

PENENTUAN JURUSAN PEMINATAN PRODI FISIKA BERDASARKAN K-MEANS

Nadiza Lediwara¹, Aulia Khamas Heikmakhtiar², Sembada Denrineksa Bimorogo³,
Muhmmad Alvito Faros⁴, Vonna Lestari Dian Subianty⁵, and Gamelia Putri Annetta⁶

^{1,2,3}Informatika, Fakultas Sains dan Teknologi Pertahanan,
Universitas Pertahanan Republik Indonesia

^{4,5}Fisika, Fakultas MIPA Militer, Universitas Pertahanan Republik Indonesia

⁶Biologi, Fakultas MIPA Militer, Universitas Pertahanan Republik Indonesia

¹nadizalediwara@gmail.com; ²auliakphd@gmail.com; ³denri5693@gmail.com;

⁴muhammadalvitofaros@gmail.com; ⁵vonnalestaridian@gmail.com;

⁶annetagamelia@gmail.com

Abstrak — Penelitian ini bertujuan untuk menentukan jurusan peminatan mahasiswa Program Studi Fisika dengan menggunakan metode K-Means pada aplikasi RapidMiner. Pemilihan jurusan peminatan dianggap langkah krusial dalam mengembangkan minat dan bakat mahasiswa di bidang ilmu fisika. Metode K-Means dipilih sebagai pendekatan pengelompokan yang efektif untuk mengidentifikasi pola kesamaan dalam data mahasiswa. Data yang digunakan mencakup rata-rata nilai ujian akhir mahasiswa pada empat mata kuliah tertentu. Hasil penelitian ini menunjukkan pembentukan empat cluster mahasiswa dengan karakteristik nilai yang berbeda. Penelitian ini diharapkan dapat memberikan kontribusi pada pengembangan sistem penentuan jurusan peminatan yang lebih personal dan sesuai dengan potensi masing-masing mahasiswa di Program Studi Fisika. Metode K-Means pada aplikasi RapidMiner terbukti menjadi pendekatan yang efisien dalam menghasilkan rekomendasi peminatan berdasarkan analisis data nilai mahasiswa.

Keywords: Cluster, K-Means, Fisika

1. PENDAHULUAN

Pemilihan jurusan peminatan pada Program Studi Fisika merupakan tahap yang krusial dalam perjalanan akademis mahasiswa. Hal ini merupakan peran penting dalam membimbing kadet mahasiswa untuk mengembangkan minat dan bakat khusus di bidang ilmu fisika. Dalam menghadapi keragaman minat dan keahlian mahasiswa, metode pengelompokan atau *clustering* menjadi pilihan strategis untuk mencapai personalisasi pendidikan yang lebih efektif. Salah satu metode *clustering* yang umum digunakan, terutama dalam konteks penentuan jurusan peminatan, adalah metode K-

Means [1], [2]. Metode K-Means sebagai teknik pengelompokan data memiliki tujuan utama untuk membentuk kelompok-kelompok homogen berdasarkan pola atau karakteristik tertentu [3]. Dalam konteks Program Studi Fisika, penerapan metode K-Means dapat memberikan gambaran yang lebih jelas mengenai kesamaan minat, bakat, atau kecenderungan mahasiswa terhadap bidang fisika tertentu. Hal ini memungkinkan penyelenggara pendidikan untuk memahami dinamika heterogenitas mahasiswa, memberikan dasar untuk menyesuaikan pendekatan pembelajaran dan menawarkan pilihan jurusan peminatan yang lebih sesuai dengan

kebutuhan individu. Latar belakang ini muncul dari realitas bahwa setiap mahasiswa memiliki keberagaman minat dan potensi yang unik. Keragaman ini, sementara menjadi aset, dapat menjadi tantangan bagi penyelenggara pendidikan untuk menyusun pengalaman belajar yang sesuai dengan kebutuhan dan harapan masing-masing mahasiswa. Dengan menerapkan metode K-Means, kita dapat menggali dan memahami pola-pola inherent dalam data mahasiswa, mengidentifikasi kelompok-kelompok yang memiliki karakteristik serupa, dan menghasilkan informasi berharga untuk penyesuaian kurikulum. Proses pengelompokan menggunakan K-Means dianggap sebagai langkah kritis untuk menciptakan kelompok-kelompok homogen [4], [5]. Parameter seperti nilai akademis, proyek penelitian, atau preferensi topik dalam fisika menjadi kriteria penting dalam membentuk kelompok-kelompok ini. Dengan menggunakan analisis berbagai dimensi ini, penyelenggara pendidikan dapat merancang strategi pengajaran yang lebih tepat sasaran. perkembangan akademis serta profesional mahasiswa di bidang fisika. Diharapkan dengan penggunaan metode K-Means tidak hanya menjadi suatu pendekatan teknis, tetapi juga menjadi landasan filosofis untuk menciptakan pengalaman pendidikan yang lebih personal dan relevan. Dengan memanfaatkan potensi mahasiswa di Program Studi Fisika melalui penentuan jurusan peminatan yang terstruktur, diharapkan pula bahwa hasilnya akan memberikan dampak positif dalam membentuk generasi mahasiswa yang terampil dan berkompoten dalam dunia ilmu fisika.

2. LANDASAN TEORI

Penelitian yang membahas mengenai pengelompokan data dengan model K-Means pernah dilakukan untuk mengelompokkan daerah rawan tanah longsor di Provinsi Jawa Barat [6]. Penelitian ini menggunakan membandingkan model K-Means dan K-Medoids. Hasil dari per-

bandingan kedua metode tersebut adalah model K-Means bekerja lebih optimal dalam hal pengelompokan daerah rawan tanah longsor. Pembagian kelompok sejumlah 6 cluster, cluster kedua mempunyai anggota cluster terbanyak, sedangkan untuk jumlah anggota cluster paling sedikit ada pada cluster 5 dengan jumlah anggota 4 daerah dan jumlah kejadian yaitu 106 kejadian. Penelitian yang kedua yang membahas pengelompokan data menggunakan K-Means untuk daerah potensi pertanian karet produktif di Sumatera Utara [7]. Penelitian ini menggunakan software rapid miner untuk mengelompokkan data. Data daerah tersebut dikelompokkan dalam tiga cluster. Dengan adanya pengelompokan ini diharapkan dapat membantu pemerintah untuk mengembankan dan meningkatkan perekonomian daerah. Pemodelan cluster itu sendiri bisa digunakan untuk segmentasi. Salah satu penelitian yaitu membahas segmentasi pelanggan pada PT Infomedia Nusantara [8]. Segmentasi ini didasarkan pada keluhan pelanggan. Data segmentasi ini dikategorikan dalam tiga kategori yaitu sulit, normal, dan mudah untuk melakukan penanganan keluhannya.

3. METODOLOGI

Dalam pelaksanaan penelitian ini, peneliti menggunakan beberapa tahapan CRISP-DM proses [8], [9]. Tahapan awal yaitu *Business Understanding*. Pada tahapan ini peneliti harus mengetahui apa yang akan dikerjakan untuk melakukan proses penelitian. Tahapan kedua yaitu *Data Understanding*. Pada tahapan ini peneliti mulai menyiapkan data-data apa saja yang akan digunakan untuk pengclus teran. Tahapan berikutnya yaitu *Data Preparation*. Pada tahapan ini data-data yang telah dikumpulkan harus dipersiapkan terlebih dahulu. Dipersiapkan dalam proses ini seperti memeriksa kelengkapan data, missing value, dan

hal lainnya sehingga data yang akan diolah dalam penelitian merupakan data yang benar-benar layak untuk diolah. Setelah melakukan *Data Preparation*, tahapan selanjutnya adalah *Modeling*. Tahapan modelling ini merupakan tahapan implementasi data ke dalam suatu model *Data Science*. Data ini dikelompokkan dengan model K-Means. Setelah melakukan *Modelling*, tahapan berikutnya adalah *Evaluation*. Tahapan *Evaluation* ini adalah tahapan untuk memeriksa sudah seoptimal apa pemodelan tersebut bekerja apakah sudah baik atau belum. Tahapan CRISP-DM terakhir adalah *Deployment*. Namun, pada penelitian ini hanya sebatas proses *Evaluation* yang peneliti lakukan.

4. HASIL PEMBAHASAN DAN PENELITIAN

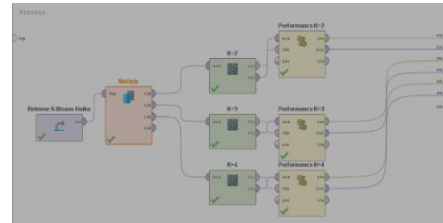
Data yang digunakan untuk pengelompokan kadet mahasiswa Fisika terdiri dari nilai mata kuliah Elektro, Termodinamika, Mekanika, dan Fisika Matematika. Keempat mata kuliah ini berdasarkan hasil wawancara ke kadet mahasiswa merupakan mata kuliah yang cukup sulit di program studi Fisika. Tampilan data nilai keempat mata kuliah tersebut dapat dilihat pada Gambar 1.

	A	B	C	D
1	elektro	termo	mekanika	fismat
2	87.000	83.000	91.000	84.000
3	42.000	78.000	73.000	86.000
4	87.000	43.000	58.000	90.000
5	82.000	77.000	83.000	82.000
6	96.000	100.000	80.000	85.000
7	95.000	82.000	84.000	90.000
8	83.000	82.000	86.000	85.000
9	90.000	96.000	91.000	90.000
10	95.000	96.000	83.000	87.000
11	78.000	71.000	72.000	81.000
12	88.000	98.000	79.000	85.000
13	95.000	90.000	100.000	90.000
14	91.000	87.000	44.000	90.000

Gambar 1. Tampilan Nilai Mata Kuliah

Data mata kuliah tersebut menggunakan tipe data integer. Hal itu dikarenakan untuk proses pemodelan dengan K-Means, data tersebut harus berupa data berbentuk angka. Pada tahapan *preprocessing*, data yang dikumpulkan telah lengkap dan bisa diproses ke tahapan berikutnya. Pemodelan K-Means ini membandingkan hasil pengelompokan dengan jumlah dua, tiga, dan

empat cluster. Dari ketiga jumlah cluster ini maka akan dipilih jumlah cluster terbaik dilihat dari hasil evaluasinya. Data diproses dengan software Rapid Miner. Tampilan pemodelan dapat dilihat pada Gambar 2.



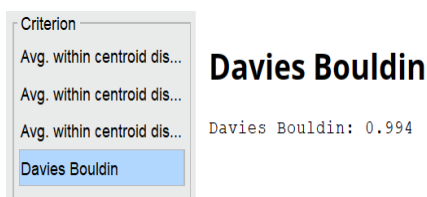
Gambar 2. Pemodelan Rapid Miner

Untuk membandingkan jumlah cluster yang terbaik maka menggunakan operator "multiply" agar bisa mengeksekusi model dan *performance* dalam satu kali eksekusi. Terlihat jumlah cluster yang akan dibandingkan adalah dua, tiga, dan empat. Begitu juga *performance* untuk hasil evaluasi, mengikuti jumlah cluster yang diberikan. Dari tampilan Gambar 2, K=2 adalah dengan jumlah cluster 2, K=3 adalah jumlah cluster 3, dan K=4 adalah jumlah cluster 4. Begitu juga *performance*, operator tersebut mengikuti sesuai dengan jumlah cluster. Sebagai contoh jumlah cluster ditampilkan pada Gambar 3.



Gambar 3. Contoh Jumlah Cluster

Dalam hal evaluasi, nilai evaluasi yang digunakan adalah skor Davies Bouldin. Skor terbaik adalah yang dapat meminimalkan skor yang paling mendekati nilai 0. Tampilan hasil evaluasi terbaik ternyata ada di jumlah cluster 4. Terlihat pada Gambar 4, Gambar 5, Gambar 6 jumlah skor evaluasi berdasarkan jumlah cluster.



Gambar 4. Jumlah Cluster Dua



Gambar 5. Jumlah Cluster Tiga



Gambar 6. Jumlah Cluster Empat

Berdasarkan perbandingan Gambar 4, 5, dan 6, maka hasil evaluasi terbaik adalah pada jumlah cluster sebanyak 4. Skor Davies Bouldin yang didapat dengan jumlah 4 adalah 0.804. Untuk proses selanjutnya maka yang akan dijadikan acuan untuk proses pengelompokan adalah dengan jumlah cluster sebanyak 4. Tampilan hasil cluster dapat dilihat pada Gambar 7.

Row No.	id	cluster	elektro	termo	mekanika	fisimat
1	1	cluster_0	87	83	91	84
2	2	cluster_1	42	78	73	86
3	3	cluster_3	87	43	58	90
4	4	cluster_0	82	77	93	82
5	5	cluster_0	96	80	80	85
6	6	cluster_0	95	82	84	90
7	7	cluster_3	82	62	86	85
8	8	cluster_0	80	96	91	90
9	9	cluster_0	95	96	83	87
10	10	cluster_3	75	71	72	81
11	11	cluster_0	85	98	79	85
12	12	cluster_1	55	80	100	90

Gambar 7. Hasil Cluster

Pembahasan mengenai anggota masing-masing cluster adalah sebagai berikut:

Cluster 0:

- Memiliki nilai yang relatif tinggi pada semua dimensi.
- Cenderung memiliki skor tinggi secara keseluruhan.
- Terlihat bahwa semua baris dalam Cluster 0 memiliki nilai yang tinggi pada setiap dimensi.

- Berisi titik-titik data dengan nilai-nilai berkisar dari 80 hingga 100 di berbagai dimensi.
- Pusat klaster ini tampaknya berada di sekitar 90 untuk sebagian besar dimensi.

Cluster 1:

- Cenderung memiliki nilai yang bervariasi di setiap dimensi, tidak konsisten seperti Cluster 0.
- Tampaknya memiliki variasi skor yang lebih besar di beberapa dimensi.
- Beberapa memiliki variasi skor yang lebih besar dibandingkan dengan Cluster 0.
- Menunjukkan rentang nilai yang lebih luas dibandingkan Klaster 0, dengan nilai-nilai berkisar dari 42 hingga 100.
- Pusat klaster ini tampaknya berada di sekitar 70 untuk dimensi pertama dan sekitar 85 untuk dimensi lainnya

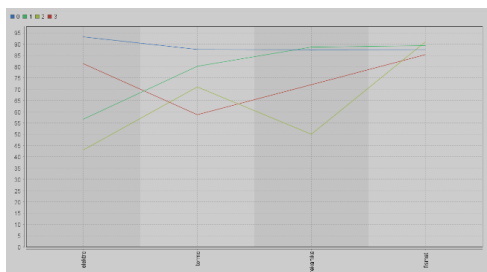
Cluster 2:

- Menunjukkan nilai yang lebih rendah pada beberapa dimensi.
- Skor cenderung lebih rendah secara keseluruhan
- Beberapa memiliki skor yang rendah
- Berisi titik-titik data dengan nilai yang lebih rendah, berkisar dari 31 hingga 67 di dimensi pertama dan 44 hingga 92 di dimensi terakhir.
- Pusat klaster ini tampaknya berada di sekitar 43 untuk dimensi pertama dan sekitar 88 untuk dimensi terakhir.

Cluster 3:

- Memiliki kombinasi skor yang bervariasi di setiap dimensi.
- Tidak menunjukkan pola yang jelas pada satu dimensi tertentu.
- memiliki kombinasi skor yang berbeda setiap dimensi.
- Menunjukkan campuran nilai menengah hingga tinggi di seluruh dimensi, dengan nilai-nilai berkisar dari 75 hingga 87 di dimensi pertama dan 81 hingga 90 di dimensi terakhir.
- Pusat klaster ini tampaknya berada di sekitar 81 untuk dimensi pertama dan sekitar 85 untuk dimensi terakhir.

Dari Gambar 8 terlihat bahwa cluster 0 memiliki nilai yang relatif tinggi dan tetap konstan, sementara cluster 1 menunjukkan kecenderungan nilai yang semakin meningkat dan tetap konstan. Tampilan pergerakan cluster dapat dilihat pada Gambar 8.



Gambar 8. Sebaran Cluster

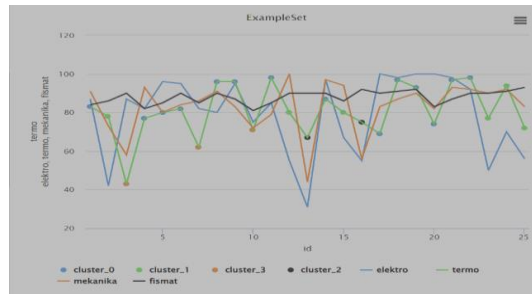
Dari Gambar 9 didapatkan bahwa rata-rata nilai elektro tertinggi diperoleh cluster 0, rata-rata nilai Fisika Matematika tertinggi diperoleh cluster 2, rata-rata nilai Mekanika tertinggi diperoleh Cluster 1, rata-rata nilai Termo tertinggi diperoleh Cluster 0, dari kesimpulan yang di dapat kan kluster 0 terbukti memiliki nilai yang relatif tinggi sesuai dengan hasil yang lainnya.

Attribute ↑	cluster_0	cluster_1	cluster_2	cluster_3
elektro	93.214	56.667	43	81.333
fisimat	87.571	89.333	91	85.333
mekanika	87.500	88.667	50	72
termo	87.643	80.167	71	56.667

Gambar 9. Statistik Cluster

Berdasarkan data yang diberikan, kita dapat menentukan cluster yang cocok untuk peminatan dengan tingkat kekompleksan yang berbeda. Peminatan kompleks dapat diidentifikasi dengan melihat cluster yang menunjukkan karakteristik nilai tinggi dan variasi yang rendah, menandakan konsistensi prestasi tinggi. Dari hasil cluster, cluster 0 muncul sebagai pilihan yang potensial karena menunjukkan nilai yang cenderung tinggi dan variasi yang relatif rendah sepanjang rentang waktu, mengindikasikan stabilitas prestasi tinggi. Untuk peminatan menengah, cluster yang menunjukkan variasi nilai yang sedang mungkin menjadi pertimbangan. Cluster 1 menonjol karena menunjukkan

variasi nilai yang lebih luas dibandingkan cluster 0, dengan nilai yang cenderung berada di tengah rentang. Hal ini dapat diartikan sebagai kemampuan siswa untuk menangani materi yang beragam, membuat cluster 1 menjadi kandidat potensial untuk peminatan menengah. Sementara itu, cluster yang menunjukkan nilai rendah dan variasi yang rendah mungkin menjadi pertimbangan. Cluster 2 dan cluster 3 menarik perhatian karena keduanya memiliki nilai yang cenderung rendah dan variasi yang relatif rendah, menunjukkan konsistensi dalam prestasi yang lebih rendah. Oleh karena itu, keduanya dapat dianggap sebagai kandidat potensial untuk peminatan sederhana. Dengan demikian, melalui analisis karakteristik nilai dari masing-masing cluster, cluster 0, cluster 1, cluster 2, dan cluster 3 dapat dianggap sebagai kandidat potensial untuk peminatan dengan tingkat kekompleksitasan yang berbeda, yaitu kompleks, menengah, dan sederhana. Visualisasi mengenai anggota cluster dapat dilihat pada Gambar 10.



Gambar 10. Visualisasi Cluster

5. KESIMPULAN

- Penelitian ini menggunakan metode K-Means pada aplikasi RapidMiner untuk menentukan peminatan mahasiswa Program Studi Fisika. Hasil analisis cluster menunjukkan pembentukan empat kelompok mahasiswa dengan karakteristik nilai yang berbeda. Cluster tersebut diidentifikasi sebagai potensial untuk peminatan kompleks, menengah, dan sederhana. Melalui evaluasi visualisasi, ditemukan pola karakteristik nilai

setiap cluster, memudahkan identifikasi peminatan yang sesuai.

- Rekomendasi peminatan diberikan berdasarkan prestasi mahasiswa dalam setiap cluster. Metode K-Means terbukti menjadi pendekatan efisien dalam memahami pola kesamaan dalam data nilai mahasiswa dan memberikan dasar yang kuat untuk penentuan jurusan peminatan yang lebih sesuai dengan potensi individu.
- Penelitian ini diharapkan dapat memberikan kontribusi pada penyelenggaraan pendidikan dalam mengoptimalkan proses penentuan jurusan peminatan, meningkatkan personalisasi pendidikan, dan mendukung perkembangan akademis serta profesional mahasiswa di bidang Fisika. Evaluasi yang teliti terhadap karakteristik nilai dan rekomendasi peminatan dapat menjadi panduan yang berharga bagi mahasiswa dan lembaga pendidikan untuk mencapai tujuan akademis yang optimal.

6. REFERENCES

- [1] A. Yudhistira and R. Andika, "Pengelompokan Data Nilai Siswa Menggunakan Metode K-Means Clustering," *Journal of Artificial Intelligence and Technology Information*, vol. 1, no. 1, pp. 20–28, Mar. 2023, doi: 10.58602/JAITI.V1I1.22.
- [2] P. Rahayu, I. Anikah, D. B. Saputra, T. Anelia, and Martanto, "Penerapan Data Mining Metode K-Means Clustering Untuk Analisa Penjualan Rotan," *KOPERTIP: Scientific Journal of Informatics Management and Computer*, vol. 4, no. 2, pp. 42–50, Jun. 2020, doi: 10.32485/KOPERTIP.V4I2.118.
- [3] H. Kurniawan, S. Defit, and Sumijan, "Data Mining Menggunakan Metode K-Means Clustering Untuk Menentukan Besaran Uang Kuliah Tunggal," *Journal of Applied Computer Science and Technology*, vol. 1, no. 2, pp. 80–89, Dec. 2020, doi: 10.52158/JACOST.V1I2.102.
- [4] A. Fikri, B. F. Hutabarat, and U. Khaira, "Komparasi K-Means Clustering Dan Complete Linkage Dalam Pengelompokan Penyaluran Pinjaman *Financial Technology*," *Jurnal Ilmiah Media Sisfo*, vol. 17, no. 2, pp. 228–239, Oct. 2023, doi: 10.33998/MEDIA SISFO. 2023.17. 2.1373
- [5] S. Izzah, S. A. Izzah, I. Zain, and E. O. Permatasari, "Pengelompokan Kabupaten/Kota Berdasarkan Faktor Penyebab Perceraian di Provinsi Jawa Barat," *Jurnal Sains dan Seni ITS*, vol. 11, no. 3, pp. D266–D273, Jul. 2022, doi: 10.12962/j23373520.v11i3.77018.
- [6] M. Herviany, S. Putri Delima, T. Nurhidayah, P. Studi Sistem Informasi, and F. H. Sains dan Teknologi UIN Sultan Syarif Kasim Riau Jl Soebrantas, "Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Daerah Rawan Tanah Longsor Pada Provinsi Jawa Barat," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 1, no. 1, pp. 34–40, Mar. 2021, doi: 10.57152/MALCOM.V1I1.60.
- [7] P. Alkhairi and A. P. Windarto, "Penerapan K-Means Cluster Pada Daerah Potensi Pertanian Karet Produktif di Sumatera Utara," *Seminar Nasional Teknologi Komputer & Sains (SAINTEKS)*, vol. 1, no. 1, Feb. 2019, Accessed: Jan. 15, 2024. [Online]. Available: <http://seminar-id.com/prosiding/index.php/sainteks/article/view/228>

- [8] F. Martinez-Plumed *et al.*, "CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories," *IEEE Trans Knowl Data Eng*, vol. 33, no. 8, pp. 3048–3061, Aug.2021,doi:10.1109/TKDE.2019.2962680.
- [9] S. N. Luqman *et al.*, "Komparasi Algoritma Klasifikasi Genre Musik pada Spotify Menggunakan CRISP-DM:," *Jurnal Sistem Cerdas*, vol. 4, no. 2, pp. 114–125, Aug. 2021, doi: 10.37396/ JSC.V4I2.162.