



Department of Digital Business

Journal of Artificial Intelligence and Digital Business (RIGGS)

Homepage: <https://journal.ilmudata.co.id/index.php/RIGGS>

Vol. 4 No.3 (2025) pp: 2719-2726

P-ISSN: 2963-9298, e-ISSN: 2963-914X

Implementasi Algoritma *K-Means* untuk Pengelompokan Risiko *Drop out* Mahasiswa

Muhammad Raikhan Al Fatah¹, Aang Alim Murtopo², Erni Unggul Sedyu Utami³

^{1,2}Teknik Informatika, STMIK YMI Tegal

³Sistem Informasi, STMIK YMI Tegal

¹ han28af@gmail.com*, ² aang.alim@stmik-tegal.ac.id, ³ eunggulsu@gmail.com

Abstrak

Fenomena mahasiswa putus studi (*drop out*) menjadi tantangan serius bagi perguruan tinggi karena berdampak pada citra institusi, efektivitas pembelajaran, dan efisiensi sumber daya. Penelitian ini bertujuan mengelompokkan risiko *drop out* mahasiswa penerima Kartu Indonesia Pintar (KIP) semester 6 di STMIK YMI Tegal untuk mendukung deteksi dini dan intervensi akademik yang lebih tepat sasaran. Metode penelitian menggunakan algoritma *K-Means Clustering* dengan data Indeks Prestasi Kumulatif (IPK) dan kehadiran mahasiswa, melalui tahapan pra-pemrosesan data, penentuan jumlah kluster optimal menggunakan metode *Elbow*, implementasi *K-Means*, serta evaluasi hasil menggunakan *silhouette score*. Hasil penelitian menunjukkan terbentuknya tiga kluster risiko yaitu rendah, sedang, dan tinggi, dengan nilai *silhouette score* 0,4101 yang mengindikasikan kualitas pemisahan kluster cukup baik; mahasiswa dengan IPK dan kehadiran tinggi tergolong berisiko rendah, sedangkan mahasiswa dengan IPK dan kehadiran rendah tergolong berisiko tinggi. Implikasi penelitian ini adalah model yang dihasilkan dapat digunakan sebagai alat bantu pihak kampus dalam mengidentifikasi mahasiswa berisiko sejak dini, merumuskan kebijakan pendampingan akademik yang lebih efektif, serta menjadi dasar pengembangan sistem pemantauan akademik berbasis data mining yang dapat diperluas dengan algoritma lain dan variabel tambahan untuk meningkatkan akurasi dan generalisasi model.

Kata kunci: Clustering, Data Mining, Drop out, K-Means, Risiko Mahasiswa

1. Latar Belakang

Pendidikan tinggi memiliki peran penting sebagai fondasi utama dalam mencetak sumber daya manusia yang berkualitas dan berdaya saing di era globalisasi serta revolusi industri 4.0. Perguruan tinggi tidak hanya dituntut menghasilkan lulusan dengan kompetensi akademik yang baik, tetapi juga membekali mereka dengan keterampilan relevan untuk menghadapi dunia kerja yang kompleks [1], [2]. Namun, kenyataan di lapangan menunjukkan bahwa tidak semua mahasiswa mampu menyelesaikan studinya tepat waktu. Fenomena mahasiswa putus studi (*drop out*) menjadi tantangan serius yang dihadapi banyak perguruan tinggi, baik negeri maupun swasta [3].

Drop out merupakan kondisi ketika mahasiswa menghentikan proses studinya sebelum menyelesaikan program pendidikan [4]. Penyebabnya beragam, mulai dari faktor ekonomi, masalah pribadi, rendahnya motivasi belajar, hingga prestasi akademik yang kurang memadai. Tingginya angka *drop out* tidak hanya merugikan mahasiswa secara individu, tetapi juga berdampak negatif bagi institusi pendidikan, seperti penurunan citra kampus, rendahnya efisiensi penggunaan sumber daya, serta terganggunya perencanaan akademik [5]. Oleh karena itu, deteksi dini mahasiswa berisiko *drop out* menjadi kebutuhan mendesak dalam manajemen pendidikan tinggi.

Penelitian sebelumnya telah menunjukkan efektivitas metode data mining dalam menganalisis risiko akademik mahasiswa. Salah satu penelitian membandingkan metode *decision tree* dan *deep learning* berbasis GRU dalam memprediksi risiko *drop out* menggunakan data IPK, SKS, dan status keaktifan, dengan hasil bahwa *decision tree* memberikan akurasi lebih tinggi [6]. Penelitian lain menerapkan algoritma *K-Means* untuk membagi mahasiswa menjadi lima kelompok berdasarkan kinerja mata kuliah dan IPK guna mengidentifikasi mahasiswa yang membutuhkan bimbingan akademik [7]. Studi lain mengelompokkan mahasiswa menjadi dua kluster berdasarkan IPK dan nilai mata kuliah untuk mendeteksi risiko *drop out* tinggi dan rendah [8]. Penelitian berikutnya menggunakan *K-Means* untuk mengklasifikasikan risiko akademik menjadi tiga kategori (rendah, sedang, tinggi) berdasarkan IPK dan SKS [9]. Selanjutnya studi yang menerapkan algoritma *K-Medoids* untuk mengatasi kelemahan *K-Means* terhadap outlier, menghasilkan kluster yang lebih stabil dengan atribut IP semester awal, IPK,

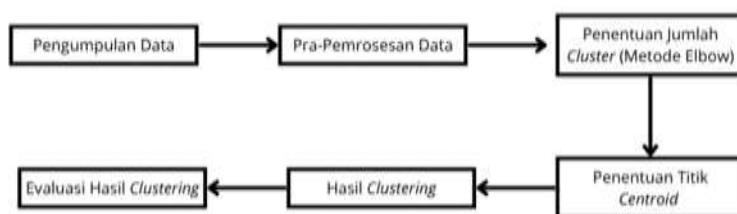
dan status beasiswa [10]. Sebagian besar penelitian tersebut berfokus pada data akademik seperti IPK atau SKS, sedangkan pemanfaatan data kehadiran sebagai indikator keterlibatan mahasiswa masih jarang dilakukan. Padahal, kehadiran merupakan sinyal awal yang mudah diakses dan relevan dalam mendeteksi risiko *drop out*.

Penelitian ini bertujuan mengimplementasikan algoritma *K-Means* untuk mengelompokkan risiko *drop out* mahasiswa penerima Kartu Indonesia Pintar (KIP) semester 6 di STMIK YMI Tegal berdasarkan kombinasi data IPK dan kehadiran. Inovasi penelitian ini terletak pada integrasi dua parameter kunci untuk membentuk tiga kluster risiko (rendah, sedang, tinggi) serta evaluasi kualitas kluster menggunakan *silhouette score*. Dengan pendekatan ini, pihak kampus diharapkan dapat mengenali mahasiswa berisiko lebih cepat dan menyusun kebijakan pendampingan akademik yang tepat sasaran.

2. Metode Penelitian

Studi ini menerapkan pendekatan kuantitatif dengan metode data mining berbasis algoritma *K-Means Clustering*. Tujuan utama dari studi ini untuk mengelompokkan mahasiswa penerima KIP-K STMIK YMI Tegal angkatan 2022 berdasarkan tiga variabel, Indeks Prestasi Kumulatif (IPK), kehadiran (absensi) mahasiswa, serta Satuan Kredit Semester (SKS) guna mengetahui potensi risiko *drop out*.

Algoritma *K-Means* dipilih karena mampu mengelompokkan data tanpa label awal (*unsupervised learning*) dan efektif dalam membagi mahasiswa berdasarkan kesamaan karakteristik akademik. Implementasi dilakukan melalui beberapa tahap, yaitu pra-pemrosesan data untuk menghilangkan nilai kosong dan normalisasi variabel menggunakan *Min-Max Scaling*, penentuan jumlah kluster optimal dengan metode *Elbow*, penerapan *K-Means* untuk membentuk tiga kluster risiko (rendah, sedang, tinggi), serta evaluasi kualitas pengelompokan menggunakan *silhouette score* [11].



Gambar 1. Proses Analisis Data

Gambar 1. Menampilkan proses analisis data pada penelitian ini, dimulai dari pengumpulan data Mahasiswa KIP semester 6, setelah data diperoleh kemudian tahapan selanjutnya pra-pemrosesan data dengan melakukan *cleaning* data dan normalisasi *Min-Max Scaling*, kemudian penentuan jumlah *cluster* menggunakan metode *elbow*, selanjutnya penentuan titik *centroid*, kemudian akan menampilkan hasil *Clustering* dari algoritma *K-Means*, setelah itu akan di evaluasi hasil *Clustering* tersebut menggunakan *silhouette score*.

2.1. Pengumpulan Data

Data diperoleh berasal dari bagian staff akademik STMIK YMI TEGAL yang berupa rekapitulasi nilai IPK dan kehadiran mahasiswa KIP angkatan 2022 selama satu semester. Data yang dikumpulkan kemudian diolah dan dibersihkan agar dapat dianalisis menggunakan metode *K-Means*.

2.2. Pra-pemrosesan Data

Tahap ini meliputi pembersihan data dari duplikasi dan nilai kosong, serta normalisasi menggunakan metode *Min-Max Scaling* agar seluruh variabel berada pada skala yang sebanding. *Min-Max Scaling* metode normalisasi data yang digunakan untuk merubah nilai suatu fitur ke dalam rentang tertentu, biasanya antara 0 dan 1, dengan tujuan menghilangkan perbedaan skala antar variabel sehingga setiap fitur memiliki kontribusi yang seimbang dalam analisis data [12].

2.3. Penentuan Jumlah *Cluster*

Salah satu metode yang umum digunakan untuk menentukan jumlah *cluster* terbaik adalah Metode *Elbow*. Metode *Elbow* dilakukan melalui pengukuran nilai *Within-Cluster Sum of Squares* (WCSS) untuk berbagai jumlah *cluster* (k) dan kemudian memplotkannya dalam grafik [13], [14]. WCSS mengukur seberapa rapat data dalam masing-masing *cluster*, di mana semakin kecil nilai WCSS, maka semakin baik hasil pengelompokan [15]. Namun, penambahan jumlah *cluster* secara terus-menerus tidak selalu memberikan peningkatan yang signifikan. Oleh

karena itu, titik "siku" (*elbow*) pada grafik menjadi indikator jumlah *cluster* yang optimal. Rumus *Within-Cluster Sum of Squares* (WCSS) dituliskan pada persamaan (1)

$$WCSS = \sum_{i=1}^k \sum_{x \in C_i} \|x - \mu_i\|^2 \quad (1)$$

Dimana k adalah jumlah kluster, C_i adalah himpunan data dalam kluster ke-1. x adalah data dalam kluster tersebut, dan μ_i adalah *centroid* kluster ke- i .

2.4. Implementasi algoritma *K-Means*

Prosesnya diawali melalui penetapan jumlah *cluster* (K), lalu menetapkan K titik awal dengan acak sebagai pusat *cluster* (*centroid*). Setiap data dihitung jaraknya ke *centroid* menggunakan rumus jarak Euclidean pada persamaan (2)

$$d(x_{i1} - c_{j1}) = \sqrt{(x_{i1} - c_{j1})^2 + (x_{i2} - c_{j2})^2 + \dots + (x_{in} - c_{jn})^2} \quad (2)$$

Dimana $x_i = (x_{i1}, x_{i2}, \dots, x_{in})$ adalah data ke- i , $c_j = (c_{j1}, c_{j2}, \dots, c_{jn})$ adalah data ke- j , dan $d = (x_i, c_j)$ adalah jarak Euclidean antara data dan *centroid*. Kemudian posisi *centroid* diperbarui menggunakan rata-rata dari semua data yang ada dalam *cluster* tersebut, rumusnya di persamaan (3)

$$c_j = \frac{1}{|c_j|} \sum_{x_i \in c} x_i \quad (3)$$

Dimana C_j sebagai *centroid* baru *cluster* ke- j , C_j adalah himpunan data yang termasuk dalam kluster ke- j . Dan $|c_j|$ adalah jumlah data dalam kluster ke- j .

2.5. Evaluasi Hasil *Clustering*

Kualitas hasil pengelompokan dinilai menggunakan *silhouette score* untuk mengukur seberapa baik pemisahan antar kluster dan homogenitas data dalam kluster.

3. Hasil dan Diskusi

3.1. Deskripsi Data

Tahapan ini menampilkan hasil implementasi algoritma *K-Means* untuk pengelompokan risiko *drop out* mahasiswa penerima KIP STMIK YMI Tegal angkatan 2022, berdasarkan data kehadiran, total tidak hadir, IPK, dan jumlah SKS. Data yang digunakan terdiri dari 23 mahasiswa sebagaimana ditampilkan pada Tabel 4.1

Table 1. Data Mahasiswa

No	Nama Mahasiswa	Hadir	Total Tidak Hadir	IPK	Jumlah SKS
1	AWALIA MUTIARA HIKMAH	92	3	3,87	20
2	ADAM	93	2	3,75	20
3	AFFA MUNTAZAH	94	1	3,72	20
4	MUFIDA ANGGRADITA	95	0	3,93	20
5	NURUL KHAFIDOH	95	0	4	20
...
...
...
23	FEBRIAN NAUFAL PUTRA PRATAMA	86	9	3,4	20

3.2. Pra-pemrosesan Data

Proses ini dilakukan dengan normalisasi menggunakan metode *Min-Max Scaling*, di mana seluruh nilai kehadiran, total tidak hadir, dan IPK diubah ke rentang 0 hingga 1. contoh perhitungan manual pada data mahasiswa bernama Awalia Mutiara Hikmah. Nilai min pada data mahasiswa pada kolom hadir bernilai 67, sedangkan nilai *max* bernilai 95 berikut perhitungan manual normalisasi *Min-Max Scaling*.

$$\frac{92-67}{95-67} = \frac{25}{28} = 0,892857$$

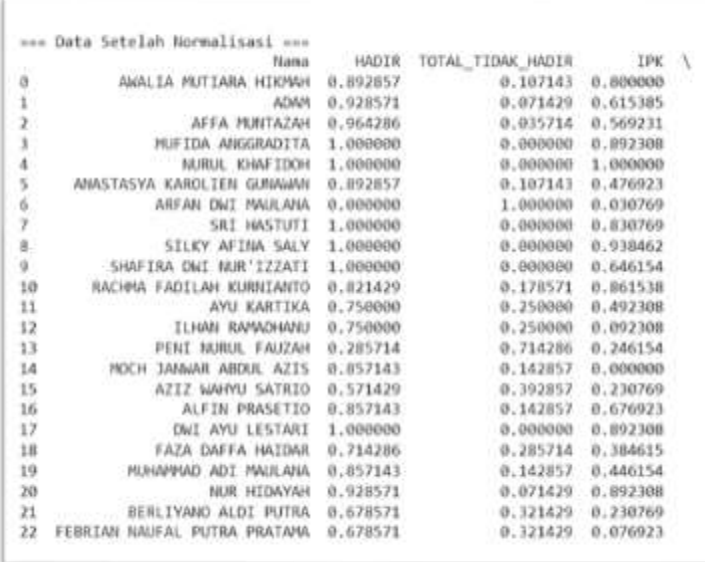
Nilai *min* pada data mahasiswa pada kolom tidak hadir bernilai 0, sedangkan nilai *max* bernilai 28 berikut perhitungan manual normalisasi *min max scaling*.

$$\frac{3-0}{28-0} = \frac{3}{28} = 0,107143$$

Nilai *min* pada data mahasiswa pada kolom IPK bernilai 3,35, sedangkan nilai *max* bernilai 4 berikut perhitungan manual normalisasi *min max scaling*.

$$\frac{3,87-3,35}{4-3,35} = \frac{0,52}{0,65} = 0,8$$

Gambar 2. menunjukkan hasil normalisasi data mahasiswa yang digunakan dalam penelitian ini

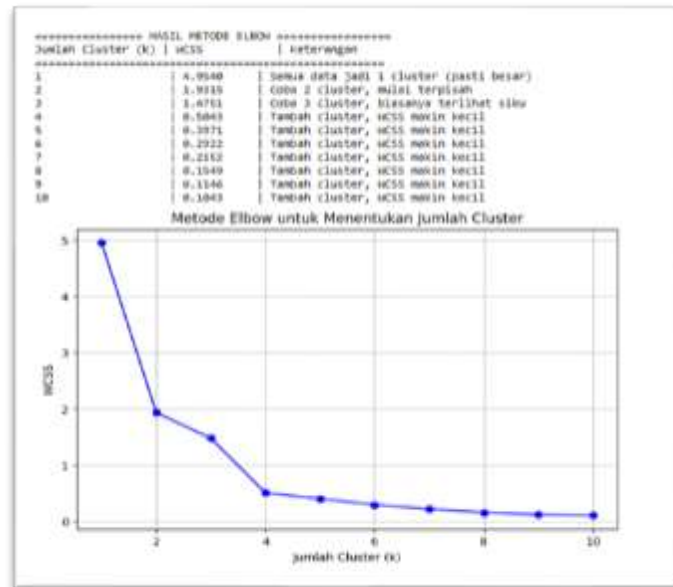


	Nama	HADIR	TOTAL_TIDAK_HADIR	IPK
0	AKALIA MUTIARA HIKMAH	0,892857	0,107143	0,800000
1	ADAM	0,928571	0,071429	0,615385
2	AFFA MUNTAZAH	0,964286	0,035714	0,569231
3	MUFIDA ANGGRADITA	1,000000	0,000000	0,892308
4	NURUL KHAFIDOH	1,000000	0,000000	1,000000
5	ANASTASYA KAROLIEN GUNAWAN	0,892857	0,107143	0,476923
6	ARFAN DWI MAULANA	0,000000	1,000000	0,030769
7	SRI HASTUTI	1,000000	0,000000	0,830769
8	SILKY AFINA SALLY	1,000000	0,000000	0,938462
9	SHAFIRA DANI MUR'IZZATI	1,000000	0,000000	0,646154
10	RACHMA FADILAH KURNIANTO	0,821429	0,178571	0,861538
11	AYU KARTIKA	0,750000	0,250000	0,492308
12	ILHAM RAMADHANU	0,750000	0,250000	0,092308
13	PENI HARUL FAUZH	0,285714	0,714286	0,246154
14	MOCH JANUAR ABDUL AZIS	0,857143	0,142857	0,000000
15	AZIZ WAHYU SATRIO	0,571429	0,392857	0,230769
16	ALFIN PRASETIO	0,857143	0,142857	0,676923
17	DWI AYU LESTARI	1,000000	0,000000	0,892308
18	FAZA DAFFA HAIDAR	0,714286	0,285714	0,384615
19	MUHAMMAD ADI MAULANA	0,857143	0,142857	0,446154
20	MUR HIDAYAH	0,928571	0,071429	0,892308
21	BERLIYANO ALDI PUTRA	0,678571	0,321429	0,230769
22	FEBRIAN NAUFAL PUTRA PRATAMA	0,678571	0,321429	0,076923

Gambar 2. Data Setelah Normalisasi

3.3. Penentuan Jumlah *Cluster* Optimal (Metode *Elbow*)

Setelah data dinormalisasi, langkah berikutnya adalah menentukan jumlah *cluster* yang optimal untuk algoritma *K-Means* menggunakan Metode *Elbow*.



Gambar 3. Hasil Metode *Elbow*

Berdasarkan Gambar 3. terlihat bahwa penurunan WCSS paling signifikan terjadi pada saat $k = 3$. Setelah titik ini, penurunan WCSS menjadi lebih landai. Maka dari itu, jumlah *cluster* optimal ditentukan sebanyak 3 *cluster*, karena pada titik tersebut terjadi perubahan yang mencolok sebelum nilai WCSS menurun secara perlahan.

Setelah mendapatkan jumlah kluster yang optimal, langkah selanjutnya adalah melakukan proses *Clustering* menggunakan algoritma *K-Means*. Tahapan ini menghasilkan tiga *centroid* yang merepresentasikan karakteristik utama dari masing-masing kelompok mahasiswa berdasarkan variabel kehadiran, total tidak hadir, dan IPK. Nilai *centroid* dari tiap kluster dalam skala asli disajikan pada Gambar 4. Berikut

==== CENTROID TIAP CLUSTER (SKALA ASLI) ====				
	HADIR	TOTAL_TIDAK_HADIR	IPK	Cluster
0	93.750	1.25	3.927500	0
1	82.875	12.00	3.455000	1
2	92.000	3.00	3.714286	2

Gambar 4. Penentuan Titik Centroid

Berdasarkan Gambar 4. hasil penentuan titik centroid menunjukkan bahwa *Cluster 0* memiliki kehadiran tertinggi (93,75%) dan IPK tinggi (3,93) yang mengindikasikan risiko *drop out* rendah, *Cluster 1* memiliki kehadiran terendah (82,88%) dengan ketidakhadiran tertinggi (12 kali) dan IPK terendah (3,45) yang menunjukkan risiko *drop out* tinggi, sedangkan *Cluster 2* berada pada kategori sedang dengan kehadiran 92%, ketidakhadiran 3 kali, dan IPK 3,71.

3.4. Hasil *Clustering*

Setelah dilakukan proses *Clustering* dengan algoritma *K-Means*, sebanyak 29 mahasiswa berhasil dikelompokkan ke dalam tiga kluster berdasarkan data kehadiran, total ketidakhadiran, dan IPK.

```
----- DATA MAHASISWA DENGAN KETERANGAN RISIKO -----
      Nama  HADIR  TOTAL_TIDAK_HADIR  IPK  Cluster \
0      ANALIA MUTIARA HIKMAH  92      3  3.87      0
1      ADAM  93      2  3.75      2
2      AFFA BUNTAZAH  94      1  3.72      2
3      MUFIDA ANGGRADITA  95      0  3.93      0
4      NURUL KHAFIDOH  95      0  4.00      0
5      ANASTASYA KAROLIEN GUNAWAN  92      3  3.66      2
6      ARFAN DHI MAULANA  67      28  3.37      1
7      SRI HASTUTI  95      0  3.89      0
8      SILKY AFINA SALY  95      0  3.96      0
9      SHAFIRA DHI NUR'IZZATI  95      0  3.77      2
10     RACHMA FADILAH KURNIANTO  90      5  3.91      0
11     AYU KARTIKA  88      7  3.67      2
12     ILHAN RAMADHANU  88      7  3.41      1
13     PENI NURUL FAUZAH  75      20  3.51      1
14     MOCH JANWAR ABDUL AZIS  91      4  3.35      1
15     AZIZ WAHYU SATRIO  83      11  3.50      1
16     ALFIN PRASETIO  91      4  3.79      2
17     DHI AYU LESTARI  95      0  3.93      0
18     FAZA DAFFA HAIDAR  87      8  3.60      1
19     MUHAMMAD ADI MAULANA  91      4  3.64      2
20     NUR HIDAYAH  93      2  3.93      0
21     BERLIYAND ALDI PUTRA  86      9  3.50      1
22     FEBRIAN NAUFAL PUTRA PRATAMA  86      9  3.40      1
```

Gambar 5. Hasil *Clustering*

Gambar 5. menampilkan hasil *Clustering* menggunakan algoritma *K-Means* menunjukkan bahwa *Cluster 0* berisi mahasiswa dengan tingkat kehadiran di atas 92% dan IPK rata-rata di atas 3,8 yang dikategorikan berisiko *drop out* rendah, *Cluster 1* berisi mahasiswa dengan kehadiran sekitar 85–87% dan IPK sekitar 3,4 yang dikategorikan berisiko tinggi, sedangkan *Cluster 2* memiliki karakteristik menengah baik dari segi kehadiran maupun IPK sehingga dikategorikan berisiko sedang.

Berdasarkan hasil klasifikasi menggunakan algoritma *K-Means*, mahasiswa penerima KIP STMIK YMI Tegal angkatan 2022 berhasil diklasifikasikan menjadi tiga kategori tingkat risiko *drop out*, antara lain: risiko tinggi, sedang, dan rendah. Pembagian ini didasarkan pada kombinasi nilai IPK, jumlah kehadiran, dan total ketidakhadiran.

```

----- DAFTAR MAHASISWA BERDASARKAN RISIKO -----

--- Risiko Tinggi ---
      Nama  TOTAL_TIDAK_HADIR  HADIR  IPK
      ARFAN DWI MAULANA          28    67 3.37
      ILHAN RAMADHANU            7    88 3.41
      PENI NURUL FAUZH          20    75 3.51
      MOCH JANHAR ABDUL AZIS      4    91 3.35
      AZIZ WAHYU SATRIO           11    83 3.50
      FAZA DAFFA HAIDAR           8    87 3.60
      BERLIYANO ALDI PUTRA        9    86 3.50
      FEBRIAN NAUFAL PUTRA PRATAMA 9    86 3.40

--- Risiko Sedang ---
      Nama  TOTAL_TIDAK_HADIR  HADIR  IPK
      ADAM          2    93 3.75
      AFFA MUNTAZAH  1    94 3.72
      ANASTASYA KAROLIEN GUNAWAN  3    92 3.66
      SHAFIRA DWI NUR'IZZATI       0    95 3.77
      AYU KARTIKA       7    88 3.67
      ALFIN PRASETIO     4    91 3.79
      MUHAMMAD ADI MAULANA         4    91 3.64

--- Risiko Rendah ---
      Nama  TOTAL_TIDAK_HADIR  HADIR  IPK
      AMALIA MUTIARA HIKMAH         3    92 3.87
      MUFIDA ANGGRADITA             0    95 3.93
      NURUL KHAFIDOH                0    95 4.00
      SRI HASTUTI                   0    95 3.89
      SILKY AFINA SALLY              0    95 3.96
      RACHMA FADILAH KURNIANTO       5    90 3.91
      DWI AYU LESTARI                0    95 3.93
    
```

Gambar 6. Daftar Mahasiswa Berdasarkan Tingkat Risiko DO

Gambar 6. hasil *Clustering* menunjukkan bahwa kelompok risiko tinggi terdiri dari mahasiswa dengan kehadiran di bawah 88% dan IPK relatif rendah sehingga memerlukan perhatian akademik lebih intensif, kelompok risiko sedang mencakup mahasiswa dengan kehadiran dan IPK pada kisaran menengah yang tetap perlu dimonitor, sedangkan kelompok risiko rendah memiliki kehadiran $\geq 93\%$ dan IPK di atas rata-rata yang menunjukkan kinerja akademik stabil serta relatif aman dari risiko *drop out*.

3.5. Hasil *Clustering*

Evaluasi terhadap hasil *Clustering* dilakukan menggunakan metode *Silhouette Coefficient* bertujuan untuk mengukur kualitas dari pembentukan *cluster*.

```

===== NILAI SILHOUETTE SCORE UNTUK 3 CLUSTER =====
Silhouette Score: 0.4101
    
```

Gambar 7. Hasil Evaluasi *Silhouette Score*

Gambar 7. Menampilkan hasil evaluasi *Silhouette Score* dan mendapatkan nilai 0.4101 untuk 3 *cluster*. Nilai ini mengindikasikan bahwa hasil *Clustering* termasuk dalam kategori cukup baik, di mana sebagian besar data sudah berada pada *cluster* yang tepat, meskipun masih terdapat kemungkinan adanya tumpang tindih antar *cluster*.

Hasil penelitian menunjukkan bahwa algoritma *K-Means* berhasil mengelompokkan mahasiswa penerima KIP STMik YMI Tegal angkatan 2022 ke dalam tiga kluster risiko *drop out* berdasarkan kombinasi data kehadiran, total ketidakhadiran, dan IPK. *Cluster 0* memiliki kehadiran $\geq 93\%$ dan IPK tinggi ($>3,8$) yang mengindikasikan risiko rendah, *Cluster 1* memiliki kehadiran sekitar 85–87% dengan IPK sekitar 3,4 yang menunjukkan risiko tinggi, sedangkan *Cluster 2* berada pada kategori sedang dengan karakteristik kehadiran dan IPK menengah. Hasil evaluasi menggunakan *silhouette score* sebesar 0,4101 menunjukkan kualitas pemisahan kluster cukup baik, di mana sebagian besar data sudah berkelompok dengan tepat meskipun masih ada potensi tumpang tindih antar kluster. Temuan ini menegaskan bahwa integrasi IPK dan kehadiran efektif dalam mendeteksi mahasiswa berisiko, sehingga pihak kampus dapat melakukan intervensi akademik lebih cepat dan tepat sasaran.

4. Kesimpulan

Algoritma *K-Means* mampu mengelompokkan mahasiswa penerima KIP semester 6 di STMIK YMI Tegal ke dalam tiga kategori risiko *drop out* (rendah, sedang, tinggi) secara efektif berdasarkan kombinasi data kehadiran dan IPK, dengan kualitas pemisahan klaster yang cukup baik. Hasil ini membuktikan bahwa indikator kehadiran dan IPK dapat digunakan sebagai dasar deteksi dini mahasiswa berisiko sehingga pihak kampus dapat melakukan intervensi akademik secara tepat sasaran. Untuk penelitian selanjutnya, disarankan penggunaan algoritma *Clustering* lain atau integrasi dengan metode optimasi seperti PSO, penambahan variabel non-akademik seperti faktor ekonomi atau psikologis, serta pengujian pada data mahasiswa dari tahun atau program studi berbeda untuk meningkatkan akurasi dan generalisasi model.

Referensi

- [1] Humiati and D. Budiarti, "Peran Perguruan Tinggi Dalam Meningkatkan Sumber Daya Manusia," *JMM (Jurnal Masyarakat Merdeka)*, vol. 3, no. 1, 2020, doi: 10.51213/jmm.v3i1.46.
- [2] A. Nurdiana, "Peran Kepemimpinan dalam Pengelolaan Sumber Daya Manusia untuk Meningkatkan Kualitas Pendidikan di STAI Siliwangi Garut," *Jurnal Ekonomi Utama*, vol. 2, no. 3, pp. 278–286, Nov. 2023, doi: 10.55903/juria.v2i3.127.
- [3] R. M. Tillah and Firman, "Peran dukungan sosial orangtua terhadap subjective well-being mahasiswa yang terancam drop out: Studi Korelasional," *JURNAL ILMIAH NUSANTARA*, vol. 2, no. 5, pp. 308–320, Jun. 2025, doi: 10.61722/jinu.v2i5.5454.
- [4] T. A. Marzuqi, E. Kristiani, and Marcel, "Prediksi Mahasiswa Drop-Out Di Universitas XYZ," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 11, no. 6, pp. 1345–1350, Dec. 2024, doi: 10.25126/jtiik.2024118689.
- [5] I. Syahputra, "ANALISIS ALGORITMA BACKPROPAGATION UNTUK MENGIDENTIFIKASI MAHASISWA YANG DROP OUT DARI KAMPUS," *Deli Sains Informatika*, vol. 1, no. 1, pp. 1–8, 2021.
- [6] D. W. Puteri, P. W. Buana, and I. M. Sukarsa, "Komparasi Metode Decision Tree dan Deep Learning dalam Meramalkan Jumlah Mahasiswa Drop Out Berdasarkan Nilai Akademik," *Journal of Internet and Software Engineering*, vol. 1, no. 2, p. 12, Mar. 2024, doi: 10.47134/pjise.v1i2.2327.
- [7] A. Fatkhudin, A. Khambali, F. A. Artanto, N. A. Putra Zade, and U. Muhammadiyah Pekajangan Pekalongan, "Implementasi Algoritma Clustering K-Means Dalam Pengelompokan Mahasiswa Studi Kasus (Prodi Manajemen Informatika)," *Jurnal Minfo Polgan*, vol. 12, no. 2, 2023, doi: 10.33395/jmp.v12i2.12494.
- [8] A. S. Lomi, A. A. Pekuwali, and R. T. Abineno, "PENGELOMPOKAN MAHASISWA BERPOTENSI DROP OUT PADA PROGRAM STUDI TEKNIK INFORMATIKA MENGGUNAKAN METODE K-MEANS CLUSTERING Clustering Students With Potential To Drop Out In The Informatics Engineering Study Program Using The K-Means Clustering Method," *Prosiding Semnas SATI*, vol. 3, no. 1, p. 340, 2024.
- [9] C. Lika Mestika Sandy, "Implementasi Metode K-Means Dalam Menentukan Mahasiswa Potensial Drop Out," *Jurnal Elektronika dan Teknologi Informasi*, vol. 3, no. 2, pp. 23–28, 2022.
- [10] S. Bahri and D. M. Midyanti, "Penerapan Metode K-Medoids untuk Pengelompokan Mahasiswa Berpotensi Drop Out," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 10, no. 1, pp. 165–172, Feb. 2023, doi: 10.25126/jtiik.2023106643.
- [11] L. Awaliyah, N. Rahaningsih, and R. D. Dana, "Implementasi Algoritma K-Means Dalam Analisis Cluster Korban Kekerasan Di Provinsi Jawa Barat," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 1, pp. 188–195, 2024, doi: 10.36040/jati.v8i1.8332.
- [12] P. Palinggik Alloorerung, A. Erna, M. Bagussahrir, and S. Alam, "Analisis Performa Normalisasi Data untuk Klasifikasi K-Nearest Neighbor pada Dataset Penyakit," *Jurnal Informatika Sunan Kalijaga*, vol. 9, no. 3, pp. 178–191, Sep. 2024, doi: 10.14421/jiska.2024.9.3.178-191.
- [13] A. Putri Riani, A. Voutama, and T. Ridwan, "Penerapan K-Means Clustering Dalam Pengelompokan Hasil Belajar Peserta Didik Dengan Metode Elbow," *Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD*, vol. 6, no. 1, 2023, doi: 10.53513/jsk.v6i1.7351.
- [14] N. Hendrastuty, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa," *Jurnal Ilmiah Informatika dan Ilmu Komputer (JIMA-ILKOM)*, vol. 3, no. 1, pp. 46–56, Mar. 2024, doi: 10.58602/jima-ilkom.v3i1.26.
- [15] F. Kurnia Rahman, S. Sanjaya, L. Handayani, and F. Insani, "Penerapan Algoritma K-Means Clustering pada Kinerja Mesin Screw press," *Bulletin of Information Technology (BIT)*, vol. 6, no. 2, pp. 59–70, 2025, doi: 10.47065/bit.v5i2.1783.