



Penerapan Algoritma K-Means Clustering pada Sistem Prediksi Kelulusan Tepat Waktu

Tati Harihayati Mardzuki^{1*}, Riani Lubis², Fakhrian Fadlia Adiwijaya³

^{1,2,3}) Program Studi Teknik Informatika, Universitas Komputer Indonesia
Jl. Dipati Ukur No. 112 - 116, Bandung, Indonesia 40132

*email: tati.harihayati@email.unikom.ac.id

(Naskah masuk: 19 September 2024; direvisi: 28 Oktober 2024; diterima untuk diterbitkan: 31 Oktober 2024)

ABSTRAK – Saat ini, Dikti menetapkan presentase kelulusan tepat waktu yang baik untuk akreditasi program studi ada dikisaran 50% untuk setiap angkatan. Oleh karena itu Kepala Program Studi memerlukan informasi mengenai jumlah mahasiswa yang diprediksi akan lulus tepat waktu di semester delapan nantinya, agar dapat diambil kebijakan sedini mungkin jika jumlahnya tidak sesuai dengan yang diharapkan. Metode yang digunakan pada penelitian ini adalah algoritma K-Means Clustering, dimana algoritma ini akan membagi data mahasiswa menjadi dua kelompok (klaster), yaitu jumlah mahasiswa yang diprediksi lulus tepat waktu dan yang tidak lulus tepat waktu. Data set yang digunakan adalah data akademik mahasiswa dari semester satu sampai semester enam dengan lima kriteria yaitu IPK sampai semester enam, jumlah sks yang lulus sampai semester enam, jumlah semester yang telah diambil sampai semester enam, jumlah cuti hingga semester enam dan jenis asal sekolah. Hasil dari penelitian ini menunjukkan bahwa jumlah mahasiswa yang diprediksikan akan lulus tepat waktu adalah sekitar 92,84% (311 mahasiswa) berdasarkan data set mahasiswa semester enam yang berjumlah 335 mahasiswa dengan enam kali iterasi pada algoritma K-Means Clustering.

Kata Kunci – Akreditasi; K-Means Clustering; Mahasiswa; Prediksi Kelulusan; Program Studi.

Implementation of K-Means Clustering Algorithm on Timely Graduation Prediction System

ABSTRACT – Currently, Dikti sets a good on-time graduation percentage for study program accreditation at around 50% for each class. Therefore, the Head of Study Program requires information on the number of students predicted to graduate on time in the eighth semester later, so that policies can be taken as early as possible if the number is not as expected. The method used in this study is the K-Means Clustering algorithm, where this algorithm will divide student data into two groups (clusters), namely the number of students predicted to graduate on time and those who do not graduate on time. The data set used is student academic data from semester one to semester six with five criteria, namely GPA up to semester six, number of credits graduated up to semester six, number of semesters taken up to semester six, number of leaves up to semester six and type of school origin. The results of this study indicate that the number of students predicted to graduate on time is around 92.84% (311 students) based on the sixth semester student data set totaling 335 students with six iterations on the K-Means Clustering algorithm.

Keywords – Accreditation; Graduation Prediction; K-Means Clustering; Student; Study Program.

1. PENDAHULUAN

Menurut data dari Kementerian Pendidikan, Kebudayaan, Riset dan Teknologi (Kemendikbud Ristek) menunjukkan bahwa jumlah mahasiswa di Indonesia mengalami kenaikan sekitar 4,02% pada tahun 2022, dengan rata-rata lulus tepat waktu dalam lima tahun (2019 - 2023) adalah sekitar 31,5%. Sebuah program studi dianggap memiliki kualitas yang baik salah satunya dapat dilihat dari jumlah mahasiswa yang lulus tepat waktu sebagai salah satu indikator lulusan saat akreditasi program studi, namun jika jumlah mahasiswa yang tidak lulus tepat waktu lebih banyak, maka akan memberikan nilai yang kurang pada indikator lulusan. Data mahasiswa yang ada di program studi dapat bermanfaat jika diolah secara maksimal, seperti untuk memprediksi jumlah mahasiswa yang akan lulus tepat waktu. Teknik dalam memprediksi jumlah mahasiswa yang lulus tepat waktu merupakan langkah awal dalam proses perencanaan dari program studi untuk mengambil keputusan yang tepat terhadap mahasiswa yang tidak lulus tepat waktu sedini mungkin.

Informasi strategis untuk mendukung pengambilan keputusan dengan cepat dan akurat dapat dihasilkan dari sistem informasi berbasis komputer yang biasanya disebut sebagai sistem pendukung keputusan (SPK) atau *decision supports system* (DSS). SPK juga dapat digunakan untuk memprediksi seperti pada penelitian yang dilakukan oleh Riani Lubis dkk. dengan judul Sistem Prediksi Jumlah Mahasiswa Baru Menggunakan Metode Simulasi Monte Carlo [1], dimana penelitian ini menggunakan pendekatan simulasi untuk memprediksi jumlah mahasiswa baru di periode mendatang.

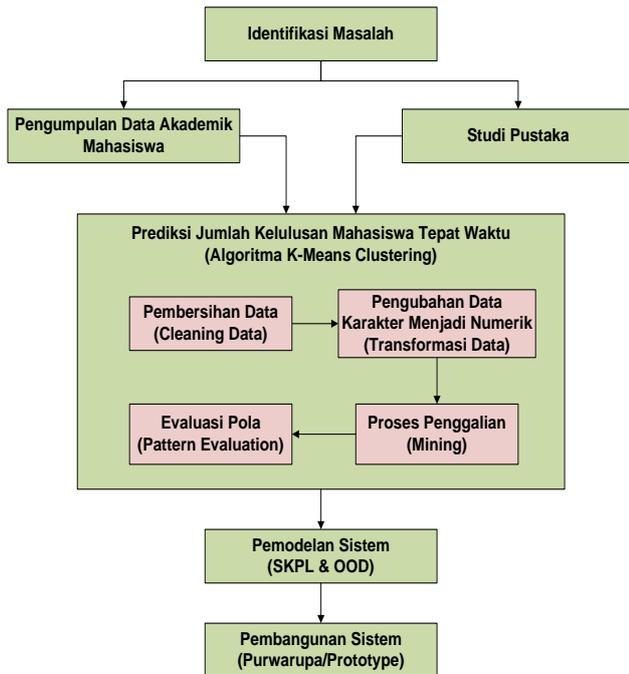
Saat ini dalam membangun sistem pendukung keputusan sudah menerapkan model keputusan strategi berdasarkan analisa data (data mining). Metode atau algoritma data mining yang digunakan pun cukup banyak seperti algoritma Agglomerative Hierarchical Clustering (ACH) yang digunakan dalam menentukan strategi promosi calon mahasiswa baru [2], algoritma Random Forest yang digunakan untuk memprediksi kanker paru [3], algoritma K-Nearest Neighbor (KNN) yang digunakan untuk memprediksi penjualan kusen terlaris [4], algoritma Frequent Pattern Growth (FP Growth) yang digunakan untuk menentukan kecenderungan mahasiswa mengambil mata kuliah pilihan [5], algoritma C5.0 yang digunakan untuk menentukan penerima bantuan sosial [6], beberapa penelitian yang menggunakan algoritma Naïve Bayes [7], [8], dan [9], beberapa penelitian menggunakan algoritma C4.5 [10], [11], [12], dan [13], algoritma K-

Means Clustering yang digunakan pada beberapa penelitian [14], [15], [16], [17], [18], [19], dan [20], serta masih banyak lagi algoritma-algoritma lainnya yang mengabungkan SPK dengan data mining. Penelitian ini menggunakan algoritma K- Means Clustering, dimana algoritma ini memberikan kemudahan implementasi, efisiensi komputasi, dan konsumsi memori yang rendah dibandingkan dengan teknik *clustering* lainnya [21].

Beberapa penelitian yang ditujukan untuk memprediksi jumlah kelulusan tepat waktu sudah pernah dilakukan, seperti penelitian yang dilakukan oleh M. Risky Qisthiano, dkk. menggunakan data mahasiswa alumni dengan kriteria jurusan, jenis kelas pilihan, nilai setiap mahasiswa mulai dari semester pertama sampai semester empat, data tahun masuk dan tahun kelulusan mahasiswa serta algoritma yang digunakan adalah algoritma Decision Tree [22] dan penelitian yang dilakukan oleh D.A. Putra & M. Kamayani yang menggunakan data mahasiswa yang sudah lulus dengan atribut jenis kelamin, IPK semester satu sampai semester empat dan sks semester satu sampai semester empat dengan menggunakan algoritma Naïve Bayes [23], selain itu penelitian yang dilakukan oleh Fahrim Irhamna Rahma, dkk. yang menggunakan algoritma yang sama dengan penelitian ini yaitu algoritma K-Means Clustering, tapi dengan menggunakan atribut/kriteria masa studi, total SKS yang diambil sampai semester tujuh, IPK, IP semester satu sampai semester tujuh [24]. Sedangkan pada penelitian ini data yang digunakan adalah data mahasiswa yang masih aktif dari semester satu sampai semester enam dengan kriteria yang diambil dari data mahasiswa adalah IPK dari semester satu sampai semester enam, jumlah sks yang lulus sampai semester enam, jumlah semester yang telah diambil pada semester yang berjalan, jumlah cuti hingga semester berjalan dan asal sekolah. Jumlah cuti dijadikan kriteria karena semakin sering mengambil cuti pasti akan berpengaruh terhadap waktu kelulusan, sedangkan asal sekolah dijadikan kriteria karena untuk melihat apakah ada pengaruhnya jumlah kelulusan tepat waktu antara mahasiswa yang berasal dari sekolah umum dengan sekolah kejuruan, dengan demikian kemungkinan hasil prediksi yang diperoleh lebih akurat.

2. METODE DAN BAHAN

Metode penelitian dalam penelitian ini menggunakan pendekatan kuantitatif deskriptif, dengan tahapan penelitian seperti terlihat pada gambar 1, dengan aplikasi yang dibangun berupa purwarupa (prototype).



Gambar 1. Tahapan Penelitian

Penjelasan dari tahapan-tahapan yang ada pada Gambar 1 sebagai berikut:

a) **Identifikasi Masalah.** Identifikasi masalah adalah suatu proses untuk pengenalan masalah yang berhubungan dengan penelitian yang dilakukan berdasarkan fakta yang ada.

b) **Pengumpulan Data Akademik Mahasiswa.** Tahap ini peneliti melakukan pengumpulan data pendukung. Data pendukung yang digunakan dalam penelitian ini adalah data akademik mahasiswa program studi Teknik Informatika angkatan tahun 2020 dari semester satu sampai semester enam (semester Genap tahun akademik 2022/2023) dengan field yang digunakan adalah NIM, nama mahasiswa, IPK sampai semester enam, jumlah sks yang lulus sampai semester enam, jumlah semester yang sudah ditempuh sampai semester berjalan, jumlah cuti dan asal sekolah.

c) **Studi Pustaka.** Tahap ini dilakukan studi pustaka yang berkaitan dengan topik penelitian yang bersumber dari jurnal-jurnal, buku, dan pustaka lainnya. Algoritma K-Means Clustering menjadi bahan untuk dipelajari terutama dalam hal memprediksi.

d) **Prediksi Jumlah Kelulusan Mahasiswa Tepat Waktu.** Tahap ini dimulai dari pembersihan data (cleaning data), kemudian data yang bentuknya karakter diubah menjadi numerik (transformasi data). Setelah proses transformasi data, akan dilakukan proses mining menggunakan algoritma k-means clustering.

Tahapan dalam algoritma K-Means adalah [24]:

1. Tentukan berapa banyak cluster k pada data set yang akan diolah
2. Tetapkan secara acak data k menjadi sebagai pusat awal lokasi cluster (centroid)
3. Jarak antara objek dengan centroid dihitung menggunakan persamaan Euclidian Distance seperti pada Persamaan 1 [25]:

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

dimana:

d = jarak

n = banyaknya atribut

x_i = atribut ke- i dari objek x

y_i = atribut ke- i dari objek y

4. Alokasikan masing-masing objek ke dalam centroid terdekat
5. Hitung ulang posisi centroid baru dengan mengambil rata-rata dari seluruh data dalam cluster dengan rumus
6. Ulangi langkah c, d dan e sampai tidak ada data lagi yang berubah cluster

Proses terakhir dari tahap ini adalah proses evaluasi pola.

e) **Pemodelan Sistem.** Tahap ini dilakukan analisis dan perancangan sistem yang akan dibangun. Analisis sistem yang akan dibangun menggunakan pendekatan Spesifikasi Kebutuhan Perangkat Lunak (SKPL) sedangkan perancangannya menggunakan pendekatan pemodelan Object Oriented Design (OOD).

f) **Pembangunan Sistem.** Sistem yang dibangun berbasis web dan berupa purwarupa (prototype), serta menggunakan bahasa pemrograman bahasa Python dalam pembuatan aplikasinya.

3. HASIL DAN PEMBAHASAN

Identifikasi Masalah

Masalah yang teridentifikasi pada penelitian ini adalah Kepala Program Studi kesulitan dalam memprediksi jumlah mahasiswa yang akan lulus tepat waktu di semester delapan nantinya.

Pengumpulan Data Akademik Mahasiswa

Proses untuk memprediksi jumlah mahasiswa yang lulus tepat waktu menggunakan algoritma *K-Means Clustering* membutuhkan data masukan berupa NIM, nama mahasiswa, IPK sampai semester 6, jumlah SKS lulus sampai semester 6, jumlah semester yang sudah diambil sampai semester 6, jumlah cuti yang pernah diambil dan jenis asal

sekolah. Data yang digunakan pada penelitian ini adalah data mahasiswa di salah satu program studi angkatan tahun 2020, maka data masukan yang nanti akan dipakai adalah data mahasiswa tersebut sampai semester 6 yaitu di semester Genap T.A. 2023/2024. Adapun data yang digunakan tersebut, melalui tahapan data mining dan diproses dengan menggunakan algoritma K-Means Clustering sehingga menghasilkan keluaran (*output*) yang diinginkan.

Studi Pustaka

Pustaka yang dipelajari adalah pustaka-pustaka yang berkaitan dengan algoritma K-Means Clustering terutama dalam hal memprediksi, yang bersumber dari jurnal-jurnal, buku, dan pustaka lainnya.

Prediksi Jumlah Kelulusan Mahasiswa Tepat Waktu

Tahapan ini terdiri dari pembersihan data, transformasi data, mining data dan evaluasi pola. Berikut ini adalah tahapan prediksi jumlah kelulusan tepat waktu dengan menggunakan algoritma *K-Means Clustering*:

1. Tahap Cleaning Data

Tahap ini dilakukan terhadap data yang akan digunakan dalam proses mining. Aktivitas yang tercakup dalam tahap ini diantaranya adalah membuang data kosong, membuang duplikat data,

memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data. Data mahasiswa yang dijadikan objek penelitian berjumlah 403 mahasiswa, namun setelah dilakukan *cleaning data*, yaitu dengan membuang data kosong maupun data mahasiswa yang telah pindah prodi ataupun pindah universitas, maka didapat data mahasiswa yang akan digunakan berjumlah 335 mahasiswa. Selanjutnya hasil dari *cleaning data* ini akan diambil sebagai sampel dalam perhitungan manual pada tahap mining.

2. Tahap Transformasi Data

Tahap transformasi data adalah proses perubahan data dari karakter menjadi numerik agar dapat diterapkan pada algoritma K-means. Data yang akan ditransformasi pada penelitian ini adalah data Asal Sekolah yang diubah ke dalam format numerik, seperti terlihat pada Tabel 1. Sedangkan Tabel 2 menunjukkan potongan hasil transformasi data mahasiswa angkatan 2020 pada Tahun Akademik 2022/2023.

Tabel 1. Transformasi Data Tipe Asal Sekolah

Asal sekolah	Format Numerik
SMA Umum/PGRI/Plus	3
SMK	2
Lainnya (MA/Pesantren/homeshooling)	1

Tabel 2. Potongan Hasil Transformasi Data Mahasiswa Angkatan 2020 pada T.A. 2022/2023

No	NIM	NAMA	IPK s/d smt. Aktif	SKS Lulus (SKS)	Jumlah smt. yang telah diambil (smt.)	Jumlah Cuti hingga smt. Genap TA 2022/2023 (smt.)	Tipe Asal Sekolah
1	10120001	Rudiansyah Adi Prasetyo	3,07	115	6	0	2
2	10120003	Bayu Rifki Alghifari	3,27	115	6	0	2
3	10120004	Muhammad Fariz Rahmat A	3,49	115	6	0	3
4	10120005	Fahrizal Zamil Mar'ipatullah	2,62	101	6	0	3
5	10120006	Ahmad Salim Asri	1,93	56	6	0	1
6	10120007	Yandi David Riyandi	1,91	76	4	1	2
7	10120008	William Marcellino Putra	3,23	115	6	0	3
8	10120009	Thoriq Hidayat	3,33	115	6	0	2
9	10120010	Irfan Choiruddin Anwar	2,69	115	6	0	3
10	10120011	Rasya Galih Eky Viryadani	3,15	115	6	0	2
...

No	NIM	NAMA	IPK s/d smt. Aktif	SKS Lulus (SKS)	Jumlah smt. yang telah diambil (smt.)	Jumlah Cuti hingga smt. Genap TA 2022/2023 (smt.)	Tipe Asal Sekolah
325	10120904	Bayu Dava Heryana Putra	2,77	47	6	0	3
326	10120905	Johndy Panca	3,73	115	6	0	1
327	10120906	Sansan Fajrian	3,80	115	6	0	2
328	10120907	Muhammad Nur Alviansyah	3,59	115	6	0	2
329	10120908	Rizka Ghinna Auliya	3,78	115	6	0	2
330	10120909	Vincentius Mickey Satya Wisesa	2,84	115	6	0	2
331	10120911	Putu Mafrizal Alif Astika	2,71	109	6	0	2
332	10120912	Syayful Hidayat Abraham	3,07	76	4	2	2
333	10120913	Wellmart Marudut Tua Siahhaan	3,17	71	4	2	3
334	10120917	Ristian Aditya	3,71	115	6	0	2
335	10120918	Dennie Fuzi Alviyanies	3,73	115	6	0	2

3. Tahap Mining dengan menerapkan algoritma K-Means Clustering

Analisis prediksi jumlah mahasiswa yang lulus tepat waktu dengan menggunakan algoritma *K-Means Clustering* melalui langkah-langkah sebagai berikut:

a. Menentukan jumlah klaster (cluster)

Langkah pertama pada algoritma *K-Means Clustering* adalah menentukan jumlah klaster (*cluster*) yang akan digunakan, dimana pada penelitian ini akan digunakan 2 klaster yaitu *Cluster* pertama (C1): Lulus Tepat Waktu (lulus pada rentang waktu 4 tahun atau setara 8 semester), dan *Cluster* kedua (C2): Lulus Tidak Tepat Waktu (lulus telat/lulus pada rentang waktu lebih dari 4 tahun atau setara lebih dari 8 semester).

b. Menentukan pusat cluster (centroid)

Centroid kriteria yang digunakan beserta notasinya adalah sebagai berikut:

X_1 = IPK terakhir pada semester genap T.A. 2022/2023

X_2 = Total SKS yang telah diambil hingga semester genap T.A. 2022/2023

X_3 = Jumlah semester yang telah diambil hingga semester genap T.A. 2022/2023

X_4 = Jumlah cuti yang pernah diambil hingga semester genap T.A. 2022/2023

X_5 = Tipe Asal sekolah

Pusat cluster awal diambil secara acak dari data yang digunakan serta mewakili setiap *cluster*. Tabel 3 menunjukkan *centroid* awal.

c. Menentukan jarak objek ke centroid dengan menggunakan persamaan jarak Euclidian

Tahap ini dilakukan untuk menghitung jarak dari setiap data ke setiap pusat cluster yang ada dengan Persamaan (1). Adapun rekapitulasi hasil perhitungan jarak objek ke centroid pada iterasi 1 terhadap data set yang sudah disiapkan dapat dilihat pada Tabel 4.

d. Menghitung pusat cluster baru

Hasil perhitungan jarak terhadap *centroid* awal (Tabel 4) menjadi dasar untuk dilakukan perhitungan *centroid* baru untuk iterasi 2 yang akan menempati *cluster* C1 dan C2. Data yang menempati masing-masing *cluster* tersebut dijumlahkan kemudian dihitung nilai rata-ratanya. Nilai inilah yang akan menjadi pusat *cluster* pada iterasi 2.

e. Menghitung jarak objek terhadap centroid baru

Tahap berikutnya adalah melakukan perhitungan jarak objek dengan *centroid* baru pada iterasi 2 seperti yang dilakukan pada langkah 3. Kemudian dilakukan pengelompokkan kembali terhadap data tersebut berdasarkan *cluster* C1 dan C2, sehingga ditemukan terjadi beberapa perubahan *cluster* pada data tersebut seperti terlihat pada Tabel 5.

Tabel 3. Centroid Awal Pada Setiap Centroid Kriteria

Data ke-	Centroid	X1	X2	X3	X4	X5	Kluster
111	Y1	3,69	115	6	0	3	Lulus Tepat Waktu
224	Y2	2,74	70	6	1	2	Lulus Tidak Tepat Waktu

Tabel 4. Potongan Hasil Perhitungan Jarak Objek Ke Centroid Pada Iterasi 1

NO	X1	X2	X3	X4	X5	Jarak ke C1	Jarak ke C2	Kluster
1	3,07	115	6	0	2	1,18	45,01	C1
2	3,27	115	6	0	2	1,08	45,01	C1
3	3,49	115	6	0	3	0,20	45,03	C1
4	2,62	101	6	0	3	14,04	31,03	C1
5	1,93	56	6	0	1	59,06	14,09	C2
6	1,91	76	4	1	2	39,12	6,38	C2
7	3,23	115	6	0	3	0,46	45,02	C1
8	3,33	115	6	0	2	1,06	45,01	C1
9	2,69	115	6	0	3	1,00	45,02	C1
10	3,15	115	6	0	2	1,14	45,01	C1
...
325	2,77	47	6	0	3	68,01	23,04	C2
326	3,73	115	6	0	1	2,00	45,03	C1
327	3,8	115	6	0	2	1,01	45,02	C1
328	3,59	115	6	0	2	1,00	45,02	C1
329	3,78	115	6	0	2	1,00	45,02	C1
330	2,84	115	6	0	2	1,31	45,01	C1
331	2,71	109	6	0	2	6,16	39,01	C1
332	3,07	76	4	2	2	39,12	6,41	C2
333	3,17	71	4	2	3	44,09	2,68	C2
334	3,71	115	6	0	2	1,00	45,02	C1
335	3,73	115	6	0	2	1,00	45,02	C1

Tabel 5. Potongan Hasil Perhitungan Jarak Objek ke Centroid pada Iterasi 2

NO	X1	X2	X3	X4	X5	Jarak Ke C1	Jarak Ke C2	Min	Kluster	Keterangan
1	3,07	115	6	0	2	19,61	49,76	19,61	C1	TETAP
2	3,27	115	6	0	2	19,60	49,76	19,60	C1	TETAP
...
21	2,78	89	6	0	2	45,57	23,77	23,77	C2	TETAP
22	2,26	99	6	0	2	35,60	33,76	33,76	C2	BERUBAH
...
334	3,71	115	6	0	2	19,59	49,77	19,59	C1	TETAP
335	3,73	115	6	0	2	19,59	49,77	19,59	C1	TETAP

Langkah c sampai langkah e terus dilakukan berulang hingga posisi data tidak mengalami perubahan. Berdasarkan data set yang dijadikan studi kasus, proses perhitungan berakhir pada iterasi ke-6. Tabel 6 menunjukkan potongan hasil perhitungan jarak objek ke *centroid* pada iterasi ke-6.

4. Tahap Evaluasi Pola (Patern Evaluation)

Tahap ini dilakukan terhadap hasil perhitungan jarak objek ke *centroid* pada iterasi terakhir (pada penelitian ini iterasi ke-6). Sehingga berdasarkan Tabel 6 dilakukan pengelompokkan objek (data mahasiswa) berdasarkan *cluster* masing-masing,

yaitu yang berada pada *cluster* C1 (*cluster* mahasiswa yang diprediksi lulus tepat waktu) sebanyak 311 mahasiswa dan data mahasiswa yang berada pada *cluster* C2 (*cluster* mahasiswa yang diprediksi lulus tidak tepat waktu) sebanyak 24 mahasiswa.

Pemodelan Sistem

Perangkat lunak yang dibangun pada penelitian ini berupa purwarupa (*prototype*), dimana model pengembangan perangkat lunaknya menggunakan pendekatan *Object Oriented Design* (OOD). Spesifikasi kebutuhan perangkat lunak (SKPL) dari sistem yang akan dibangun terdiri dari SKPL fungsional seperti

Tabel 6. Potongan Hasil Perhitungan Jarak Objek ke Centroid pada Iterasi 6

NO	X1	X2	X3	X4	X5	Jarak Ke C1	Jarak Ke C2	MIN	KLASTER	KET.
1	3,07	115	6	0	2	19,61	49,76	19,61	C1	TETAP
2	3,27	115	6	0	2	19,60	49,76	19,60	C1	TETAP
...
21	2,78	89	6	0	2	45,57	23,77	23,77	C2	TETAP
22	2,26	99	6	0	2	35,60	33,76	33,76	C1	TETAP
...
334	3,71	115	6	0	2	19,59	49,77	19,59	C1	TETAP
335	3,73	115	6	0	2	19,59	49,77	19,59	C1	TETAP

terlihat pada Tabel 7 dan SKPL non fungsional seperti terlihat pada Tabel 8.

Tabel 7. SKPL Kebutuhan Fungsional

Kode SKPL	Spesifikasi Kebutuhan Fungsional
SKPL-F-001	Sistem menyediakan fasilitas untuk menggugah dataset mahasiswa dalam format Excel
SKPL-F-002	Sistem menyediakan fasilitas untuk menampilkan dataset mahasiswa
SKPL-F-003	Sistem menyediakan fasilitas untuk pembentukan kelompok mahasiswa
SKPL-F-004	Sistem menyediakan fasilitas untuk menampilkan karakteristik kelompok mahasiswa berdasarkan ipk, jumlah sks yang lulus, jumlah semester yang telah diambil, jumlah cuti, dan tipe asal sekolah
SKPL-F-005	Sistem menyediakan fasilitas untuk mengunduh hasil pembentukan kelompok kedalam format excel
SKPL-F-006	Sistem menyediakan fasilitas untuk memberikan label pada kelompok yang telah terbentuk
SKPL-F-007	Sistem menyediakan fasilitas untuk melihat proporsi serta representasi dari kelompok yang terbentuk

Tabel 8. SKPL Kebutuhan Non Fungsional

Kode SKPL	Keterangan
SKPL-NF-001	Sistem yang dibangun menggunakan bahasa pemrograman python
SKPL-NF-002	Sistem bekerja menggunakan web browser
SKPL-NF-003	Antarmuka pengguna menggunakan pustaka streamlit yang berjalan di platform berbahasa python

Pemodelan dari sistem yang dibangun menggunakan diagram *usecase* (Gambar 2), skenario *usecase*, dan *class diagram* aplikasi (Gambar 3).

Diagram *usecase* menggambarkan seluruh fungsionalitas yang dapat dijalankan oleh sistem yang dibangun serta bagaimana hubungan antara fungsionalitas dengan entitas luar sistem. Definisi

aktor pada Gambar 2 merupakan deskripsi singkat yang menguraikan aktor yang terlibat didalam *Usecase Diagram*. Definisi aktor dapat dilihat pada Tabel 9 dan penjelasan dari fungsionalitas yang ada pada *usecase diagram* Gambar 2 dapat dilihat pada Tabel 10.

Tabel 9. Definisi Aktor

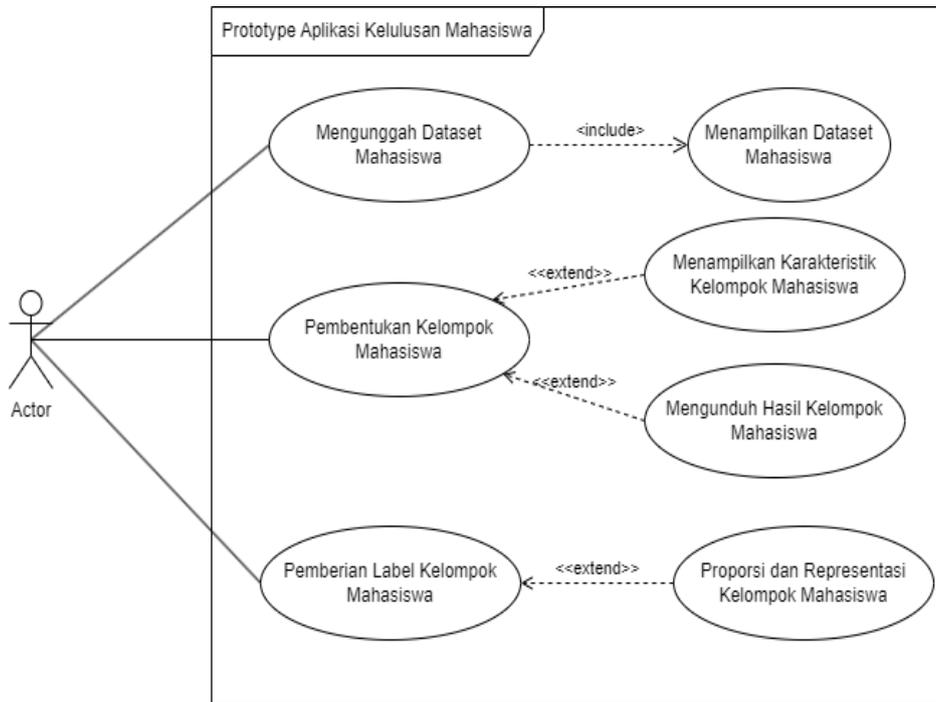
No	Aktor	Deskripsi
1	Kaprodi	Aktor ini memiliki tanggung jawab memantau kondisi mahasiswa dibawah prodi yang dinaungi

Tabel 10. Definisi Usecase Sistem Prediksi Kelulusan Tepat Waktu Menggunakan Algoritma K-Means Clustering

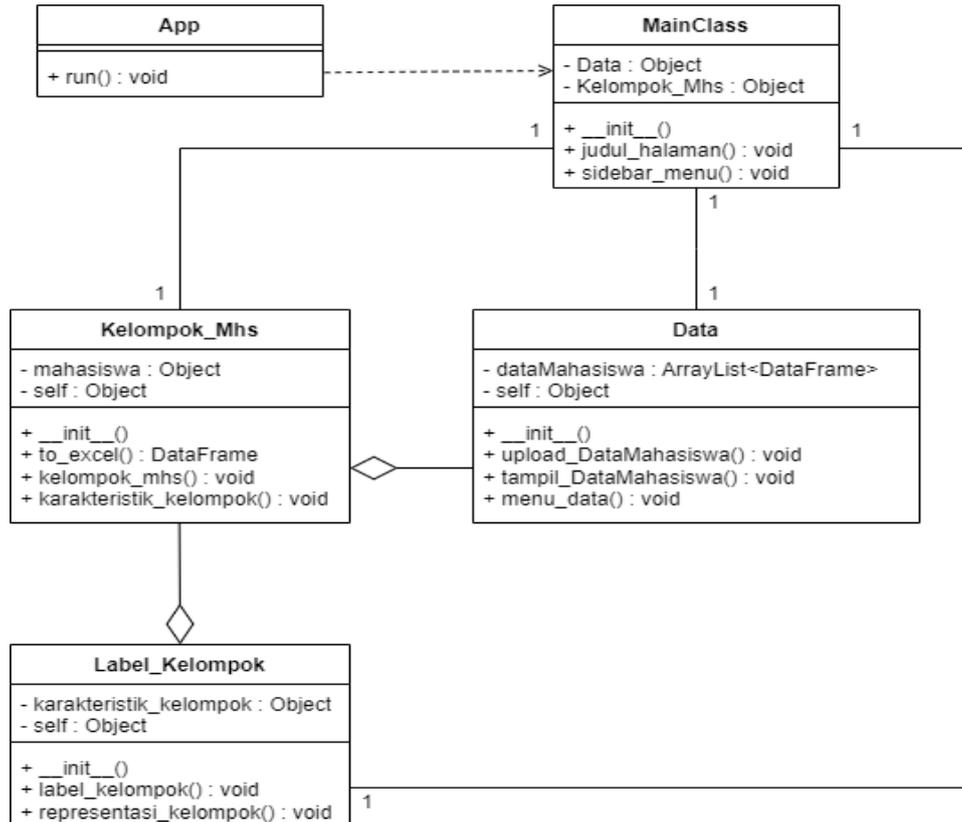
Usecase	Spesifikasi
Menggugah Dataset Mahasiswa	Sistem mempunyai fungsionalitas untuk dapat menerima file dataset mahasiswa dalam format Excel yang diunggah oleh kaprodi
Menampilkan Dataset Mahasiswa	Sistem mempunyai fungsionalitas untuk dapat menampilkan isi file Dataset Mahasiswa yang diunggah oleh Kaprodi.
Pembentukan Kelompok Mahasiswa	Sistem mempunyai fungsionalitas untuk dapat membentuk kelompok mahasiswa berdasarkan kebutuhan ipk, sks yang telah diambil, lama cuti, dan tipe sekolah
Menampilkan Karakteristik Kelompok Mahasiswa	Sistem mempunyai fungsionalitas untuk dapat menampilkan karakteristik kelompok mahasiswa yang terbentuk berdasarkan ipk, sks yang telah diambil, lama cuti, dan tipe sekolah
Mengunduh Hasil Kelompok Mahasiswa	Sistem mempunyai fungsionalitas untuk dapat mengunduh hasil pembentukan kelompok kedalam format file excel
Pemberian Label Kelompok Mahasiswa	Sistem mempunyai fungsionalitas untuk dapat memberikan label untuk setiap kelompok mahasiswa yang terbentuk
Proporsi dan Representasi Kelompok Mahasiswa	Sistem mempunyai fungsionalitas untuk dapat menampilkan proporsi kelompok yang terbentuk serta memberikan representasi pengetahuan berdasarkan karakteristik setiap kelompok

Class diagram digunakan untuk menggambarkan seluruh rancangan class beserta relasi yang terbentuk antar class tersebut didalam sistem.

Gambar 3 merupakan model class diagram dari sistem yang sedang dibangun.



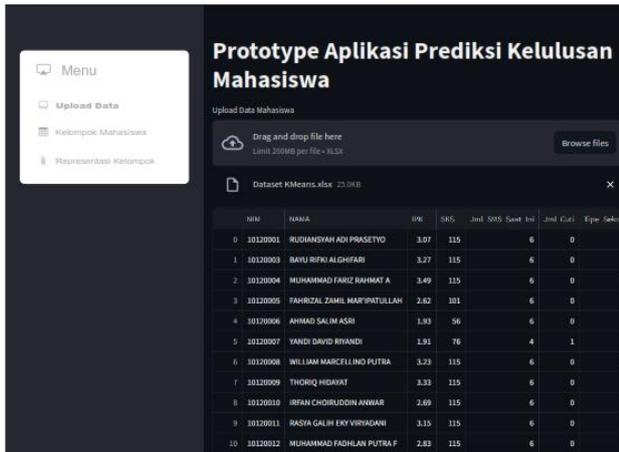
Gambar 2. Usecase Diagram Sistem Prediksi Kelulusan Tepat Waktu



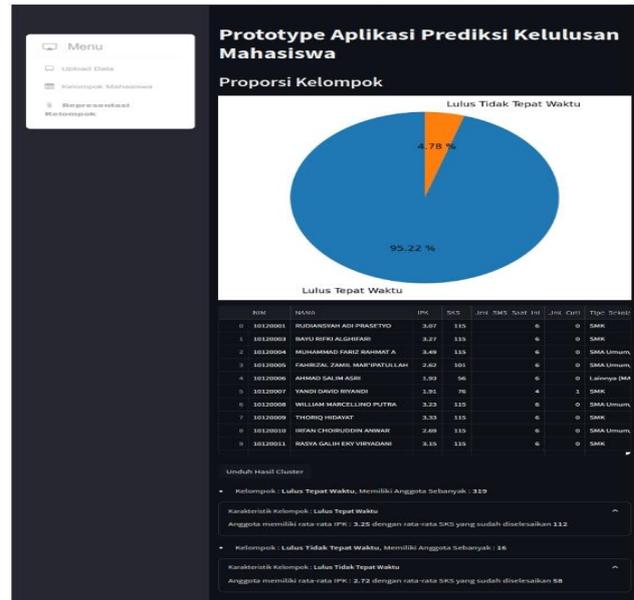
Gambar 3. Class Diagram Prototype Sistem Prediksi Kelulusan Tepat Waktu

Pembangunan Sistem

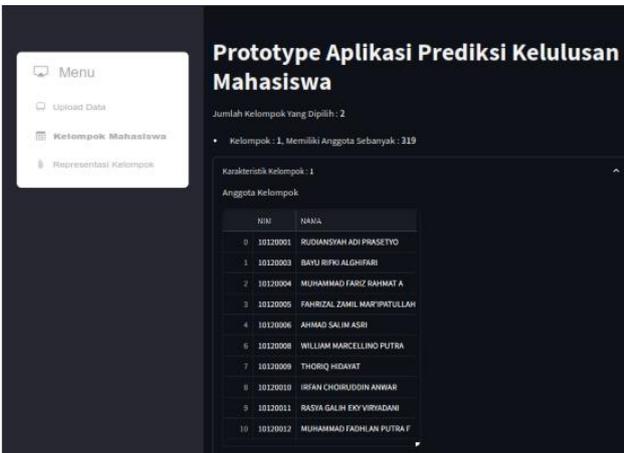
Antarmuka dari sistem yang dibangun terlihat pada Gambar 4 sampai Gambar 7.



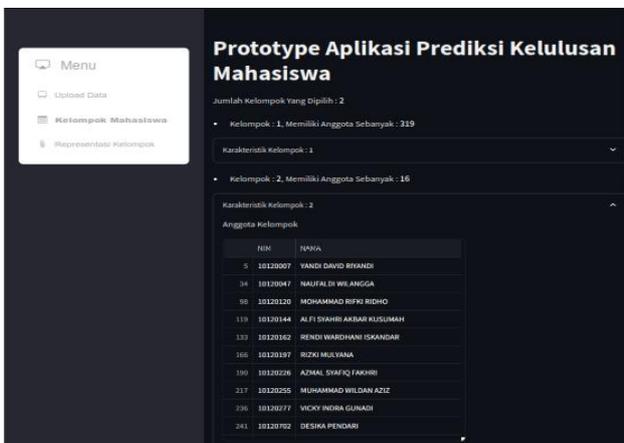
Gambar 4. Antarmuka Mengunggah Dataset Mahasiswa



Gambar 7. Antarmuka Menampilkan Proporsi Antara Lulus Tepat Waktu Dengan Lulus Tidak Tepat Waktu



Gambar 5. Antarmuka Menampilkan Karakteristik Kelompok Kesatu



Gambar 6. Antarmuka Menampilkan Karakteristik Kelompok Kedua

4. KESIMPULAN

Penelitian ini dapat memprediksi jumlah mahasiswa yang kemungkinan lulus tepat waktu sebesar 92,84% atau sejumlah 311 mahasiswa dari 335 mahasiswa angkatan 2020 yang dihitung menggunakan algoritma K-Means Clustering pada semester 6 (tahun akademik 2022/2023), tapi pada kenyataannya mahasiswa angkatan 2020 yang saat ini sedang mengambil mata kuliah skripsi di semester delapan tahun akademik 2023/2024 hanya 194 mahasiswa atau sekitar 57,91% berarti ada selisih sekitar 37,62% dibandingkan hasil yang diprediksikan. Hal ini kemungkinan dipengaruhi oleh adanya mata kuliah prasyarat yang diterapkan oleh prodi tersebut dalam pengambilan mata kuliah skripsi sebagai syarat kelulusan.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih yang terhingga khususnya kepada Direktorat DP3M Unikom yang telah memberikan motivasi dan dorongan atas penelitian ini.

DAFTAR PUSTAKA

- [1] R. Lubis, F. F. Adiwijaya dan C. Hardyanto, "Sistem Prediksi Jumlah Mahasiswa Baru Menggunakan Metode," *Jurnal Sistem dan Teknologi Informasi (Justin)*, vol. 11, no. 3, pp. 466-472, 2023.

- [2] F. F. Adiwijaya, C. Hardyanto dan R. Lubis, "Analisa Data Mining Menggunakan Algoritma Agglomerative Hierarchical Clustering untuk Menentukan Strategi Promosi Calon Mahasiswa Baru," *Jurnal Sistem dan Teknologi Informasi (Justin)*, vol. 12, no. 3, pp. 549-558, 2024.
- [3] L. Sari, A. Romadloni dan R. Listyaningrum, "Penerapan Data Mining dalam Analisis Prediksi Kanker Paru Menggunakan Algoritma Random Forest," *Infotekmesin*, vol. 14, no. 1, pp. 155-162, 2023.
- [4] R. Bahtiar, "Implementasi Data Mining Untuk Prediksi Penjualan Kusen Terlaris Menggunakan Metode K-Nearest Neighbor," *Jurnal Informatika MULTI (JIM)*, vol. 1, no. 3, pp. 203-214, 2023.
- [5] T. Muhlis dan N. Sitompul, "Penerapan Algoritma Fp-Growth dalam menentukan Kecenderungan Mahasiswa Mengambil Mata Kuliah Pilihan," *Jurnal Ilmiah NERO*, vol. 6, no. 1, pp. 59-66, 2021.
- [6] F. S. Batubara, A. H. Hasugian dan H. Santoso, "Aplikasi Sistem Pendukung Keputusan Menggunakan Algoritma C5 untuk Menentukan Penerima Bantuan Sosial," *Jurnal Ilmiah Binary*, vol. 5, no. 2, pp. 223-231, 2023.
- [7] M. Sura, D. Syaifullah dan N. Damastuti, "Sistem Pendukung Keputusan Seleksi Atlet Beladiri Jiu Jitsu Unesa Berdasarkan Data Fisik Menggunakan Naive Bayes," *Jurnal Intra-Tech*, vol. 6, no. 2, pp. 1-10, 2022.
- [8] P. S. Dewi, C. Kurnia dan G. Dudih, "Sistem Pendukung Keputusan Kenaikan Jabatan Menggunakan Metode Algoritma Naïve Bayes Classifier," *Jurnal Teknologi dan Informasi (JATI)*, vol. 11, no. 1, pp. 66-80, 2021.
- [9] N. Widiastuti, A. Hermawan dan D. Avianto, "Komparasi Algoritma Data Mining untuk Prediksi Minat Pencari Kerja," *Jurnal TEKNOINFO*, vol. 17, no. 1, pp. 219-227, 2023.
- [10] A. Ramadhan, "Sistem Pendukung Keputusan Evaluasi Problematika Pendampingan Pembelajaran Daring dengan Algoritma C4.5," *Jurnal Sistem Informasi dan Teknologi (JSISFOTEK)*, vol. 4, no. 2, pp. 58-63, 2022.
- [11] M. F. Hidayattullah, T. Abidin, I. Nururriqzi dan I. Khoeriyah, "Sistem Pendukung Keputusan Latihan Memanah Menggunakan Algoritma Decision Tree," *Jurnal Sistem Komputer (Komputika)*, vol. 10, no. 2, pp. 93-99, 2021.
- [12] A. Merdekawati dan L. K. Rahayu, "Sistem Pendukung Keputusan Pemberian Bantuan Ekonomi Menggunakan Algoritma C4.5," *Jurnal Serambi Engineering*, vol. 6, no. 1, pp. 1452-1464, 2021.
- [13] Z. Gustiana dan A. N. Sari, "Sistem Pendukung Keputusan Penentuan Kelulusan Mahasiswa Menggunakan Kombinasi Algoritma C 4.5 dan Profile Matching," *Jurnal Teknologi Informasi (JTIULM)*, vol. 6, no. 2, pp. 61-70, 2021.
- [14] S. Sari dan J. N. Utamajaya, "Sistem Pendukung Keputusan Penerima Bantuan Langsung Tunai Dana Desa Menggunakan Metode Algoritma K-Means Clustering," *Jurnal Penelitian Ilmu dan Teknologi Komputer (JUPITER)*, vol. 14, no. 1, pp. 150-160, 2022.
- [15] C. S. Dinta Br Sembiring, L. Hanum dan S. P. Tamba, "Penerapan Data Mining Menggunakan Algoritma K-Means untuk Menentukan Judul Skripsi dan Jurnal Penelitian (Studi Kasus FTIK UNPRI)," *Jurnal Sistem Informasi dan Ilmu Komputer (JUSIKOM) Prima*, vol. 5, no. 2, pp. 80-85, 2022.
- [16] O. Herdiana, S. Maulani dan E. A. Firdaus, "Strategi Pemasaran Produk Industri Kreatif Menggunakan Algoritma K-Means Clustering Berbasis Particle Swarm Optimization," *Jurnal Nuansa Informatika*, vol. 15, no. 2, pp. 1-13, 2021.
- [17] F. A. Dinata, A. Nazir, M. Fikry dan I. Afrianty, "Implementasi Data Mining Untuk Prediksi Stok Penjualan Keramik," *Journal of Computer System and Informatics (JoSYC)*, vol. 5, no. 3, pp. 701-708, 2024.
- [18] R. Kurniawan, S. Defit dan S. Sumijan, "Prediksi Tingkat Kerugian Peternak Akibat Penyakit pada Sapi Menggunakan Algoritma K-Means Clustering," *Jurnal Infomasi dan Teknologi (JIDT)*, vol. 3, no. 1, pp. 29-35, 2021.
- [19] M. P. Agus Ariawan, I. B. A. Peling dan G. B. Subiksa, "Prediksi Nilai Akhir Matakuliah Mahasiswa Menggunakan Metode K-Means Clustering (Studi Kasus : Matakuliah Pemrograman Dasar)," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 9, no. 2, pp. 122-131, 2023.
- [20] M. M. Effendi, I. Inka dan A. Siswandi, "Analysis Prediksi Wilayah Rawan Banjir dengan Algoritma K-Means," *Journal of Information System Research (JOSH)*, vol. 5, no. 2, pp. 697-703, 2024.

- [21] L. Morissette dan S. Chartier , “The k-means clustering technique: General considerations and implementation in Mathematica,” *Tutorials in Quantitative Methods for Psychology*, vol. 9, no. 1, pp. 15-24, 2013 .
- [22] M. R. Qisthiano, P. A. Prayesy dan I. Ruswita, “Penerapan Algoritma Decision Tree dalam Klasifikasi Data Prediksi Kelulusan,” *Jurnal Teknologi Terapan (G-Tech)*, vol. 7, no. 1, pp. 21-28, 2023.
- [23] D. A. Putra dan M. Kamayani, “Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Naive Bayes di Program Studi Teknik Informatika UHAMKA,” dalam *Seminar Nasional Teknoka*, Jakarta, 2021.
- [24] F. I. Rahman, S. Mujadilah, T. Wahyuni dan L. Anas, “Prediksi Tingkat Kelulusan Menggunakan K-Means Pada Program Studi Informatika Unismuh Makassar,” *Jurnal FASILKOM*, vol. 13, no. 3, pp. 504-510, 2023.
- [25] Z. Fang dan C. Chiao, “Research on Prediction and Recommendation of Financial Stocks Based on K-Means Clustering Algorithm Optimization,” *Journal of Computational Methods in Sciences and Engineering*, vol. 21, no. 5, pp. 1081-1089, 2021.