

# Identifikasi Ketepatan Waktu Studi Mahasiswa Dengan Algoritma K-Means Clustering Sebagai Rekomendasi Kebijakan Akademik Perguruan Tinggi

Agung Adi Firdaus<sup>1</sup>, Yudha Andriano Rismawan<sup>2</sup>, Farhan Taufikurrahman  
Setiyawan<sup>3</sup>, Arya Kresna Abiaska<sup>4</sup>, Anan Nugroho<sup>5</sup>

Universitas Negeri Semarang  
Sekaran, Gunungpati, Kota Semarang, Jawa Tengah 50229  
agungfirdaus119@gmail.com, yudhaandriano32@students.unnes.ac.id,  
farhants123@students.unnes.ac.id, aryakresnaab@gmail.com,  
anannugroho@mail.unnes.ac.id

## Abstract

*The on-time graduation is one of the important aspects that becomes a benchmark in the success of higher education. Therefore, universities need to analyze the duration of study to build better academic development strategies. The data analysis technique used is clustering using the K-Means Clustering algorithm. The research problem is how to apply and what knowledge is obtained in clustering the duration of studies. The results are divided into 3 clusters of study duration: fast, on time, and late. It was found that the late cluster had the largest frequency, followed by the fast cluster, and the lowest was the on-time cluster. The 3 clusters are distributed based on work status, UKM (activity unit), organizations, and faculties. It is concluded that there is no correlation or difference between students who study while working and students who do not work. There is no correlation and difference between students who do not participate in UKM, participate in UKM 1, participate in UKM 2, participate in UKM 3, or participate in UKM 4. Different results are shown based on the campus organization, it is concluded that following the organization and not joining the organization have a strong correlation and significant differences in the duration of the study. Meanwhile, based on the origin of the faculty, there is no significant correlation or difference, but the clusters of students who graduate on time with the least frequency are FISIP, FIKOM, and DKV.*

**Keywords:** Clustering, Study Duration, K-Means Clustering

## Abstrak

*Perguruan tinggi mempunyai tugas melaksanakan pendidikan yang berkualitas. Ketepatan waktu studi merupakan salah satu aspek penting yang menjadi tolok ukur dalam keberhasilan perguruan tinggi. Oleh karena itu perguruan tinggi perlu menganalisis ketepatan waktu studi mahasiswa untuk membangun strategi pengembangan akademik yang lebih baik. Teknik analisis data yang digunakan adalah clustering menggunakan algoritma K-Means Clustering. Rumusan masalah penelitian adalah bagaimana penerapan dan apa saja pengetahuan yang didapatkan pada clustering ketepatan waktu studi mahasiswa. Prosedur dalam analisis data terdiri dari data understanding, studi literatur, data cleaning, data selection, data transformation, normalisasi data, clustering, data visualization serta pembahasan dan*

*interpretasi hasil. Hasil yang diperoleh adalah terbagi 3 klaster ketepatan waktu studi yang terdiri dari cepat, tepat waktu, dan telat. Diperoleh pengetahuan bahwa klaster telat mempunyai frekuensi terbesar, diikuti dengan klaster cepat, dan terendah adalah klaster tepat waktu. Ketiga klaster tersebut didistribusikan persebarannya berdasarkan status kerja, UKM, organisasi, dan fakultas. Disimpulkan bahwa tidak terdapat hubungan dan perbedaan antara mahasiswa yang berkuliah sambil bekerja dan mahasiswa yang tidak bekerja. Tidak terdapat hubungan dan perbedaan antara mahasiswa yang tidak mengikuti UKM, mengikuti UKM 1, mengikuti UKM 2, mengikuti UKM 3, ataupun mengikuti UKM 4. Hasil berbeda ditunjukkan berdasarkan organisasi kampus, disimpulkan bahwa mengikuti organisasi dan tidak mengikuti organisasi mempunyai hubungan yang kuat dan perbedaan yang signifikan terhadap ketepatan waktu studi. Sedangkan berdasarkan asal fakultas, tidak terdapat hubungan dan perbedaan yang signifikan, namun klaster mahasiswa yang lulus tepat waktu dengan frekuensi paling sedikit adalah FISIP, FIKOM dan DKV.*

**Kata kunci:** *Clustering, Ketepatan waktu studi, K-Means Clustering*

## **1. PENDAHULUAN**

Akreditasi perguruan tinggi merupakan bagian penting untuk menjamin kualitas institusi [1]. Perguruan tinggi mempunyai tugas melaksanakan pendidikan yang berkualitas untuk mahasiswa sehingga mampu melahirkan mahasiswa yang berkompentensi di bidangnya [2]. Salah standar penilaian bagi perguruan tinggi sebagai institusi pendidikan adalah kinerja dari mahasiswa yang terdiri dari *input*, *process* dan *output* [8]. Kelulusan mahasiswa merupakan salah satu aspek penting yang menjadi tolok ukur dalam rekomendasi keberhasilan penyelenggaraan program studi pada perguruan tinggi [3]. Selain berdampak bagi mahasiswa, persentase lulus tepat waktu adalah salah satu penilaian penting dalam akreditasi program studi oleh Badan Akreditasi Nasional untuk Pendidikan Tinggi (BAN-PT). Jika dibiarkan dan tidak ditangani, persentase lulus tepat waktu yang rendah juga menyebabkan penumpukan mahasiswa yang tidak seimbang sehingga berakibat buruk pada kegiatan akademik [4]. Oleh karena itu perguruan tinggi perlu mengidentifikasi ketepatan waktu studi mahasiswa untuk membangun suatu strategi pengembangan lebih lanjut, serta tindakan-tindakan yang dapat dilakukan untuk masa mendatang.

Faktor-faktor yang mempengaruhi ketepatan waktu studi ditentukan oleh faktor internal mahasiswa dan faktor eksternal [5]. Faktor eksternal dapat menjadi bahan rekomendasi serta pertimbangan kampus untuk meningkatkan kinerja akademik jika dianalisis dengan baik. Beberapa contoh faktor eksternal yang dapat berpengaruh diantaranya sarana dan prasarana pendidikan, karakteristik dosen, proses pembelajaran dan sistem administrasi tugas akhir [4] [5]. Bagi seorang mahasiswa dapat dikatakan telah menyelesaikan studi tepat waktu jika mampu menyelesaikan studinya kurang dari atau sama dengan waktu yang telah ditentukan, pada umumnya di Indonesia yaitu 4 tahun [6]. Sedangkan bagi mahasiswa yang menyelesaikan studi lebih dari waktu yang telah ditentukan tersebut

dikatakan tidak lulus tepat waktu. Banyaknya tingkat kelulusan dan ketepatan waktu lulus mahasiswa setiap tahunnya menjadi indikator penting dari keberhasilan akademik suatu perguruan tinggi [10].

Pertumbuhan yang pesat dari penambahan data mahasiswa ini telah menciptakan kondisi dimana suatu lembaga perguruan tinggi memiliki tumpukan data kelulusan mahasiswa. Namun pada saat ini, tumpukan data tersebut belum dimanfaatkan secara maksimal. Padahal tumpukan data tersebut dapat menjadi sebuah informasi yang bermanfaat untuk membantu pihak universitas dalam membuat perencanaan, pengawalan, kebijakan dan bimbingan lebih intensif dengan mengaplikasikan suatu teknik analisis data. *Data mining* sering disebut juga sebagai *Knowledge Discovery in Database* (KDD). KDD adalah kegiatan yang meliputi pengumpulan, pemakaian data, historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar [11]. Secara umum KDD meliputi tiga tahapan, yaitu *pre-processing*, *process (data mining)* dan *post-processing*. Pada tahapan data mining, informasi dari sekumpulan data dapat ditentukan dengan menggunakan kemiripan data satu dengan yang lainnya. Teknik yang sering digunakan adalah asosiasi, klasifikasi dan pengklasteran data [9].

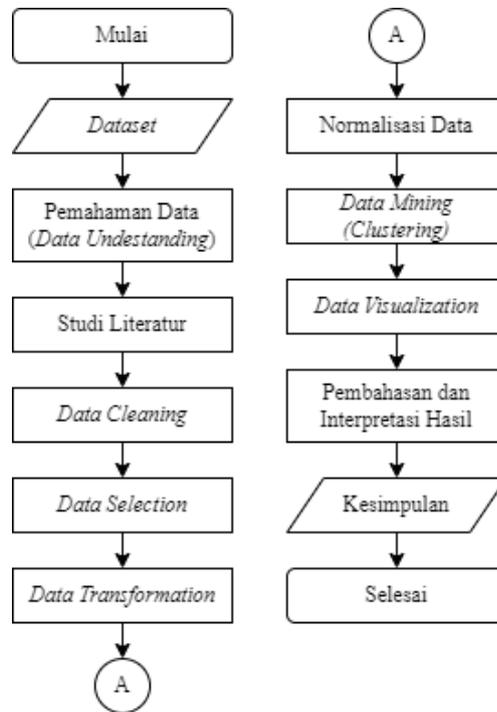
Salah satu teknik analisis data yang dapat dilakukan adalah melalui *clustering*. *Clustering* merupakan teknik pengelompokan data yang tidak mempunyai informasi berlabel khusus [12]. Melalui *clustering*, data ketepatan studi dapat dikelompokkan menjadi beberapa grup berdasarkan kemiripan anggota datanya. Salah satu algoritma *clustering* yang memiliki performa yang paling baik adalah K-Means. Selain memiliki performa yang baik, waktu yang dibutuhkan untuk menjalankan algoritma K-Means juga relatif cepat serta mudah untuk diadaptasi [7]. Manfaat penelitian ini adalah membantu pihak universitas dapat meminimalisasi kegagalan kelulusan mahasiswa dengan membuat perencanaan, pengawalan, kebijakan dan bimbingan lebih intensif.

Berdasarkan latar belakang yang telah disebutkan, tujuan penelitian ini diantaranya adalah: 1) Menerapkan teknik *clustering* ketepatan waktu studi mahasiswa sebagai rekomendasi kebijakan akademik kampus dengan mengaplikasikan algoritma K-Means, dan 2) Mengetahui apa saja informasi yang didapatkan pada penerapan teknik *clustering* ketepatan waktu studi mahasiswa sebagai rekomendasi kebijakan akademik kampus dengan mengaplikasikan algoritma K-Means. Informasi yang dihasilkan bermanfaat bagi pihak universitas sebagai penunjang dalam membuat rekomendasi, perencanaan, pengawalan, kebijakan dan bimbingan lebih intensif untuk kegiatan akademik yang lebih baik dan metode yang ditawarkan menjadi referensi untuk penelitian di masa depan.

## 2. METODOLOGI PENELITIAN

*Software* utama yang digunakan pada penelitian ini diantaranya: 1) Jupyter Notebook, 2) R Studio, 3) Microsoft Excel, dan 4) SPSS. Penelitian dilakukan dengan *coding* untuk analisis data dengan bahasa pemrograman

Python dan R. Metode penelitian yang dilakukan merupakan teknik *clustering* algoritma K-Means *Clustering* dengan tahapan proses dijelaskan secara lengkap dalam diagram alur pada Gambar 1 berikut ini.



Gambar 1. Metode penelitian

## 2.1 Dataset Penelitian

*Dataset* penelitian yang digunakan merupakan data mahasiswa di perguruan tinggi x yang telah lulus yang terdiri dari 8 kolom dan 11.449 baris. Kumpulan data yang berformat .csv dideskripsikan dengan *data frame* menggunakan program python dan ditampilkan pada gambar berikut.

```
# Input raw dataset
df = pd.read_csv('Data - Kualifikasi.csv', sep=';')
df
```

	Nama	Gender	Tinggal_Dengan	Status_Kerja	Biaya	Tgl_Daftar_Kuliah	Alamat	UKM	Organisasi_Kampus	Lama_Kuliah	Fakultas
0	Abbey Butler	Pria	Kos	Belum	Beasiswa	2008	Tangerang	UKM_3	Ya	3,5	DKV
1	Abbey Carpenter	Wanita	NaN	Belum	NaN	2009	Bekasi	UKM_2	Tidak	4	FIKOM
2	Abbey Dubois	Wanita	NaN	Bekerja	Beasiswa	2008	Serang	UKM_4	Tidak	4,5	FIKOM
3	Abbey Gosling	Wanita	Orang Tua	Belum	Beasiswa	2009	Serang	UKM_1	Tidak	5,5	DKV
4	Abbey Logan	Wanita	Orang Tua	Bekerja	NaN	2009	Bekasi	Tidak	Ya	6,5	DKV
...	...	...	...	...	...	...	...	...	...	...	...
11494	Carter Downing	Pria	Orang Tua	Belum	Orang Tua	2009	Tangerang	UKM_3	Ya	4,5	FIKOM
11495	Carter Driscoll	Wanita	NaN	Bekerja	Beasiswa	2008	Karawang	Tidak	Tidak	3,5	DKV
11496	Carter Drummond	Wanita	Kos	Bekerja	Beasiswa	2007	Bogor	UKM_2	Ya	5	FT
11497	Carter Dubois	Wanita	Kos	Bekerja	NaN	2007	Karawang	UKM_4	Tidak	4,5	DKV
11498	Carter Edley	Wanita	Kos	Belum	NaN	2008	Bekasi	UKM_2	Ya	4,5	FIKOM

11499 rows × 11 columns

Gambar 2. Dataset penelitian

## 2.2 Data Understanding

Data yang telah diberikan dikaji dan dipahami jenis dan deskripsi datanya sebelum diolah. Proses ini penting untuk menentukan langkah selanjutnya terkait teknik analisis dan visualisasi data yang akan digunakan.

## 2.3 Studi Literatur

Studi literatur dilakukan untuk mendapatkan dasar dan pengetahuan terkait hal yang diteliti. Dalam kasus ini literatur yang dikaji diambil dari beberapa *keywords*: 1) kinerja akademik perguruan tinggi, 2) *data mining*, 3) *clustering*, dan 4) *K-means clustering*. Informasi yang didapatkan dijadikan sebagai bahan pertimbangan untuk analisis data yang akan dilakukan untuk mendapatkan pengetahuan baru yang sesuai dan bermanfaat terkait topik yang dibahas.

## 2.4 Data Cleaning

Pada tahap ini dilakukan proses pembersihan data yang tidak relevan. Beberapa atribut atau fitur yang dianggap tidak penting dan tidak berpengaruh pada proses *clustering* akan dihilangkan (*cleaning*). Pada proses *clustering* ketepatan waktu studi mahasiswa ini, beberapa atribut tidak digunakan ditampilkan pada tabel di bawah ini.

Tabel 1. Atribut yang tidak digunakan

No.	Atribut	Alasan <i>Cleaning</i>
1	Nama	Atribut ini tidak relevan dan tidak berpengaruh
2	Gender	Atribut ini tidak relevan dan tidak berpengaruh
3	Tinggal_Dengan	Atribut berisi <i>value</i> yaitu kos dan orang tua. Atribut ini tidak terlalu relevan dan tidak berpengaruh. Selain itu banyak sekali data kosong ( <i>missing value</i> ) pada atribut ini mengganggu analisis data jika diikutsertakan
4	Alamat	Atribut ini tidak relevan dan tidak berpengaruh
5	Biaya	Banyak sekali data kosong ( <i>missing value</i> ) pada atribut ini sehingga akan mengganggu analisis data jika diikutsertakan
6	Tgl_Daftar_Kuliah	Atribut ini tidak relevan dan tidak berpengaruh

## 2.5 Data Selection

Data-data yang sesuai dengan kebutuhan analisis diambil dari *database* (.csv) untuk dilakukan analisis guna mendapatkan penemuan pengetahuan dari data (*knowledge discovery*). Atribut yang penting dipilih dan digunakan pada proses selanjutnya ditampilkan pada tabel berikut.

Tabel 2. Atribut yang digunakan

Status_Kerja	UKM	Organisasi_Kampus	Lama_Kuliah	Fakultas
--------------	-----	-------------------	-------------	----------

### 2.6 Data Transformation

Setelah penyeleksian data telah dilakukan, maka dilanjutkan dengan mengubah atau menggabungkan data ke dalam format yang sesuai untuk kemudian diproses dalam data mining (*clustering*). Adapun proses transformasi data ketepatan waktu studi mahasiswa yang telah diperoleh mulai dari tahap *data cleaning* sampai *data selection* adalah sebagai berikut.

Tabel 3. Atribut yang dipakai dan transformasinya

No.	Atribut	Data Transformation
1	Status_Kerja	Belum = 1 Bekerja = 2
2	UKM	Tidak = 1 UKM_1 = 2 UKM_2 = 3 UKM_3 = 4 UKM_4 = 5
3	Organisasi_Kampus	Tidak = 1 Ya = 2
4	Lama_Kuliah	>4,5 tahun = 1 4-4,5 tahun = 2 <4 tahun = 3
5	Fakultas	DKV = 1 FIKOM = 2 FISIP = 3 FTI = 4 FT = 5

### 2.7 Normalisasi Data

Normalisasi data adalah proses membuat beberapa variabel memiliki rentang nilai yang sama, tidak ada yang terlalu besar maupun terlalu kecil sehingga dapat membuat analisis menjadi lebih mudah. Metode normalisasi data yang digunakan adalah *Z-score* atau disebut juga *standard score*. Dengan formula ini, masing-masing nilai pada fitur dikurangi dengan  $\mu$  yang merupakan nilai rata-rata fitur, kemudian dibagi dengan  $\sigma$  yang merupakan standar deviasi. Hasil normalisasi data yang telah dilakukan ditampilkan pada Gambar 3. Sedangkan untuk persamaan normalisasi dengan metode *Z-score* digambarkan pada rumus berikut ini.

$$x_{new} = \frac{x_{old} - \mu}{\sigma} \quad (1)$$

	Status_Kerja	UKM	Organisasi_Kampus	Lama_Kuliah	Fakultas
0	-1.009700	0.715504	1.008121	2.132799	-1.420177
1	-1.009700	0.003406	-0.991945	0.711304	-0.713499
2	0.990393	1.427602	-0.991945	0.711304	-0.713499
3	-1.009700	-0.708692	-0.991945	-0.710191	-1.420177
4	0.990393	-1.420790	1.008121	-0.710191	-1.420177
...	...	...	...	...	...
11494	-1.009700	0.715504	1.008121	0.711304	-0.713499
11495	0.990393	-1.420790	-0.991945	2.132799	-1.420177
11496	0.990393	0.003406	1.008121	-0.710191	1.406534
11497	0.990393	1.427602	-0.991945	0.711304	-1.420177
11498	-1.009700	0.003406	1.008121	0.711304	-0.713499

11499 rows × 5 columns

Gambar 3. Hasil normalisasi data

## 2.8 Data Mining (Clustering)

Data bersih hasil normalisasi yang terdiri dari 5 atribut lalu dilakukan *clustering* menjadi 3 kelompok berdasarkan ketepatan waktu studi yaitu CEPAT (<4 tahun), TEPAT WAKTU (4-4,5 tahun), dan TELAT (>4,5 tahun) menggunakan program Python dengan aplikasi Jupyter Notebook. Teknik *clustering* yang digunakan adalah K-Means. K-Means Clustering membagi objek ke dalam kelompok-kelompok data dengan melihat titik tengah (*centroid*) yang didapatkan. *Clustering* objek ditentukan dari jarak objek dengan titik tengah yang paling dekat. Setelah mengetahui titik tengah terdekat, objek tersebut akan diklasifikasikan sebagai anggota dari kelompok tersebut [7]. Penentuan nilai tengah menggunakan teori jarak *Euclidean* [13] yang digambarkan pada rumus berikut ini.

$$d(i, j) = \left[ \sum_{k=1}^p |x_{ik} - y_{jk}|^2 \right]^{1/2} \quad (2)$$

Dimana:

$d(i, j)$  = Jarak antara objek i dan objek j

$p$  = Dimensi data

$x_{ik}$  = Koordinat dari objek i pada dimensi k

$y_{jk}$  = Koordinat dari obyek j

Setelah dilakukan *clustering*, data hasil diberikan label untuk mendeskripsikan hasil *clustering* agar lebih mudah dipahami. Sedangkan hasil *clustering* ketepatan waktu studi yang didapat ditampilkan pada Gambar 4.

	Status_Kerja	UKM	Organisasi_Kampus	Lama_Kuliah	Fakultas	label
0	1	4	2	3	1	2
1	1	3	1	2	2	1
2	2	5	1	2	2	2
3	1	2	1	1	1	1
4	2	1	2	1	1	0
...	...	...	...	...	...	...
11494	1	4	2	2	2	1
11495	2	1	1	3	1	2
11496	2	3	2	1	5	0
11497	2	5	1	2	1	2
11498	1	3	2	2	2	1

11499 rows × 6 columns

(a)

	Status_Kerja	UKM	Organisasi_Kampus	Lama_Kuliah	Fakultas	label
0	1	4	2	3	1	Cepat
1	1	3	1	2	2	Tepat Waktu
2	2	5	1	2	2	Cepat
3	1	2	1	1	1	Tepat Waktu
4	2	1	2	1	1	Telat
...	...	...	...	...	...	...
11494	1	4	2	2	2	Tepat Waktu
11495	2	1	1	3	1	Cepat
11496	2	3	2	1	5	Telat
11497	2	5	1	2	1	Cepat
11498	1	3	2	2	2	Tepat Waktu

11499 rows × 6 columns

(b)

Gambar 4. Hasil *clustering* ketepatan waktu studi: a) sebelum dilabeli, dan b) setelah dilabeli

## 2.9 Data Visualization

*Data Visualization* berfungsi untuk menampilkan data agar lebih mudah dipahami dan mendapat pengetahuan dari hasil analisis data yang telah dilakukan. Penelitian ini menggunakan *data visualization* dengan bahasa pemrograman R. Jenis visualisasi data yang digunakan pada penelitian ini adalah *scatter plot* dan *bar chart* (diagram batang). *Scatter plot* digunakan untuk mengamati hubungan antara kluster yang terbentuk. Sedangkan diagram batang digunakan untuk menunjukkan perbedaan frekuensi kluster terhadap 4 atribut lain.

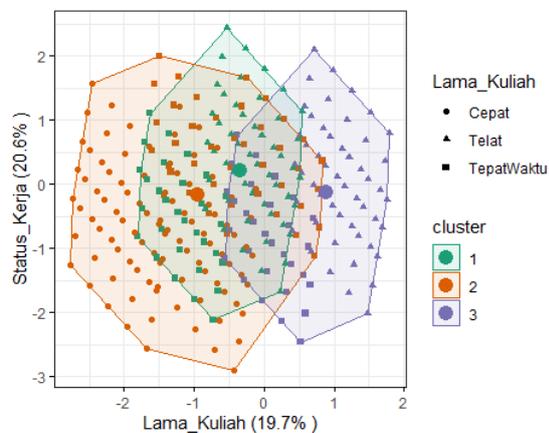
## 3. HASIL DAN PEMBAHASAN

### 3.1 Analisis *Cluster* Ketepatan Waktu Studi Berdasarkan *Scatter Plot*

*Cluster* yang dibentuk berkaitan dengan ketepatan waktu studi terdiri dari 3 kelompok ketepatan waktu studi yaitu CEPAT (*cluster* 3), TEPAT

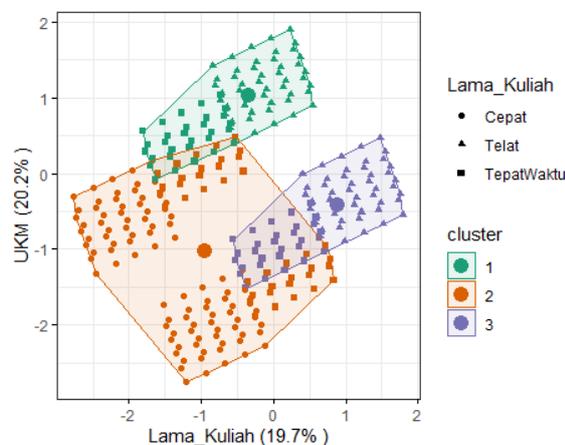
WAKTU (*cluster 2*) dan TELAT (*cluster 1*). Visualisasi kalster menggunakan *scatter plot* dan grafik batang. Penafsiran *scatter plot* dilakukan untuk melihat hubungan antar atribut atau variabel dalam klaster yang terbentuk. Hubungan yang membentuk garis lurus disebut dengan hubungan linier. Hubungan linier dapat membentuk hubungan yang positif dan negatif. Sedangkan jika plotnya menyebar, maka bisa disimpulkan bahwa hubungan antara variabel 1 dengan variabel 2 sangatlah kecil atau tidak ada hubungan sama sekali.

Visualisasi data menggunakan *scatter plot* ditampilkan pada Gambar 5 sampai dengan Gambar 8. Hasil dari visualisasi data 3 kelompok ketepatan waktu studi berdasarkan status kerja ditampilkan pada *scatter plot* berikut ini.



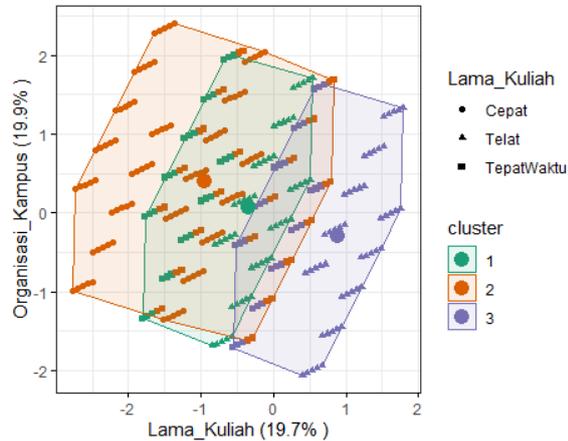
Gambar 5. *Cluster* ketepatan waktu studi berdasarkan status kerja

*Plot* yang terjadi pada Gambar 5 di atas menyebar dan tidak membentuk garis lurus, maka variabel 1 (lama kuliah) dan variabel 2 (status kerja) menunjukkan hubungan yang ada sangatlah lemah. *Cluster* ketepatan waktu studi berdasarkan UKM yang diikuti digambarkan pada Gambar 6 berikut.



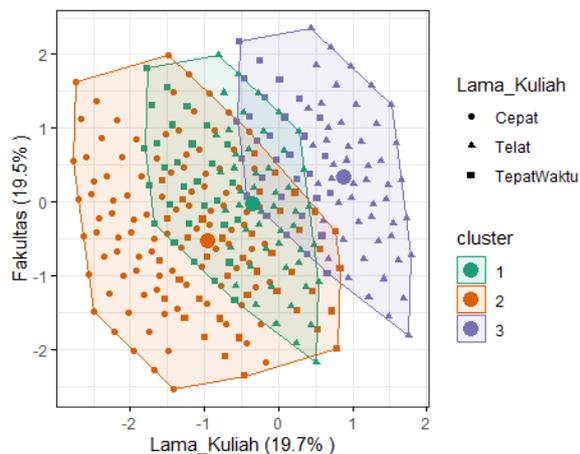
Gambar 6. *Cluster* ketepatan waktu studi berdasarkan UKM

Berdasarkan *plot* yang terbentuk menunjukkan bahwa ketepatan waktu studi berdasarkan UKM yang diikuti memiliki hubungan yang lemah. Namun jika dibandingkan dengan berdasarkan status kerja, UKM memiliki hubungan yang lebih kuat dengan ketepatan waktu studi.



Gambar 7. *Cluster* ketepatan waktu studi berdasarkan organisasi kampus

Hasil berbeda ditunjukkan oleh Gambar 7 yaitu hubungan yang membentuk garis lurus atau hubungan linier. Hubungan linier dapat membentuk hubungan yang positif dan negatif. Jika *plot* yang terjadi seperti pada Gambar 7, maka variabel 1 dan variabel 2 menunjukkan hubungan yang positif. Peningkatan yang terjadi pada variabel 1 diikuti peningkatan oleh variabel 2. Jika variabel 1 mengalami penurunan, maka variabel 2 juga akan mengalami penurunan. Hal tersebut berarti jika satu variabel naik yang lain akan juga akan naik dan kalau satu variabel turun maka yang lain akan turun. Oleh karena itu disimpulkan bahwa klaster ketepatan waktu studi berdasarkan organisasi kampus yang diikuti menunjukkan hubungan yang kuat.

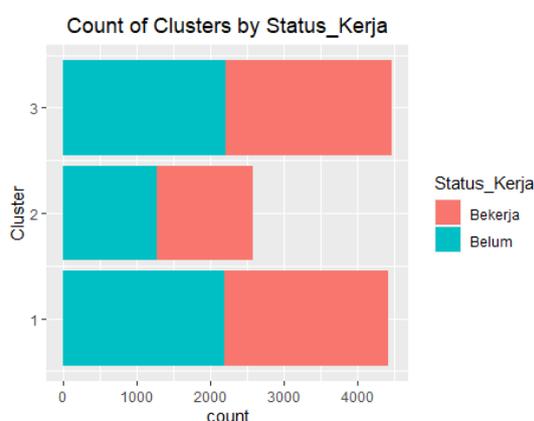


Gambar 8. *Cluster* ketepatan waktu studi berdasarkan fakultas

Sedangkan hubungan yang terbentuk antara ketepatan waktu studi dengan asal fakultas mahasiswa menunjukkan hubungan yang tidak linier atau menyebar. Hal tersebut berarti hubungan antara ketepatan waktu studi dengan fakultas asal mahasiswa memiliki hubungan yang lemah.

### 3.2 Analisis *Cluster* Ketepatan Waktu Studi Berdasarkan Grafik Batang

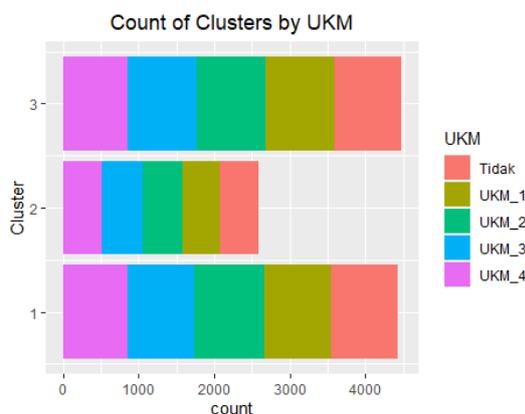
Analisis *cluster* ketepatan waktu studi berdasarkan grafik batang (*bar chart*) ditampilkan pada Gambar 9 sampai dengan Gambar 12. Klaster ketepatan waktu studi berdasarkan status kerja mahasiswa yang terbentuk ditampilkan pada Gambar 9 berikut ini.



Gambar 9. *Cluster* ketepatan waktu studi berdasarkan status kerja

Klaster yang terbentuk pada gambar di atas menunjukkan bahwa tidak terdapat perbedaan yang signifikan pada klaster ketepatan waktu studi berdasarkan status kerja mahasiswa antara mahasiswa yang sambil bekerja dan mahasiswa yang tidak bekerja. Berdasarkan hal tersebut disimpulkan bahwa tidak terdapat perbedaan antara mahasiswa yang berkuliah sambil bekerja dan mahasiswa yang tidak bekerja terhadap ketepatan waktu studinya. Dari pengetahuan tersebut dapat menjadi pertimbangan kampus bahwa status kerja memang tidak memiliki perbedaan yang signifikan. Sehingga rekomendasi yang diberikan adalah mahasiswa diperbolehkan untuk kuliah sambil bekerja, karena selain menambah biaya untuk perkuliahan, pengalaman bekerja juga penting untuk karir mahasiswa setelah lulus.

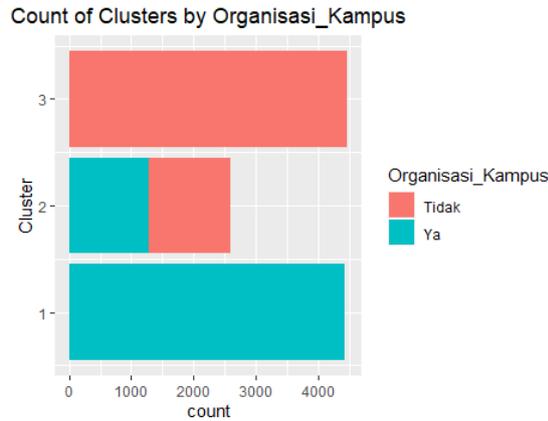
Klaster ketepatan waktu studi berdasarkan jenis Unit Kegiatan Mahasiswa (UKM) yang diikuti oleh mahasiswa yang terbentuk ditampilkan pada grafik batang pada Gambar 10.



Gambar 10. *Cluster* ketepatan waktu studi berdasarkan UKM

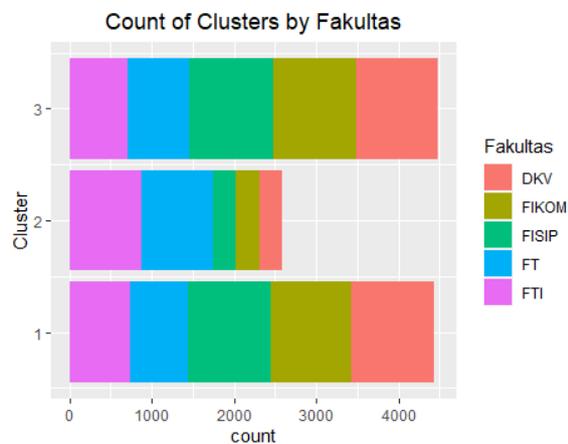
Klaster yang terbentuk pada Gambar 10 di atas menunjukkan bahwa tidak terdapat perbedaan yang signifikan pada klaster ketepatan waktu studi berdasarkan jenis UKM yang diikuti oleh mahasiswa antara mahasiswa yang tidak mengikuti UKM, mengikuti UKM 1, mengikuti UKM 2, mengikuti UKM 3, dan mengikuti UKM 4. Berdasarkan hal tersebut disimpulkan bahwa tidak terdapat perbedaan antara mahasiswa yang tidak mengikuti UKM, mengikuti UKM 1, mengikuti UKM 2, mengikuti UKM 3, ataupun mengikuti UKM 4 terhadap ketepatan waktu studinya. Data tersebut menjadi pertimbangan kampus untuk memberikan rekomendasi kepada mahasiswa bebas untuk tidak mengikuti ataupun mengikuti UKM sesuai dengan minat dan bakat yang dimiliki karena tidak mempengaruhi ketepatan waktu studi mahasiswa.

Sama seperti pada *scatter plot*, perbedaan signifikan juga terjadi berdasarkan organisasi kampus yang diikuti. *Cluster* ketepatan waktu studi berdasarkan keikutsertaan mahasiswa di organisasi kampus menghasilkan perbedaan yang signifikan. Klaster lulus telat (*cluster* 1) disumbang semuanya oleh mahasiswa aktif organisasi dengan frekuensi lebih dari 4.000 mahasiswa. Sedangkan klaster lulus cepat (*cluster* 3) semuanya disumbang oleh mahasiswa tidak ikut organisasi dengan frekuensi juga lebih dari 4.000 mahasiswa. Sedangkan untuk klaster tepat waktu tidak terdapat perbedaan yang signifikan antara mahasiswa yang ikut organisasi dan tidak mengikuti organisasi mempunyai hubungan yang kuat terhadap ketepatan waktu studinya. Oleh karena itu dibutuhkan kebijakan kampus terkait hal ini agar tidak terjadi persentase kelulusan telat yang semakin banyak. Klaster ketepatan waktu studi berdasarkan keikutsertaan mahasiswa di organisasi kampus yang terbentuk ditampilkan pada grafik batang di Gambar 11.



Gambar 11. *Cluster* ketepatan waktu studi berdasarkan organisasi kampus

Kebijakan yang dapat diterapkan yaitu pembatasan jumlah maksimal organisasi yang dapat diikuti, pelatihan *time management* untuk anggota organisasi, ataupun sosialisasi strategi studi tepat waktu bagi organisasi terkait. Hal ini perlu didukung dengan data riset bahwa organisasi apa yang paling banyak berpengaruh terhadap keterlambatan kelulusan pada mahasiswa. Sedangkan pada kluster ketepatan waktu studi berdasarkan fakultas asal mahasiswa yang terbentuk ditampilkan pada Gambar 12.



Gambar 12. *Cluster* ketepatan waktu studi berdasarkan fakultas

Berdasarkan persebaran *cluster* ketepatan waktu studi berdasarkan fakultas asal mahasiswa yang didapat, frekuensi mahasiswa lulus telat tidak berbeda jauh dengan mahasiswa lulus cepat dan frekuensinya seimbang diantara 5 fakultas yang ada. Namun, kluster mahasiswa yang lulus tepat waktu mempunyai frekuensi paling sedikit adalah FISIP, FIKOM dan DKV. Hal tersebut dapat menjadi bahan rekomendasi kepada ketiga fakultas tersebut untuk meningkatkan kinerja akademik terkait dengan ketepatan waktu studi mahasiswa sehingga persentase angka lulus tepat waktu dapat dinaikkan pada masa mendatang.

#### 4. SIMPULAN

Telah diterapkan teknik *data mining clustering* dengan algoritma K-Means terhadap data kelulusan mahasiswa. Informasi yang dihasilkan yaitu *cluster* yang dibentuk berkaitan dengan ketepatan waktu studi terdiri dari 3 kelompok ketepatan waktu studi yaitu cepat, tepat waktu dan telat dengan frekuensi kluster ketepatan waktu studi telat dan ketepatan waktu studi cepat mempunyai hubungan yang lemah dan tidak terdapat perbedaan yang signifikan. Sedangkan ketepatan waktu studi tepat waktu mempunyai frekuensi lebih rendah dari kedua kluster lain. Sedangkan untuk analisis berdasarkan setiap variabel, disimpulkan bahwa hubungan yang ada lemah dan tidak terdapat perbedaan signifikan antara mahasiswa yang berkuliah sambil bekerja dan mahasiswa yang tidak sambil bekerja terhadap ketepatan waktu studinya. Tidak terdapat hubungan dan perbedaan juga antara mahasiswa yang tidak mengikuti UKM, mengikuti UKM 1, mengikuti UKM 2, mengikuti UKM 3, ataupun mengikuti UKM 4 terhadap ketepatan waktu studinya. Hasil berbeda ditunjukkan berdasarkan organisasi kampus, disimpulkan bahwa mengikuti organisasi dan tidak mengikuti organisasi mempunyai hubungan yang kuat dan perbedaan yang signifikan terhadap ketepatan waktu studi. Sedangkan berdasarkan asal fakultas, juga mempunyai hubungan yang lemah dan tidak terdapat perbedaan yang signifikan namun kluster mahasiswa lulus tepat waktu mempunyai yang frekuensi paling sedikit adalah FISIP, FIKOM dan DKV. Penelitian ini menggunakan data kelulusan mahasiswa pada universitas x, sehingga belum tentu representatif pada perguruan tinggi lain. Sehingga untuk kesimpulan yang lebih luas dibutuhkan sampel data dari perguruan tinggi lain. Sedangkan penelitian ini mengaplikasikan algoritma K-Means dikarenakan cocok dengan tipe data yang tersedia dan K-Means mempunyai performa yang baik berdasarkan penelitian relevan. Oleh karena itu dibutuhkan uji implementasi algoritma *clustering* lain yang berpotensi mempunyai performa yang lebih baik.

#### DAFTAR PUSTAKA

- [1] N. D. Larasati and W. S. Jatiningrum, "Analisis Faktor pada Keterlambatan Studi Mahasiswa Teknik Industri Universitas Ahmad Dahlan," *Manaj. Pendidik.*, vol. 16, no. 2, pp. 83–96, 2021, doi: 10.23917/jmp.v16i2.12134.
- [2] F. Firzada and Y. Yuhandri, "Klasterisasi Tingkat Masa Studi Tepat Waktu Mahasiswa Menggunakan Algoritma K-Medoids," *J. Sistim Inf. dan Teknol.*, vol. 3, pp. 162–168, 2021, doi: 10.37034/jsisfotek.v3i3.60.
- [3] Sumartini and Disman, "Analisis Faktor-Faktor yang Mempengaruhi Penyelesaian Studi Tepat Waktu serta Implikasinya terhadap Kualitas Lulusan," *Indones. J. Econ. Educ.*, vol. 1, no. 1, pp. 43–54, 2018, doi: 10.17509/jurnal.
- [4] D. A. Pratiwi, "Faktor-Faktor Yang Mempengaruhi Masa Studi Mahasiswa Pendidikan Teknik Bangunan Universitas Negeri Jakarta."

- 2020.
- [5] I. N. A. J. Putra, P. K. Nitiasih, N. Adil, and G. Gunatama, "Identifikasi Faktor-Faktor Yang Mempengaruhi Lama Masa Studi Mahasiswa Di Fakultas Bahasa Dan Seni," 2014.
  - [6] F. N. Al Amin, Indahwati, and Y. Angraini, "Analisis Ketepatan Waktu Lulus Berdasarkan Karakteristik Mahasiswa FEM dan Faperta Menggunakan Metode Chart," *Xplore*, vol. 2, no. 1, pp. 2–8, 2013, Available: <http://repository.ipb.ac.id:8080/handle/123456789/67157?show=full>
  - [7] Syafnidawaty, "K-means Clustering," *Universitas Raharja*, 2020.
  - [8] Direktorat Jenderal Pendidikan Tinggi, "Buku Panduan Indikator Kinerja Utama," no. 021, p. 73, 2021.
  - [9] E. Prasetyowati and N. Ramadhani, "Sistem Rekomendasi Dan Klasifikasi Kinerja Akademik Mahasiswa Universitas Madura Menggunakan Naive Bayes Dengan Dirichlet Smoothing," *JUTI J. Ilm. Teknol. Inf.*, vol. 16, no. 2, p. 192, 2018, doi: 10.12962/j24068535.v16i2.a688.
  - [10] N. Nasution, K. Djahara, and A. Zamsuri, "Rekomendasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes (Studi Kasus : Fasilkom Unilak)," *J. Fak. Ilmu Komput.*, vol. 1, no. 1, pp. 1–11, 2015. Available: <https://repository.unilak.ac.id/305/1/91-Article-Text-163-1-10-20180211%281%29.pdf>
  - [11] Muhammad Nasir and Verawaty, "Penerapan Algoritma Naive Bayes Classifier Untuk Rekomendasi Kinerja Akademik Mahasiswa Universitas Bina Darma," *JIKO (Jurnal Inform. dan Komputer)*, vol. 5, no. 2, pp. 81–88, 2021.
  - [12] V. Kumar *et al.*, *DATA Algorithms and Applications*, vol. 12, no. 2. 2015. Available: <http://www.sthda.com>
  - [13] P. S. Hasugian, "Analisa Dan Implementasi Metode K-Means Clustering Dalam Prediksi Persediaan Alat Kontrasepsi (Studi Kasus: Kabupaten Deliserdang)," *Jurnal Mantik Penusa*, vol. 1, no. 2, pp. 139-143, 2017.