

SEGMENTASI TINGGI BADAN DAN BERAT BADAN KADET MAHASISWA MENGGUNAKAN K-MEANS CLUSTERING

Nadiza Lediwara¹, Aulia Khamas Heikmakhtiar², Sembada Denrineksa Bimorogo³, Army Kanaya⁴, Almas Shafwan⁵, and Audrey Nur Aisyah⁶

^{1,2,3}Informatika, Fakultas Sains dan Teknologi Pertahanan, ⁴Matematika, Fakultas MIPA Militer, ⁵Fisika, Fakultas MIPA Militer, ⁶Kimia, Fakultas MIPA Militer Universitas Pertahanan Republik Indonesia, ¹nadizalediwara@gmail.com; ²auliakphd@gmail.com; ³denri5693@gmail.com; ⁴armykanayaazzahra@gmail.com; ⁵almasshafwan10@gmail.com; ⁶audreyaisyahh@gmail.com

Abstrak — Penelitian ini merupakan langkah awal yang digunakan sebagai syarat utama serta bekal dan persiapan dalam rangka menentukan Kadet mahasiswa UNHAN RI yang akan bekerja pada instansi pertahanan yang memiliki pengetahuan akademik dan militer. Penelitian ini bertujuan untuk segmentasi kadet mahasiswa berdasarkan tinggi badan dan berat badan yang nantinya akan membantu pembuat keputusan dalam hal pembinaan fisik. Untuk segmentasi kadet mahasiswa ini peneliti menggunakan metode K-Means Clustering. Dari hasil segmentasi didapat tiga cluster yaitu cluster 0, cluster 1, dan cluster 2. Cluster 0 menunjukkan adanya potensi untuk dilakukan pembinaan lebih lanjut, sedangkan cluster 1 dan cluster 2 juga bisa dilakukan pembinaan tapi dengan level sedang dan sederhana.

Kata Kunci: Cluster, K-Means, Berat Badan, Tinggi Badan

Abstract — This research is the first step used as the main requirement as well as provision and preparation in order to determine Indonesian UNHAN student cadets who will work in defense agencies who have academic and military knowledge. This research aims to segment student cadets based on height and weight which will later help decision makers in terms of physical development. To segment student cadets, researchers used the K-Means Clustering method. From the segmentation results, three clusters were obtained, namely cluster 0, cluster 1, and cluster 2. Cluster 0 shows the potential for further coaching, while cluster 1 and cluster 2 can also be coached but at a medium and simple level.

Keywords: Cluster, K-Means, Body Weight, Height

1. PENDAHULUAN

Setiap mahasiswa atau yang kerap dikenal sebagai kadet mahasiswa Universitas Pertahanan Republik Indonesia (UNHAN RI) tidak hanya dibekali nilai-nilai akademik sesuai dengan program studinya masing-masing, namun juga dibekali nilai-nilai kemiliteran dan menjalankan sistem kehidupan asrama dalam peraturan kemiliteran. Setelah menyelesaikan pendi

dikan sesuai dengan program studinya para kadet mahasiswa akan diarahkan untuk melanjutkan Pendidikan Prajurit Karir (PaPK) sebagai seorang perwira TNI sesuai dengan kemampuan intelektualnya dibidang masing-masing. Para kadet mahasiswa disiapkan secara matang untuk menjadi seorang sarjana yang memiliki kecerdasan intelektual dan juga kemampuan bela negara. Sesuai dengan sistem yang mengarahkan kadet mahasiswa menjadi seorang

perwira TNI maka dijalankan beragam program untuk mendukung persiapan para kadet mahasiswa untuk mengikuti tes Perwira TNI. Salah satu program yang saat ini dijalankan adalah pengelompokan status Kesehatan (stakes) kadet mahasiswa berdasarkan tinggi badan dan berat badan yang selalu diperbarui setiap bulannya. Pengelompokan stakes dilakukan secara berkala guna mengawasi postur kadet mahasiswa untuk mempersiapkan para kadet mahasiswa mengikuti tes perwira TNI. Dalam pengelompokan pada data science, algoritma yang digunakan adalah algoritma untuk clustering. Algoritma yang banyak digunakan untuk clustering ini salah satunya adalah K-Means. Algoritma ini merupakan algoritma yang mudah dan telah dibanyak digunakan dalam beberapa penelitian [1]. Oleh karena itu, untuk mengetahui segmentasi pengelompokan data status kesehatan (stakes) kadet mahasiswa, maka dapat digunakan metode K-Means clustering.

LANDASAN TEORI

Penelitian yang membahas mengenai pengelompokan data yaitu penelitian yang telah dilakukan oleh Caesaria Dewi et al. [2]. Penelitian ini membahas perbandingan kinerja antara model DBSCAN dan K-Means. Kedua model ini merupakan model yang sama untuk cluster untuk pengelompokan status desa di Jawa Tengah pada tahun 2020. Hasil cluster yang didapat dari kedua model ini berupa DBSCAN dengan 4 cluster, sedangkan K-Means menggunakan 3 cluster. Untuk nilai evaluasi dari dua model ini didapatkan DBSCAN memiliki nilai silhouette disbanding dengan K-Means. Penelitian kedua yang membahas mengenai performa DBSCAN dan K-Means pernah dilakukan untuk pengelompokan data gangguan PT PLN UID Kalselteng [3]. Sama seperti penelien pertama uji performa menggunakan nilai silhouette. Hal yang berbeda dari hasil uji performa, ternyata K-

Means memiliki kinerja yang cukup unggul dibandingkan DBSCAN. Nilai silhouette terbaik yaitu 0,581 dengan jumlah cluster adalah enam. Hal ini tentunya bertolak belakang terhadap hasil pada penelitian pertama. Peneltian ketiga yang membahas clustering data yaitu penerapan K-Means untuk klasterisasi kasus stunting Balita di Desa Tegalwangi [4]. Penelitian ini didasari adanya kenaikan jumlah stunting pada tahun 2019 ke 2020. Metode yang digunakan untuk penelitian ini adalah Knowledge Discovery in Databases (KDD). Hasil dari penelien ini adalah jumlah cluster terbaik adalah dua cluster. Dengan nilai DBI yang paling optimal yaitu 0,007 yang hampir mendekati nilai 0. Penelitian keempat juga membahas mengenai cluster. Penelitian ini membahas segmentasi mengenai Infeksi Saluran Pernafasan (ISPA) [5]. Pada penelitian ini dibagi menjadi tiga cluster yaitu C1 (ISPA Biasa), C2 (ISPA sedang), dan C3 (ISPA Berat).

2. METODE PENELITIAN

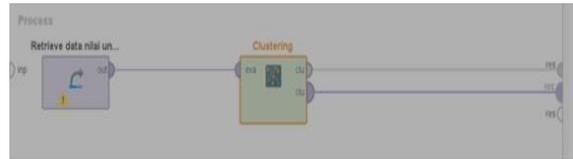
Dalam pelaksanaan penelitian ini, peneliti menggunakan beberapa tahapan *Knowledge Discovery in Database Process* (KDD) [6], [7]. Tahapan ini terdiri dari beberapa tahap untuk proses pengerjaan data science.

- Tahap pertama yaitu *data cleaning*. Pada tahapan ini data yang telah diperoleh dari hasil observasi harus dibersihkan terlebih dahulu. Maksud dibersihkan disini seperti memilih data yang dapat digunakan, mengurangi kesalahan input data yang dilakukan.
- Tahapan kedua yaitu *data integration*. Pada tahapan ini jika ada data yang telah dibersihkan ada yang berulang, maka data tersebut dapat disatukan untuk mengurangi redudansi data.
- Tahapan yang ketiga adalah *selection*. Pada tahapan ini data yang ada diseleksi apakah sudah sesuai dan relevan untuk

proses pemodelan.

- Tahapan keempat yaitu *data transformation*. Pada tahapan ini terbagi menjadi dua proses lagi yaitu *data mapping* dan *code generation*.
- Tahapan kelima yaitu *data mining*. Proses ini bisa juga dikatakan proses pemodelan Data yang telah siap digunakan diproses untuk dimodelkan. Pada penelitian ini data menggunakan algoritma clustering yaitu K-Means.
- Tahapan keenam yaitu *pattern evolution*. Dalam proses ini dilakukan evaluasi dari performa algoritma yang digunakan.
- Tahapan terakhir yaitu *Knowledge Presentation*. Proses paling akhir ini yaitu melakukan visualisasi. Salah satu manfaat proses ini adalah memudahkan pengambil keputusan untuk melihat visualisasi dari pemodelan. Visualisasi ini bisanya berupa dalam bentuk grafik ataupun diagram.

Data yang akan dimodelkan ini menggunakan tiga cluster dengan maximum run yaitu 10. Tampilan pemodelan dapat dilihat pada gambar 2.



Gambar 2. Pemodelan Cluster

Untuk Tampilan jumlah cluster dari maximum run terdapat pada gambar 3.

Gambar 3. Input Cluster dan Max Runs

Terdapat tiga cluster untuk pengelompokan berat badan dan tinggi badan kadet mahasiswa. Adapun tampilan salah satu hasil dari cluster dapat dilihat pada gambar 4.

3. HASIL PENELITIAN

Data yang digunakan untuk segmentasi penelitian adalah data tinggi dan berat badan kadet mahasiswa. Tinggi dan berat badan kadet mahasiswa diperoleh dari proses observasi. Untuk kepentingan penelitian maka data observasi diblur, dapat dilihat pada gambar 1.

Row No.	id	cluster	No.	Tinggi Badan	Berat Badan
1	1	cluster_2	1		
2	2	cluster_2	2		
3	3	cluster_2	3		
4	4	cluster_2	4		
5	5	cluster_2	5		
6	6	cluster_2	6		
7	7	cluster_2	7		
8	8	cluster_2	8		
9	9	cluster_2	9		
10	10	cluster_2	10		
11	11	cluster_2	11		
12	12	cluster_2	12		
13	13	cluster_2	13		

Gambar 4. Hasil Clusterisasi Kadet Mahasiswa

	A	B	C
1	No.	Tinggi Badan	Berat Badan
2	1		
3	2		
4	3		
5	4		
6	5		
7	6		
8	7		
9	8		
10	9		
11	10		
12	11		
13	12		
14	13		
15	14		
16	15		
17	16		
18	17		
19	18		
20	19		
21	20		
22	21		

Gambar 1. Beberapa Tampilan Data Tinggi dan Berat Badan Kadet Mahasiswa

Pembagian data menjadi tiga cluster. Hasil rata-rata dari pembagian cluster dapat dilihat pada gambar 5.

Attribute	cluster_0	cluster_1	cluster_2
No.	51.500	84.500	17.500
Tinggi Badan	170.485	168.609	169.294
Berat Badan	65.071	62.555	62.409

Gambar 5. Rata-Rata Pembagian Cluster

Adapun penjabaran dari tiga cluster tersebut adalah sebagai berikut:

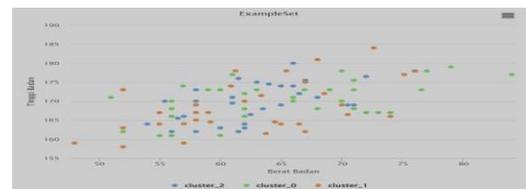
- Cluster 0 bisa dikatakan cluster yang punya potensi *overweight*. Individu pada cluster ini cenderung memiliki tinggi badan yang mungkin lebih rendah dari

rata-rata, namun berat badan yang relatif tinggi. Namun jika dihitung dengan standar BMI, anggota cluster ini masih tergolong normal. Hanya saja pada cluster ini harus diantisipasi agar tidak ada kelebihan berat badan. Pengelompokan ini dapat memberikan informasi tentang individu yang perlu mendapatkan perhatian khusus terkait pola makan dan aktivitas fisik untuk mengelola berat badan.

- Cluster 1 ini bisa dikatakan sebagai cluster yang anggotanya kombinasi normal dan *underweight*. Cluster 1 mencakup individu dengan variasi dalam tinggi badan dan berat badan, menunjukkan bahwa beberapa dari mereka memiliki berat badan normal, sementara yang lain mungkin cenderung kurang dari normal. Hal ini dapat menunjukkan keragaman dalam pola fisik di antara anggota klaster ini. Pemahaman lebih lanjut tentang gaya hidup dan faktor-faktor kesehatan juga perlu dilakukan untuk memahami variasi ini.
- Cluster 2 ini mirip dengan Cluster 1. cluster ini juga mencakup individu dengan variasi tinggi badan dan berat badan, tetapi dengan karakteristik yang mungkin sedikit berbeda. Pengelompokan ini dapat memberikan wawasan tambahan tentang seberapa bervariasi kondisi fisik kadet mahasiswa.

Hasil dari segmentasi berdasarkan tinggi badan dan berat badan pada Kadet Mahasiswa menggunakan metode K-Means Clustering memberikan pemahaman yang mendalam tentang variasi kondisi Kesehatan. Cluster 0, yang mencirikan individu dengan kelebihan berat badan tetapi jika dihitung berdasarkan BMI masih tergolong normal, menunjukkan tingkat konsistensi yang tinggi dengan variasi yang rendah, memberikan indikasi bahwa mahasiswa dalam cluster ini memiliki stabilitas dalam masalah punya potensi kelebihan

berat badan. Sementara itu, cluster 1 dan cluster 2 yang mencakup kadet mahasiswa dengan berat badan normal dan kurang dari normal, menampilkan variasi yang lebih tinggi, mencerminkan kompleksitas kondisi kesehatan di dalamnya. Evaluasi ini menyoroti relevansi cluster dengan tujuan peminatan kesehatan, di mana cluster 0 dapat menjadi kandidat potensial untuk dilakukan pembinaan lebih lanjut. Cluster 1 mewakili pembinaan menengah dan cluster 2 dapat dikatakan dengan pembinaan sederhana. Untuk sebaran anggota cluster dapat dilihat pada gambar 6.



Gambar 6. Visualisasi Cluster

Dari visualisasi yang dihasilkan beberapa kadet mahasiswa memang berada pada cluster 0. Cluster 0 ini punya potensi untuk kelebihan berat badan, tetapi jika dihitung berdasarkan skor BMI masih masuk dalam kategori normal. Namun hal ini tentunya harus menjadi pertimbangan lebih lanjut. Dilakukannya pembinaan secara simultan terutama untuk cluster 0. Untuk cluster 1 dan 2 sebaiknya juga dilakukan pembinaan karena ada beberapa kadet mahasiswa yang mempunyai berat badan yang kurang. Namun untuk cluster 1 dan 2 ini juga masih tergolong normal jika dihitung berdasarkan skor BMI.

4. KESIMPULAN DAN REKOMENDASI

Dari penelitian yang telah dilakukan dapat dikelompokkan kadet mahasiswa berdasarkan tinggi badan dan berat badan dengan metode K-Means Clustering. Metode K-Means Clustering ini menjadi metode yang efektif untuk dataset kadet mahasiswa. Adapun proses clustering yang

dilakukan menghasilkan tiga cluster. Hasil dari penelitian ini dapat digunakan untuk memahami suatu segmentasi perilaku kadet mahasiswa dalam penilaian status kesehatan yang didasarkan oleh tinggi badan dan berat badan.

REFERENSI

- [1] T. A. Aria, Y. Yuliadi, M. Julkarnain, and F. Hamdani, "Penerapan Algoritma K-Means Clustering untuk Data Obat," *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 4, no. 1, pp. 649–657, Aug.2023,doi:10.30865/KLIK.V4I1.1117.
- [2] M. M. Putri, C. Dewi, E. Permata Siam, G. Asri Wijayanti, N. Aulia, and R. Nooraeni, "Comparison of DBSCAN and K-Means Clustering for Grouping the Village Status in Central Java 2020," *Jurnal Matematika, Statistika dan Komputasi*, vol. 17, no. 3, pp. 394–404, May 2021,doi: 10.20956/j.v17i3.11704.
- [3] T. A. Wijaya, E. Utami, and H. Al Fatta, "Perbandingan Algoritma DBSCAN dan K-Means Clustering untuk Pengelompokan Data Gangguan PT. PLN UID Kalselteng," *Innovative: Journal Of Social Science Research*, vol. 4, no. 1, pp. 8846–8854, Feb. 2024, doi: 10.31004/INNOVATIVE.V4I1.8920.
- [4] P. Apriyani, A. R. Dikananda, and I. Ali, "Penerapan Algoritma K-Means dalam Klasterisasi Kasus Stunting Balita Desa Tegalwangi," *Hello World Jurnal Ilmu Komputer*, vol. 2, no. 1, pp. 20–33, Mar. 2023, doi:10.56211/HELLOWORLD.V2I1.230.
- [5] D. Ariyanto, "Data Mining Menggunakan Algoritma K-Means untuk Klasifikasi Penyakit Infeksi Saluran Pernafasan Akut," *Jurnal Sistim Informasi dan Teknologi*, vol. 4, no. 1, pp. 13–18, Feb. 2022,doi:10.37034/JSISFOTEK.V4I1.117
- [6] M. Volk, D. Staegemann, S. Bosse, R. Häusler, and K. Turowski, "Approaching The (Big) Data Science Engineering Process," in *5th International Conference on Internet of Things, Big Data and Security (IoTBDS)*, SCITEPRESS–Science and Technology Publications, 2020, pp. 428–435. doi: 10.5220/0009569804280435.
- [7] C. J. Costa and J. T. Aparicio, "POST-DS: A Methodology to Boost Data Science," *Iberian Conference on Information Systems and Technologies, CISTI*, vol. 2020-June, Jun. 2020,Doi: 10.23919/CISTI49556.2020.9140932.