

Penerapan Metode Clustering dengan Algoritma K-Means Pada Pengelompokan Indeks Prestasi Akademik Mahasiswa

Suraya¹, Muhammad Sholeh^{2*}, Dina Andayati³

¹Fakultas Teknologi Informasi dan Bisnis, Rekayasa Sistem Komputer, IST AKPRIND Yogyakarta, Indonesia

²Fakultas Teknologi Informasi dan Bisnis, Informatika, IST AKPRIND Yogyakarta, Indonesia

³Fakultas Teknologi Industri, Teknik Mesin, IST AKPRIND Yogyakarta, Indonesia

E-mail: ¹Suraya@akprind.ac.id, ^{2*}muhash@akprind.ac.id, ³dina_asnawi@yahoo.com

Abstrak

Proses evaluasi hasil prestasi akademik mahasiswa sangat diperlukan sebagai upaya untuk memantau perkembangan akademik mahasiswa. Pengelompokan mahasiswa sesuai dengan hasil perolehan prestasi akademik sangat diperlukan. Salah satu kegunaan adalah sebagai proses pemetaan mahasiswa yang dapat diperkirakan dapat lulus tepat waktu dan mahasiswa yang perolehan nilainya masih rendah sehingga perlu adanya pendampingan. Tujuan pengelompokan ini dapat digunakan sebagai pemetaan indeks prestasi mahasiswa. Penelitian yang dilakukan bertujuan memberikan alternatif pengelompokan prestasi akademik berdasar pada proses data mining dengan menggunakan model klusterisasi. Metode penelitian menggunakan metode data mining yaitu KDD (*Knowledge Discovery in Database*.) Metode data mining dengan KDD terdiri dari selection, pre-processing, transformation, model dan evaluation. Datasheet yang digunakan merupakan kumpulan data indeks prestasi mahasiswa yang diolah dari perolehan indeks prestasi mahasiswa angkatan 2018, program studi Informatika- Institut Sains & Teknologi AKPRIND Yogyakarta. Datasheet diolah mulai dari semester 1 sampai 6. Model klustering dibangun dengan menggunakan salah satu algoritma metode clustering yaitu K-Means. Hasil penelitian menghasilkan pengelompokan atau cluster yang terbaik adalah pengelompokan sebanyak 2, Proses untuk mendapatkan pengelompokan yang paling baik dilakukan dengan menguji model dengan 6 pengelompokan. Hasil Pengelompokan yang terbaik dilakukan dengan pengujian Davies Bouldin. Kesimpulan penelitian dengan hasil 2 kelompok tersebut dapat diberi kategori, cluster 0 dengan nama kategori bagus dan cluster 1 dengan kategori tidak bagus.

Kata kunci: akademik, indeks prestasi, mahasiswa, model, klusterisasi

Abstract

The process of evaluating student academic achievement results is needed as an effort to monitor student academic development. Grouping students according to the results of academic achievement is needed. One of the uses is as a mapping process for students who can be expected to graduate on time and students whose grades are still low so that assistance is needed. The purpose of this grouping can be used as a student achievement index mapping. The research conducted aims to provide an alternative grouping of academic achievement based on the data mining process using a clustering model. The research method uses the Knowledge Discovery in Database method. The data mining method with Knowledge Discovery in Database consists of selection, pre-processing, transformation, model and evaluation. The datasheet used is a collection of student achievement index data processed from the acquisition of the 2018 student achievement index, Informatics study program - AKPRIND Institute of Science & Technology Yogyakarta. Datasheets are processed from semester 1 to 6. The clustering model is built using the K-Means algorithm. The results of the study resulted in the best grouping is grouping as much as 2, The process of getting the best grouping is done by testing the model with 6 groupings. The best grouping results are done with Davies Bouldin testing. The conclusion of the study with the results of the 2 groups can be given a category, cluster 0 with the name of the good category and cluster 1 with the category not good.

Keywords: academic, grade point average, student, model, clustering

1. PENDAHULUAN

Indeks prestasi mahasiswa dalam proses akademik menjadi salah satu tolok ukur dalam proses akademik. Indeks prestasi merupakan sistem atau cara yang digunakan dalam penilaian yang diterapkan dalam proses akademik. Indeks prestasi (IP) merupakan hasil dari penjumlahan semua nilai mata kuliah yang ditempuh mahasiswa dan telah diselesaikan dalam kurun waktu

satu semester.[1], [2]. Hasil indeks prestasi mahasiswa ini dapat menjadi salah satu acuan dalam untuk menentukan keberhasilan mahasiswa.

Penentuan indikator yang sederhana diantaranya melihat indeks prestasi kumulatif (IPK). IPK merupakan rata-rata dari IP dalam kurun waktu tertentu. Semakin tinggi IPK mahasiswa tentunya semakin tinggi perkiraan untuk lulus tepat waktu. Model lain yang dapat digunakan untuk membuat pengelompokan IP dari mahasiswa adalah dengan membuat data mining model clustering. Model clustering merupakan model yang menghasilkan kelompok-kelompok yang dapat dibuat kategori/ pengelompokan. Algoritma yang dapat digunakan dalam pembuatan model clustering adalah K-Means.[3],[4],[5]

Data mining dapat digunakan dalam proses pembuatan model baik untuk model prediksi, klasifikasi maupun clustering. Penelitian dengan objek mahasiswa dan diolah dengan data mining model klasifikasi dilakukan [6],[7], [8],[9]. Penelitian [6], menggunakan data mahasiswa sebanyak 377 dan terdapat 72 attribute. Model yang digunakan dalam pembuatan model adalah model klasifikasi dengan menggunakan algoritma K-Nearest Neighbour. Penelitian [10], melakukan penelitian dengan fokus untuk melakukan evaluasi kinerja akademik bagi mahasiswa yang menempuh kuliah di STMIK Dipanegara Makassar. Data mahasiswa yang diolah merupakan data pada dua tahun pertama. Model klasifikasi menggunakan algoritma Naive Bayes Classifier (NBC). Penelitian [7], membuat model klasifikasi dengan Metode Algoritma Naïve Bayes. Model klasifikasi digunakan dengan tujuan untuk melakukan prediksi dari kelulusan mahasiswa. Datasheet yang digunakan terdiri dari beberapa attribute diantaranya IPK.

Salah satu model data mining yang sering digunakan dalam membuat model adalah clustering. Penelitian dengan menggunakan algoritma K-means dilakukan oleh [11],[12],[13],[14]. Penelitian [11], dalam penelitian yang dilakukan membuat model clustering untuk membuat pemetaan hasil tri darma dosen khususnya dosen di Universitas medan area, Data yang digunakan diantara data yang wajib disiapkan dosen seperti membuat dokumen yang digunakan dalam proses pembelajaran seperti Silabus, rencana pembelajaran dan lainnya. Penelitian [12], membuat pemetaan data-data kecamatan yang ada di kabupaten Blora Jawa Tengah. Dasar pembuatan pengelompokan adalah data kesehatan dan keberhasilan dalam memberikan pelayanan kesehatan. Luaran dari penelitian adalah menghasilkan pengelompokan data yang dapat digunakan pendistribusian data kecamatan sesuai profil penanganan dan pelayanan kesehatan masyarakat. Penelitian [13], dengan datasheet data buku di perpustakaan di salah satu perpustakaan di Batam. Model yang dibuat menghasilkan pengelompokan buku yang sering dipinjam ada di cluster 1 dan terdapat 9 item. Cluster 2, buku yang paling sedikit dipinjam terdapat 15 item, dan jarang dipinjam ada pada cluster 0, sebanyak 12 item.

Penelitian yang menggunakan model *Clustering* yang terkait dengan datasheet mahasiswa dilakukan [15],[16],[17],[18]. Penelitian [15], melakukan penelitian menggunakan algoritma K-Means, hasil penelitian diharapkan dapat membantu pihak kampus maupun mahasiswa untuk melakukan prediksi kelulusan tepat waktu. Penelitian [16], memanfaatkan data pendaftaran dan data lulus tes mahasiswa baru. Data yang diolah diharapkan dapat memberikan informasi pengelompokan dan keterkaitan antara program studi atau jurusan dengan jalur masuk mahasiswa. Algoritma yang digunakan dalam penelitian menggunakan algoritma K-Means. Penelitian [16], membuat model clustering dengan menggunakan datasheet penerimaan mahasiswa baru. Model ini diharapkan dapat memberikan informasi pengelompokan dan keterkaitan hubungan jurusan yang ditempuh mahasiswa dengan jalur masuk yang dipilih mahasiswa. Proses pembuatan model menggunakan algoritma K-Means, model menghasilkan grafik berupa cluster data mahasiswa.

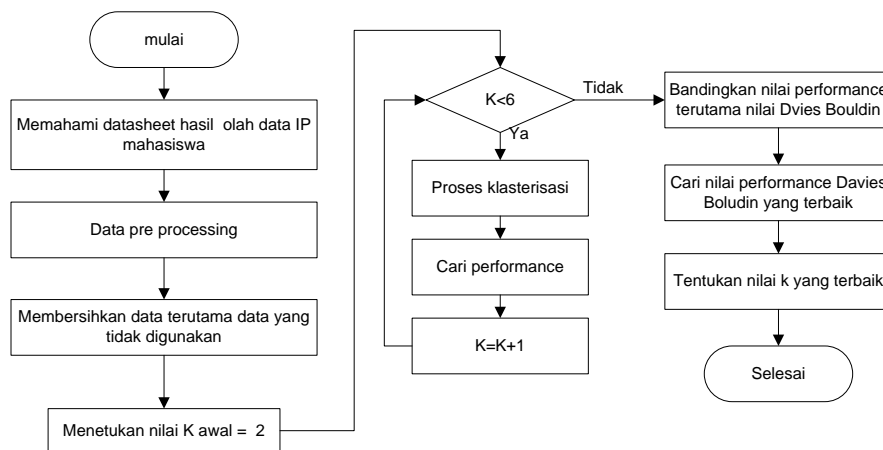
Implementasi data mining dapat menggunakan berbagai aplikasi. Aplikasi yang dapat digunakan dapat menggunakan Python, Rapid miner, Orange, R dan lainnya. Implementasi data mining dengan menggunakan Rapid Miner dilakukan [19],[20],[21], menggunakan python [22],[23],[24],[25],[26]

Berdasarkan pada pendahuluan dan tinjauan pustaka [15],[16],[17],[18] tersebut menjadi salah satu pertimbangan dalam melakukan penelitian. Penelitian dengan topik data mining masih relevan untuk dilakukan. Penelitian dilakukan dengan Batasan *datasheet* hanya menggunakan

data mahasiswa Angkatan 18 dan diolah dengan menggunakan model *clustering*. Algoritma yang digunakan dalam proses pembuatan model adalah algoritma K-means dan untuk evaluasi model diuji dengan melihat nilai Davies Bouldin.

2. METODE PENELITIAN

Metodologi penelitian dilakukan dengan melakukan proses olah data yang menjadi proses pengolahan data mining model clustering. Datasheet diambil dari kumpulan data IP mahasiswa mulai dari semester 1 sampai 6. Proses penelitian disajikan pada gambar 1.



Gambar 1. Proses Tahapan Dalam Penelitian

2.1. Teknik Pengumpulan Data

1) Observasi Metode

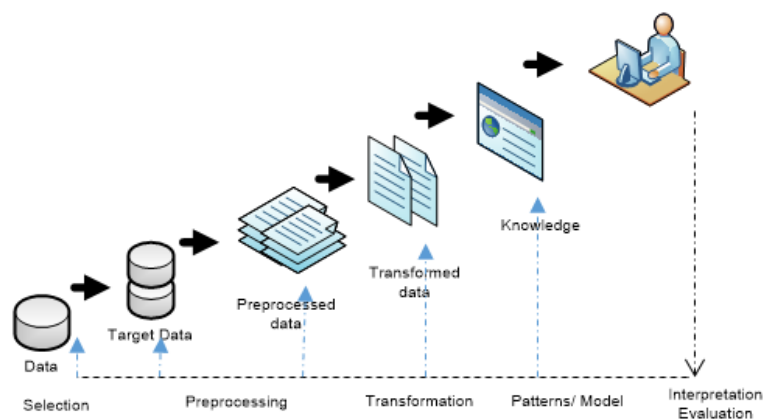
Metode observasi digunakan untuk pengumpulan data. Data diolah dan diambil portal akademik dan dilakukan proses pembersihan data.

2) Studi Literatur

Studi literatur dilakukan untuk memperoleh data yang sesuai untuk penelitian. Cara ini dilakukan dengan cara mencari dan memahami literatur dari beberapa jurnal yang berkaitan proses pembuatan model clustering

2.2. Teknik Analisis Data

Dalam melakukan penelitian ini metode analisis data menggunakan metode KDD. KDD merupakan aktivitas yang mencakup pengumpulan, pemakaian data historis buat memilih keteraturan, pola atau interaksi pada sebuah set data yang ukuran besar. Keluaran berdasarkan data mining pola dipakai buat dalam proses penentuan keputusan. Gambar 2, proses KDD.



Gambar 2. Tahapan dalam Metode *Knowledge Discovery in Database*

2.3. Algoritma K-Means

Algoritma K-Means dapat digunakan dalam proses pembuatan model clustering dan algoritma yang masuk dalam unsupervised learning. Algoritma K-Means digunakan untuk melakukan pengelompokan dataset ke dalam kluster-kluster yang tidak sama. Proses pembuatan atau banyaknya cluster dilambangkan dengan Simbol K. $K = 2$, menandakan ada 2 kluster dan jika $K = 3$ menandakan ada 3 kluster [27].

2.4 Evaluasi Model

Evaluasi model clustering digunakan untuk memilih nilai K atau banyaknya cluster (pengelompokan) yang paling baik. Teknik evaluasi mode cluster dapat menggunakan metode Elbow dengan *sum of square error* (SSE), Davis Bouldin Index (DBI) dan Silhouette Indeks (SI) [28].

Davis Bouldin Index (DBI) merupakan salah satu cara metode untuk melakukan mengevaluasi model clustering. Pemilihan hasil evaluasi dilakukan dengan membandingkan nilai yang terkecil dari model cluster yang dibuat. Jumlah kluster yang dipilih adalah dengan melihat nilai BDI yang terkecil.

2.4. Datasheet

Datasheet diolah dari portal perguruan tinggi, data merupakan perolehan prestasi akademik mahasiswa yang masuk pada tahun 2018. Data indeks prestasi yang diolah mulai dari perolehan IP semester 1 sampai semester 6. Datasheet terdiri dari 208 record dan 21 attribute. Datasheet yang diolah ditampilkan pada gambar 3.

...	↑	Nomhs	N...	sks1	mk1	ip1	sks2	mk2	ip2	sks3	mk3	ip3	sks4	mk4	ip4	sks5	mk5	ip5	sks6	mk6	ip6
1		1051001	M...	20	7	3.280	21	7	3.320	24	8	3.250	23	7	3.510	23	7	3.510	18	6	4
2		1051003	Tr...	20	7	2.530	21	7	2.700	22	7	2.090	19	6	2.470	19	6	2.470	20	7	2...
3		1051004	C...	20	7	3.630	21	7	3.700	24	8	3.600	23	7	3.860	23	7	3.860	18	6	3...
4		1051006	D...	20	7	2.870	21	7	3.030	22	7	2.820	22	7	3.090	22	7	3.090	18	6	3...
5		1051007	Y...	20	7	3.570	21	7	3.130	24	8	3.400	23	7	3.550	23	7	3.550	18	7	3...
6		1051008	Ir...	20	7	3.720	21	7	3.620	24	8	3.830	23	7	3.830	23	7	3.830	18	6	4
7		1051009	A...	20	7	3.450	21	7	3.560	24	8	3.720	23	7	3.960	23	7	3.960	18	6	3...
8		1051011	M...	20	7	3.250	21	7	3.160	24	8	3.110	20	6	2.170	23	7	3.550	18	7	3...
9		1051012	A...	20	7	2.470	21	7	2.830	20	6	0.250	22	7	2.880	20	6	2.170	20	6	0...
10		1051014	N...	20	7	2.580	21	7	2.480	22	7	2.730	23	7	3.670	22	7	2.880	22	8	3...
11		1051018	Sr...	20	7	3.680	21	7	3.250	24	8	3.600	23	7	3.900	23	7	3.670	19	6	0...

ExampleSet (203 examples, 0 special attributes, 21 regular attributes)

Gambar 3. Datasheet Perolehan Indeks Prestasi Mahasiswa

3. HASIL DAN PEMBAHASAN

3.1. Data Cleaning and Integration

Proses *data cleaning* dilakukan dengan memeriksa *datasheet* yang digunakan. Proses ini dilakukan dengan pengecekan data kosong, data ganda dan data yang diluar batas jangkauan (*outlier*)

a. Pengecekan data kosong

Pengecekan data kosong dilakukan dengan melihat kolom missing value pada *datasheet*. Gambar 4 hasil pengecekan data *missing value*. Hasil pengecekan *datasheet* tidak mengandung data kosong

Name	Type	Missing
Nomhs	Polynomial	0
Nama	Polynomial	0
sks1	Integer	0
mk1	Integer	0
ip1	Real	0
sks2	Integer	0
mk2	Integer	0
ip2	Real	0

Gambar 4. Pengecekan data kosong

b. Pengecekan data yang menyimpang jauh (*outlier*)

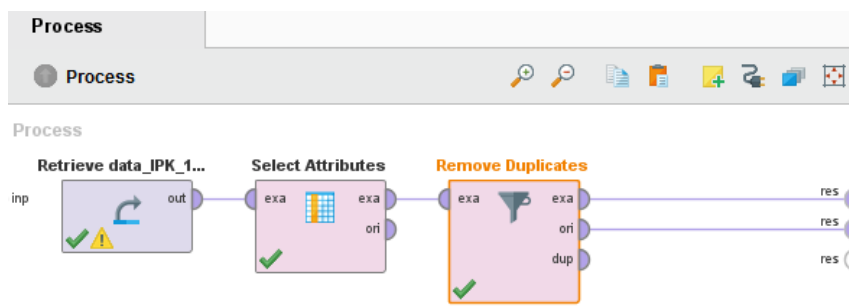
Pengecekan data pencilan dilihat dari nilai minimal dan maksimal dari setiap attribute. Gambar 5 informasi nilai minimal dan maksimal dari datasheet. Dari gambar tersebut, nilai minimal dan maksimal masih sesuai dengan data yang diperlukan, misal data IP masih berkisar 0-4

Name	Type	Missing	Statistics
Nomhs	Polynomial	0	Least: 171057039 (1) Most: 171021001 (1)
Nama	Polynomial	0	Least: Yulita Hatina (1) Most: Abdul Muis (1)
sks1	Integer	0	Min: 20 Max: 23
mk1	Integer	0	Min: 7 Max: 9
ip1	Real	0	Min: 0 Max: 4
sks2	Integer	0	Min: 15 Max: 24
mk2	Integer	0	Min: 5 Max: 9
ip2	Real	0	Min: 0 Max: 4

Gambar 5. Pengecekan data yang menyimpang jauh (pencilan)

c. Pengecekan data ganda

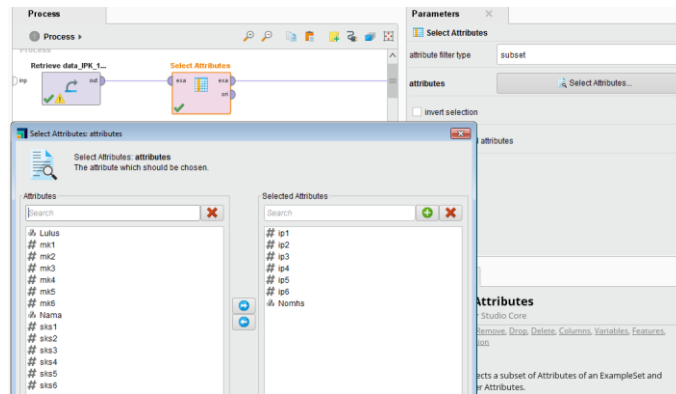
Pengecekan data ganda dilakukan dengan menggunakan *operator remove duplicate*. Operator ini akan menghapus jika ada data dalam satu baris yang mempunyai data yang sama. Gambar 5 penggunaan *operator remove duplicate* untuk menghapus data ganda. Hasil dari proses ini, *datasheet* tidak mengandung data kembar.



Gambar 5. Pengecekan data Ganda

3.2. Data Selection and Transformation

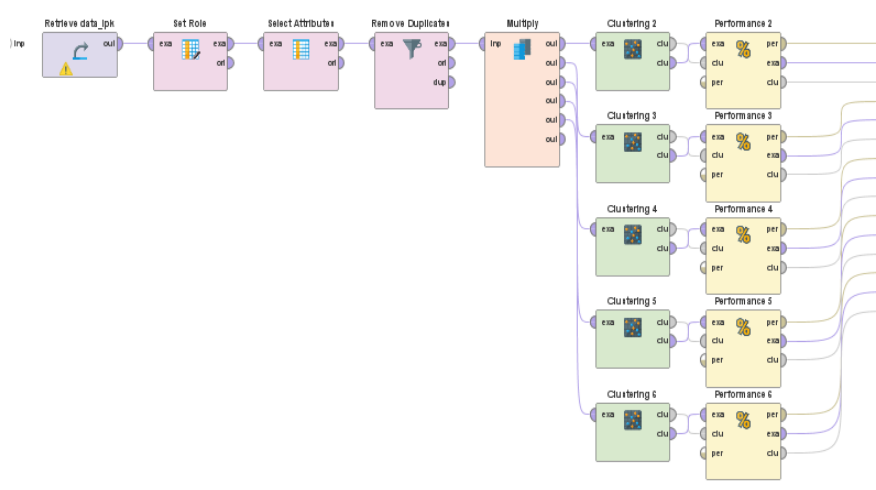
Proses ini dilakukan dengan melakukan seleksi pada attribute yang digunakan. Tidak semua attribute yang ada pada datasheet digunakan. Dalam penelitian ini attribute yang digunakan hanya attribute IP1, IP2, IP3, IP4, IP5 dan IP6. Proses pemilihan seleksi menggunakan operator select attribute. Gambar 7 penggunaan operator select attribute dalam proses pemilihan attribute.



Gambar 7. Penggunaan Operator Select Attribute Dalam Proses Pemilihan Attribute.

3.3. Data Mining

Pembuatan model clustering menggunakan algoritma K-Means. Proses pembuatan model di Rapid miner menggunakan operator K-Means dan proses evaluasi untuk uji performance menggunakan dengan cluster distance performance. Proses untuk mendapatkan cluster yang baik dilakukan pengujian dari cluster 2 sampai cluster sebanyak 6. Masing-masing hasil cluster dilakukan proses uji performance. Gambar 8, proses pembuatan model clustering dengan algoritma K-Means.

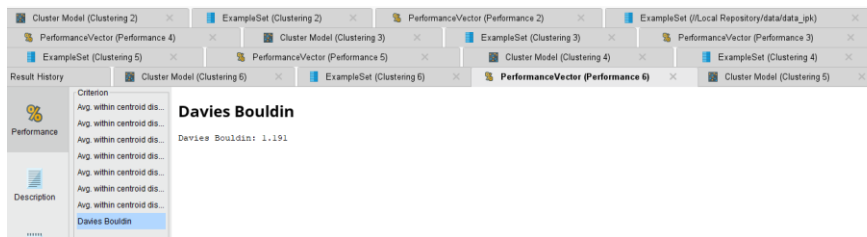


Gambar 8. Proses Pembuatan Model Clustering dengan Algoritma K-Means.

3.4. Evaluasi Model

Evaluasi model dilakukan untuk menemukan cluster yang baik dalam proses pembuatan pengelompokan data. Proses untuk mendapatkan cluster terbaik dilakukan dengan melihat hasil nilai evaluasi dengan Davies Bouldin. Hasil nilai Davies Bouldin, cluster yang paling baik adalah nilai evaluasi yang paling kecil. Gambar 9, merupakan hasil dari model K-Means yang

dikembangkan. Masing-masing algoritma K-means dari nilai K=1 sampai K=6 dilihat dan dievaluasi nilai *performance*.



Gambar 9. Hasil dari Proses Model yang dibuat

Hasil Davies Bouldin yang paling kecil adalah cluster sebanyak 2, sehingga model yang dipilih adalah model clustering dengan nilai K=2. Tabel 1, Hasil Davies Bouldin masing-masing nilai K dari 2 sampai K=6.

Tabel 1. Perbandingan nilai Davies Bouldin

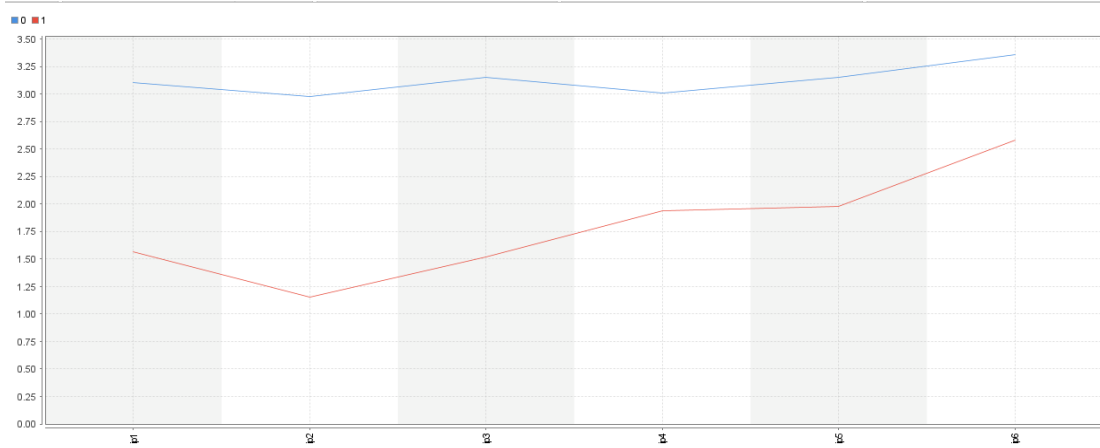
Banyak Cluster	Nilai Davies Bouldin
2	0.909
3	1.087
4	1.234
5	1.044
6	1.191

Pengelompokan *datasheet* dengan nilai K=2 terbagi menjadi cluster 0 sebanyak 160 data dan cluster 1 sebanyak 43 data. Tabel 2 menunjukkan banyak *cluster* untuk masing-masing nilai K.

Tabel 2. Banyak *Cluster* untuk Masing-Masing Nilai K.

Banyak Cluster	Pengelompokan data
2	kelompok 0: 160 data kelompok 1: 43 data
3	kelompok 0: 102 data kelompok 1: 74 data kelompok 2: 27 data
4	kelompok 0: 102 data kelompok 1: 7 data kelompok 2: 20 data kelompok 3: 74 data
5	kelompok 0: 100 data kelompok 1: 72 data kelompok 2: 23 data kelompok 3: 6 data kelompok 4: 2 data
6	kelompok 0: 5 data kelompok 1: 99 data kelompok 2: 6 data kelompok 3: 18 data kelompok 4: 68 data kelompok 5: 7 data

Pengelompokan *datasheet* menjadi 2 dapat dibuat kategori penamaan untuk masing-masing *cluster*. Pembuatan nama kategori dapat dilihat pada plot yang ada pada gambar 10. Pada gambar 10, grafik paling atas merupakan *cluster* 0 dan paling bawah merupakan *cluster* 1. Dari gambar 10, *cluster* 0 merupakan data mahasiswa yang mempunyai nilai indeks prestasi yang bagus dan *cluster* 1 merupakan data yang mempunyai indeks prestasi tidak bagus.



Gambar 10. Plot Hasil dari Cluster 0 dan Cluster 1

Hasil pengelompokan dapat dilihat mahasiswa yang masuk dalam cluster 0 dan cluster 1. Cluster 0 merupakan hasil pengelompokan data mahasiswa yang nilai IP nya di atas 3 dan cluster 1 berisi data mahasiswa yang berisi IP di bawah 3. Gambar 11, hasil cluster untuk masing-masing data.

Row No.	Na...	...	cluster	ip1	ip2	ip3	ip4	ip5	ip6
10	Nan...	...	cluster_0	2.580	2.480	2.730	3.670	2.880	3.800
11	Sri...	...	cluster_0	3.680	3.250	3.600	3.900	3.670	0.960
12	Her...	...	cluster_0	3.430	3.270	3.790	3.600	3.900	3.890
13	Sugi...	...	cluster_0	3.650	3.220	3.750	3.110	3.600	3.780
14	Bis...	...	cluster_0	3.080	3.330	3.030	2.990	3.110	3.950
15	Muh...	...	cluster_0	2.450	2.790	2.570	3.320	2.990	3.610
16	Ther...	...	cluster_0	2.800	3.050	3.290	3.120	3.320	3.200
17	Esa...	...	cluster_0	2.780	3.100	3.060	3.230	3.120	3.330
18	Mar...	...	cluster_0	2.720	2.860	3	3.150	3.230	3.330
19	Rez...	...	cluster_0	3.300	3.030	2.750	1.870	3.150	3.670
20	Rind...	...	cluster_0	2.900	3.050	2.800	2.090	1.870	3.890
21	Nur...	...	cluster_1	0.900	1.540	1.430	3.540	2.090	2.630
22	Selly...	...	cluster_0	3.550	3.250	3.720	3.170	3.540	3.370

Gambar 11. Hasil Cluster untuk Masing-Masing Data

4. KESIMPULAN DAN SARAN

Model *clustering* dapat menjadi salah satu model yang digunakan untuk membuat pengelompokan suatu datasheet. Pengelompokan data ini dapat digunakan dalam pengambilan keputusan dalam menentukan kelompok yang perlu ada kebijakan atau strategi untuk meningkatkan hasil agar data yang masuk kelompok yang kurang bagus menjadi atau masuk dalam kelompok yang bagus. Hasil penelitian dengan datasheet indeks prestasi mahasiswa dari semester 1 sampai semester 6 menghasilkan 2 kelompok Cluster 0 diberi nama kategori bagus dan cluster 1 diberi nama kategori tidak bagus. Dari kelompok ini, perlu strategi dalam pengambilan kebijakan, data mahasiswa yang masuk dalam kategori bagus harus dipertahankan dan dipantau agar dapat lulus tepat waktu. Mahasiswa yang masuk kategori 1, harus dilakukan pemantauan dan dipacu agar dapat menyelesaikan studi serta jangan sampai masa studi lewat dari 7 tahun.

DAFTAR PUSTAKA

- [1] S. S. Novel, *Kampuspedia Direktori PTN & PTS di Indonesia*. Jakarta: Grasindo, 2017.
- [2] M. Anangkota, *Jangan Takut Kuliah Sambil Berorganisasi*. Yogyakarta: Deepublish, 2019.
- [3] D. Cielen, A. D. B. Meysman, and M. Ali, *Introducing Data Science*. 2016.
- [4] M. Arhami and M. Nasir, *Data Mining - Algoritma dan Implementasi*. Yogyakarta: Penerbit Andi, 2020.
- [5] B. Santosa and A. Umam, *Buku Data Mining dan Big Data Analytics*. Bantul: Penebar Media Pustaka, 2018.
- [6] I. A. Nikmatun and I. Waspada, "Implementasi Data Mining untuk Klasifikasi Masa Studi Mahasiswa Menggunakan Algoritma K-Nearest Neighbor," *Jurnal SIMETRIS*, vol. 10, no. 2, pp. 421–432, 2019.
- [7] M. Syukri Mustafa, M. Rizky Ramadhan, and A. P. Thenata, "Implementasi Data Mining untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier," *Citec Journal*, vol. 4, no. 2, pp. 151–162, 2017.
- [8] D. F. Shiddieq and Patricia, "Implementasi Algoritma Data Mining Naive Bayes untuk Prediksi Kelulusan Mahasiswa," *Jurnal Komputer Bisnis*, no. 456, pp. 8–13, 2020.
- [9] C. Nas, "Data Mining Prediksi Minat Calon Mahasiswa Memilih Perguruan Tinggi Menggunakan Algoritma C4.5," *Jurnal Manajemen Informatika (JAMIKA)*, vol. 11, no. 2, pp. 131–145, 2021, doi: 10.34010/jamika.v11i2.5506.
- [10] M. N. Yatimah, "Implementasi Data Mining untuk Prediksi Kelulusan Tepat Waktu Mahasiswa STIMIK ESQ Menggunakan Decision Tree C4.5," *JUMANJI (Jurnal Masyarakat Informatika Unjani)*, vol. 5, no. 2, p. 89, 2021, doi: 10.26874/jumanji.v5i2.95.
- [11] R. Muliono and Z. Sembiring, "Data Mining Clustering Menggunakan Algoritma K-Means Untuk Klasterisasi Tingkat Tridarma Pengajaran Dosen," *CESS (Journal of Computer Engineering, System and Science)*, vol. 4, no. 2, pp. 2502–714, 2019.
- [12] N. Cahyana and A. Aribowo, "Metode Data Mining K-Means Untuk Klasterisasi Data Penanganan Dan Pelayanan Kesehatan Masyarakat," *Seminar Nasional Informatika Medis*, no. 5, pp. 24–31, 2018.
- [13] J. Nasir, "Penerapan Data Mining Clustering Dalam Mengelompokkan Buku Dengan Metode K-Means," *SIMETRIS*, vol. 11, no. 2, 2020.
- [14] Parjito and Permata, "Penerapan Data Mining untuk Clustering Data Penduduk Miskin Menggunakan Metode K-Means," *Jurnal Informatika*, vol. 3, no. 1, p. 7, 2021.
- [15] H. Priyatman, F. Sajid, and D. Haldivany, "Klasterisasi Menggunakan Algoritma K-Means Clustering untuk Memprediksi Waktu Kelulusan Mahasiswa," *Jurnal Edukasi dan Penelitian Informatika (JEPIN)*, vol. 5, no. 1, p. 62, 2019, doi: 10.26418/jp.v5i1.29611.
- [16] D. R. Yusian, "Analisa Penerapan Data Mining Pada Penerimaan Mahasiswa Politeknik Negeri Lhokseumawe Menggunakan Algoritma K-Means," *Journal of Informatics and Computer ...*, vol. 4, no. 2, pp. 208–216, 2018.
- [17] W. Lestari, "Clustering Data Mahasiswa Menggunakan Algoritma K-Means Untuk Menunjang Strategi Promosi (Studi Kasus : STMIK Bina Bangsa Kendari)," *Simkom*, vol. 4, no. 2, pp. 35–48, 2019, doi: 10.51717/simkom.v4i2.37.
- [18] A. K. Turkhamun, B. Panjaitan, and R. Guntara, "Implementasi Data Mining Clustering Data Mahasiswa Teknik Informatika Menggunakan Algoritma K-Means," *Seminar Nasional Cendekiawan*, vol. 1, no. ISSN (P) : 2460-8696 ISSN (E) : 2540-7589, pp. 805–810, 2018.
- [19] R. W. Sari, H. Dedy, I. dan Gunawan, and W. P. Agus, "Aplikasi RapidMiner dalam Pengelompokkan Kasus Penyakit AIDS berdasarkan Provinsi dengan Data Mining K-means Clustering," *Regional Development Industry & Health Science, Technology and Art of Life*, pp. 59–69, 2018.
- [20] R. W. Sari, A. Wanto, and A. P. Windarto, "Implementasi Rapidminer Dengan Metode K-Means (Study Kasus: Imunisasi Campak Pada Balita Berdasarkan Provinsi)," *KOMIK*

- (*Konferensi Nasional Teknologi Informasi dan Komputer*), vol. 2, no. 1, pp. 224–230, 2018, doi: 10.30865/komik.v2i1.930.
- [21] Amanda and M. Veronica Sitorus, “Jurnal Ilmiah MIKA AMIK Al Muslim Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Konsumsi Produk Kosmetik milik PT Cedefindo,” *Jurnal Ilmiah MIKA AMIK Al Muslim*, vol. V, no. 2, pp. 63–68, 2021.
- [22] D. A. Manalu, G. Gunadi, and T. Informatika, “Implementasi Metode Data Mining K-Means Clustering Terhadap Data Pembayaran Transaksi Menggunakan Bahasa Pemrograman Python Pada CV Digital Dimensi,” *Infotech: Journal of Technology Information*, vol. 8, no. 1, pp. 45–54, 2022.
- [23] B. Suharjo, “Application of K-Means Cluster and Spatial Statistics using Python to Analyze the Indicators of Indonesia Information Technology,” *Digital Zone: Jurnal Teknologi Informasi & Komunikasi*, vol. 12, no. 1, pp. 11–18, 2021.
- [24] R. M. Sagala, “Prediksi Kelulusan Mahasiswa Menggunakan Data mining Prediction of college subject using K-means Algorithm in Data mining,” *Jurnal TelKa*, vol. 11, no. 2, pp. 131–142, 2021.
- [25] M. Sholeh and D. Andayati, “Machine Linear untuk Analisis Regresi Linier Biaya Asuransi Kesehatan dengan Menggunakan Python Jupyter Notebook,” *Jurnal Edukasi dan Penelitian Informatika*, vol. 8, no. 1, pp. 20–27, 2022.
- [26] H. Cuesta and S. Kumar, *Practical Data Analysis Second Edition*. Birmingham: Packt Publishing Ltd, 2016.
- [27] S. E. Deny Jollyta , Muhammad Siddik , Herman Mawengkang, *Teknik Evaluasi Cluster Solusi Menggunakan Python Dan Rapidminer*. Yogyakarta: Deepublish Publisher, 2021.