



Analisis Tingkat Kelulusan Mahasiswa di Unisba dengan menggunakan Algoritma *K-Means Clustering*

Ayunia Khaerunnisa*

Prodi Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Bandung, Indonesia.

ARTICLE INFO

Article history :

Received : 16/4/2022
Revised : 6/7/2022
Published : 9/7/2022



Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

Volume : 2
No. : 1
Halaman : 67 - 76
Terbitan : Juli 2022

ABSTRAK

Dalam lingkungan akademik di perguruan tinggi, khususnya pada program studi matematika di Universitas Islam Bandung, masih terdapat permasalahan mengenai mahasiswa yang tidak dapat menyelesaikan studinya dalam waktu yang telah ditentukan. Adanya permasalahan tersebut maka membutuhkan suatu metode atau teknik yang dapat mengelompokkan mahasiswa berdasarkan SKS (Satuan Kredit Semester) yang telah ditempuh dan nilai IPK agar waktu kelulusan dapat diprediksi. Teknik yang digunakan dalam penelitian ini yaitu dengan teknik *Clustering* dengan menggunakan algoritma *K-Means* yang bertujuan untuk menemukan pola kelulusan mahasiswa yang masih menempuh pendidikan maupun sudah lulus tepat waktu sesuai tahun kelulusan yang sudah ditentukan. Hasil penelitian dengan menggunakan data mahasiswa aktif pada program studi matematika 2015-2016 diketahui dapat dikelompokkan menjadi tiga *Cluster* diantaranya *Cluster* pertama adalah mahasiswa yang lulus lebih awal, *Cluster* kedua adalah mahasiswa yang lulus tepat waktu, dan *Cluster* ketiga adalah mahasiswa yang lulus lebih dari empat tahun. Sehingga prodi dapat mengetahui mahasiswa mana yang termasuk dalam ketiga *Cluster* tersebut dan dapat diusulkan untuk dijadikan masukan bagi pengambil keputusan.

Kata Kunci : Kelulusan; *Clustering* ; Algoritma K-Means.

ABSTRACT

In the academic environment at universities, especially in the mathematics study program at the Islamic University of Bandung, there are still problems regarding students who cannot complete their studies within the allotted time. The existence of these problems requires a method or technique that can classify students based on the SKS (Semester Credit Units) that have been taken and the IPK value so that the graduation time can be predicted. The technique used in this research is the *Clustering* technique using the *K-Means* algorithm which aims to find the graduation pattern of students who are still studying or have graduated on time according to the predetermined graduation year. The results of the study using data from active students in the 2015-2016 mathematics study program are known to be can grouped into three *Clusters* including the first *Cluster* is students who graduate early, the second *Cluster* is students who graduate on time, and the third *Cluster* is students who graduate more than four years. So the study program can find out which students are included in the three *Clusters* and can be proposed to be used as input for decision makers.

Keywords : Graduation; *Clustering* ; *K-Means* Algorithm

© 2022 Jurnal Riset Matematika Unisba Press. All rights reserved.

A. Pendahuluan

Tingkat kelulusan sangat penting untuk melihat suatu keberhasilan akademik dalam proses pendidikan disuatu perguruan tinggi. Kelulusan tidak hanya penting bagi perguruan tinggi tetapi penting juga bagi mahasiswa karena kelulusan merupakan awal dalam menentukan karir di masa depan atau menuju jenjang lebih tinggi.

Mahasiswa dikatakan berhasil dalam kinerja akademik apabila mahasiswa tersebut dapat menyelesaikan studi sesuai dengan waktu yang telah ditentukan. Normalnya, mahasiswa (strata-1) lulus dalam jangka waktu 4 tahun, namun dalam kenyataannya pada lingkungan akademik dalam perguruan tinggi khususnya di program studi matematika Universitas Islam Bandung masih ditemukan permasalahan mengenai mahasiswa yang tidak dapat menuntaskan studinya dalam jangka waktu yang ditentukan. Untuk memecahkan masalah tersebut, maka perlu adanya metode atau teknik yang dapat mengelompokkan mahasiswa berdasarkan jumlah SKS (Satuan Kredit Semester) yang telah ditempuh dan nilai IPK agar waktu kelulusan dapat diprediksi.

Teknik yang digunakan dalam penelitian ini yaitu dengan teknik *Clustering* yang bertujuan untuk menemukan pola kelulusan mahasiswa yang masih menempuh pendidikan maupun sudah lulus tepat waktu sesuai tahun kelulusan yang sudah ditentukan. *Clustering* yaitu mengelompokkan objek-objek data hanya berdasarkan pada informasi yang terdapat pada data yang menjelaskan objek dan relasinya. Proses inti dari *Clustering* adalah pengelompokan data menjadi beberapa kelompok sehingga objek di dalam satu kelompok memiliki banyak kesamaan dan memiliki perbedaan dengan objek dikelompok lain [1][2]. Teknik *Clustering* merupakan teknik yang sudah cukup dikenal dan banyak digunakan. Teknik *Clustering* sederhana dan umum digunakan adalah algoritma *k-means*, dimana algoritma *K-Means* ini mengelompokkan objek berdasarkan jarak. Rumus jarak yang digunakan dalam algoritma *K-Means* ini adalah *euclidean distance space*. Dengan algoritma *k-means*, karakteristik mahasiswa dapat dianalisa berdasarkan tingkat kelulusannya. Maka penelitian ini perlu dilakukan karena dari hasil pengelompokan dan hasil prediksi dengan menggunakan *K-Means Clustering* prodi dapat mengetahui banyaknya mahasiswa yang dapat lulus 3,5 tahun yaitu lulus lebih cepat, 4 tahun yaitu lulus tepat waktu dan lebih dari 4 tahun yaitu lulus lambat. Sehingga dapat diusulkan untuk dijadikan sebagai masukan pengambil keputusan dalam mengantisipasi mahasiswa yang diprediksi akan lulus lebih dari 4 tahun.

Berdasarkan latar belakang yang telah diuraikan, maka perumusan masalah dalam penelitian ini sebagai berikut: “Bagaimana *Clustering* masa studi mahasiswa menggunakan algoritma *K-Means Clustering*?”, “Bagaimana prediksi tingkat kelulusan mahasiswa?”, “Bagaimana tingkat akurasi prediksi kelulusan mahasiswa di Universitas Islam Bandung?”

Selanjutnya, tujuan dalam penelitian ini diuraikan dalam pokok-pokok sebagai berikut: (1) Mengetahui hasil pengelompokkan masa studi mahasiswa menggunakan algoritma *K-Means Clustering*; (2) Mengetahui hasil prediksi tingkat kelulusan mahasiswa; (3) Mengetahui tingkat akurasi prediksi kelulusan mahasiswa di Universitas Islam Bandung.

B. Metode Penelitian

Analisis Data

Analisis data merupakan upaya mencari dan menata secara sistematis catatan hasil observasi, wawancara, dan lainnya untuk meningkatkan pemahaman peneliti tentang kasus yang diteliti dan menyajikannya sebagai temuan bagi orang lain. Sedangkan untuk meningkatkan pemahaman tersebut analisis perlu dilanjutkan dengan berupaya mencari makna [3].

Peneliti menggunakan metode teknik analisis data kuantitatif, dimana data yang digunakan merupakan data mahasiswa aktif matematika Universitas Islam Bandung angkatan tahun 2015-2016 yang terdiri dari 58 mahasiswa. Adapun metode analisis data yang digunakan yaitu analisis deskriptif dan teknik *Clustering* menggunakan algoritma *k-means*.

Analisis data kuantitatif merupakan salah satu jenis analisis data yang berupa angka atau bisa diangkakan. Analisis data kuantitatif yaitu mengolah atau mengelola data-data bersifat angka-angka atau

statistik. Pada teknik analisis data kuantitatif, data-data yang digunakan ialah data-data angka atau data numerik yang dapat dihitung secara tepat dengan perhitungan rumus statistik[4].

Analisis deskriptif merupakan metode pada data kuantitatif. Cara ini dilakukan untuk melihat performa data di masa lalu agar dapat mengambil kesimpulan dari hal tersebut. Analisis deskriptif memiliki dua proses yang berbeda di dalamnya berupa deskripsi dan interpretasi. Jenis metode ini biasa digunakan dalam menyajikan data statistik [5].

Clustering

Clustering yaitu suatu teknik pengelompokkan objek-objek data hanya berdasarkan pada informasi yang terdapat pada data yang menjelaskan objek dan relasinya. Proses inti dari *Clustering* adalah pengelompokan data menjadi beberapa kelompok sehingga objek di dalam satu kelompok memiliki banyak kesamaan dan memiliki perbedaan dengan objek dikelompok lain.

Ada beberapa pendekatan yang digunakan dalam mengembangkan teknik *Clustering* yang sudah dikembangkan oleh para ahli dimana masing-masing pendekatan mempunyai karakter, kelebihan, dan kekurangan. *Clustering* dapat dibedakan menurut struktur *Cluster*, keanggotaan data dalam *Cluster* dan kekompakan data dalam *Cluster* [6]. Pengelompokan data ini dapat dilakukan dengan dua pendekatan yaitu metode *hierarichal Clustering* dan *non-hierarichal Clustering*.

Clustering dengan pendekatan *non-hierarichal Clustering* bertujuan untuk mengelompokkan n objek kedalam k *Cluster*, dimana nilai k sudah ditentukan sebelumnya. Setelah jumlah *Cluster* diketahui, baru proses *Cluster* dilakukan tanpa mengikuti proses hierarki. Biasanya dilakukan dengan menentukan pusat *Cluster* awal, lalu dilakukan realokasi objek berdasarkan kriteria tertentu sampai dicapai pengelompokkan yang optimal. Contoh *Clustering* dengan pendekatan *non-hierarichal Clustering* adalah *K-Means Clustering* [7].

Algoritma K-Means

Algoritma *K-Means* merupakan algoritma klasterisasi yang mengelompokan data berdasarkan titik pusat (*Centroid*) yang terdekat dengan data. *K-Means* merupakan salah satu algoritma *Clustering* yang masuk ke dalam kelompok *unsupervised learning* yang digunakan untuk membagi data menjadi beberapa kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu *Cluster* yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lain [8].

K-Means termasuk dalam metode *partitioning Clustering* yaitu setiap data harus masuk dalam *Cluster* tertentu dan memungkinkan bagi setiap data yang termasuk dalam *Cluster* tertentu pada suatu tahapan proses, pada tahapan berikutnya berpindah ke *Cluster* lain. Berikut langkah-langkah penyelesaian dengan algoritma *K-Means* [9]:

Menentukan jumlah *Cluster* (*k*) yang diinginkan,

Menentukan nilai pusat awal *Cluster* (*Centroid*) dilakukan secara acak. Sedangkan untuk *Centroid* pada iterasi, digunakan rumus sebagai berikut:

$v_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} x_{kj}$	(1)
--	-----

Dimana:

v_{ij} adalah *Centroid* atau rata-rata *Cluster* ke-*i* untuk variabel ke-*j*;

N_i adalah jumlah data yang menjadi anggota *Cluster* ke-*i*;

i, k adalah indeks dari *Cluster*;

j adalah indeks dari variabel;

x_{kj} adalah nilai data ke-*k* yang terdapat pada *Cluster* tersebut untuk variable ke-*j*.

Menempatkan setiap data atau objek ke *Cluster* terdekat. Kedekatan dua objek ditentukan berdasarkan jarak. Jarak yang digunakan dalam algoritma *K-Means* adalah *euclidean distance space*. *Euclidean distance space* ini sering digunakan karena hasil yang diperoleh merupakan jarak terpendek antara dua titik yang diperhitungkan [10]. Secara literatur *euclidean distance space* dapat digunakan untuk tipe data seperti pada penelitian ini. Namun secara literatur dengan menggunakan *euclidean distance space* ini akan memiliki resiko

nilai galat yang besar. Jarak terpendek antara dua titik yang diperhitungkan merupakan jarak yang akan menentukan sebuah data masuk ke dalam *Cluster* tersebut.

$$d_{Euclidean} = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2} \tag{2}$$

Dimana:

$d_{Euclidean}$ adalah ukuran kemiripan atau jarak

i adalah objek;

(x, y) adalah koordinat objek;

(s, t) adalah koordinat pusat;

Dalam tahap ini hitung jarak tiap data dengan pusat *Cluster*.

Pengelompokan objek berdasarkan jarak minimum. Jarak paling dekat antara satu data dengan suatu *Cluster* tertentu maka akan menentukan suatu data masuk dalam *Cluster* mana.

Menghitung kembali pusat *Cluster* dengan keanggotaan *Cluster* yang baru. Titik pusat pada setiap *Cluster* diambil dari rata-rata (mean) dengan rumus seperti pada point 2 dari semua nilai data pada setiap fiturnya.

Ulangi langkah ke 2 sampai 5 hingga pusat *Cluster* pada iterasi sebelumnya sama dengan pusat *Cluster* pada iterasi selanjutnya.

Tahap terakhir didapat hasil akhir *Clustering*.

Akurasi Hasil *Cluster*

Akurasi hasil *Cluster* bertujuan untuk mengetahui suatu ketepatan dari hasil pengelompokan dengan menggunakan rumus *Confusion Matrix* 3x3 dengan 3 tingkat kelulusan yaitu lulus cepat 3,5 tahun, lulus tepat waktu 4 tahun dan lulus lambat > 4 tahun. Tabel dibawah ini merupakan bentuk *Confusion Matrix* 3x3 dengan A adalah lulus cepat 3,5 tahun, B adalah lulus tepat waktu 4 tahun dan C adalah lulus lambat > 4 tahun [11]:

Tabel 1. *Confusion Matrix* 3x3

Aktual	Prediksi		
	Lulus Cepat (3,5 tahun)	Tepat Waktu (4 tahun)	Lambat (> 4 Tahun)
Lulus Cepat (3,5 tahun)	P_{AA}	P_{AB}	P_{AC}
Tepat Waktu (4 tahun)	P_{BA}	P_{BB}	P_{BC}
Lambat (> 4 Tahun)	P_{CA}	P_{CB}	P_{CC}

$Accuracy = \frac{P_{AA} + P_{BB} + P_{CC}}{P_{AA} + P_{AB} + P_{AC} + P_{BA} + P_{BB} + P_{BC} + P_{CA} + P_{CB} + P_{CC}} \times 100\%$	(3)
$Galat = \frac{P_{CA} + P_{CB} + P_{AB} + P_{AC} + P_{BA} + P_{BC}}{P_{AA} + P_{AB} + P_{AC} + P_{BA} + P_{BB} + P_{BC} + P_{CA} + P_{CB} + P_{CC}} \times 100\%$	(4)

C. Hasil dan Pembahasan

Data yang digunakan dalam penelitian ini merupakan data sekunder. Data sekunder ini adalah data yang diperoleh dari SISFO Universitas Islam Bandung yang terdiri dari 58 data mahasiswa aktif matematika angkatan tahun 2015-2016. Dalam penelitian ini ada dua atribut yang akan digunakan diantaranya yaitu total satuan kredit semester (SKS) yang telah ditempuh dan nilai IPK. Data yang akan digunakan dalam penelitian ini sebagai berikut:

Tabel 2. Data Mahasiswa Aktif Matematika Angkatan Tahun 2015-2016

No	NAMA	SKS	IPK
1	M1	152	3,57
2	M2	147	3,68
3	M3	147	3,34
4	M4	147	3,47
5	M5	152	3,45
6	M6	142	2,81
7	M7	147	3,88
8	M8	152	3,78
9	M9	149	3,75
10	M10	152	3,56
...
58	M58	143	3,43

Sumber: SISFO Universitas Islam Bandung

Analisa Cluster

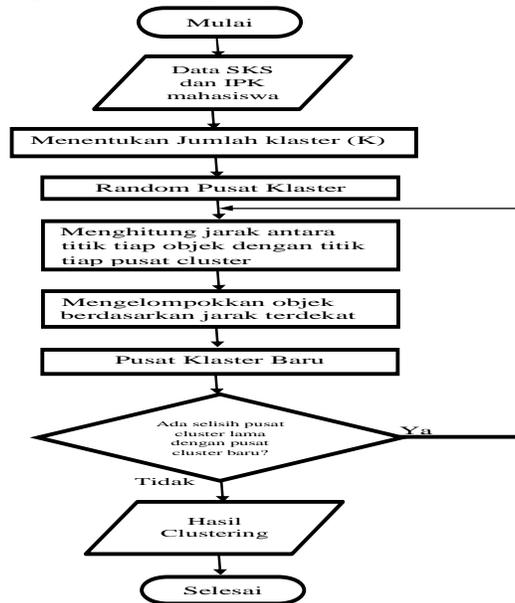
Berdasarkan data mahasiswa aktif angkatan tahun 2015-2016 yang terdiri dari 58 mahasiswa akan dibagi menjadi tiga tingkat kelulusan yaitu mahasiswa dengan kategori lulus lebih cepat yaitu 3,5 tahun, mahasiswa lulus dengan waktu normal yaitu 4 tahun atau lulus tepat waktu dan mahasiswa lulus lebih dari 4 tahun. Teknik *Clustering* akan digunakan untuk mengelompokkan mahasiswa dengan menggunakan algoritma *K-Means Clustering*. Adapun tingkat kelulusan ini dibagi menjadi tiga *Cluster* yang dapat dilihat pada tabel berikut ini:

Tabel 3. Pembagian Tiga *Cluster*

<i>Cluster 1 (C1)</i>	<i>Cluster 1 (lulus 3,5 tahun)</i>
<i>Cluster 2 (C2)</i>	<i>Cluster 2 (lulus 4 tahun)</i>
<i>Cluster 3 (C3)</i>	<i>Cluster 3 (lulus > 4 tahun)</i>

Syarat kelulusan berdasarkan kebijakan prodi matematika Universitas Islam Bandung yaitu setiap mahasiswa memiliki beban studi sekurang-kurangnya 147 SKS (satuan kredit semester) yang harus ditempuh dalam 7 semester dan paling lama 14 semester dengan IPK minimal 2,00.

Proses dalam teknik *Clustering* menggunakan algoritma *K-Means* terdapat beberapa langkah yang harus dilalui, dimana langkah-langkah tersebut seperti pada *flowchart* berikut ini:



Gambar 1. *Flowchart Clustering* dengan Algoritma *k-means*

Dalam tahap ini akan dijelaskan langkah-langkah perhitungan menggunakan algoritma *K-Means Clustering* secara manual. Diketahui jumlah *Cluster* sebanyak 3, jumlah sampel data sebanyak 10, jumlah atribut sebanyak 2.

Iterasi 1

Menentukan jumlah *Cluster* (K). Data akan dibagi ke dalam tiga *Cluster* (k=3) dan akan terbentuk 3 pusat *Cluster*.

Menentukan pusat awal *Cluster* dilakukan dengan cara *random* (pemilihan acak) secara manual atau dapat menggunakan bantuan sistem dari data total SKS dan nilai IPK yang telah ditempuh oleh mahasiswa. Nilai *Cluster* 1 diambil dari baris ke-37, nilai *Cluster* 2 diambil dari baris ke-9 dan nilai *Cluster* 3 diambil dari baris ke-20. Ditentukan pusat *Cluster* sebagai berikut:

Tabel 4. Pusat *Cluster* Awal

Kategori	<i>Cluster</i> 1	<i>Cluster</i> 2	<i>Cluster</i> 3
SKS	153	149	139
IPK	3,56	3,75	3,15

Melakukan perhitungan jarak dengan rumus *euclidean distance* dari masing-masing item dari pusat *Cluster*. Cara menghitung jarak sebagai berikut:

$d_{Euclidean} = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2}$	(5)
--	-----

Perhitungan jarak terdekat tiap objek dihitung sebanyak tiga kali sesuai banyaknya *Cluster* yaitu C1, C2, C3 berdasarkan SKS dan nilai IPK yang telah di tempuh dengan titik pusatnya dan dilakukan untuk semua mahasiswa di iterasi 1. Berikut hasil iterasi 1 untuk 10 mahasiswa:

Tabel 5. Perhitungan Jarak Iterasi 1

No	NAMA	SKS	IPK	C1	C2	C3
1	M1	152	3,57	1,000	3,005	13,007
2	M2	147	3,68	6,001	2,001	8,018
3	M3	147	3,34	6,004	2,042	8,002
4	M4	147	3,47	6,001	2,020	8,006
5	M5	152	3,45	1,006	3,015	13,003
6	M6	142	2,81	11,026	7,063	3,019
7	M7	147	3,88	6,009	2,004	8,033
8	M8	152	3,78	1,024	3,000	13,015
9	M9	149	3,75	4,005	0,000	10,018
10	M10	152	3,56	1,000	3,006	13,006

Sumber: Data Penelitian yang Sudah Diolah

Pengelompokan objek berdasarkan jarak minimum. Dalam keputusannya dari hasil yang diperoleh sebelumnya semakin minimum nilai jaraknya maka semakin besar kemiripannya terhadap suatu *Cluster*. Sehingga jarak terdekat itulah yang menentukan setiap mahasiswa termasuk ke dalam *Cluster* tersebut. Pengelompokan data dilakukan untuk semua mahasiswa di iterasi 1. Hasil pengelompokan objek berdasarkan jarak minimum dari hasil perhitungan sebelumnya dapat dilihat pada **tabel 6**.

Tabel 6. Pengelompokan Iterasi 1

No	NAMA	SKS	IPK	C1	C2	C3	Jarak Terdekat ke <i>Cluster</i>
1	M1	152	3,57	1,000	3,005	13,007	1
2	M2	147	3,68	6,001	2,001	8,018	2
3	M3	147	3,34	6,004	2,042	8,002	2
4	M4	147	3,47	6,001	2,020	8,006	2
5	M5	152	3,45	1,006	3,015	13,003	1
6	M6	142	2,81	11,026	7,063	3,019	3
7	M7	147	3,88	6,009	2,004	8,033	2
8	M8	152	3,78	1,024	3,000	13,015	1
9	M9	149	3,75	4,005	0,000	10,018	2
10	M10	152	3,56	1,000	3,006	13,006	1

Sumber: Data Penelitian yang Sudah Diolah

Menghitung kembali pusat *Cluster* baru dengan menghitung rata-rata dari data yang ada di masing-masing *Cluster* untuk proses iterasi berikutnya. Dibawah ini merupakan tabel pusat *Cluster* untuk iterasi ke-2.

Tabel 7. Centroid Baru

Kategori	<i>Cluster 1</i>	<i>Cluster 2</i>	<i>Cluster 3</i>
SKS	152	148	138
IPK	3,58	3,56	2,99

Sumber: Data Penelitian yang Sudah Diolah

Dari titik pusat yang diperoleh diatas diketahui bahwa pusat *Cluster* awal dengan pusat *Cluster* baru memiliki nilai yang berbeda sehingga perlu dilakukan iterasi kedua yaitu menghitung kembali jarak dan mengelompokan data hingga pusat *Cluster* pada iterasi sebelumnya sama dengan pusat *Cluster* pada iterasi selanjutnya.

Iterasi 2

Perhitungan jarak pada iterasi kedua dan iterasi selanjutnya menggunakan proses yang sama seperti pada iterasi 1. Hasil perhitungan jarak yang diperoleh pada iterasi 2 untuk 10 mahasiswa sebagai berikut:

Tabel 8. Perhitungan Jarak Iterasi 2

No	NAMA	SKS	IPK	C1	C2	C3
1	M1	152	3,57	1,000	3,005	13,007
2	M2	147	3,68	6,001	2,001	8,018
3	M3	147	3,34	6,004	2,042	8,002
4	M4	147	3,47	6,001	2,020	8,006
5	M5	152	3,45	1,006	3,015	13,003
6	M6	142	2,81	11,026	7,063	3,019
7	M7	147	3,88	6,009	2,004	8,033
8	M8	152	3,78	1,024	3,000	13,015
9	M9	149	3,75	4,005	0,000	10,018
10	M10	152	3,56	1,000	3,006	13,006

Sumber: Data Penelitian yang Sudah Diolah

Pengelompokan objek berdasarkan jarak minimum. Hasil yang diperoleh sebelumnya merupakan nilai jarak masing-masing objek ke dalam tiap pusat *Cluster*. Dalam keputusannya semakin minimum nilai jaraknya maka semakin besar kemiripannya terhadap suatu *Cluster*. Sehingga jarak terdekat itulah yang menentukan setiap mahasiswa termasuk ke dalam *Cluster* tersebut. Pengelompokan data dilakukan untuk semua mahasiswa di iterasi Hasil pengelompokan objek berdasarkan jarak minimum dari hasil perhitungan sebelumnya dapat dilihat pada tabel 9 berikut ini:

Tabel 9. Pengelompokan Iterasi 2

No	NAMA	SKS	IPK	C1	C2	C3	Jarak Terdekat ke <i>Cluster</i>
1	M1	152	3,57	1,000	3,005	13,007	1
2	M2	147	3,68	6,001	2,001	8,018	2
3	M3	147	3,34	6,004	2,042	8,002	2
4	M4	147	3,47	6,001	2,020	8,006	2
5	M5	152	3,45	1,006	3,015	13,003	1
6	M6	142	2,81	11,026	7,063	3,019	3
7	M7	147	3,88	6,009	2,004	8,033	2
8	M8	152	3,78	1,024	3,000	13,015	1
9	M9	149	3,75	4,005	0,000	10,018	2
10	M10	152	3,56	1,000	3,006	13,006	1

Sumber: Data Penelitian yang Sudah Diolah

Selanjutnya menghitung kembali pusat *Cluster* dan didapatkan pusat *Cluster* baru untuk iterasi ke-3 yaitu sebagai berikut:

Tabel 10. Centroid Baru

Kategori	<i>Cluster 1</i>	<i>Cluster 2</i>	<i>Cluster 3</i>
SKS	152	148	138
IPK	3,58	3,56	2,99

Setelah menghitung kembali pusat *Cluster* baru untuk iterasi 3 diketahui bahwa dari hasil penelitian untuk proses perhitungan iterasi berhenti pada iterasi 3 karena pada pusat *Cluster* iterasi 2 dan iterasi 3 diketahui sudah sama (tidak berubah) dan tidak perlu lagi dilakukan iterasi. Sehingga hasil yang diperoleh dari perhitungan diatas untuk mengelompokan mahasiswa ke dalam 3 *Cluster* dengan menggunakan algoritma *K-Means* yang dimulai dari iterasi 1 sampai dengan 3 didapatkan *Cluster* sebagai berikut:

Tabel 11. Hasil Akhir *Cluster* dengan Perhitungan Manual

No	Lama Studi	Anggota <i>Cluster</i>	SKS	IPK
1	<i>Cluster 1</i>	8	152	3,58
2	<i>Cluster 2</i>	37	148	3,56
3	<i>Cluster 3</i>	13	140	3,10
Total		31		

Sumber: Data Penelitian yang Sudah Diolah

Cluster 1 memiliki 8 data yang diartikan bahwa kelompok pertama adalah kategori mahasiswa yang menyelesaikan masa studinya dengan waktu lebih cepat yaitu 3,5 tahun dengan nilai IPK tinggi. *Cluster 2* memiliki 37 data yang kategori mahasiswa yang menyelesaikan masa studinya dengan tepat waktu yaitu 4 tahun dengan nilai IPK sedang. *Cluster 3* memiliki 13 data yang diartikan bahwa kelompok ketiga adalah mahasiswa yang menyelesaikan masa studinya lebih dari 4 tahun dengan nilai IPK rendah.

Analisis Akurasi Cluster

Berdasarkan data sebanyak 58 mahasiswa, didapatkan bahwa mahasiswa yang lulus lebih cepat sebanyak 8 mahasiswa, lulus tepat waktu sebanyak 37 mahasiswa dan mahasiswa yang belum lulus sebanyak 13 mahasiswa.

Tabel 12. Hasil Pengelompokan

Aktual	Prediksi			Total
	Lulus Cepat (3,5 tahun)	Tepat Waktu (4 tahun)	Lambat (> 4 Tahun)	
Lulus Cepat (3,5 tahun)	7	0	0	7
Tepat Waktu (4 tahun)	0	36	0	36
Lambat (> 4 Tahun)	1	1	13	15
Total	8	37	13	58

Sumber: Data Penelitian yang Sudah Diolah

$$Accuracy = \frac{7 + 36 + 13}{7 + 0 + 0 + 0 + 36 + 0 + 1 + 1 + 13} = \frac{56}{58} \times 100\% = 96,55\% \tag{3}$$

$$Galat = \frac{1 + 1 + 0 + 0}{7 + 0 + 0 + 0 + 36 + 0 + 1 + 1 + 13} = \frac{2}{58} \times 100\% = 3,45\% \tag{4}$$

D. Kesimpulan

Berdasarkan pembahasan dalam penelitian ini, peneliti menyimpulkan beberapa hasil penelitian sebagai berikut:

Hasil pengelompokan yang diperoleh dengan menggunakan algoritma *K-Means Clustering* yaitu *Cluster 1* berjumlah 8 mahasiswa *Cluster 2* berjumlah 37 mahasiswa dan *Cluster 3* berjumlah 13 mahasiswa.

Hasil prediksi yang diperoleh dari hasil pengelompokan yaitu *Cluster 1* berjumlah 8 mahasiswa yang merupakan mahasiswa yang berpotensi menyelesaikan masa studi dalam waktu lebih cepat yaitu 3,5 tahun dengan nilai IPK tinggi. *Cluster 2* berjumlah 37 mahasiswa yang merupakan mahasiswa yang berpotensi menyelesaikan masa studi dalam waktu normal 4 tahun atau lulus tepat waktu dengan nilai IPK sedang dan *Cluster 3* berjumlah 13 mahasiswa yang merupakan mahasiswa yang menyelesaikan masa studinya lebih dari 4 tahun dengan nilai IPK rendah.

Nilai akurasi prediksi dari pengelompokan kelulusan yaitu sebesar 96,55% dan untuk nilai *error* dari pengelompokan sebesar 3,45%. Sehingga metode ini dapat digunakan untuk memprediksi kelulusan mahasiswa.

Daftar Pustaka

- [1] P. Suwirmayanti, I. K. G. . Putra, and I. N. . Kumara, “Optimasi Pusat Cluster K-Prototype dengan Algoritma Genetika,” *J. Teknol. Elektro*, pp. 16–23, 2014.
- [2] Rahadatul Aisyi, “Implementasi Pemilihan Siswa Berprestasi Menggunakan Metode Preferences Selection Index,” *J. Ris. Mat.*, vol. 1, no. 2, pp. 145–153, Feb. 2022, doi: 10.29313/jrm.v1i2.487.
- [3] A. Rijali, “Analisis Data Kualitatif Ahmad Rijali UIN Antasari Banjarmasin,” vol. 17, no. 33, pp. 81–95, 2018.
- [4] I. Nurdin and S. Hartati, *Metodologi Penelitian Sosial*. Surabaya: Media Sahabat Cendekia, 2019.
- [5] Sugiyono, *Metode Penelitian Kuantitatif, Kualitatif dan R&D*. Bandung: Alfabeta, 2012.
- [6] M. Ad. W. K. Murti, “Penerapan Metode *K-Means Clustering* Untuk Mengelompokan Potensi Produksi Buah – Buah Di Provinsi Daerah Istimewa Yogyakarta,” 2017.
- [7] S. Santoso, *Statistik multivariat dengan SPSS*, 1st ed. Jakarta: PT Elex Media Komputindo, 2017.
- [8] W. Mochamad and dkk, *Data Mining: Penerapan Algoritma K-Means Clustering dan K-Medoids Clustering*, 1st ed. Medan: Yayasan Kita Menulis, 2020.
- [9] R. . Ariesanto and dkk, *Modul Belajar Data Mining dan RapidMiner*, 1st ed. Klaten: Lekeisha, 2020.

- [10] D. Retno Lestari, "Implementasi Data Mining Untuk Pengaturan Layout Swalayan Delimas Lestari Kencana Dengan Menggunakan Dengan Menggunakan Algoritma *K-Means Clustering* ," *Jurnal Ris. Komput.*, vol. 6, no. 2, pp. 112–117, 2019, [Online]. Available: <http://ejurnal.stmik-budidarma.ac.id/index.php/jurikom%7CPage%7C112>
- [11] I. Kurniawan and A. Susanto, "Implementasi Metode *K-Means* dan Naïve Bayes Classifier untuk Analisis Sentimen Pemilihan Presiden (Pilpres) 2019," *J. Eksplora Inform.*, vol. 9, no. 1, pp. 1–10, 2019, doi: 10.30864/eksplora.v9i1.237.