



CLUSTERING DATA NILAI UJIAN AKHIR SEMESTER MENGUNAKAN ALGORITMA DATA MINING K-MEANS

^{1)*} **Indra Maulana**

Institut Pendidikan dan Bahasa Invada Cirebon
indra@stkipinvada.ac.id

²⁾ **Utami Rosalina**

Institut Pendidikan dan Bahasa Invada Cirebon
utamirosalina@stkipinvada.ac.id

Artikel history

Diterima : 7 Januari 2021
Direvisi : 15 April 2021
Disetujui : 12 Mei 2021

Kata Kunci: *Clustering, K-Means, Akademik*

Keywords: *Clustering, K-Means, Academic*

Abstrak

Menganalisis kinerja siswa dalam pendidikan merupakan area aktif dan cukup baru dalam penelitian pendidikan. Pemetaan kinerja siswa berguna bagi guru dan siswa. Namun, faktor-faktor yang mempengaruhi kinerja siswa perlu diidentifikasi terlebih dahulu untuk membangun model prediksi awal tersebut. Klastering nilai UAS bertindak sebagai indikator apakah penyebaran pengetahuan dari guru ke siswa sudah merata atau belum. Ini sangat relevan dengan keadaan pandemic covid yang sedang terjadi di Indonesia dan dunia, kita bisa mengetahui klaster siswa yang benar benar belajar mandiri ataupun tidak. *K-Means* dapat digunakan untuk memetakan kemampuan akademik mahasiswa, agar dapat mengoptimalkan pembelajaran dan pemahaman mahasiswa.

Abstract

Analyzing student performance in education is an active and fairly new area of educational research. Mapping student performance is useful for both teachers and students. However, the factors that affect student performance need to be identified first to build the initial predictive model. The clustering of UAS scores acts as an indicator of whether the spread of knowledge from teachers to students is evenly distributed or not. This is very relevant to the state of the covid pandemic that is happening in Indonesia and the world, we can know which clusters of students are truly self-study or not. *K-Means* can be used to map students' academic abilities, in order to optimize student learning and understanding.

Koresponden: indra@stkipinvada.ac.id

artikel dengan akses terbuka di bawah lisensi
CC BY SA
2021



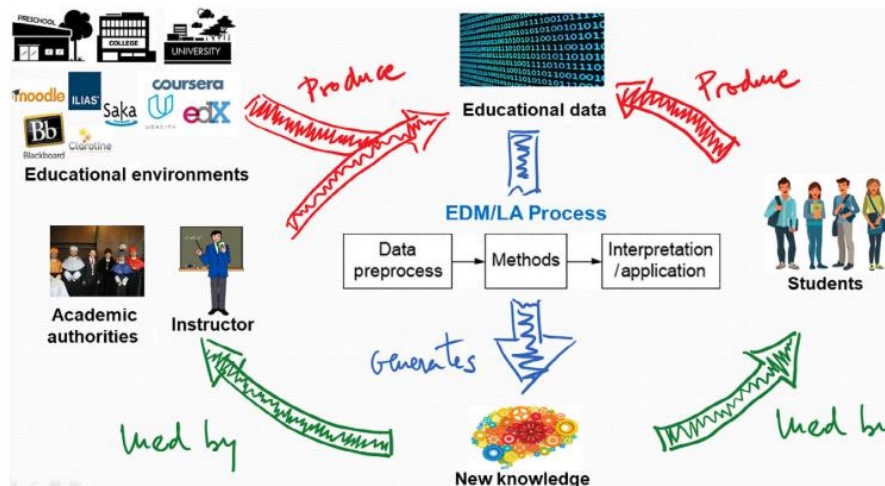
PENDAHULUAN

Perkembangan ilmu pengetahuan dan teknologi akan mempermudah pekerjaan manusia. Namun, di lain sisi hal ini akan meningkatkan persaingan baik itu persaingan lokal maupun global Untuk menghadapi persaingan, perlu dipersiapkan sumber daya manusia. Dalam hal ini mahasiswa yang kompeten dalam bidang akademik berupa pengetahuan dan keterampilan. Salah satu komponen yang menjadi kunci tinggi rendahnya mutu pendidikan adalah baik tidaknya teknik evaluasi hasil belajar yang digunakan oleh guru. Salah satu cara untuk mengukur kompetensi akademik adalah dengan nilai mata pelajaran. Untuk mengetahui tingkat kompetensi akademik, diperlukan suatu prosedur penilaian. Ada beberapa komponen penilaian dalam menentukan nilai akhir semester di sekolah, yang salah satunya adalah nilai Ujian Akhir Semester (UAS). Tes dapat didefinisikan sebagai suatu pernyataan atau tugas atau seperangkat tugas yang direncanakan untuk memperoleh informasi tentang trait(sifat) atau atribut pendidikan atau psikologik yang setiap butir pertanyaan atau tugas tersebut mempunyai jawaban atau ketentuan yang dianggap benar (Djemari Mardapi, 2008). Dengan mengetahui nilai UAS, Dosen dapat mengetahui sebaran siswa dalam hal kompetensi akademik. Penilaian adalah proses pengumpulan dan pengolahan informasi untuk mengukur pencapaian hasil belajar peserta didik. Kegiatan Penilaian memerlukan instrumen penilaian dan teknik penilaian. Penilaian tidak hanya difokuskan pada hasil belajar tetapi juga pada proses belajar (Wiwik Setiawati, Oktavia Asmira, Yoki Ariyana, Reisky Bestary, & Pudjiastuti, 2019). Berdasarkan pendapat tersebut dapat disimpulkan bahwa evaluasi merupakan proses yang sistematis dan berkelanjutan untuk mengumpulkan, mendeskripsikan, menginterpretasikan dan menyajikan informasi untuk dapat digunakan sebagai dasar membuat keputusan dan atau menyusun kebijakan. memberikan informasi kepada dosen untuk mengobservasi atau mendeskripsikan satu atau lebih karakteristik siswa dengan menggunakan skala numerik

Dengan pertimbangan tersebut, perlu diadakan pengelompokan mahasiswa berdasarkan nilai UAS agar guru dapat menerapkan metode pembelajaran yang tepat dan menjadi bahan pertimbangan untuk menentukan kelompok belajar sesuai dengan sebaran nilai ujian siswa. Salah satu metode untuk mengelompokan data adalah dengan menggunakan algoritma *K-Means*. Pekerjaan seperti ini disebut dengan *data mining*. Menganalisis kinerja siswa dalam pendidikan merupakan area aktif dan cukup baru dalam penelitian pendidikan. Prediksi kinerja siswa berguna bagi guru dan siswa. Namun, faktor-faktor yang mempengaruhi kinerja siswa perlu diidentifikasi terlebih dahulu untuk membangun model prediksi awal tersebut. Klastering nilai UAS bertindak sebagai indikator apakah penyebaran pengetahuan dari guru ke siswa sudah errata atau belum. Ini sangat relevan dengan keadaan pandemic covid yang sedang terjadi di Indonesia dan dunia, kita bisa mengetahui klaster siswa yang benar benar belajar mandiri ataupun tidak. Penambangan data dalam dunia pendidikan berguna untuk mengekstraksi hubungan tersembunyi dari kumpulan data yang cukup besar, seperti nilai UTS, UAS, Ulangan Harian, absensi dll. Menurut Anupam Khan (Khan & Ghosh, 2018), Data mining dapat membantu meningkatkan pemahaman proses pembelajaran dalam bentuk Klaster, Asosiasi, dan Klasifikasi.

Data mining merupakan proses dalam menemukan hubungan yang berarti, pola dan tren dengan memeriksa data berukuran besar dalam suatu penyimpanan dengan menggunakan teknologi pengenalan pola, misalnya statistik dan matematika. Data mining adalah kegiatan mengekstrak informasi atau pengetahuan penting dari suatu set data berukuran besar dengan menggunakan Teknik tertentu. Informasi atau *knowledge* yang dihasilkan dari data mining bisa dipakai untuk memperbaiki keputusan. Dinamakan data mining atau penambangan data karena proses penemuan informasi dalam set data dilakukan seperti orang melakukan kegiatan penambangan. Data mining melibatkan penggunaan meted atau *tools* untuk mendeteksi pola dan melakukan tugas prediksi ataupun segmentasi. Ketepatan prediksi ataupun segmentasi

inilah keuntungan yang diharapkan dalam bisnis dari penggunaan data mining (Santosa & Umam, 2018) (Rutkauskiene, Gudoniene, & Maskeliunas, 2016).

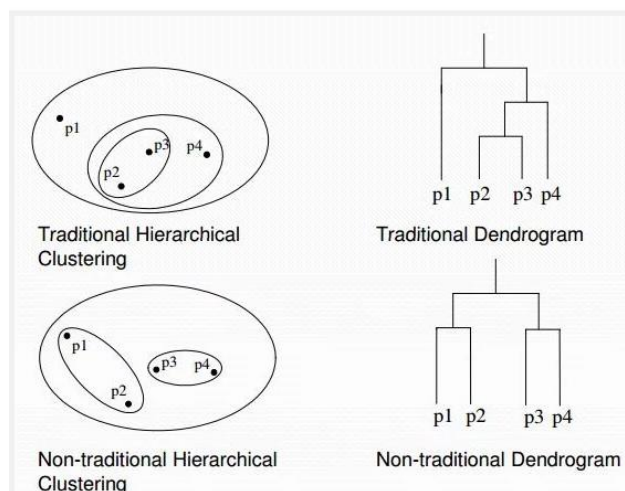


Gambar 1. Data mining Education

(Romero & Ventura, 2020)

Penambangan data dalam dunia pendidikan berguna untuk mengekstraksi hubungan tersembunyi dari kumpulan data yang cukup besar, seperti nilai UTS, UAS, Ulangan Harian, absensi dll. Menurut Anupam Khan (Khan & Ghosh, 2018), Data mining dapat membantu meningkatkan pemahaman proses pembelajaran dalam bentuk Klaster, Asosiasi, dan Klasifikasi.

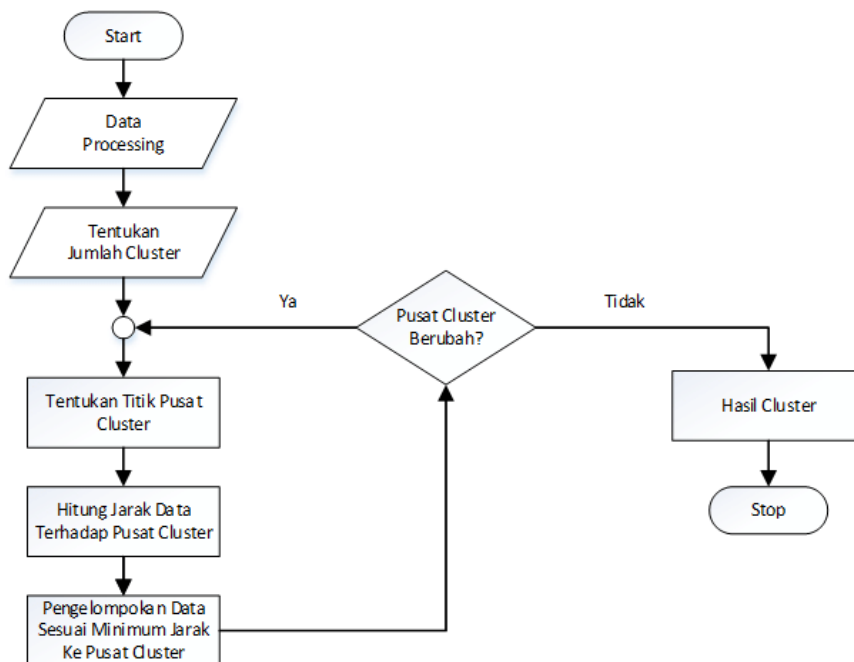
K-means merupakan salah satu metode data cluster non hirarki yang berusaha



mempartisi data yang ada ke dalam bentuk satu atau lebih cluster/ kelompok Algoritma K-means merupakan salah satu algoritma clustering (pengelompokan). K-means clustering merupakan metode clustering non-hirarki yang mengelompokkan data dalam bentuk satu atau lebih cluster/kelompok. Data-data yang memiliki karakteristik yang sama dikelompokkan dalam satu cluster/kelompok dan data yang memiliki karakteristik yang berbeda dikelompokkan dengan cluster/kelompok yang lain, sehingga data yang berada dalam satu cluster/kelompok memiliki tingkat variasi kecil. Metode **K-Means** berusaha mengelompokkan data yang ada ke dalam beberapa kelompok, dimana data dalam satu kelompok tersebut mempunyai karakteristik yang sama satu sama lainnya dan mempunyai karakteristik yang berbeda dengan data yang ada di dalam kelompok yang lain.

Gambar 2. Partitional clustering

Dengan kata lain, metode K-Means Clustering bertujuan untuk meminimalisasikan objective function yang diset dalam proses clustering dengan cara meminimalkan variasi antar data yang ada di dalam suatu cluster dan memaksimalkan variasi dengan data yang ada di cluster lainnya. Data clustering menggunakan metode K-Means ini secara umum dilakukan dengan algoritma dasar sebagai berikut: (1) tentukan jumlah cluster; (2) alokasikan data ke dalam cluster secara random; (3) hitung centroid/rata-rata dari data yang ada di masing-masing cluster; (4) alokasikan masing-masing data ke centroid/rata-rata terdekat; (5) kembali ke Step 3, apabila masih ada data yang berpindah cluster atau apabila perubahan nilai centroid, ada yang di atas nilai threshold yang ditentukan atau apabila perubahan nilai pada objective function yang digunakan di atas nilai threshold yang ditentukan.

**Gambar 3.** Flowchart K-Means

Metode clustering menggunakan Algoritma K-means, ukuran kedekatan data dihitung menggunakan jarak Euclidean. Algoritma K-means bertujuan untuk meminimumkan jarak total Euclidean diantara setiap titik x_i , dan cluster terdekat yakni c_j (Wu et al., 2008).

$$d_{ij} = \sqrt{\sum_{k=1}^n \{x_{ik} - x_{jk}\}^2}$$

Keterangan:

d_{ij} = jarak antara data ke-i dan data ke- j

n = dimensi data

x_{ik} = koordinat data ke-i pada dimensi k

x_{jk} = koordinat data ke-j pada dimensi k

Penelitian tentang penerapan Klastering dengan menggunakan K-means dilakukan oleh beberpa peneliti sebelumnya. Khoiriyatus Sya'iyah (Sya'iyah, Yuliansyah, & Arfiani, 2019) melakukan penelitian dengan data siswa, Dengan pengelompokan data siswa, maka institusi pendidikan mendapatkan potensi ilmu yang bermanfaat. Metodologi dalam

penelitian ini dibagi menjadi lima tahap, yaitu pembersihan data, pemilihan data, transformasi data, pengelompokan menggunakan K-Means Algorithms, dan presentasi pengetahuan. Peneliti membagi cluster data siswa menjadi tiga kelompok. Peneliti ingin mendapatkan karakteristik siswa yang berprestasi, berprestasi standar, dan berprestasi rendah, menggunakan 724 data siswa dan empat variable. Dan peneliti menemukan hasil pengelompokan data yang dilakukan memiliki titik tengah yang berdekatan sehingga tidak dapat merepresentasikan karakteristik siswa yang ingin melakukan pencarian. Peneliti tersebut memprediksi bahwa jumlah data yang kami gunakan terlalu kecil dan perlu mengevaluasi distribusi data. Perbaikan ini akan berhasil di masa depan penelitian kami.

Dedi Kuswandi dkk (Kuswandi, Surahman, Thaariq, & Muthmainnah, 2018), melakukan penelitian yang bertujuan untuk menggambarkan Persepsi mahasiswa terhadap penerapan model pembelajaran berbasis proyek pada mata kuliah perancangan program pelatihan dan strategi dengan mengelompokkan data menggunakan metode K-Means. Hasil penelitian menunjukkan bahwa dari keempat klaster sangat setuju, setuju, tidak setuju dan sangat tidak setuju klaster 1 (setuju) memiliki data paling banyak yaitu 18 data yang menunjukkan bahwa siswa setuju model pembelajaran berbasis proyek mampu memberikan kepuasan ketika proyek selesai, model pembelajaran yang inovatif, praktek keterampilan manajemen proyek, melatih budaya kerjasama, meminimalkan dominasi dosen dalam pembelajaran, berorientasi pada optimalisasi potensi mahasiswa, melatih keterampilan berpikir kreatif, melatih keterampilan kognitif dalam kebutuhan materi analitik, merancang, mengembangkan, melaksanakan dan mengevaluasi program pendidikan dan pelatihan, membekali keterampilan kognitif, afektif dan psikomotorik serta keterampilan belajar yang menantang, sehingga cocok untuk diterapkan dalam mata kuliah desain dan strategi perkuliahan.

METODE PENELITIAN

Penelitian ini mengambil data mahasiswa dari STKIP Invada Cirebon, dan menggunakan nilai Ujian Akhir Semester pada jurusan Prodi Pendidikan Teknik Informatika dan Komputer. Preprocessing dilakukan dengan tujuan untuk menghapus data yang mengandung outlier/noise, data yang tidak konsisten serta data yang tidak lengkap atau missing value. Pemrosesan awal data adalah tugas yang sulit dan rumit, dan terkadang pemrosesan awal data itu sendiri menghabiskan lebih dari setengah dari total waktu yang dihabiskan untuk menyelesaikan masalah data mining. Data dari proses pembelajaran yang tersedia untuk proses klastering tidak dalam bentuk yang sesuai, oleh karena itu perlu dilakukan pembersihan data, kedalam data yang sesuai dan siap untuk digunakan.

Pada tahap ini, metode clustering digunakan untuk mengelompokkan data menjadi tiga kelompok, yaitu sangat baik, baik dan kurang baik. Proses klastering data ujian Akhir Semester ini menggunakan algoritma K-means.

Uji validitas cluster digunakan untuk mengevaluasi hasil dari analisis cluster secara kuantitatif sehingga dihasilkan kelompok optimal. Kelompok optimal adalah kelompok yang mempunyai jarak yang padat antar individu dalam kelompok dan terisolasi dari kelompok lain dengan baik. Evaluasi penelitian ini menggunakan metode Davies Bouldin Index, yang merupakan metrik untuk mengevaluasi algoritma klastering dengan skema evaluasi internal. Berfungsi untuk memvalidasi seberapa baik klastering yang telah dilakukan. Pengukuran ini bertujuan untuk memaksimalkan jarak intercluster antara tiap klaster. Semakin kecil nilai Davies Bouldin Index menunjukkan skema klaster yang optimal dan itu menunjukkan kualitas klaster terbaik (Rinaldi, Surlanto, Sudrajat, & Kurnia, 2019).

Davies Bouldin Index (DBI) adalah menghitung rata-rata nilai setiap titik pada

himpunan data. Perhitungan nilai setiap titik merupakan jumlah nilai compactness yang dibagi dengan jarak antara kedua titik pusat cluster sebagai separation. DBI digunakan untuk mengoptimalkan jarak diluar cluster dan meminimalkan jarak didalam cluster dengan persamaannya.

HASIL DAN PEMBAHASAN

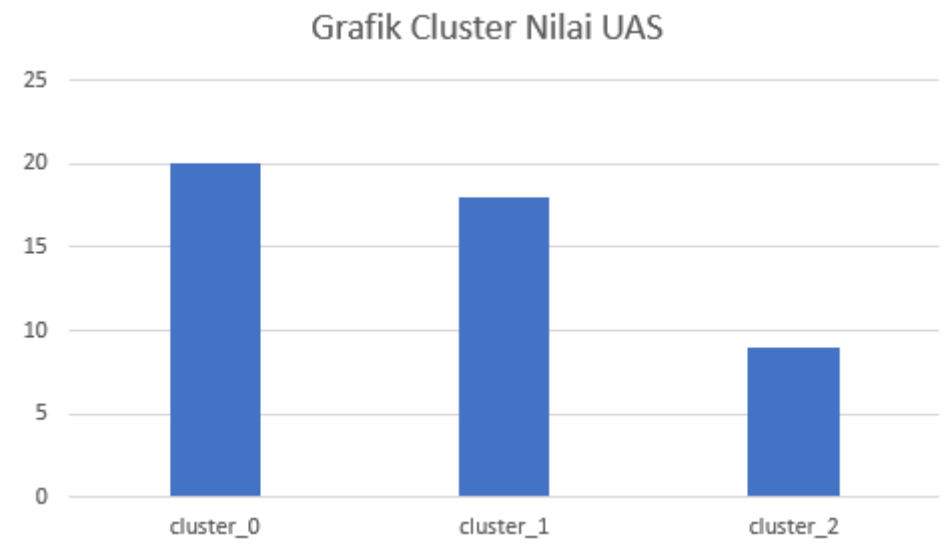
Pada tahap ini penulis akan menampilkan hasil implementasi dari tools data mining yang telah digunakan, berikut dapat dilihat pada gambar-gambar tampilan dibawah ini :

Tabel 1. Centroid masing-masing cluster

Attribute	cluster_0	cluster_1	cluster_2
att_1	-0.875	-0.738	-0.949
att_2	-0.574	-0.220	-0.475
att_3	-0.474	0.207	-0.518
att_4	1.318	-0.550	-0.930
att_5	0.820	-0.178	-0.762
att_6	-0.375	0.527	-0.339
att_7	-1.470	-0.277	-0.305
att_8	0.123	0.250	0.711
att_9	1.019	0.049	0.041
att_10	-0.574	-0.160	-0.549
att_11	0.820	-0.114	-0.719
att_12	0.820	-0.170	0.895
att_13	0.422	-0.665	-0.414
att_14	1.318	-0.499	0.718
att_15	-1.470	-1.127	-0.056
att_16	0.123	-0.032	-0.614
att_17	0.621	-0.127	0.254
att_18	0.820	-0.251	-1.160
att_19	-1.768	-0.212	0.857
att_20	-1.470	0.057	0.895
att_21	1.318	0.386	-0.128
att_22	1.318	-0.195	-0.230
att_23	-1.370	-0.875	0.441
att_24	-1.270	-0.554	0.929
att_25	-0.320	0.361	0.505
att_26	-0.867	0.720	-0.235

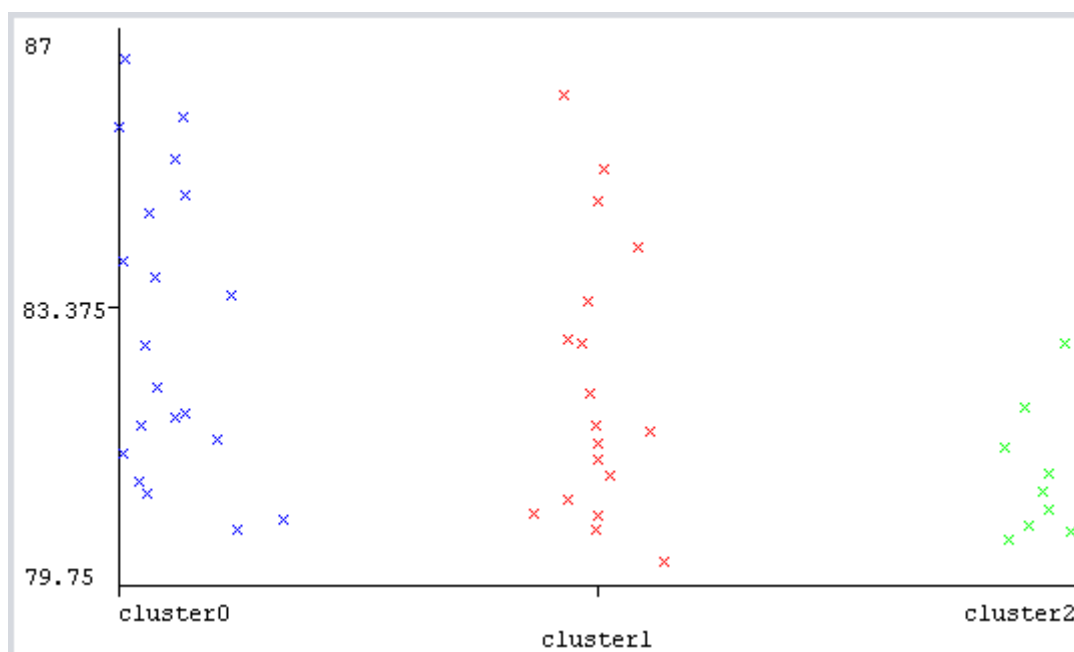
Hasil dari tabel diatas menunjukkan bahwa data berdistribusi normal. Berdasarkan hasil perhitungan normalitas diatas menunjukkan bahwa Asymp. Sig. (2-tailed) 0,321 lebih besar dari pada 0,05 dan hal tersebut dapat disimpulkan bahwa sebaran data kuisisioner berdistribusi normal.

Dari table diatas, diketahui bahwa cluster 1, diisi oleh Pengantar Pendidikan dan Kewarganegaraan, cluster 2, diisi oleh Pengantar mata kuliah Bahasa Inggris dan Kesenian, sedangkan cluster 3 diisi oleh mata kuliah Fisika dan Matematika.

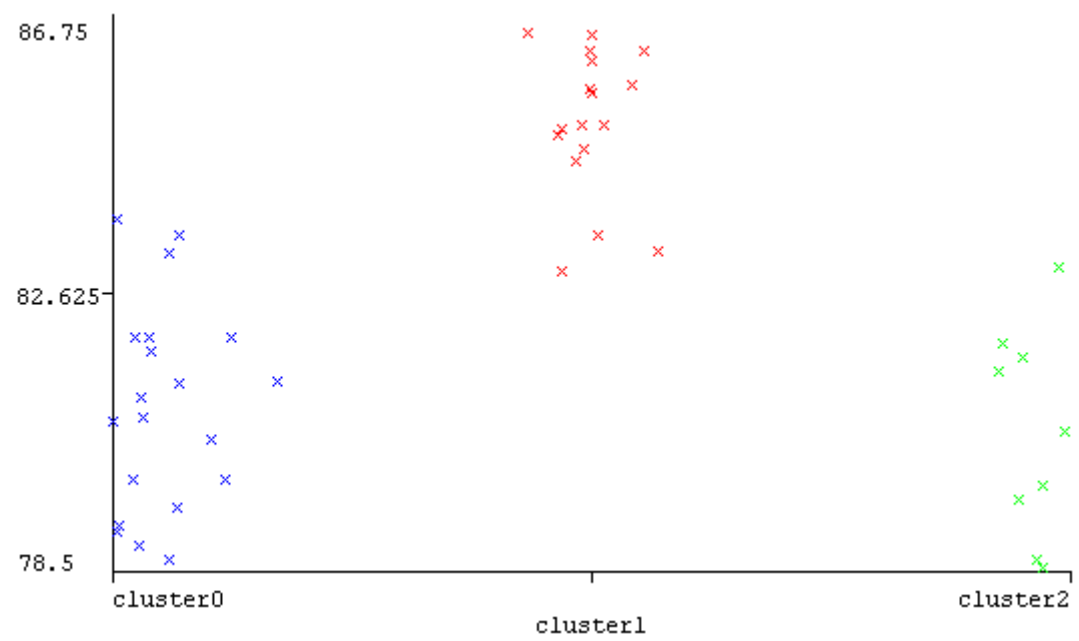


Gambar 4. Grafik Cluster

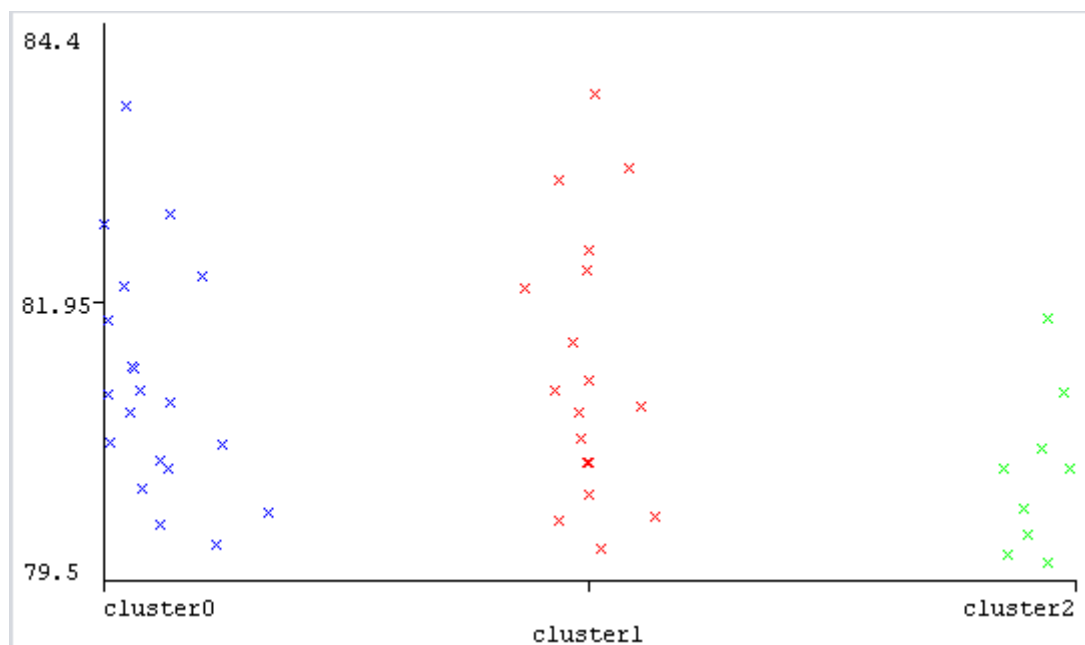
Dari grafik diatas dapat kita lihat jumlah masing-masing kluster, kluster 1 (cluster_0) terdapat 20siswa, kluster 2 (cluster_1) terdapat 18siswa, kluster 3 (cluster_2) terdapat 9siswa. Pada kluster 1 memiliki nilai rata-rata yang tinggi, sehingga disimpulkan bahwa cluster_0 sangat baik, cluster_1 baik, dan cluster_2 adalah kluster yang kurang.



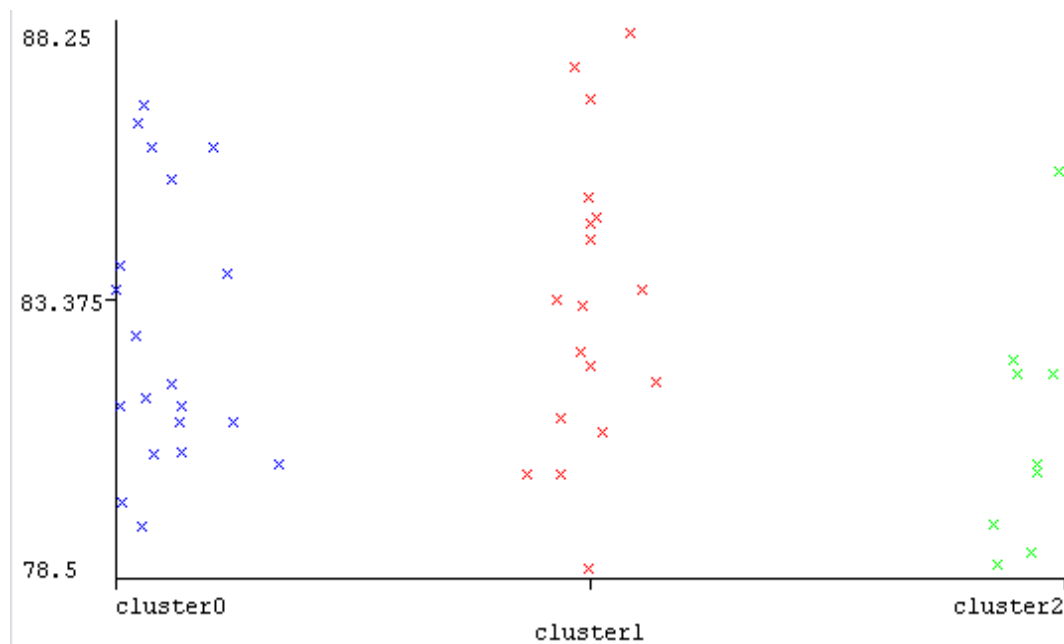
Gambar 5. Cluster MataKuliah Kewarnageraan



Gambar 6. Cluster Mata Kuliah Bahasa Inggris



Gambar 7. Cluster Mata Kuliah Fisika Dasar



Gambar 8. Cluster Mata Kuliah Matematika Dasar

Pada gambar diatas diketahui klaster 1 memiliki nilai yang cukup tinggi, ini menunjukkan bahwa mahasiswa pada klaster 1 memiliki nilai yang tinggi dengan kata lain pada klaster 1 adalah kumpulan mahasiswa dengan nilai akademis yang sangat baik.

Table 2. Performance K-means

Avg. within centroid distance:	-14.389
Avg. within centroid distance_cluster_0	-12.741
Avg. within centroid distance_cluster_1	-16.079
Avg. within centroid distance_cluster_2	-14.347
Davies Bouldin	-1.117

KESIMPULAN

Adapun beberapa kesimpulan yang dapat diambil dari penelitian ini yaitu, antara lain: (1) dengan menerapkan metode k-means clustering pada pengolahan data dari keembat jenis data yang penulis gunakan sebagai bahan penelitian ini dan penulis gabungkan menjadi 1 jenis data menghasilkan sebuah hasil (output) yang berbeda-beda pula seperti yang telah dijelaskan sebelumnya; (2) dari hasil (output) yang telah diperoleh dari proses penerapan metode k-means clustering, maka hasil dari analisis dan evaluasi yang di dapat berupa sebuah kelompok data yang telah ter-cluster dari masing-masing kelompok, dan hasil (output) tersebut dapat dijadikan sebuah informasi yang berguna bagi Prodi dan juga Institusi , serta dapat menjadi acuan untuk dunia Pendidikan dapat menerapkan metode k-means clustering dalam membantu proses klaster nilai Ujian mereka selanjutnya; (3) K-Means dapat digunakan untuk memetakan kemampuan akademik mahasiswa, agar dapat mengoptimalkan pembelajaran dan pemahaman mahasiswa.

REFRENSI

Djemari Mardapi. (2008). *Teknik penyusunan Instrumen tes dan nontes* (Vol. 127). Yogyakarta: Mitra Cendikia.

- Khan, Anupam, & Ghosh, Soumya K. (2018). Data mining based analysis to explore the effect. *Springer*.
- Kuswandi, Dedi, Surahman, Ence, Thaariq, Zahid Zufar At, & Muthmainnah, Mahmuda. (2018). K-Means Clustering of Student Perceptions on Project-Based Learning Model Application. *2018 4th International Conference on Education and Technology, ICET 2018*, 9–12. <https://doi.org/10.1109/ICEAT.2018.8693932>
- Rinaldi, Ade Rizki, Surlanto, Lana, Sudrajat, Dadang, & Kurnia, Dian Ade. (2019). AnalisaTingkat Kepuasan Mahasiswa Terhadap Layanan Pembelajaran Menggunakan K-Means dan Algoritma Genetika. *Jurnal ICT: Information Communication & Technology*, 18(1), 60–64. <https://doi.org/10.36054/jict-ikmi.v18i1.55>
- Romero, Cristobal, & Ventura, Sebastian. (2020). Educational data mining and learning analytics: An updated survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 10(3), 1–21. <https://doi.org/10.1002/widm.1355>
- Rutkauskiene, Danguole, Gudoniene, Daina, & Maskeliunas, Rytis. (2016). *Smart Education and e-Learning 2016* (Vol. 59). <https://doi.org/10.1007/978-3-319-39690-3>
- Santosa, Budi, & Umam, Ardian. (2018). Data mining dan bigdata analysis, Teori dan implementasi menggunakan pyton dan apache spark. In Isya (Ed.), *Penebar Media Pustaka* (2nd ed., Vol. 2). Yogyakarta: Penebar Media Pustaka.
- Sya'iyah, Khoiriyatus, Yuliansyah, Herman, & Arfiani, Ika. (2019). Clustering student data based on K-means algorithms. *International Journal of Scientific and Technology Research*, 8(8), 1014–1018.
- Wiwik Setiawati, M. P., Oktavia Asmira, MT, Yoki Ariyana, MT, Reisky Bestary, M. P., & Pudjiastuti, Dr. Ari. (2019). Implementasi Program Pengembangan Keprofesian Berkelanjutan Guru melalui Peningkatan Kompetensi Pembelajaran Berbasis Zonasi. In *Direktorat Jenderal Guru dan Tenaga Kependidikan*. <https://doi.org/10.37411/pedagogika.v10i2.60>
- Wu, Xindong, Kumar, Vipin, Ross, Quinlan J., Ghosh, Joydeep, Yang, Qiang, Motoda, Hiroshi, McLachlan, Geoffrey J., Ng, Angus, Liu, Bing, Yu, Philip S., Zhou, Zhi Hua, Steinbach, Michael, Hand, David J., & Steinberg, Dan. (2008). Top 10 algorithms in data mining. *Knowledge and Information Systems*, 14(1), 1–37. <https://doi.org/10.1007/s10115-007-0114-2>