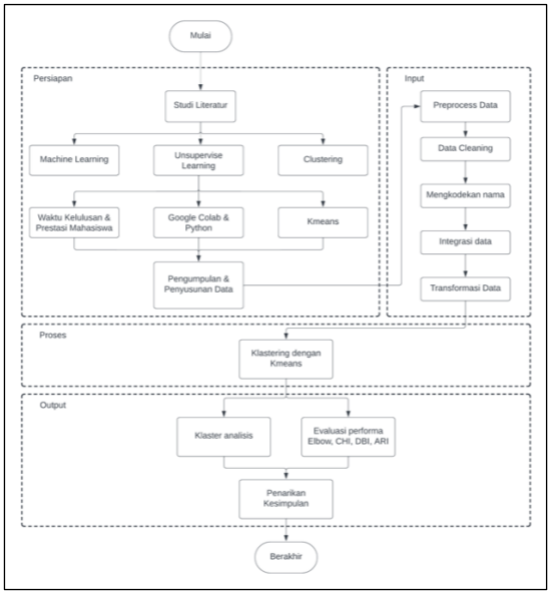
Setelah pembentukan cluster, analisis dilakukan untuk memahami karakteristik masing-masing kelompok. Rata-rata nilai, distribusi SKS, dan parameter akademik lainnya dianalisis secara statistik untuk menggali informasi mendalam.

c. Evaluasi Internal dan Eksternal

Tahap ini mencakup evaluasi hasil clustering. Evaluasi internal melibatkan metode seperti metode elbow, calinski-harabasz, davies-bouldien index untuk mengevaluasi sejauh mana objek dalam satu cluster serupa dan berbeda dari cluster lainnya. Evaluasis eksternal, seperti menggunakan Adjusted Rand Index (ARI), dapat digunakan untuk membandingkan hasil clustering dengan pemetaan referensi yang mungkin ada.

2.4. Teknologi yang Digunakan

Teknologi yang digunakan pada penelitian ini adalah laptop atau komputer, web browser,s google collab, python.



Gambar 1. Alur Penelitian

3. HASIL DAN PEMBAHASAN

Penelitian pengembangan dan data analisis ini menghasilkan sebuah program dan metode analisis dengan berbahasa pemrograman Python pada Google Collaboratory. Hasil penelitian dan pengembangan dijelaskan di bawah ini.

3.1 Pengumpulan dan Penyusunan Data

Data yang menjadi fokus utama penelitian mencakup informasi tentang mahasiswa, termasuk data identitas, hasil akademik, dan data kelulusan.
Data yang diperlukan untuk penelitian ini mencakup tiga aspek utama, yaitu:

- Data Identitas Mahasiswa: Informasi dasar mengenai identitas mahasiswa, seperti nama, NIM (Nomor Induk Mahasiswa), dan informasi pribadi lainnya yang relevan.
- Data Hasil Akademik: Informasi mengenai pencapaian akademik mahasiswa, indeks prestasi, dan data terkait lainnya.
- Data Kelulusan: Informasi mengenai status kelulusan mahasiswa, mencakup tanggal kelulusan dan keterangan lainnya yang relevan.

3.2 Data *Preprocessing*

Proses *preprocessing* diperlukan untuk memastikan kualitas data yang akan digunakan dan kevalidan sebelum menganalisis *clustering* waktu kelulusan potensial siswa berdasarkan hasil akademik. Langkah pertama adalah membersihkan data (Data *Cleaning*) untuk menghilangkan data yang kosong (*missing*) dan redundan (duplikat), karena hal ini dapat mempengaruhi keakuratan analisis *clustering*. Setelah pembersihan data dilakukan, dilanjutkan dengan penkodean identitas mahasiswa untuk menjaga kerahasiaan data. Selanjutnya, dilakukan integrasi data untuk menyatukan dataset dari program studi tiap angkatan menjadi satu dataset yang komprehensif. Terakhir, dilakukan transformasi data seperti standarisasi atau normalisasi untuk menjaga kesamaan skala antar variabel dalam analisis *clustering*. Preprocessing ini penting untuk memastikan kevalidan dan kualitas data, agar hasil analisis klaster dapat memberikan pemahaman yang lebih baik dan membantu dalam pengambilan keputusan.

3.2.1. *Data Cleaning*

Tahap Data *Cleaning* bertujuan mengidentifikasi dan mengatasi data yang kosong (*missing*) atau redundan (duplikat) dalam dataset. Data kosong mengacu pada nilai yang hilang, yang dapat mempengaruhi hasil analisis *clustering*. Langkah pertama adalah mengidentifikasi dan menangani nilai kosong, baik dengan menghapus baris/kolom yang kosong atau menggunakan teknik imputasi. Data redundan, yang muncul lebih dari sekali dalam dataset, juga perlu diatasi karena dapat memengaruhi hasil *clustering* dan menyebabkan bias interpretasi.

Berikut adalah hasil *Preprocessing* data pada setiap angkatan:

Tabel 1. Hasil Data *Cleaning* Setiap Angkatan

Tahun	Jumlah Entri Data		Selisih Data
	Sebelum Data *Cleaning*	Sesudah Data *Cleaning*	
2016	305	286	19
2017	441	371	70
2018	483	420	63
2019	386	318	68

Pada angkatan 2016, terdapat 19 entri data yang dihapus setelah proses pembersihan, mengurangi total entri data dari 305 menjadi 286. Angkatan 2017 mengalami penurunan dari 441 menjadi 371 entri data setelah pembersihan, menandakan penghapusan sejumlah data yang tidak valid atau duplikat. Sementara itu, angkatan 2018 mengalami penurunan dari 483 menjadi 420 entri data setelah pembersihan, dengan 63 entri data yang dihapus. Pada angkatan 2019, terdapat 68 entri data yang dihapus, mengurangi total dari 386 menjadi 318 entri data setelah pembersihan.

3.2.2.Pengkodean identitas Mahasiswa

Identitas mahasiswa dilakukan dengan tujuan utama menjaga kerahasiaan dan privasi data yang sensitif, seperti nama-nama mahasiswa yang terkait. Pada tahap ini, identitas mahasiswa yang terdapat dalam dataset akan diubah menjadi kode atau angka yang hanya dapat diidentifikasi oleh peneliti atau pihak yang berwenang. Dengan adanya pengkodean ini, keamanan data serta anonimitas mahasiswa dapat terjaga dengan baik, sehingga tidak ada kemungkinan data mahasiswa dapat dikaitkan langsung dengan identitas personal mereka. Hal ini memberikan perlindungan yang penting terhadap privasi dan kerahasiaan informasi, serta memastikan integritas dan etika dalam penggunaan data dalam konteks analisis yang dilakukan.

Berikut adalah kode yang digunakan untuk menjaga identitas mahasiswa dalam data:

Tabel 2. Kode Identitas Mahasiswa

No	Program Studi	Tahun Angkatan Mahasiswa MIPA			
		2016	2017	2018	2019
1	Pendidikan Biologi	MHSBIO16001	MHSBIO17001	MHSBIO18001	MHSBIO19001
2	Pendidikan Fisika	MHSFIS16001	MHSFIS17001	MHSFIS18001	MHSFIS19001
3	Pendidikan IPA	MHSIPA16001	MHSIPA17001	MHSIPA18001	MHSIPA19001
4	Pendidikan Kimia	MHSKIM16001	MHSKIM17001	MHSKIM18001	MHSKIM19001
5	Pendidikan Komputer	MHSKOM16001	MHSKOM17001	MHSKOM18001	MHSKOM19001
6	Pendidikan Matematika	MHSMTK16001	MHSMTK17001	MHSMTK18001	MHSMTK19001

3.2.3.*Data Integration*

Langkah berikutnya dalam preprocessing adalah menggabungkan dataset dari berbagai program studi seperti IPA, Ilmu Komputer (Ilkom), Matematika (Matik), Kimia, Biologi, dan Fisika menjadi satu dataset terintegrasi. Proses integrasi data ini sangat penting karena memungkinkan penggabungan informasi dari berbagai program studi menjadi satu dataset yang lebih lengkap dan komprehensif. Dengan data yang terintegrasi, analisis *clustering* atau analisis lainnya dapat dilakukan dengan lebih akurat dan dapat diandalkan.

Berikut hasil data integration:

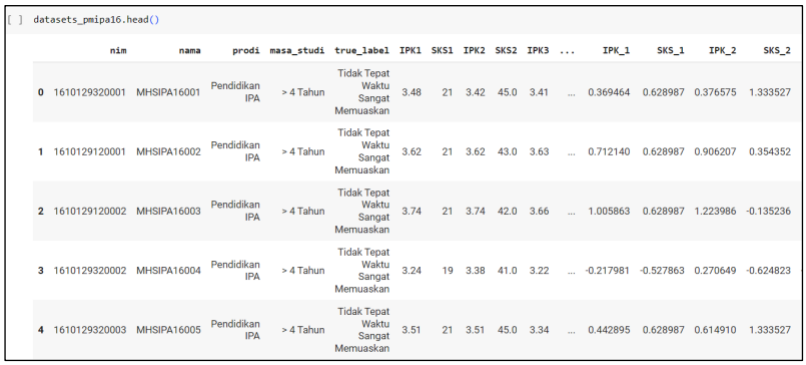
Tabel 3. Hasil Data Integration

No	Tahun	Program Studi Pendidikan MIPA						Jumlah Data
		Biologi	Fisika	IPA	Kimia	Komputer	Matematika	
1	2016	40	46	71	42	44	43	286
2	2017	44	59	81	66	75	46	371
3	2018	60	62	91	71	87	49	420
4	2019	57	27	70	87	49	38	318

Pada tabel di atas, diperlihatkan bahwa datasets tahun 2016 memiliki entri data sebanyak 286 data, lalu pada tahun 2017 memiliki entri data sebanyak 371 data, tahun 2018 memiliki entri data sebanyak 420 data, dan tahun 2019 memiliki entri data sebanyak 318 data, ini membuktikan bahwa pengintegrasian data telah berhasil, mengingat bahwa hasil dari Data *Cleaning* sama banyaknya dengan data yang telah di *Cleaning*.

3.2.4.*Data Transformation*

Setelah tahap integrasi data, langkah berikutnya dalam preprocessing adalah transformasi data. Pada tahap ini, kita mengubah skala atribut atau fitur yang akan digunakan dalam analisis *clustering* untuk memastikan konsistensi. Transformasi data dilakukan scaling pada data menggunakan objek StandardScaler dari library scikit-learn. Scaling ini penting untuk mengatasi perbedaan skala antar atribut dalam dataset. Dengan melakukan scaling, nilai-nilai dalam dataset diubah sehingga memiliki mean nol dan deviasi standar satu. Hasil dari proses scaling adalah dataset yang sudah diubah, dan data ini diubah menjadi DataFrame baru dengan nama kolom yang sama seperti data asli.



```
[ ] datasets_pmipa16.head()
```

	nim	nama	prodi	masa_studi	true_label	IPK1	SKS1	IPK2	SKS2	IPK3	...	IPK_1	SKS_1	IPK_2	SKS_2
0	1610129320001	MHSIPA16001	Pendidikan IPA	> 4 Tahun	Tidak Tepat Waktu Sangat Memuaskan	3.48	21	3.42	45.0	3.41	...	0.369464	0.628987	0.376575	1.333527
1	1610129120001	MHSIPA16002	Pendidikan IPA	> 4 Tahun	Tidak Tepat Waktu Sangat Memuaskan	3.62	21	3.62	43.0	3.63	...	0.712140	0.628987	0.906207	0.354352
2	1610129120002	MHSIPA16003	Pendidikan IPA	> 4 Tahun	Tidak Tepat Waktu Sangat Memuaskan	3.74	21	3.74	42.0	3.66	...	1.005863	0.628987	1.223986	-0.135236
3	1610129320002	MHSIPA16004	Pendidikan IPA	> 4 Tahun	Tidak Tepat Waktu Sangat Memuaskan	3.24	19	3.38	41.0	3.22	...	-0.217981	-0.527863	0.270649	-0.624823
4	1610129320003	MHSIPA16005	Pendidikan IPA	> 4 Tahun	Tidak Tepat Waktu Sangat Memuaskan	3.51	21	3.51	45.0	3.34	...	0.442895	0.628987	0.614910	1.333527

Gambar 2. Hasil *Data Transformation*

3.3 Klasterisasi Dengan *K-Means*

Pada bagian ini, akan dibahas tentang metode klasterisasi dengan menggunakan algoritma *K-means*. Langkah-langkah utama dalam proses klasterisasi dengan Kmeans adalah sebagai berikut:

3.3.1 *Penentuan Jumlah Klaster*

Penentuan jumlah klaster (K) adalah langkah awal yang penting dalam algoritma Kmeans. Hal ini dapat dilakukan dengan menggunakan metode seperti metode siku (*elbow method*) untuk menentukan jumlah klaster yang optimal (Nainggolan et al., 2019).

Berikut adalah hasil SSE menggunakan metode elbow:

a. Angkatan 2016

Tabel 4. Hasil SSE Menggunakan Metode Elbow Angkatan 2016

Semester	K klaster								
	2	3	4	5	6	7	8	9	10
1	292.91	119.48	71.01	54.55	45.00	36.32	30.16	24.47	20.66
2	300.16	182.10	131.91	94.66	74.45	60.23	51.27	44.20	40.34
3	253.30	160.06	116.50	85.16	64.86	55.26	48.75	43.50	39.68
4	236.25	140.64	100.51	82.28	66.86	57.02	47.18	40.87	36.80
5	241.24	151.98	111.53	93.44	76.25	65.98	56.34	47.89	42.82

Berdasarkan nilai SSE yang diperoleh dari perhitungan inertia pada angkatan 2016, jumlah klaster yang optimal tiap semester adalah 4 klaster sebagai jumlah yang optimal dalam pengklasteran dengan rata-rata SSE sebesar 94,35.

b. Angkatan 2017

Tabel 5. Hasil SSE Menggunakan Metode Elbow Angkatan 2017

Semester	K klaster								
	2	3	4	5	6	7	8	9	10
1	416.08	188.58	139.25	112.50	90.28	70.66	53.94	45.68	38.87
2	431.63	288.11	159.47	117.13	92.99	78.00	65.15	55.31	45.72
3	341.11	210.81	162.08	127.19	109.41	94.07	78.97	69.00	59.40
4	322.98	192.61	137.97	109.59	93.47	81.49	71.69	65.63	57.83
5	334.03	186.51	142.25	112.53	95.03	83.16	72.51	66.80	60.14

Berdasarkan nilai SSE yang diperoleh dari perhitungan inertia pada angkatan 2017 jumlah klaster yang optimal tiap semester adalah 4 klaster sebagai jumlah yang optimal dalam pengklasteran dengan rata-rata SSE sebesar 150,44.

c. Angkatan 2018

Tabel 6. Hasil SSE Menggunakan Metode Elbow Angkatan 2018

Semester	K klaster								
	2	3	4	5	6	7	8	9	10
1	433.61	195.02	148.12	124.51	99.28	76.11	60.57	50.44	41.96
2	544.23	355.90	261.64	195.00	153.58	101.32	78.12	64.89	54.49
3	466.18	280.30	187.17	144.17	113.17	95.59	79.90	68.35	58.27
4	394.13	236.58	177.37	148.64	124.90	107.11	90.10	77.36	70.00
5	391.44	227.23	171.42	136.64	110.21	95.58	85.26	75.83	66.95

Berdasarkan nilai SSE yang diperoleh dari perhitungan inertia pada angkatan 2018 jumlah klaster yang optimal tiap semester adalah 4 klaster sebagai jumlah yang optimal dalam pengklasteran dengan rata-rata SSE sebesar 199,4.

d. Angkatan 2019

Tabel 7. Hasil SSE Menggunakan Metode Elbow Angkatan 2019

Semester	K klaster								
	2	3	4	5	6	7	8	9	10
1	317.96	163.74	116.19	95.56	74.62	57.04	43.94	35.80	29.27
2	368.66	250.29	185.26	140.51	109.68	82.95	69.44	58.67	51.09
3	356.81	228.75	162.79	123.42	97.50	82.76	71.61	61.73	52.31
4	350.83	225.27	166.09	135.91	100.98	80.21	67.27	59.50	52.27
5	326.49	200.08	136.53	104.05	83.28	69.08	57.10	49.71	43.54

Berdasarkan nilai SSE yang diperoleh dari perhitungan inertiapada angkatan 2019 jumlah klaster yang optimal tiap semester adalah 4 klaster sebagai jumlah yang optimal dalam pengklasteran dengan rata-rata SSE sebesar 145,4.

Berdasarkan hasil penentuan jumlah klaster yang optimal dengan menggunakan elbow method, maka jumlah klaster optimal yang dipilih sebanyak 4 klaster dengan total rata-rata SSE sebesar 147,39.

3.3.2. Penamaan Label *Klastering*

Pada tahap Penamaan label untuk klaster, peneliti mempertimbangkan beberapa hal. Pertama, dengan menggunakan data IPK dan SKS untuk klastering, nama label harus mencerminkan ketepatan waktunya dan predikat kelulusannya. Kedua, mengacu pada Permendikbud No. 03 Tahun 2020, program sarjana memiliki batas studi maksimal 7 tahun dengan 144 SKS (Kemendikbud, 2014). Oleh karena itu, jumlah SKS mahasiswa menjadi indikator ketepatan waktu kelulusan. Ketiga, panduan akademik ULM tahun 2018 memberikan arahan mengenai predikat kelulusan, yang terdiri dari tiga predikat: Memuaskan, Sangat Memuaskan, dan Pujian (TIM, 2018). Dengan demikian, nama label yang dihasilkan adalah: Tepat Waktu Cumlaude (TWC), Tepat Waktu Sangat Memuaskan (TWSM), Tidak Tepat Waktu Sangat Memuaskan (TTWSM), dan Tidak Tepat Waktu Memuaskan (TTWM).

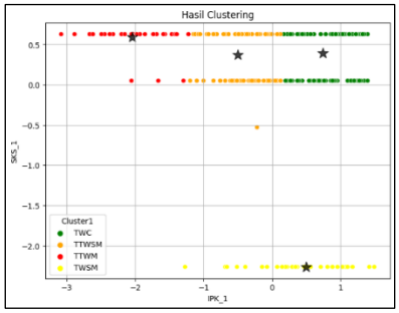
3.3.3. Klasterisasi

Setelah jumlah klaster ditentukan, algoritma *K-means* akan membagi data ke dalam K klaster yang sesuai. Setiap data akan diklasterkan ke dalam kelompok yang memiliki pusat klaster (*centroid*) terdekat . Proses ini akan menghasilkan klaster yang mengelompokkan data berdasarkan kesamaan fitur atau atributnya.

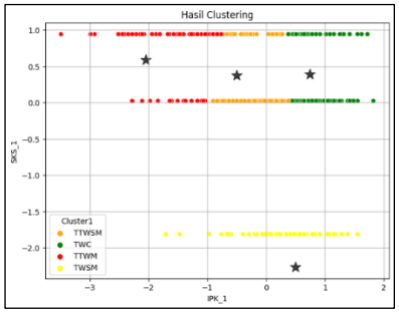
Dalam proses klasterisasi ini, prosesnya akan dibagi menjadi dua yaitu klasterisasi tanpa training untuk data tahun 2016 S.d. 2018 dan klasterisasi dengan *training* untuk data tahun 2019. Hal ini bertujuan untuk meningkatkan efektivitas prediksi untuk *machine learning*, yang dimana penggunaan data terdahulu sebagai *data training* untuk meningkatkan kualitasi prediksi *data test.*

Berikut adalah hasil dari klasterisasi algoritma *k-means* tanpa *training method* dan menggunakan *training method*:

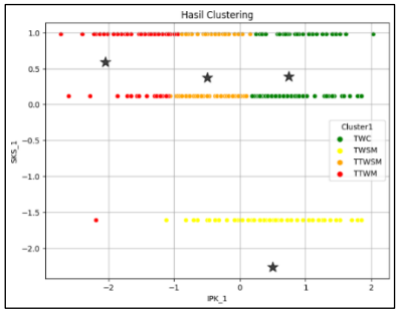
1. Hasil *clustering* pada semester 1



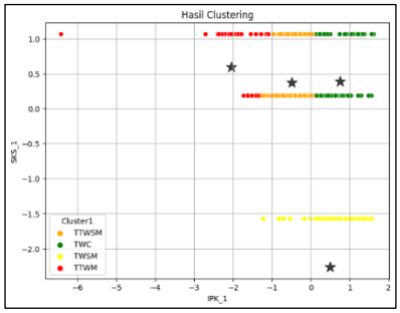
Gambar 3.A - Tahun 2016 Semester 1



Gambar 3.B - Tahun 2017 Semester 1



Gambar 3.C - Tahun 2018 Semester 1

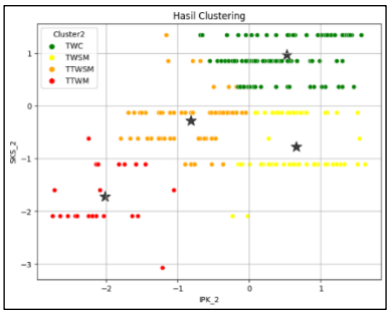


Gambar 3.D - Tahun 2019 Semester 1

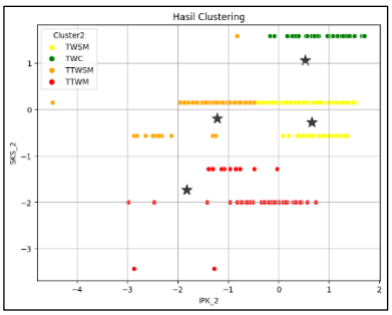
Gambar 3. Scatter Plot Label Hasil Clustering Pada Semester 1

Pada Gambar 6 tentang peningkatan jumlah mahasiswa jurusan PMIPA FKIP ULM dari tahun ke tahun, dengan fluktuasi pada tahun 2019. Potensi waktu kelulusan mahasiswa terlihat dari jumlah data antar label, dengan label Tidak Tepat Waktu Memuaskan memiliki rata-rata 56,75 mahasiswa dan label Tepat Waktu Cumlaude sebesar 116 mahasiswa. Visualisasi scatter plot menunjukkan tidak ada hubungan linier yang signifikan antara IPK dan SKS, disebabkan oleh sistem paketan mata kuliah.

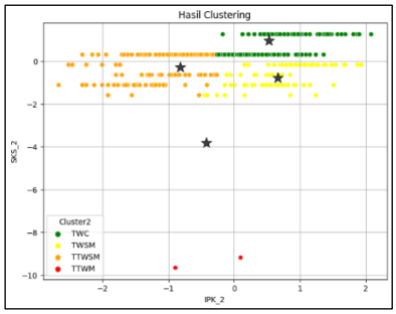
2. Hasil *clustering* pada semester 2



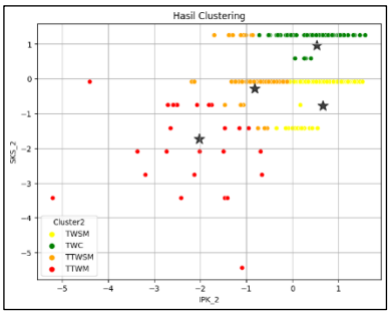
Gambar 4.A - Tahun 2016 Semester 2



Gambar 4.B - Tahun 2017 Semester 2



Gambar 4.C - Tahun 2018 Semester 2

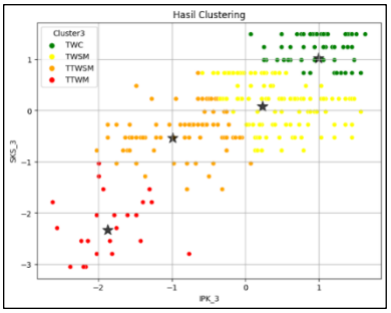


Gambar 4.D - Tahun 2019 Semester 2

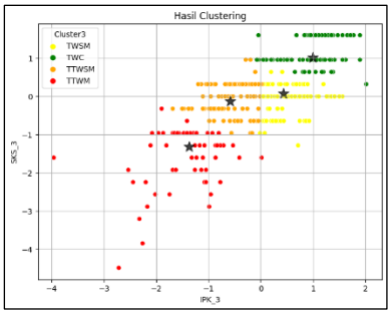
Gambar 4. Scatter Plot Label Hasil Clustering Pada Semester 2

Pada Gambar 7 potensi waktu kelulusan mahasiswa tercermin dari jumlah data antar label. Terjadi perubahan SKS dan IPK, yang mempengaruhi pergeseran jumlah label. Visualisasi scatter plot menunjukkan adanya pola tren dengan variasi, mengindikasikan dinamika kemampuan perkuliahan mahasiswa.

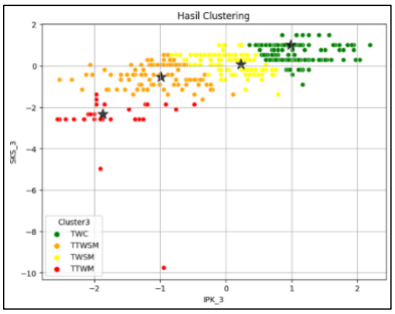
3. Hasil *clustering* pada semester 3



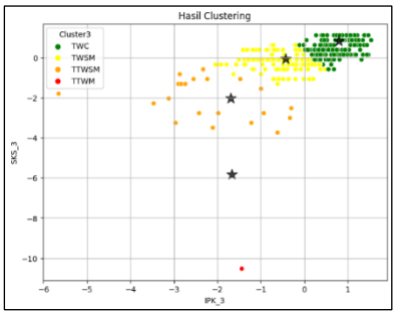
Gambar 5.A - Tahun 2016 Semester 3



Gambar 5.B - Tahun 2017 Semester 3



Gambar 5.C - Tahun 2018 Semester 3

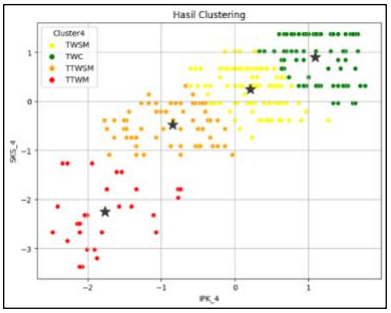


Gambar 5.D - Tahun 2019 Semester 3

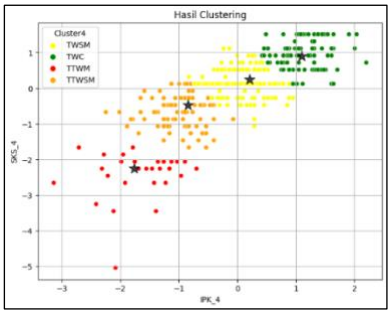
Gambar 5. Scatter Plot Label Hasil Clustering Pada Semester 3

Pada Gambar 8 potensi waktu kelulusan mahasiswa dipengaruhi oleh dinamika SKS dan IPK. Visualisasi scatter plot menunjukkan pola tren, menunjukkan variasi dan hubungan antar IPK dan SKS, serta dinamika perkuliahan mahasiswa.

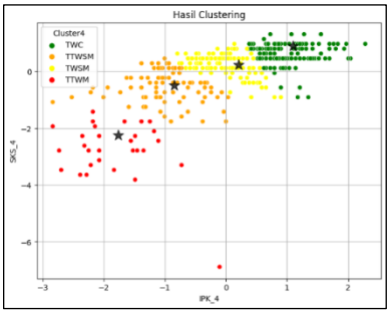
4. Hasil *clustering* pada semester 4



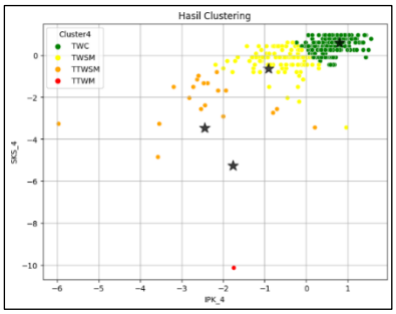
Gambar 6.A - Tahun 2016 Semester 4



Gambar 6.B - Tahun 2017 Semester 4



Gambar 6.C - Tahun 2018 Semester 4

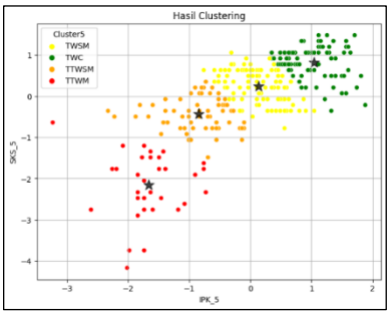


Gambar 6.D - Tahun 2019 Semester 4

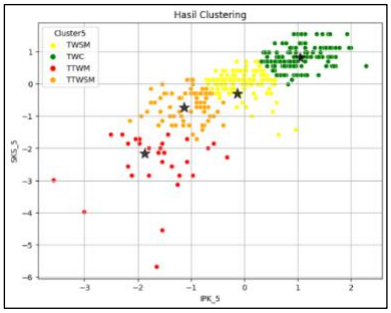
Gambar 6. Scatter Plot Label Hasil Clustering Pada Semester 4

Pada Gambar 9 terjadi perubahan jumlah label yang dapat dijelaskan oleh perubahan rentang SKS. Visualisasi scatter plot menunjukkan pola tren dengan variasi, menunjukkan adanya dinamika perkuliahan mahasiswa.

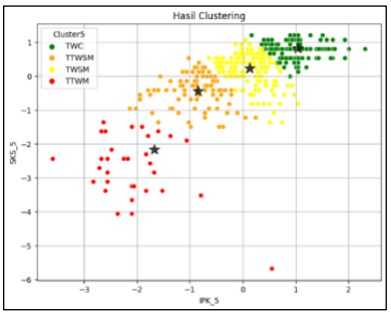
5. Hasil *clustering* pada semester 5



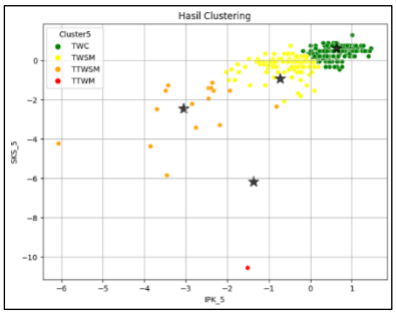
Gambar 7.A - Tahun 2016 Semester 5



Gambar 7.B - Tahun 2017 Semester 5



Gambar 7.C -Tahun 2018 Semester 5



Gambar 7.D - Tahun 2019 Semester 5

Gambar 7. Scatter Plot Label Hasil Clustering Pada Semester 5

Pada semester 10, terlihat adanya variasi dalam potensi waktu kelulusan mahasiswa. Terjadi fluktuasi SKS dan IPK, yang mempengaruhi perubahan jumlah label. Visualisasi scatter plot menunjukkan pola tren dengan variasi, menunjukkan dinamika perkuliahan mahasiswa.

3.4. Evaluasi Performa Klastering Potensi Waktu Kelulusan Mahasiswa

1. Davies Bouldin Index

Tabel 11. Skor Evaluasi Davies-Bouldien Index

Tahun	Semester					Rata-Rata DBI
	1	2	3	4	5	
2016	0.59	0.74	0.82	0.83	0.87	0.77
2017	0.82	0.59	0.83	0.76	0.78	0.75
2018	0.81	0.72	0.89	0.83	0.84	0.81
2019	0.75	0.79	0.64	0.61	0.56	0.67

Analisis DBI pada dataset menunjukkan variasi performa klasterisasi selama empat tahun terakhir. Pada 2016, DBI mencapai puncak pada Semester 5 (0.87), menandakan kemungkinan kurangnya optimalitas klaster. Namun, pada 2017, nilai terendah terjadi pada Semester 2 (0.59), menunjukkan efektivitas pembentukan klaster. Tahun 2018 menunjukkan variasi kecil, dengan nilai tertinggi pada Semester 3 (0.89). Pada 2019, terjadi peningkatan signifikan pada Semester 5 (0.56), menandakan peningkatan dalam pembentukan klaster.

Berdasarkan skala dari Davies-Bouldien Index yang dinyatakan dalam tabel (Muningsih et al., 2021) berikut:

Tabel 12. Tabel Skala Nilai Davies-Bouldien

Nilai	Keterangan
0	Klaster terpisah dengan baik dan kompak
1	Klaster tidak terpisah dengan baik dan kompak

Pada hasil DBI data uji angkatan 2019 menghasilkan rata-rata sebesar 0,67 atau dalam persentase sebesar 33%. Hasil ini dapat dinyatakan tidak memuaskan karena dataset yang terdistribusi secara acak tersebut, tidak terpartisi atau terpisah ke dalam klaster alaminya (Fu et al., 1977).

2. Calinski Harabasz Index

Tabel 13. Skor Evaluasi Calinski-Harabasz Index

Tahun	Semester					Rata-rata CHI
	1	2	3	4	5	
2016	664.12	313.60	366.71	440.93	387.99	434.67
2017	529.72	446.85	437.61	535.54	515.76	493.10
2018	645.86	344.78	388.30	516.02	541.45	487.28
2019	468.08	254.12	304.23	292.48	372.47	338.28

Pada tahun 2016, terjadi variasi signifikan dalam skor CHI. Semester 1 mencapai nilai tertinggi (664.12), menandakan perbedaan yang signifikan antar klaster. Sementara Semester 2 memiliki nilai terendah (313.60), menunjukkan homogenitas yang lebih tinggi dalam klastering pada periode tersebut. Tahun 2017, Semester 4 mencatat nilai CHI tertinggi (535.54), menunjukkan perbedaan yang tinggi antar klaster. Semester 1 dan 5 juga memiliki skor yang cukup tinggi, menunjukkan variasi yang signifikan dalam klastering. Tahun 2018 menunjukkan fluktuasi kecil dalam skor CHI. Semester 1 dan 5 mencapai nilai tertinggi (645.86 dan 541.45), menunjukkan perbedaan yang signifikan antar klaster. Pada tahun 2019, Semester 1 mencapai nilai CHI tertinggi (468.08), menunjukkan variasi yang signifikan dalam klastering. Meskipun Semester 5 memiliki nilai lebih rendah, tetapi masih menunjukkan perbedaan yang cukup besar antar klaster. Pada data uji tahun 2019, CHI memiliki nilai rata-rata sebesar 338,28, yang dapat dikategorikan sebagai baik dalam klastering, karena semakin tinggi nilai CHI, semakin cocok partisi untuk dataset (Saitta et al., 2008).

3. Adjusted Rand Index Score

Tabel 14. Skor Evaluasi Adjusted Rand Index

Tahun	Semester					Rata-rata Skor ARI
	1	2	3	4	5	
2016	0.25	0.17	0.14	0.14	0.15	0.17
2017	0.18	0.07	0.09	0.12	0.14	0.12
2018	0.13	0.11	0.22	0.21	0.3	0.2

Pada tabel ARI, ditampilkan nilai-nilai evaluasi pada masing-masing kombinasi Semester dan Tahun. Sejumlah nilai mencapai puncak pada Semester 5 tahun 2018 dengan ARI sebesar 0.3, menunjukkan tingkat kesamaan yang lebih tinggi antara dua metode *clustering* pada periode tersebut.

Dalam konsep evaluasi menggunakan ARI, skor ARI memiliki skala dari negatif satu (-1) S.d. positif satu (+1). Skor ARI dinyatakan dalam (Osamor & Osamor, 2020) tabel berikut:

Tabel 15. Tabel Skala Nilai *Adjusted Rand Index*

Nilai	Keterangan
≥0.90	Baik Sekali
≥0.80	Baik
≥0.65	Sedang
< 0.65	Buruk

Jadi dapat disimpulkan bahwa nilai-nilai ARI yang didapat dalam rata-rata setiap tahun berada di kisaran 0.1 S.d. 0.2 atau berkisar 50% S.d. 60%, mengindikasikan bahwa kualitas *clustering* yang dihasilkan adalah buruk, sehingga masih membutuhkan banyak ruang untuk perbaikan dalam metode klasteringnya atau data yang diamati.

3.5. Pembahasan

Hasil penelitian menunjukkan bahwa penggunaan metode *machine learning* dengan algoritma klastering K-Means untuk memprediksi waktu kelulusan mahasiswa dari Jurusan Pendidikan MIPA FKIP ULM belum optimal. Data yang digunakan meliputi Indeks Prestasi Kumulatif (IPK) dan jumlah SKS mahasiswa dari angkatan 2016 hingga 2019 pada semester 1 hingga 5, yang kemudian dipreprocessing untuk mempersiapkannya bagi algoritma K-Means. Output dari pengolahan data ini adalah klaster mahasiswa yang diberi label seperti Tepat Waktu Cumlaude, Tepat Waktu Sangat Memuaskan, Tidak Tepat Waktu Sangat Memuaskan, dan Tidak Tepat Waktu Memuaskan.

Evaluasi terhadap algoritma K-Means menggunakan metode *Davies-Bouldien Index* dan *Adjusted Rand Index* menunjukkan kinerja yang tidak memuaskan, dengan skor 0,67 dan 0,17 secara berturut-turut. Namun, evaluasi dengan menggunakan *Calinski-Harabasz Index* memberikan skor sebesar 338,28, menunjukkan potensi penggunaan K-Means dalam klastering data hasil akademik secara umum.

Menggarisbawahi bahwa K-Means belum optimal dalam memprediksi waktu kelulusan mahasiswa, mungkin karena kompleksitas dan keragaman data mahasiswa serta karakteristik masalah yang dihadapi. Diperlukan pengembangan metode atau penggunaan algoritma klastering lain yang lebih sesuai dengan karakteristik data dan masalah yang dihadapi. Meskipun demikian, hasil *skor Calinski-Harabasz Index* menunjukkan potensi penggunaan K-Means untuk klastering data hasil akademik secara umum, yang dapat menjadi dasar untuk pengembangan lebih lanjut atau aplikasi pada konteks lain dalam dunia akademik.

4. KESIMPULAN

Penelitian ini berhasil menerapkan analisis klastering menggunakan algoritma K-Means clustering dengan Python pada data hasil akademik untuk mengetahui potensi waktu kelulusan mahasiswa di Jurusan Pendidikan MIPA FKIP ULM. Penelitian ini menghasilkan 4 klaster yaitu Tepat Waktu Cumlaude, Tepat Waktu Sangat Memuaskan, Tidak Tepat Waktu Sangat Memuaskan, dan Tidak Tepat Waktu Memuaskan. Evaluasi internal menunjukkan hasil yang baik, dengan metode elbow mencapai nilai rata-rata 147,9 untuk 4 label klaster, dan calinski-harbaz index mencapai nilai rata-rata 338,28 untuk data uji. Dalam hasil evaluasi algoritmanya, K-means dinyatakan tidak dapat memecahkan permasalahan keterlambatan waktu kelulusan mahasiswa. Evaluasi dilakukan dengan menggunakan metode Davies-Bouldien Index yang menghasilkan skor sebesar 0,67 (tidak memuaskan). Lalu Adjusted Rand Index sebesar 0,17 (buruk). Namun dengan menggunakan metode Calinski-Harabasz Index, K-means memperoleh skor sebesar 338,28 yang berarti bahwa K-means cocok digunakan untuk klastering data hasil akademik. Hasil penelitian memberikan informas tentang distibusi jumlah data mahasiswa, tren Indeks Prestasi Kumulatif Mahasiswa, daftar mahasiswa setiap klaster potensi waktu kelulusannya.

DAFTAR PUSTAKA

Agwil, W., Fransiska, H., & Hidayati, N. (2020). Analisis Ketepatan Waktu Lulus Mahasiswa Dengan Menggunakan BAGGING CART. *FIBONACCI: Jurnal Pendidikan Matematika Dan Matematika*, *6*(2), 155–166. https://doi.org/10.24853/fbc.6.2.155-166

Al Husaini, Y., Syufiza, N., Shukor, A., Said, Y. N., & Husaini, A. (2022). Factors Affecting Students' Academic Performance: A review. *Social Science Journal*, *12*(6), 284–294. https://doi.org/https://doi.org/10.1145/321796.321811

Firman Ashari, I., Dwi Nugroho, E., Baraku, R., Yanda, I. N., & Liwardana, R. (2023). Analysis of Elbow, Silhouette, Davies-Bouldin, Calinski-Harabasz, and Rand-Index Evaluation on K-Means Algorithm for Classifying Flood-Affected Areas in Jakarta. In *Journal of Applied Informatics and Computing (JAIC)* (Vol. 7, Issue 1). http://jurnal.polibatam.ac.id/index.php/JAIC

Fu, S., Lu, S. Y., Davies, D. L., & Bouldin, D. W. (1977). The string-to-string correction problem. In *J. Ass. Comput. Mach* (Vol. 1, Issue 2).

Helilintar, R., & Farida, I. N. (2018). Penerapan Algoritma K-Means Clustering Untuk Prediksi Prestasi Nilai Akademik Mahasiswa. *Jurnal Sains Dan Informatika*, *4*(2), 80–87. http://repository.unpkediri.ac.id/id/eprint/2399

Mughnyanti, M., Efendi, S., & Zarlis, M. (2020). Analysis of determining centroid clustering x-means algorithm with davies-bouldin index evaluation. *IOP Conference Series: Materials Science and Engineering*, *725*(1). https://doi.org/10.1088/1757-899X/725/1/012128

Muningsih, E., Maryani, I., & Handayani, V. R. (2021). Penerapan Metode K-Means dan Optimasi Jumlah Cluster dengan Index Davies Bouldin untuk Clustering Propinsi Berdasarkan Potensi Desa. *Jurnal Sains Dan Manajemen*, *9*(1). https://doi.org/https://doi.org/10.31294/evolusi.v9i1.10428

Nur Addini, S., & Rosian Adhy, D. (2021). Aplikasi Prediksi Waktu Kelulusan Mahasiswa Menggunakan Algoritma Kmeans Clustering. *Saintesa*, *1*(1), 14–20.

Osamor, I. P., & Osamor, V. C. (2020). OsamorSoft: clustering index for comparison and quality validation in high throughput dataset. *Journal of Big Data*, *7*(1). https://doi.org/10.1186/s40537-020-00325-6

Othman, S., Steen, M., & Fleet, J.-A. (2020). A sequential explanatory mixed methods study design: An example of how to integrate data in a midwifery research project. *Journal of Nursing Education and Practice*, *11*(2), 75. https://doi.org/10.5430/jnep.v11n2p75

Peraturan Menteri Pendidikan Dan Kebudayaan Republik Indonesia Nomor 49 Tahun 2014 Tentang Standar Nasional Pendidikan Tinggi, 1 (2014).

Rahmawati, E., Chrisnanto, Y. H., & Maspupah, A. (2019). Identifikasi Kemampuan Akademik Mahasiswa Menggunakan K-Means Clustering. *Prosiding SEMNAS Inovasi Teknologi*, 87–92. https://doi.org/https://doi.org/10.29407/inotek.v3i1.518

Rusdiana, L., & Sam'ani, D. (2016). Pemodelan K-Means Pada Penentuan Predikat Kelulusan Mahasiswa STMIK PALANGKARAYA. *Saintekom*, *6*(1), 1–15. https://core.ac.uk/download/pdf/233889338.pdf

Saitta, S., Raphael, B., & Smith, I. F. C. (2008). A Comprehensive Validity Index for Clustering. *Intelligent Data Analysis*, *12*(6), 529–548. https://doi.org/10.3233/IDA-2008-12602

Santos, J. M., & Embrechts, M. (2009). On the Use of the Adjusted Rand Index as a Metric for Evaluating Supervised Classification. In *International Conference on Artificial Neural Networks*.

TIM. (2018). *Pedoman Akademik 2018 Universitas Lambung Mangkurat*. Universitas Lambung Mangkurat.