

HASIL CEK_Zahrotun, Jones_Data Mining; Clustering; K-Means

by Lisna Zahrotun, Anna Hendri Soleliza Jones Pengelompokan
Mahasiswa Akademik Keperawatan

Submission date: 20-Sep-2022 12:28PM (UTC+0700)

Submission ID: 1904311947

File name: swa_Akademik_Keperawatan_Berdasarkan_Asal_menggunakan_k-means.pdf (435.98K)

Word count: 2581

Character count: 15218



Pengelompokan Mahasiswa Akademik Keperawatan Berdasarkan Asal Sekolah dan Nilai Akademik Menggunakan Metode *Clustering K-Means*

Lisna Zahrotun^{1,*}, Yunus Fajri¹, Anna Hendri Soleliza Jones¹, Eni Purwaningsih²

¹ Teknik Informatika, Fakultas Teknologi Industri, Universitas Ahmad Dahlan, Yogyakarta, Indonesia

² Akademi Perawatan Karya Bakti Husada, Yogyakarta, Indonesia

Email: ¹*lisna.zahrotun@tif.uad.ac.id, ²yunusf42@gmail.com, ³anahendri@tif.uad.ac.id, ⁴enipurwaningsih1905@gmail.com

Email Penulis Korespondensi: Lisna.zahrotun@tif.uad.ac.id

Submitted: 21/12/2021; Accepted: 29/12/2021; Published: 31/12/2021

Abstrak—Akademik Keperawatan Karya Bakti Husada (AKPER KBH) Bantul merupakan salah satu akademik yang membuka jurusan keperawatan. Berdasarkan wawancara dengan Direktur AKPER KBH syarat pendaftaran menjadi mahasiswa keperawatan saat ini adalah siswa lulusan dari semua jurusan dan semua sekolah menengah. AKPER KBH belum pernah melakukan analisa terhadap data mahasiswa apakah ada hubungan antara riwayat pada waktu sekolah menengah dengan nilai kelulusan IPK sebagai bahan evaluasi dalam proses pembelajaran, meskipun untuk saat ini dengan variasinya mahasiswa baru menyebabkan tingkat kesulitan dalam proses pembelajaran lebih sulit dibandingkan dengan sebelumnya, sedangkan capaian IPK sangat penting dalam pencarian pekerjaan setelah lulus nanti. Tujuan dari penelitian ini adalah mengelompokkan data akademik mahasiswa AKPER data mahasiswa berdasarkan data asal sekolah, nilai IPK, dan nilai Keperawatan Medical Bedah II (KMB II), Keperawatan Jiwa II (Kep Jiwa II), Keperawatan Anak II (Kep Anak II), Keperawatan Maternitas II (Kep Maternitas II), dan Keperawatan Medical Bedah (KMB V). Tahapan dalam penelitian ini meliputi pengambilan data, pembersihan data, seleksi data, transformasi data, pengelompokan data menggunakan metode *K-Means* dan representasi pengetahuan, pengujian yang digunakan dalam penelitian ini adalah purity tes. Dari percobaan yang dilakukan nilai akurasi tertinggi yaitu 0.924 dengan jumlah cluster 3.

Kata kunci: Data Mining; Clustering; K-Means

Abstract—Nursing Academic of Karya Bakti Husada (AKPER KBH) Bantul is one of the academics that opened the 2000 department. Based on an interview with the Director of AKPER KBH, the registration requirements to become a student of the Academy are currently graduates from all majors and all high schools. AKPER KBH has not analyzed student data whether there is a relationship between high school history and passing grades of GPA as an evaluation material in the learning process, although at this time with the variation of new students causing difficulty in learning difficult compared to before, while GPA achievement is very important in finding job after graduation. The purpose of this study is to classify academic data of AKPER students based on data on school origin, GPA scores, and Medical Surgical Nursing II (KMB II), Mental Nursing II (Kep Jiwa II), Child Nursing II (Kep Anak II), Maternity Nursing II (Kep Maternitas II), and Medical Surgical Nursing (KMB V). The stages in this study include data collection, data search, data selection, data transformation, data grouping using the *K-Means* method and knowledge representation, the test used in this study is the purity test. From the experiments conducted, the accuracy value is 0.924 with the number of clusters 3.

Keywords: Data Mining; Clustering; K-Means

1. PENDAHULUAN

Pendidikan tenaga medis di Indonesia merupakan salah satu pendidikan yang memiliki banyak peminat, diantaranya jurusan keperawatan. Banyak sekolah-sekolah keperawatan yang ada di Indonesia baik yang tergabung dalam Politeknik maupun Universitas. Secara umum syarat untuk dapat memasuki sekolah keperawatan ini adalah lulusan IPA baik dari sekolah SMA/MA, tinggi badan dan kesehatan jasmani dan rohani. Selama 1 tahun terakhir di seluruh dunia mengalami wabah covid-19 tidak terkecuali Indonesia. Keadaan ini menyebabkan kebutuhan tenaga medis semakin banyak. Salah satu tenaga medis yang dibutuhkan adalah perawat, karena perawat merupakan bagian garda depan dalam penanganan pasien covid-19. Salah satu persyaratan untuk dapat mengabdikan menjadi tenaga medis sebagai perawat adalah Indeks Prestasi Kumulatif (IPK).

Akademik Keperawatan Karya Bakti Husada (AKPER KBH) Bantul merupakan salah satu akademik yang membuka jurusan keperawatan. Berdasarkan wawancara dengan Direktur AKPER KBH syarat pendaftaran menjadi mahasiswa keperawatan saat ini adalah siswa lulusan dari semua jurusan dan semua sekolah menengah. AKPER KBH belum pernah melakukan analisa terhadap data mahasiswa apakah ada hubungan antara riwayat pada waktu sekolah menengah dengan nilai kelulusan IPK sebagai bahan evaluasi dalam proses pembelajaran, meskipun untuk saat ini dengan variasinya mahasiswa baru menyebabkan tingkat kesulitan dalam proses pembelajaran lebih sulit dibandingkan dengan sebelumnya, sedangkan capaian IPK sangat penting dalam pencarian pekerjaan setelah lulus nanti.

Salah satu bidang yang digunakan dalam penggalian data adalah *data mining*. Diantara beberapa teknik yang ada dalam *data mining* ini adalah Teknik *data mining* yang lain adalah pengelompokan, dimana dalam pengelompokan ini digunakan untuk mencari kelompok data yang memiliki karakteristik tertentu. Pengelompokan karakteristik data rumah tangga [1], pemetaan kualifikasi pendidikan di Indonesia [2], pengelompokan karakteristik pelanggan telepon [3], pengelompokan dokumen berbasis teks [4]. Penelitian sebelumnya adalah analisis dua metode dimana hasil pengelompokan dengan metode *K-means* lebih baik dari pengelompokan menggunakan metode *Agglomerative Hierarchical Clustering* (AHC). Analisis ini dilakukan pada pengelompokan data

6 penumpang bus trans jogja [5]. Metode K-Means juga efektif dan efisien dalam mengelompokkan dalam mengelompokkan data berbentuk text [4].

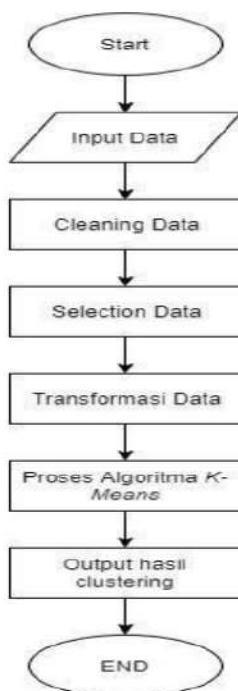
Penelitian tentang pengelompokan data riwayat mahasiswa sebelum kuliah dan masa studi juga pernah dilakukan sebelumnya dengan menggunakan metode K-Medoids [6] dan Aglomeratif Hierarchical Clustering (AHC) [7].

Dengan adanya teknik-teknik *data mining* dalam penggalian data, maka dalam penelitian ini akan dilakukan analisis data mahasiswa AKPER KBH dengan melakukan pengelompokan data akademik mahasiswa dan mencari pola asosiasi antara variable data sebelum kuliah dan nilai IPK. Tujuan penelitian yang dilakukan ini adalah mendapatkan hubungan asosiasi antara variable data mahasiswa berdasarkan proses pengelompokan sebelumnya. Hasil analisis ini diharapkan mampu memberikan rekomendasi terhadap syarat penerimaan mahasiswa baru pada AKPER KBH di tahun-tahun berikutnya.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

11 Tahap proses penelitian merupakan tahap awal dan sangat penting karena kesalahan ditahap ini akan menyebabkan kesalahan-kesalahan ditahap selanjutnya. Tahapan proses penelitian ditunjukkan dalam Gambar 1.



Gambar 1. Tahap Penelitian

1. *Input Data*
Tahapan yang pertama adalah tahapan *input* data, tahap input data yaitu menginputkan data yang sudah diperoleh sebelumnya¹¹, yaitu data mahasiswa AKPER KBH.
2. *Cleaning Data*
Cleaning data atau *pembersihan data* merupakan *proses* dimana *data mahasiswa yang* tidak lengkap akan dihapus.
3. *Selection Data*
Selection data atau *seleksi data* merupakan tahapan yang memutuskan atribut-atribut yang akan digunakan dalam penelitian ini.
4. *Transformasi Data*
Transformasi data adalah tahapan dimana data diubah menjadi nilai yang memiliki format tertentu. Pada tahapan ini, data yang berbentuk *alphabet* akan diubah menjadi data *numeric* (angka) dan data kategorikan diubah dalam teknik *One Hot Encoding*.
5. *Proses Algoritma K-Means*



Tahapan ini merupakan tahapan yang amat penting, dikarenakan algoritma *K-Means* mengolah data yang nantinya akan diterapkan kedalam sistem. *Clustering* merupakan salah satu teknik dalam data mining. Dimana *data mining* merupakan proses agar mendapatkan suatu pola atau pengetahuan yang bermanfaat dari sebuah data yang besar [8]. *K-Means clustering* adalah salah satu metode pengelompokan yang begitu populer dan banyak digunakan dikarenakan kecepatan dalam pengklasternya. *K-Means* termasuk dalam pengelompokan data non-hirarki (sekatan) yang berusaha membagi data kedalam dua atau lebih cluster. Metode *K-Means* membagi data menjadi beberapa cluster, sehingga data mempunyai karakteristik atau ciri khas yang sama maka akan dimasukkan kedalam satu kelompok yang mempunyai ciri khas sama, jika data memiliki ciri khas yang berbeda atau lain akan dijadikan satu kedalam kelompok dengan ciri khas yang lain. *Clustering* menggunakan metode *K-Means* mempunyai beberapa tahapan, menurut [9] yaitu:

- a. Menentukan nilai *k*, nilai *k* merupakan jumlah cluster yang akan dibentuk (jumlah cluster harus lebih kecil dari jumlah data).
- b. Menginisialisasi *k* yang akan digunakan sebagai centroid yang nantinya akan dibangkitkan secara acak.
- c. Menggunakan persamaan *Euclidean Distance* untuk menghitung jarak dari setiap data terhadap masing-masing centroid, persamaan *Euclidean Distance*:

$$d(P, C) = \sqrt{\sum_{i=1}^n (P_i - C_i)^2} \tag{1}$$

Keterangan:

N : Jumlah Data

P : Data

C : Centroid

I : Iterasi

- d. Mengelompokan setiap data berdasarkan jarak yang paling dekat, yaitu antara data dengan centroid.
 - e. Menentukan posisi centroid baru (*k*).
 - f. Jika centroid yang baru dengan centroid yang lama sama, maka proses *clustering* dihentikan. Bila masih terdapat perubahan nilai centroid yang baru dengan yang lama, maka Kembali pada Langkah ke 3.
6. Output hasil clustering atau representase pengetahuan
 Dalam tahap ini merupakan analisis hasil pengelompokkan yang akan menjadi rekomendasi bagi pihak AKPER KBH

3. HASIL DAN PEMBAHASAN

Dataset Mahasiswa AKPER KBH memiliki jumlah data sebanyak 386 data mahasiswa dengan atribut NIM, Nama Mahasiswa, Tanggal Lulus, Asal Sekolah, Nilai IPK, Keperawatan Medical Bedah II (KMB II), Keperawatan Jiwa II (Kep Jiwa II), Keperawatan Anak II (Kep Anak II), Keperawatan Maternitas II (Kep Maternitas II), dan Keperawatan Medical Bedah (KMB V). Pada penelitian ini, data yang akan diolah dapat dilihat pada gambar 1.

No	Nim	Nama Mahasiswa	Tanggal Lulus	Asal Sekolah	Nilai IPK	KMB II	Kep Jiwa II	Kep Anak II	Kep Maternitas II	KMB IV
1	15106800	Ety Rita Astuti	2009-08-11 00:00:00	NaN	2.80	NaN	NaN	NaN	NaN	NaN
2	16107150	Agus Rina Tri Ariyani	2009-08-11 00:00:00	NaN	2.80	NaN	NaN	NaN	NaN	NaN
3	16107160	Ahmad Lutfi Hariri	2009-08-11 00:00:00	NaN	2.60	NaN	NaN	NaN	NaN	NaN
4	16107170	Andanni Anggi Kisworoati Udit	2009-08-11 00:00:00	NaN	2.70	NaN	NaN	NaN	NaN	NaN
5	16107180	Dedi Ika Prasetya	2009-08-11 00:00:00	NaN	3.40	NaN	NaN	NaN	NaN	NaN
...
382	201612610	Kristiana	2019	SMKS	3.16	B	B	B	A	A
383	201612620	Nur Eka Kumiawan	2019	SMAN	3.04	B	B	B	A	B
384	201612630	Septiningrum Nur Cahyati	2019	SMAS	2.62	B	B	B	A	B
385	201612640	Trisna Nurhakim	2019	Paket C	3.57	B	B	B	A	A
386	201612650	Ulfa Nur 'Aini	2019	MAN	3.29	B	B	B	A	A

Gambar 2. Tabel Data Mahasiswa AKPER KBH

Dari data mentah tersebut, kemudian melakukan proses *preprocessing*. *Preprocessing* merupakan tahapan pembersihan data. Data yang tidak konsisten, noise dan atribut yang tidak digunakan akan dihapus. Dari 306 data setelah dilakukan proses pembersihan data hanya menjadi 218 data yang siap untuk dilakukan proses pengelompokkan. Sedangkan untuk proses seleksi data



Dari banyaknya atribut, hanya beberapa atribut saja yang akan digunakan, diantaranya adalah Asal Sekolah, Nilai IPK, KMB 2, Kep Jiwa 2, Kep Anak 2, Kep Maternitas 2, dan KMB 4. *Preprocessing* dapat dilihat pada gambar 3 berikut:

	Asal_Sekolah	Ipk	KMB II	Kep Jiwa II	Kep Anak II	Kep Maternitas II	KMB IV
0	SMAN	2.80	B	B	B	B	B
1	SMAN	3.29	B	B	B	B	B
2	SMAN	2.83	B	B	B	B	B
3	SMAN	2.98	B	B	B	B	B
4	SMAN	2.97	B	B	B	B	B
...
214	SMKS	3.16	B	B	B	A	A
215	SMAN	3.04	B	B	B	A	B
216	SMAS	2.62	B	B	B	A	B
217	Paket C	3.57	B	B	B	A	A
218	MAN	3.29	B	B	B	A	A

Gambar 3. Tabel hasil dari pembersihan dan seleksi data

Tahapan selanjutnya adalah transformasi data. Data yang akan di transformasi adalah atribut KMB 2, Kep Jiwa 2, Kep Anak 2, Kep Maternitas 2, KMB 4 dan atribut Asal Sekolah yang akan di transformasi menggunakan metode *One Hot Encoding*. Nilai Matakuliah di transformasi dari nilai *alphabet* menjadi nilai *numeric* dengan ketentuan nilai A menjadi 4, B menjadi 3, C menjadi 2, D menjadi 1, dan E menjadi 0. Hasil dari transformasi data dapat dilihat pada gambar 4.

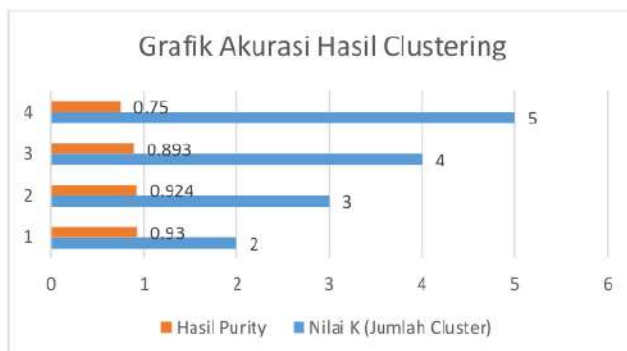
	Asal_Sekolah	Ipk	KMB II	Kep Jiwa II	Kep Anak II	Kep Maternitas II	KMB IV	MA	MAK	MAN	PKBM
0	SMAN	2.80	3	3	3	3	3	0	0	0	0
1	SMAN	3.29	3	3	3	3	3	0	0	0	0
2	SMAN	2.83	3	3	3	3	3	0	0	0	0
3	SMAN	2.98	3	3	3	3	3	0	0	0	0
4	SMAN	2.97	3	3	3	3	3	0	0	0	0
...
214	SMKS	3.16	3	3	3	4	4	0	0	0	0
215	SMAN	3.04	3	3	3	4	3	0	0	0	0
216	SMAS	2.62	3	3	3	4	3	0	0	0	0
217	Paket C	3.57	3	3	3	4	4	0	0	0	0
218	MAN	3.29	3	3	3	4	4	0	0	1	0

Gambar 4. Tabel hasil transformasi data

Setelah mendapatkan hasil dari transformasi data, tahap selanjutnya adalah penerapan Algoritma *K-Means*. Langkah awal algoritma *K-Means* adalah menentukan nilai *k* dimana nilai *k* merupakan jumlah *cluster* yang akan dibentuk. Nilai *k* yang digunakan adalah 2, 3, 4, 5. Setelah mempunyai hasil *clustering* dari *k*: 2, 3, 4, dan 5 dilakukan pengujian *clustering* menggunakan metode *purity* untuk mendapatkan jumlah *cluster* yang paling cocok. Hasil dari pengujian *purity* setiap *k* dapat dilihat pada Tabel 1 dan Gambar grafik hasil akurasi ditunjukkan dalam Gambar 2.

Tabel 1. Hasil Pengujian Purity

Nilai K (Jumlah Cluster)	Hasil Purity
2	0.93
3	0.924
4	0.893
5	0.75



Gambar 5. Grafik pengujian hasil clustering

Dari hasil perhitungan tersebut didapatkan hasil 3 pusat cluster, dimana titik pusat cluster pertama dapat dilihat pada Gambar 3.

2.05128205e-01 3.84615385e-02 6.41025641e-02 2.82051282e-01
 6.47435897e-01 2.56410256e-02 1.28205128e-02 2.17948718e-01
 1.28205128e-02 5.12820513e-02 1.28205128e-02 6.66133815e-16
 2.43589744e-01 1.92307692e-01 2.17948718e-01 1.28205128e-02

Gambar 6. Titik Pusat Cluster 1

Sedangkan untuk titik pusat cluster kedua dapat dilihat pada Gambar 4.

1.75824176e-01 2.19780220e-02 0.00000000e+00 3.73626374e-01
 5.93406593e-01 0.00000000e+00 0.00000000e+00 1.52655666e-16
 0.00000000e+00 0.00000000e+00 0.00000000e+00 1.00000000e+00
 -4.16333634e-17 0.00000000e+00 1.38777878e-16 0.00000000e+00

Gambar 7. Titik Pusat Cluster 2

Titik pusat cluster ketiga dapat dilihat pada Gambar 5.

7.60000000e-01 8.60000000e-01 8.20000000e-01 3.80000000e-01
 8.30000000e-01 0.00000000e+00 0.00000000e+00 6.00000000e-02
 0.00000000e+00 0.00000000e+00 0.00000000e+00 5.80000000e-01
 4.00000000e-02 1.60000000e-01 1.60000000e-01 0.00000000e+00

Gambar 8. Titik Pusat Cluster 3

Hasil dari pengelompokkan 3 cluster ditunjukkan dalam gambar 9.

	Nama	Asal_Sekolah	Ipik	KMB II	Kep Jiwa II	Kep Anak II	Kep Maternitas II	KMB IV	Kluster
0	Adik Aprilia Putri Primawati	SMAN	2.80	B	B	B	B	B	C2
1	Amin Nurhidayah Zunita Ferdana	SMAN	3.29	B	B	B	B	B	C2
2	Anom Prihasmoro	SMAN	2.83	B	B	B	B	B	C2
3	Arfa Larasati	SMAN	2.98	B	B	B	B	B	C2
4	Charul Nina	SMAN	2.97	B	B	B	B	B	C2
...
214	Kristiana	SMKS	3.16	B	B	B	A	A	C1
215	Nur Eka Kurniawan	SMAN	3.04	B	B	B	A	B	C2
216	Septiningrum Nur Cahyani	SMAS	2.62	B	B	B	A	B	C1
217	Trisna Nurhakim	Paket C	3.57	B	B	B	A	A	C1
218	Ulfah Nur'Aini	MAN	3.29	B	B	B	A	A	C1

Gambar 9. Tabel hasil clustering

Berdasarkan hasil pengujian purity pada tabel diatas, dapat ditarik kesimpulan bahwa jumlah cluster yang optimal adalah 3. Representasi pengetahuan dari hasil cluster dengan nilai k adalah 3 sebagai berikut:

- Cluster 0 mempunyai anggota sebanyak 50, dengan mempunyai rata-rata nilai IPK sebesar 3.26 dan terdiri dari asal sekolah MAN sebanyak 3 siswa, SMAN sebanyak 29 siswa, SMAS sebanyak 2 siswa, SMKN sebanyak 8 siswa, dan SMKS sebanyak 8 siswa.



2. Cluster 1 mempunyai anggota sebanyak 78, dengan mempunyai rata-rata nilai IPK sebesar 3.05 dan terdiri dari asal sekolah MA sebanyak 2 siswa, MAK sebanyak 1 siswa, MAN sebanyak 17 siswa, Paket C sebanyak 4 siswa, PPKBM sebanyak 1 siswa, SIMK sebanyak 1 siswa, SMAS sebanyak 19 siswa, SMKN sebanyak 15 siswa, SMKS sebanyak 17 siswa, dan SNIK sebanyak 1 siswa.

Cluster 2 mempunyai anggota sebanyak 91 dengan mempunyai rata-rata nilai IPK sebesar 3.08 dan asal sekolah hanya terdiri dari SMAN dengan jumlah siswa 91.

4. KESIMPULAN

Dari hasil pengelompokan diperoleh rekomendasi untuk AKPER KBH bahwa mahasiswa yang berasal dari SMAN, SMKN dan SMKS cenderung memiliki nilai IPK yang cukup bagus, hal ini dapat dibuktikan selisih nilai IPK antara cluster 0, cluster 1 dan cluster 2 tidak terlalu jauh. Sehingga dari hasil pengelompokan ini pihak AKPER dapat membuka kesempatan bagi siswa SMAN, SMKN dan SMKS untuk tetap mendaftar sebagai mahasiswa AKPER KBH.

UCAPAN TERIMA KASIH

Penelitian ini dapat terlaksana dengan dukungan dari Lembaga Penelitian dan Pengabdian Kepada Masyarakat Universitas Ahmad Dahlan dalam hibah dana penelitian Dasar Tahun Anggaran 2021/2022

REFERENCES

- [1] N. T. Septioko, H. A. Parhusip, and T. Mahatma, "Aplikasi K-Means Untuk pengelompokan Rumah tangga di Salatiga berdasarkan Data Susenas 2011," in *Pekan Imiah Dosen FEB-UKSW*, 2012, pp. 353–372.
- [2] G. S. Nugraha, Hairani, and R. F. P. Ardi, "Aplikasi Pemetaan Kualitas Pendidikan Di Indonesia Menggunakan Metode K-Means," *J. Matrik*, vol. 17, no. 2, pp. 13–23, 2018.
- [3] M. A. Permatadevi, R. A., M. E. Hendrawan, S.Kom, and M. S. Irmasari Hafidz, S.Kom, "Karakteristik pelanggan telepon kabel menggunakan clustering som dan k-means untuk mengurangi kesalahan klasifikasi pelanggan perusahaan telekomunikasi (studi kasus : pt. Telkom Mojokerto)," *J. Tek. pomits*, vol. 1, no. 1, pp. 1–7, 2013.
- [4] L. Zahrotun, N. hutami Putri, and A. N. Khusna, "The Implementation of K-Means Clustering Method in Classifying Undergraduate Thesisi Titles," in *12th International Conference on Telecommunication Systems, Services, and Applications (TSSA)*, 2018.
- [5] L. Zahrotun, "Analisis Pengelompokan Jumlah Penumpang Bus Trans Jogja Menggunakan metode Clustering K-Means dan Agglomerative Hierarchical Clustering (AHC)," *J. Inform.*, vol. 9, no. 1, pp. 1039–1047, 2015.
- [6] H. Kumia, L. Zahrotun, and U. Linarti, "InfoTekJar : Jurnal Nasional Informatika dan Pengelompokan Mahasiswa Berdasarkan Data Akademik Sebelum Kuliah dan Masa Studi Menggunakan K-Medoids," *J. Nas. Inform. dan Teknol. Jar.*, vol. 2, no. 2, pp. 265–272, 2021.
- [7] B. H. T. Suandi and L. Zahrotun, "PENERAPAN DATA MINING DALAM MENGELOMPOKKAN DATA RIWAYAT AKADEMIK SEBELUM KULIAH DAN DATA KELULUSAN (Implementation Of Data Mining In Grouping Academic History Data Before Students And Student Graduation Data Using Method Agglomerative Hierarchical Cluste," *J. Teknol. Informasi, Komput. dan Apl.*, vol. 3, no. 1, pp. 62–71, 2021.
- [8] J. Han, Kamber M. Jian Pie, *Data Mining: Concepts and Techniques*. 2011.
- [9] A. Asroni, H. Fitri, and E. Prasetyo, "Penerapan Metode Clustering dengan Algoritma K-Means pada Pengelompokan Data Calon Mahasiswa Baru di Universitas Muhammadiyah Yogyakarta (Studi Kasus: Fakultas Kedokteran dan Ilmu Kesehatan, dan Fakultas Ilmu Sosial dan Ilmu Politik)," *Semesta Tek.*, vol. 21, no. 1, pp. 60–64, 2018.

HASIL CEK_Zahrotun, Jones_Data Mining; Clustering; K-Means

ORIGINALITY REPORT

9%

SIMILARITY INDEX

9%

INTERNET SOURCES

5%

PUBLICATIONS

1%

STUDENT PAPERS

PRIMARY SOURCES

1	jtika.if.unram.ac.id Internet Source	1%
2	conference.binadarma.ac.id Internet Source	1%
3	jurnal.ustjogja.ac.id Internet Source	1%
4	repository.ub.ac.id Internet Source	1%
5	Badrul Wajdi. "PEMANFAATAN MEDIA POWER POINT DALAM PEMBELAJARAN FISIKA DENGAN PENDEKATAN KONSTRUKTIVISME", Jurnal Pendidikan Fisika, 2015 Publication	1%
6	jurnal.uisu.ac.id Internet Source	1%
7	jurnal.umsb.ac.id Internet Source	1%
8	Harisman Nizar, Somakim Somakim, Muhammad Yusuf. "Pengembangan LKS	1%

dengan Model Discovery Learning pada
Materi Irisan Dua Lingkaran", Jurnal Elemen,
2016

Publication

9	Submitted to UIN Sunan Ampel Surabaya Student Paper	1 %
10	ejournal3.undip.ac.id Internet Source	1 %
11	widuri.raharja.info Internet Source	1 %
12	digilib.uinsby.ac.id Internet Source	1 %
13	www.scribd.com Internet Source	1 %

Exclude quotes On

Exclude matches < 1%

Exclude bibliography On

Pengelompokan Mahasiswa Akademik Keperawatan Berdasarkan Asal Sekolah dan Nilai Akademik Menggunakan Metode *Clustering K-Means*

Lisna Zahrotun^{1*}, Yunus Fajri¹, Anna Hendri Soleliza Jones¹, Ani Purwaningsih²

¹ Teknik Informatika, Fakultas Teknologi Industri, Universitas Ahmad Dahlan, Yogyakarta, Indonesia

² Akademi Perawatan Karya Bakti Husada, Yogyakarta, Indonesia

Email: ^{1*}Lisna.zahrotun@tif.uad.ac.id, ²yunusf42@gmail.com, ³annahendri@tif.uad.ac.id, ⁴enipurwaningsih1905@gmail.com

Email Penulis Korespondensi: Lisna.zahrotun@tif.uad.ac.id

Submitted: 21/12/2021; Accepted: 29/12/2021; Published: 31/12/2021

Abstrak—Akademik Keperawatan Karya Bakti Husada (AKPER KBH) Bantul merupakan salah satu akademik yang membuka jurusan keperawatan. Berdasarkan wawancara dengan Direktur AKPER KBH syarat pendaftaran menjadi mahasiswa keperawatan saat ini adalah siswa lulusan dari semua jurusan dan semua sekolah menengah. AKPER KBH belum pernah melakukan analisa terhadap data mahasiswa apakah ada hubungan antara riwayat pada waktu sekolah menengah dengan nilai kelulusan IPK sebagai bahan evaluasi dalam proses pembelajaran, meskipun untuk saat ini dengan variasinya mahasiswa baru menyebabkan tingkat kesulitan dalam proses pembelajaran lebih sulit dibandingkan dengan sebelumnya, sedangkan capaian IPK sangat penting dalam pencarian pekerjaan setelah lulus nanti. Tujuan dari penelitian ini adalah mengelompokkan data akademik mahasiswa AKPER data mahasiswa berdasarkan data asal sekolah, nilai IPK, dan nilai Keperawatan Medical Bedah II (KMB II), Keperawatan Jiwa II (Kep Jiwa II), Keperawatan Anak II (Kep Anak II), Keperawatan Maternitas II (Kep Maternitas II), dan Keperawatan Medical Bedah (KMB V). Tahapan dalam penelitian ini meliputi pengambilan data, pembersihan data, seleksi data, transformasi data, pengelompokan data menggunakan metode K-Means dan representasi pengetahuan, pengujian yang digunakan dalam penelitian ini adalah purity tes. Dari percobaan yang dilakukan nilai akurasi tertinggi yaitu 0.924 dengan jumlah cluster 3.

Kata kunci: Data Mining; Clustering; K-Means

Abstract—Nursing Academic of Karya Bakti Husada (AKPER KBH) Bantul is one of the academics that opened the 2000 department. Based on an interview with the Director of AKPER KBH, the registration requirements to become a student of the Academy are currently graduates from all majors and all high schools. AKPER KBH has not analyzed student data whether there is a relationship between high school history and passing grades of GPA as an evaluation material in the learning process, although at this time with the variation of new students causing difficulty in learning difficult compared to before, while GPA achievement is very important in finding job after graduation. The purpose of this study is to classify academic data of AKPER students based on data on school origin, GPA scores, and Medical Surgical Nursing II (KMB II), Mental Nursing II (Kep Jiwa II), Child Nursing II (Kep Anak II), Maternity Nursing II (Kep Maternitas II), and Medical Surgical Nursing (KMB V). The stages in this study include data collection, data search, data selection, data transformation, data grouping using the K-Means method and knowledge representation, the test used in this study is the purity test. From the experiments conducted, the accuracy value is 0.924 with the number of clusters 3.

Keywords: Data Mining; Clustering; K-Means

1. PENDAHULUAN

Pendidikan tenaga medis di Indonesia merupakan salah satu pendidikan yang memiliki banyak peminat, diantaranya jurusan keperawatan. Banyak sekolah-sekolah keperawatan yang ada di Indonesia baik yang tergabung dalam Politeknik maupun Universitas. Secara umum syarat untuk dapat memasuki sekolah keperawatan ini adalah lulusan IPA baik dari sekolah SMA/MA, tinggi badan dan kesehatan jasmani dan rohani. Selama 1 tahun terakhir di seluruh dunia mengalami wabah covid-19 tidak terkecuali Indonesia. Keadaan ini menyebabkan kebutuhan tenaga medis semakin banyak. Salah satu tenaga medis yang dibutuhkan adalah perawat, karena perawat merupakan bagian garda depan dalam penanganan pasien covid-19. Salah satu persyaratan untuk dapat mengabdikan menjadi tenaga medis sebagai perawat adalah Indek Prestasi Kumulatif (IPK).

Akademik Keperawatan Karya Bakti Husada (AKPER KBH) Bantul merupakan salah satu akademik yang membuka jurusan keperawatan. Berdasarkan wawancara dengan Direktur AKPER KBH syarat pendaftaran menjadi mahasiswa keperawatan saat ini adalah siswa lulusan dari semua jurusan dan semua sekolah menengah. AKPER KBH belum pernah melakukan analisa terhadap data mahasiswa apakah ada hubungan antara riwayat pada waktu sekolah menengah dengan nilai kelulusan IPK sebagai bahan evaluasi dalam proses pembelajaran, meskipun untuk saat ini dengan variasinya mahasiswa baru menyebabkan tingkat kesulitan dalam proses pembelajaran lebih sulit dibandingkan dengan sebelumnya, sedangkan capaian IPK sangat penting dalam pencarian pekerjaan setelah lulus nanti.

Salah satu bidang yang digunakan dalam penggalian data adalah *data mining*. Diantara beberapa teknik yang ada dalam *data mining* ini adalah Teknik *data mining* yang lain adalah pengelompokan, dimana dalam pengelompokan ini digunakan untuk mencari kelompok data yang memiliki karakteristik tertentu. Pengelompokan karakteristik data rumah tangga [1], pemetaan kualifikasi pendidikan di Indonesia [2], pengelompokan karakteristik pelanggan telepon [3] dan pengelompokan dokumen berbasis teks [4]. Penelitian sebelumnya adalah analisis dua metode dimana hasil pengelompokan dengan metode *K-means* lebih baik dari pengelompokan menggunakan metode *Agglomerative Hierarchical Clustering* (AHC). Analisis ini dilakukan pada pengelompokan data

penumpang bus trans jogja [5]. Metode K-Means juga efektif dan efisien dalam mengelompokkan dalam mengelompokkan data berbentuk text [4].

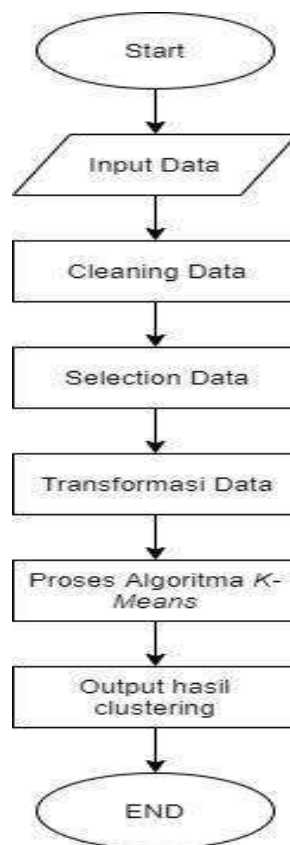
Penelitian tentang pengelompokkan data riwayat mahasiswa sebelum kuliah dan masa studi juga pernah dilakukan sebelumnya dengan menggunakan metode K-Medoids [6] dan Aglomeratif Hierarchical Clustering (AHC) [7].

Dengan adanya teknik-teknik *data mining* dalam penggalian data, maka dalam penelitian ini akan dilakukan analisis data mahasiswa AKPER KBH dengan melakukan pengelompokkan data akademik mahasiswa dan mencari pola asosiasi antara variable data sebelum kuliah dan nilai IPK. Tujuan penelitian yang dilakukan ini adalah mendapatkan hubungan asosiasi antara variable data mahasiswa berdasarkan proses pengelompokkan sebelumnya. Hasil analisis ini diharapkan mampu memberikan rekomendasi terhadap syarat penerimaan mahasiswa baru pada AKPER KBH di tahun-tahun berikutnya.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Tahap proses penelitian merupakan tahap awal dan sangat penting karena kesalahan ditahap ini akan menyebabkan kesalahan-kesalahan ditahap selanjutnya. Tahapan proses penelitian ditunjukkan dalam Gambar 1.



Gambar 1. Tahap Penelitian

1. *Input Data*
Tahapan yang pertama adalah tahapan *input* data, tahap input data yaitu menginputkan data yang sudah diperoleh sebelumnya, yaitu data mahasiswa AKPER KBH.
2. *Cleaning Data*
Cleaning data atau pembersihan data merupakan proses dimana data mahasiswa yang tidak lengkap akan di hapus.
3. *Selection Data*
Selection data atau seleksi data merupakan tahapan yang memutuskan atribut-atribut yang akan digunakan dalam penelitian ini.
4. *Transformasi Data*
Transformasi data adalah tahapan dimana data diubah menjadi nilai yang memiliki format tertentu. Pada tahapan ini, data yang berbentuk *alphabet* akan diubah menjadi data *numeric* (angka) dan data kategorikan diubah dalam teknik *One Hot Encoding*.
5. *Proses Algoritma K-Means*



Tahapan ini merupakan tahapan yang amat penting, dikarenakan algoritma *K-Means* mengolah data yang nantinya akan diterapkan kedalam sistem. *Clustering* merupakan salah satu teknik dalam data mining. Dimana *data mining* merupakan proses agar mendapatkan suatu pola atau pengetahuan yang bermanfaat dari sebuah data yang besar [8]. *K-Means clustering* adalah salah satu metode pengelompokan yang begitu populer dan banyak digunakan dikarenakan kecepatan dalam pengklasterannya. *K-Means* termasuk dalam pengelompokan data non-hirarki (sekatan) yang berusaha membagi data kedalam dua atau lebih cluster. Metode *K-Means* membagi data menjadi beberapa *cluster*, sehingga data mempunyai karakteristik atau ciri khas yang sama maka akan dimasukkan kedalam satu kelompok yang mempunyai ciri khas sama, jika data memiliki ciri khas yang berbeda atau lain akan dijadikan satu kedalam kelompok dengan ciri khas yang lain. *Clustering* menggunakan metode *K-Means* mempunyai beberapa tahapan, menurut [9] yaitu:

- a. Menentukan nilai k, nilai k merupakan jumlah *cluster* yang akan dibentuk (jumlah *cluster* harus lebih kecil dari jumlah data).
- b. Menginisialisasi k yang akan digunakan sebagai centroid yang nantinya akan dibangkitkan secara acak.
- c. Menggunakan persamaan *Euclidean Distance* untuk menghitung jarak dari setiap data terhadap masing-masing *centroid*, persamaan *Euclidean Distance*:

$$d(P, C) = \sqrt{\sum_{i=1}^n (P_i - C_i)^2} \tag{1}$$

Keterangan:

N : Jumlah Data

P : Data

C : Centroid

I : Iterasi

- d. Mengelompokan setiap data berdasarkan jarak yang paling dekat, yaitu antara data dengan *centroid*.
 - e. Menentukan posisi *centroid* baru (k).
 - f. Jika centroid yang baru dengan *centroid* yang lama sama, maka proses *clustering* dihentikan. Bila masih terdapat perubahan nilai *centroid* yang baru dengan yang lama, maka Kembali pada Langkah ke 3.
6. Output hasil clustering atau representase pengetahuan
 Dalam tahap ini merupakan analisis hasil pengelompokan yang akan menjadi rekomendasi bagi pihak AKPER KBH

3. HASIL DAN PEMBAHASAN

Dataset Mahasiswa AKPER KBH memiliki jumlah data sebanyak 386 data mahasiswa dengan atribut NIM, Nama Mahasiswa, Tanggal Lulus, Asal Sekolah, Nilai IPK, Keperawatan Medical Bedah II (KMB II), Keperawatan Jiwa II (Kep Jiwa II), Keperawatan Anak II (Kep Anak II), Keperawatan Maternitas II (Kep Maternitas II), dan Keperawatan Medical Bedah (KMB V). Pada penelitian ini, data yang akan diolah dapat dilihat pada gambar 1.

No	Nim	Nama Mahasiswa	Tanggal Lulus	Asal Sekolah	Nilai IPK	KMB II	Kep Jiwa II	Kep Anak II	Kep Maternitas II	KMB IV
1	1510680.0	Ety Rita Astuti	2009-08-11 00:00:00	NaN	2.80	NaN	NaN	NaN	NaN	NaN
2	1610715.0	Agus Rina Tri Ariyani	2009-08-11 00:00:00	NaN	2.80	NaN	NaN	NaN	NaN	NaN
3	1610716.0	Ahmad Lutfi Harini	2009-08-11 00:00:00	NaN	2.60	NaN	NaN	NaN	NaN	NaN
4	1610717.0	Andarini Anggi Kisworowati Udit	2009-08-11 00:00:00	NaN	2.70	NaN	NaN	NaN	NaN	NaN
5	1610718.0	Dedi Ika Prasetya	2009-08-11 00:00:00	NaN	3.40	NaN	NaN	NaN	NaN	NaN
...
382	20161261.0	Kristiana	2019	SMKS	3.16	B	B	B	A	A
383	20161262.0	Nur Eka Kurniawan	2019	SMAN	3.04	B	B	B	A	B
384	20161263.0	Septiningrum Nur Cahyati	2019	SMAS	2.62	B	B	B	A	B
385	20161264.0	Trisna Nurhakim	2019	Paket C	3.57	B	B	B	A	A
386	20161265.0	Ulfah Nur Aini	2019	MAN	3.29	B	B	B	A	A

Gambar 2. Tabel Data Mahasiswa AKPER KBH

Dari data mentah tersebut, kemudian melakukan proses *preprocessing*. *Preprocessing* merupakan tahapan pembersihan data. Data yang tidak konsisten, noise dan atribut yang tidak digunakan akan dihapus. Dari 306 data setelah dilakukan proses pembersihan data hanya menjadi 218 data yang siap untuk dilakukan proses pengelompokan. Sedangkan untuk proses seleksi data



Dari banyaknya atribut, hanya beberapa atribut saja yang akan digunakan, diantaranya adalah Asal Sekolah, Nilai IPK, KMB 2, Kep Jiwa 2, Kep Anak 2, Kep Maternitas 2, dan KMB 4. *Preprocessing* dapat dilihat pada gambar 3 berikut:

	Asal_Sekolah	IpK	KMB II	Kep Jiwa II	Kep Anak II	Kep Maternitas II	KMB IV
0	SMAN	2.80	B	B	B	B	B
1	SMAN	3.29	B	B	B	B	B
2	SMAN	2.83	B	B	B	B	B
3	SMAN	2.98	B	B	B	B	B
4	SMAN	2.97	B	B	B	B	B
...
214	SMKS	3.16	B	B	B	A	A
215	SMAN	3.04	B	B	B	A	B
216	SMAS	2.62	B	B	B	A	B
217	Paket C	3.57	B	B	B	A	A
218	MAN	3.29	B	B	B	A	A

Gambar 3. Tabel hasil dari pembersihan dan seleksi data

Tahapan selanjutnya adalah transformasi data. Data yang akan di transformasi adalah atribut KMB 2, Kep Jiwa 2, Kep Anak 2, Kep Maternitas 2, KMB 4 dan atribut Asal Sekolah yang akan di transformasi menggunakan metode *One Hot Encoding*. Nilai Matakuliah di transformasi dari nilai *alphabet* menjadi nilai *numeric* dengan ketentuan nilai A menjadi 4, B menjadi 3, C menjadi 2, D menjadi 1, dan E menjadi 0. Hasil dari transformasi data dapat dilihat pada gambar 4.

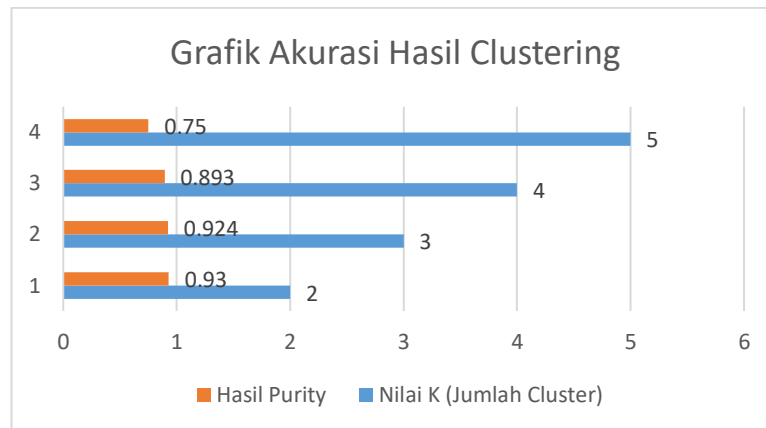
	Asal_Sekolah	IpK	KMB II	Kep Jiwa II	Kep Anak II	Kep Maternitas II	KMB IV	MA	MAK	MAN	PRBM
0	SMAN	2.80	3	3	3	3	3	0	0	0	0
1	SMAN	3.29	3	3	3	3	3	0	0	0	0
2	SMAN	2.83	3	3	3	3	3	0	0	0	0
3	SMAN	2.98	3	3	3	3	3	0	0	0	0
4	SMAN	2.97	3	3	3	3	3	0	0	0	0
...
214	SMKS	3.16	3	3	3	4	4	0	0	0	0
215	SMAN	3.04	3	3	3	4	3	0	0	0	0
216	SMAS	2.62	3	3	3	4	3	0	0	0	0
217	Paket C	3.57	3	3	3	4	4	0	0	0	0
218	MAN	3.29	3	3	3	4	4	0	0	1	0

Gambar 4. Tabel hasil transformasi data

Setelah mendapatkan hasil dari transformasi data, tahap selanjutnya adalah penerapan Algoritma *K-Means*. Langkah awal algoritma *K-Means* adalah menentukan nilai k dimana nilai k merupakan jumlah *cluster* yang akan dibentuk. Nilai k yang digunakan adalah 2, 3, 4, 5. Setelah mempunyai hasil *clustering* dari k: 2, 3, 4, dan 5 dilakukan pengujian *clustering* menggunakan metode *purity* untuk mendapatkan jumlah *cluster* yang paling cocok. Hasil dari pengujian *purity* setiap k dapat dilihat pada Tabel 1 dan Gambar grafik hasil akurasi ditunjukkan dalam Gambar 2.

Tabel 1. Hasil Pengujian Purity

Nilai K (Jumlah Cluster)	Hasil Purity
2	0.93
3	0.924
4	0.893
5	0.75



Gambar 5. Grafik pengujian hasil *clustering*

Dari hasil perhitungan tersebut didapatkan hasil 3 pusat *cluster*, dimana titik pusat *cluster* pertama dapat dilihat pada Gambar 3.

2.05128205e-01 3.84615385e-02 6.41025641e-02 2.82051282e-01
 6.47435897e-01 2.56410256e-02 1.28205128e-02 2.17948718e-01
 1.28205128e-02 5.12820513e-02 1.28205128e-02 6.66133815e-16
 2.43589744e-01 1.92307692e-01 2.17948718e-01 1.28205128e-02

Gambar 6. Titik Pusat Cluster 1

Sedangkan untuk titik pusat cluster kedua dapat dilihat pada Gambar 4.

1.75824176e-01 2.19780220e-02 0.00000000e+00 3.73626374e-01
 5.93406593e-01 0.00000000e+00 0.00000000e+00 1.52655666e-16
 0.00000000e+00 0.00000000e+00 0.00000000e+00 1.00000000e+00
 -4.16333634e-17 0.00000000e+00 1.38777878e-16 0.00000000e+00

Gambar 7. Titik Pusat Cluster 2

Titik pusat cluster ketiga dapat dilihat pada Gambar 5.

7.60000000e-01 8.60000000e-01 8.20000000e-01 3.80000000e-01
 8.30000000e-01 0.00000000e+00 0.00000000e+00 6.00000000e-02
 0.00000000e+00 0.00000000e+00 0.00000000e+00 5.80000000e-01
 4.00000000e-02 1.60000000e-01 1.60000000e-01 0.00000000e+00

Gambar 8. Titik Pusat Cluster 3

Hasil dari pengelompokkan 3 cluster ditunjukkan dalam gambar 9.

	Nama	Asal_Sekolah	Ipk	KMB II	Kep Jiwa II	Kep Anak II	Kep Maternitas II	KMB IV	Kluster
0	Adik Aprilia Putri Primawati	SMAN	2.80	B	B	B	B	B	C2
1	Amin Nurhidayah Zunita Ferdiana	SMAN	3.28	B	B	B	B	B	C2
2	Anom Prihasmoro	SMAN	2.83	B	B	B	B	B	C2
3	Arla Larasati	SMAN	2.98	B	B	B	B	B	C2
4	Chairul Nina	SMAN	2.97	B	B	B	B	B	C2
...
214	Kristiana	SMKS	3.16	B	B	B	A	A	C1
215	Nur Eka Kurniawan	SMAN	3.04	B	B	B	A	B	C2
216	Septiningrum Nur Cahyadi	SMAS	2.62	B	B	B	A	B	C1
217	Trisna Nurhakim	Paket C	3.57	B	B	B	A	A	C1
218	Ulfah Nur'Ani	MAN	3.28	B	B	B	A	A	C1

Gambar 9. Tabel hasil clustering

Berdasarkan hasil pengujian *purity* pada tabel diatas, dapat ditarik kesimpulan bahwa jumlah *cluster* yang optimal adalah 3. Representasi pengetahuan dari hasil *cluster* dengan nilai k adalah 3 sebagai berikut:

1. *Cluster 0* mempunyai anggota sebanyak 50, dengan mempunyai rata-rata nilai IPK sebesar 3.26 dan terdiri dari asal sekolah MAN sebanyak 3 siswa, SMAN sebanyak 29 siswa, SMAS sebanyak 2 siswa, SMKN sebanyak 8 siswa, dan SMKS sebanyak 8 siswa.



2. *Cluster 1* mempunyai anggota sebanyak 78, dengan mempunyai rata-rata nilai IPK sebesar 3.05 dan terdiri dari asal sekolah MA sebanyak 2 siswa, MAK sebanyak 1 siswa, MAN sebanyak 17 siswa, Paket C sebanyak 4 siswa, PPKBM sebanyak 1 siswa, SIMK sebanyak 1 siswa, SMAS sebanyak 19 siswa, SMKN sebanyak 15 siswa, SMKS sebanyak 17 siswa, dan SNIK sebanyak 1 siswa.

Cluster 2 mempunyai anggota sebanyak 91 dengan mempunyai rata-rata nilai IPK sebesar 3.08 dan asal sekolah hanya terdiri dari SMAN dengan jumlah siswa 91.

4. KESIMPULAN

Dari hasil pengelompokan diperoleh rekomendasi untuk AKPER KBH bahwa mahasiswa yang berasal dari SMAN, SMKN dan SMKS cenderung memiliki nilai IPK yang cukup bagus, hal ini dapat dibuktikan selisih nilai IPK antara cluster 0, cluster 1 dan cluster 2 tidak terlalu jauh. Sehingga dari hasil pengelompokan ini pihak AKPER dapat membuka kesempatan bagi siswa SMAN, SMKN dan SMKS untuk tetap mendaftar sebagai mahasiswa AKPER KBH.

UCAPAN TERIMA KASIH

Penelitian ini dapat terlaksana dengan dukungan dari Lembaga Penelitian dan Pengabdian Kepada Masyarakat Universitas Ahmad Dahlan dalam hibah dana penelitian Dasar Tahun Anggaran 2021/2022

REFERENCES

- [1] N, T. Septioko, H. A. Parhusip, and T. Mahatma, "Aplikasi K-Means Untuk pengelompokan Rumah tangga di Salatiga berdasarkan Data Susenas 2011," in *Pekan Imiah Dosen FEB-UKSW*, 2012, pp. 353–372.
- [2] G. S. Nugraha, Hairani, and R. F. P. Ardi, "Aplikasi Pemetaan Kualitas Pendidikan Di Indonesia Menggunakan Metode K-Means," *J. MATRIK*, vol. 17, no. 2, pp. 13–23, 2018.
- [3] M. A. Permatadevi, R. A., M. E. Hendrawan, S.Kom, and M. S. Irmasari Hafidz, S.Kom, "Karakteristik pelanggan telepon kabel menggunakan clustering som dan k-means untuk mengurangi kesalahan klasifikasi pelanggan perusahaan telekomunikasi (studi kasus : pt. Telkom Mojokerto)," *J. Tek. pomits*, vol. 1, no. 1, pp. 1–7, 2013.
- [4] L. Zahrotun, N. hutami Putri, and A. N. Khusna, "The Implementation of K-Means Clustering Method in Classifying Undergraduate Thesis Titles," in *12th International Conference on Telecommunication Systems, Services, and Applications (TSSA)*, 2018.
- [5] L. Zahrotun, "Analisis Pengelompokan Jumlah Penumpang Bus Trans Jogja Menggunakan metode Clustering K-Means dan Agglomerative Hierarchical Clustering (AHC)," *J. Inform.*, vol. 9, no. 1, pp. 1039–1047, 2015.
- [6] H. Kurnia, L. Zahrotun, and U. Linarti, "InfoTekJar : Jurnal Nasional Informatika dan Pengelompokan Mahasiswa Berdasarkan Data Akademik Sebelum Kuliah dan Masa Studi Menggunakan K-Medoids," *J. Nas. Inform. dan Teknol. Jar.*, vol. 2, no. 2, pp. 265–272, 2021.
- [7] B. H. T. Suandi and L. Zahrotun, "PENERAPAN DATA MINING DALAM MENGELOMPOKKAN DATA RIWAYAT AKADEMIK SEBELUM KULIAH DAN DATA KELULUSAN (Implementation Of Data Mining In Grouping Academic History Data Before Students And Student Graduation Data Using Method Agglomerative Hierarchical Cluste)," *J. Teknol. Informasi, Komput. dan Apl.*, vol. 3, no. 1, pp. 62–71, 2021.
- [8] J.Han, Kamber.M.,Jian Pie, *Data Mining: Concepts and Techniques*. 2011.
- [9] A. Asroni, H. Fitri, and E. Prasetyo, "Penerapan Metode Clustering dengan Algoritma K-Means pada Pengelompokan Data Calon Mahasiswa Baru di Universitas Muhammadiyah Yogyakarta (Studi Kasus: Fakultas Kedokteran dan Ilmu Kesehatan, dan Fakultas Ilmu Sosial dan Ilmu Politik)," *Semesta Tek.*, vol. 21, no. 1, pp. 60–64, 2018.