

Penggunaan Algoritma K-Means dan K-Medoids untuk Pengembangan Strategi Promosi Penerimaan Mahasiswa Baru

Devita Maulina Putri^{1*}
Asri Samsiar Ilmananda²
Nadita Prisanta³

^{1,2,3}Sistem Informasi, Universitas Merdeka Malang, Jalan Terusan Dieng No. 62-64, Klojen, Pisang Candi, Sukun, Jawa Timur, 65146, Indonesia

¹devita.maulina@unmer.ac.id, ²asri.ilmananda@unmer.ac.id, ³2208100012@student.unmer.ac.id

***Penulis Korespondensi:**
Devita Maulina Putri
devita.maulina@unmer.ac.id

Abstrak

Penerimaan mahasiswa baru merupakan aktivitas krusial bagi perguruan tinggi, terutama perguruan tinggi swasta, dalam memperoleh mahasiswa baru. FTI Unmer Malang telah menerapkan berbagai teknik promosi, namun masih mengalami kendala dalam mencapai target penerimaan mahasiswa. Jumlah mahasiswa baru mengalami fluktuasi, dengan puncak penerimaan pada tahun 2019 dan penurunan signifikan sebesar 23% dalam tiga tahun terakhir. Salah satu masalah utama adalah kurangnya penyebaran informasi ke daerah terpencil. Untuk mengatasi masalah tersebut, pada penelitian ini diterapkan metode data mining dengan klusterisasi untuk mengelompokkan data mahasiswa baru berdasarkan daerah asal. Dua algoritma clustering, yaitu K-Means dan K-Medoids, digunakan untuk membandingkan hasil klusterisasi guna menemukan strategi promosi yang optimal. Data yang digunakan mencakup mahasiswa baru dari tahun akademik 2016 hingga 2022. Hasil penelitian menunjukkan bahwa algoritma K-Means memperlihatkan kinerja yang lebih baik daripada algoritma K-Medoids dengan tingkat akurasi indeks DBI sebesar 0,344. Hasil penelitian diharapkan dapat membantu FTI Unmer Malang dalam menentukan strategi promosi yang lebih efektif berdasarkan daerah asal mahasiswa.

Kata Kunci: K-Means; K-Medoids; Strategi Promosi; Penerimaan Mahasiswa Baru

Abstract

New student admission is a crucial activity for universities, especially private universities, in obtaining new students. FTI Unmer Malang has implemented various promotional techniques, but still experiences obstacles in achieving student admission targets. The number of new students fluctuates, with a peak in admission in 2019 and a significant decrease of 23% in the last three years. One of the main problems is the lack of information dissemination to remote areas. To overcome this problem, this study applies a data mining method with clustering to group new student data based on their area of origin. Two clustering algorithms, namely K-Means and K-Medoids, are used to compare clustering results to find the optimal promotion strategy. The data used includes new students from the 2016 to 2022 academic years. The results of the study show that the K-Means algorithm shows better performance than the K-Medoids algorithm with DBI index accuracy level of 0.344. The results of the study are expected to help FTI Unmer Malang in determining a more effective promotion strategy based on the student's area of origin.

Keywords: K-Means; K-Medoids; Promotion Strategies; New Student Admissions

1. Pendahuluan

Penerimaan mahasiswa baru merupakan kegiatan ujung tombak perguruan tinggi dalam mendapatkan mahasiswa baru khususnya untuk perguruan tinggi swasta. Keberhasilan dalam penyelenggaraan lembaga perguruan tinggi bergantung pada pengelolaan komponen-komponen pendukung pelaksana kegiatan. Banyak cara maupun teknik promosi yang digunakan FTI Unmer Malang dalam hal mendapatkan mahasiswa baru, antara lain melalui promosi, media cetak, baliho, hingga *website*. Media promosi melalui media sosial memberikan dapat pengaruh positif terhadap penerimaan mahasiswa [1].

Selama 10 tahun terakhir, jumlah mahasiswa baru di FTI Unmer Malang mengalami perubahan secara fluktuatif [2]. Jumlah mahasiswa baru tertinggi terjadi pada tahun 2019 yaitu sebanyak 285

mahasiswa. Sementara itu, pada 3 tahun terakhir yaitu pada tahun 2021 hingga 2022 jumlah mahasiswa baru relatif stabil di kisaran angka 220. Prosentasi jumlah mahasiswa baru turun sebesar 23% dan hingga saat ini masih belum mencapai target. Dari beberapa teknik promosi yang telah dilakukan, masih ada beberapa kendala yang dihadapi oleh FTI Unmer Malang dalam hal mendapatkan mahasiswa baru, diantaranya kurang akurat informasi yang didapatkan calon mahasiswa baru karena informasi tidak sampai kepada masyarakat khususnya yang tinggal di daerah pelosok. Hal ini disebabkan karena teknik promosi yang dilakukan tidak memberikan informasi sampai ke seluruh daerah yang ada di kota Malang ataupun luar daerah Malang dan sekitarnya.

Pada masing-masing daerah membutuhkan teknik promosi yang berbeda tergantung dari lokasi dan kondisi yang ada pada daerah tersebut. Disebutkan dalam [3], bahwa strategi promosi dengan menggunakan dimensi *positioning* dapat memberikan penjelasan secara terperinci untuk kemudian mengetahui apa yang diinginkan konsumen. Berdasarkan dari data mahasiswa baru yang ada, tempat asal mahasiswa terdiri dari berbagai daerah di seluruh Indonesia. Proses melakukan pengelompokan mahasiswa baru FTI Unmer Malang masih dilakukan dengan perhitungan manual, sehingga membutuhkan waktu yang tidak sedikit. Agar dapat dibaca sesuai dengan kebutuhan, diperlukan penerapan *data mining* yaitu dengan melakukan klasterisasi terhadap data dari seluruh mahasiswa baru [4]. *Data mining* adalah proses menganalisis kumpulan data yang sangat besar untuk menemukan hubungan yang tidak terduga. Selain itu, *data mining* memiliki kemampuan untuk merangkum data dengan cara yang baru sehingga dapat berguna dan dipahami oleh pengguna [5].

Dalam analisis dan *data mining*, *clustering* adalah jenis penelitian di mana objek data dengan atribut yang sama dikelompokkan dalam satu kelompok, dan objek data dengan atribut yang berbeda dikelompokkan dalam kelompok yang berbeda [6]. Ada berbagai macam *clustering* yang ada, tetapi pada penelitian ini menggunakan K-Means dan K-Medoids sebagai pembanding untuk mencari tahu hasil yang terbaik.

K-Medoids adalah salah satu metode *data mining* yang dapat menyelesaikan masalah pengelompokan atau *clustering*. Algoritma ini menggunakan objek untuk menunjukkan *cluster* pada sekumpulan objek. Sangat terkenal bahwa algoritma K-Medoids bekerja dengan baik pada *dataset* kecil. Selain itu, hasil proses *clustering* K-Medoids tidak bergantung pada urutan masuk *dataset*. K-Algoritma K-Medoids cocok digunakan untuk data yang memiliki *outlier* dan bentuk *cluster* yang lebih kompleks, akan tetapi membutuhkan waktu komputasi yang lebih tinggi [7].

K-Means adalah metode *clustering* yang terkenal dibandingkan dengan algoritma *clustering* lainnya karena efisiensi dan kesederhanaannya [8]. K-Means adalah salah satu metode pengelompokan data non-hierarki (sekatan) yang dapat membagi data ke dalam dua kelompok atau lebih. Kelompok ini menggabungkan data dengan karakteristik yang sama ke dalam satu kelompok, dan kelompok lain menggabungkan data dengan karakteristik yang berbeda. Tujuan dari pengelompokan adalah meminimalkan variasi dalam suatu kelompok dan memaksimalkan variasi antar kelompok [9]. Secara umum, algoritma K-Means lebih cepat secara komputasional, serta lebih cocok digunakan untuk *dataset* yang besar dan tidak terpengaruh oleh *outlier*.

Penelitian ini bertujuan untuk menentukan strategi promosi yang sesuai berdasarkan daerah asal mahasiswa baru melalui metode klasterisasi. Algoritma *clustering* yang digunakan adalah K-Means dan K-Medoids, dimana akan dilakukan perbandingan klastering untuk menilai tingkat keakuratan evaluasi. Pada penelitian ini, proses pencarian klasifikasi dilakukan dengan menentukan klaster nama, jurusan, asal kota, provinsi, dan asal sekolah mahasiswa berdasarkan data dari tahun akademik 2016-2022. Melalui penelitian ini, diharapkan dapat membantu FTI Unmer Malang dalam menentukan strategi promosi yang tepat berdasarkan daerah asal mahasiswa.

2. Metode Penelitian

Pada tahap pengumpulan data diperoleh data mahasiswa baru dari tahun 2016 sampai dengan 2022 sebagai *dataset*. *Data training* yang digunakan yaitu sebanyak 1492 dengan atribut terdiri dari nama, jurusan, provinsi, asal kota, dan asal sekolah. Namun semua nama atribut ini akan diganti menggunakan angka dari masing-masing atribut melalui pengolahan data awal. Adapun proses pengolahan data awal menggunakan *software* Microsoft Excel terlebih dahulu.

Setelah pengumpulan data dan bukti yang mendukung penelitian, selanjutnya dilakukan proses persiapan data awal (*pre-processing*). Tahap ini dilakukan untuk mengidentifikasi dan menghilangkan data yang kosong (*missing value*) [10]. Normalisasi data dilakukan menggunakan operator *Filter Examples* pada *software* RapidMainer. Kemudian setelah itu dilanjutkan dengan proses klusterisasi dalam tahap analisis data.

Analisis data dilakukan dengan mencari proses serta menyusun secara sistematis data yang telah didapat dari hasil observasi dengan cara mengorganisasikan data tersebut ke dalam kategori dan menjabarkan ke dalam unit-unit, menyusun kedalam pola, hingga memilih data mana yang penting. Kemudian dari hasil analisis akan dibuat kesimpulan agar penelitian dapat dipahami.

Pada tahap analisis data, kegiatan yang dilakukan yaitu menganalisa kebutuhan apa saja yang diperlukan dalam melakukan klusterisasi teknik promosi dengan menggunakan algoritma K-mens dan K-medoids. Untuk itu, proses transformasi dilakukan dengan tujuan untuk menyamakan skala atribut data ke dalam *range* yang lebih spesifik. Sehingga data dapat diolah dengan mudah menggunakan dua metode algoritma *clustering*. Pada tahap ini, *dataset* akan ditransformasi menjadi angka sehingga didapatkan nilai yang akan diolah.

Adapun variabel yang digunakan pada data mahasiswa baru yaitu program studi, asal kota, dan asal sekolah. Hasil dari transformasi data ke angka dapat dilihat pada Tabel 1. Metode yang digunakan untuk klusterisasi antara lain algoritma K-Means dan K-Medoids. Perbedaan antara kedua algoritma tersebut adalah algoritma K-Medoids menggunakan objek sebagai perwakilan (Medoid) pusat *cluster* untuk setiap *cluster*. Sementara algoritma K-Means membutuhkan nilai rata-rata (Mean) sebagai pusat *cluster*. K-Means merupakan salah satu dari beberapa metode data *clustering* non-hierarki dengan sistem kerja mempartisi data yang ada ke dalam bentuk satu atau lebih *cluster*/kelompok. Algoritma K-Medoids dapat meminimalkan jumlah perbedaan antar titik data dalam *cluster* dengan titik data terpilih di setiap *cluster* sebagai pusat (medoid).

Tabel 1. Data Hasil Transformasi

ID	Provinsi	Asal Sekolah	Kota	Jurusan	ID	Provinsi	Asal Sekolah	Kota	Jurusan
K1	1	1	1	1	K14	1	3	1	1
K2	1	2	1	1	K15	1	1	1	1
K3	1	2	1	1	K16	1	1	1	1
K4	1	3	1	1	K17	2	1	1	1
K5	1	1	1	1	K18	1	1	1	1
K6	1	1	1	1	K19	2	2	3	1
K7	1	2	1	1	K20	2	2	3	1
K8	1	2	2	1	K21	2	1	4	1
K9	1	4	1	1	K22	2	5	5	1
K10	2	1	3	1	K23	2	2	6	1
K11	1	1	1	1	K24	1	2	1	1
K12	1	1	1	1	K25	2	5	7	1
K13	1	1	1	1					

Sumber: Data jadi Jumlah Mahasiswa, 2022

Selanjutnya, hasil *clustering* dari kedua metode ini akan dilanjutkan ke tahap pengujian, baik secara manual maupun menggunakan *tools* RapidMiner. Hasil pengujian kemudian akan dibandingkan, yaitu berupa penentuan nilai *centroid* dan hasil *clustering*. Evaluasi kualitas hasil *clustering* diukur dengan menggunakan metrik *Davies Bouldin Index* (DBI). DBI adalah metode yang digunakan untuk mengukur kesamaan antar anggota dalam suatu *cluster* atau yang memiliki kemiripan [11]. Semakin kecil nilai DBI, menunjukkan semakin baik kualitas hasil *cluster*.

3. Hasil

Pada hasil transformasi didapatkan atribut provinsi sebanyak 24 data, atribut asal sekolah sebanyak 11 data, sedangkan atribut kota asal sebanyak 73 data. Frekuensi dan kode untuk atribut provinsi disajikan pada Tabel 2. Frekuensi dan kode untuk atribut asal sekolah disajikan pada Tabel 3.

Pada Tabel 2 diperlihatkan bahwa frekuensi terbanyak ditempati oleh provinsi Jawa Timur dengan angka 807, disusul dengan Nusa Tenggara Timur dengan angka 354. Frekuensi terendah yaitu pada provinsi Kalimantan Selatan dan Sulawesi Tengah, masing-masing dengan angka 1 dan 2.

Tabel 2. Inisialisasi Data Provinsi

Provinsi	Frekuensi	Kode
Jawa Timur	807	1
Nusa Tenggara Timur	354	2
Papua Barat	27	3
Papua	62	4
Nusa Tenggara Barat	21	5
Kalimantan Utara	17	6
Maluku	50	7
Banten	2	8
Luar Negeri	4	9
Kalimantan Selatan	1	10
DKI Jakarta	10	11
Sumatera Selatan	20	13
Kalimantan Utara	11	14
Maluku Utara	5	15
Sumatera Selatan	6	16
Bali	15	17
Kepulauan Riau	6	18
Kalimantan Barat	7	19
Sulawesi Tengah	2	20
Sulawesi Tenggara	8	24

Pada Tabel 3 diperlihatkan bahwa frekuensi terbanyak berasal dari SMA sebanyak 475. Meskipun demikian, asal sekolah SMK dan SMKN apabila dijumlahkan memiliki nilai yang lebih besar yaitu 706. Sementara itu, frekuensi terendah yaitu dengan asal sekolah SLB dan ST dengan nilai 1.

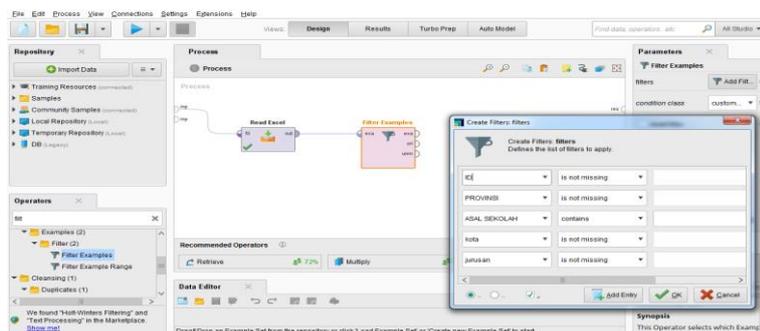
Tabel 3. Inisial Data Sekolah

Asal Sekolah	Frekuensi	Kode
SMK	379	1
SMKN	327	2
MA	22	3
PKBM	6	4
SMA	475	5
SMAN	242	6
MAN	31	7

Madrasah	3	8
SLB	1	9
ST	1	10
SLA	2	11

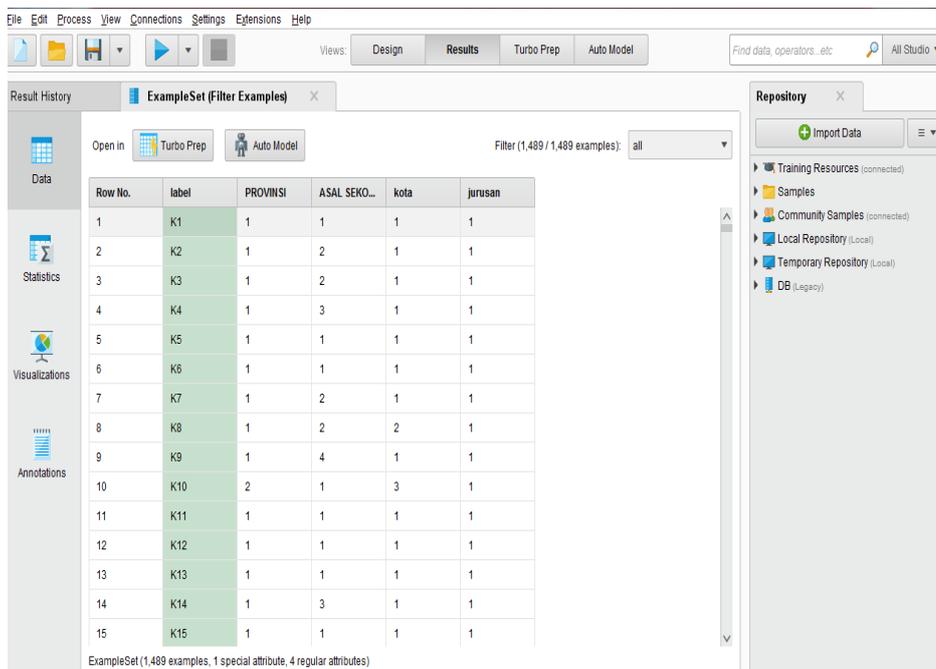
4. Pembahasan

Sebelumnya pada tahap *pre-processing* dilakukan seleksi atribut dan pengkategorian data menggunakan *software* Microsoft Excel, serta penghapusan data yang kosong. Untuk memaksimalkan hasil akurasi, dilakukan pengecekan ulang data supaya mengantisipasi adanya data yang kosong dengan menggunakan operator *Filter Examples* pada *software* RapidMainer seperti yang diperlihatkan pada Gambar 1.



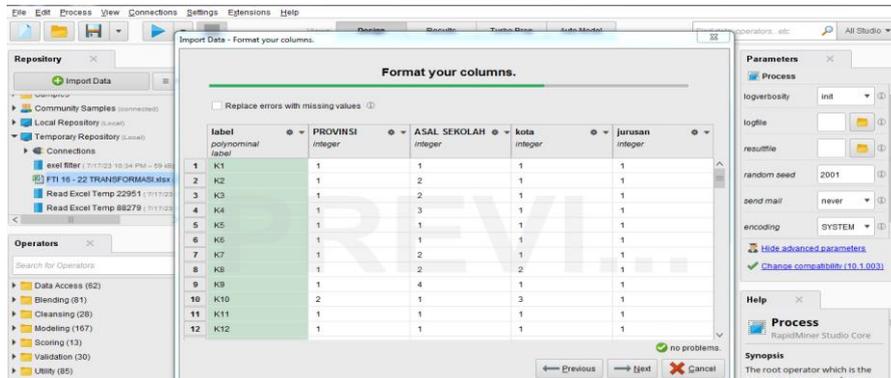
Gambar 1. Pemodelan Menggunakan *Filter Examples*

Setelah membuat pemodelan pada *Filter Examples* pada *software* RapidMiner, langkah selanjutnya adalah dilakukan pengaturan pada parameter-parameter. Dilakukan pembersihan atribut antara lain ID, provinsi, asal sekolah, kota, dan jurusan pada *condition class* dengan *Add Filter*. Hasil normalisasi data diperlihatkan pada Gambar 2. Sebelum dilakukan pembersihan *missing value*, total data sebanyak 1.492. Setelah dihilangkan *missing value*, total data menjadi 1489.



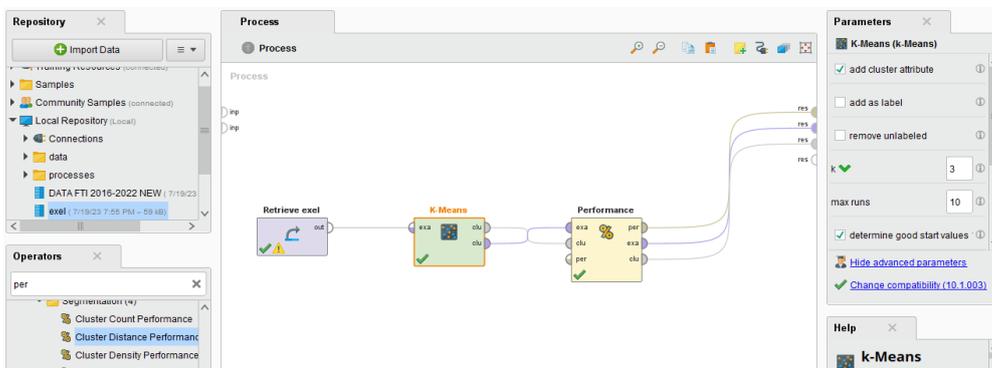
Gambar 2. Hasil Normalisasi Data

Tahap selanjutnya adalah pengujian dengan menggunakan *software* RapidMiner. RapidMiner adalah perangkat lunak untuk pengolahan data dengan menggunakan prinsip dan algoritma *data mining* [12]. Dalam melakukan pengujian, digunakan data yang sudah dinormalisasikan menggunakan *tools* RapidMiner. Kemudian dilakukan pemrosesan data serta pengujian yang dilakukan dua kali yaitu untuk algoritma K-Means dan K-Medoids. Pada Gambar 3 diperlihatkan proses mengimpor data ke RapidMiner.

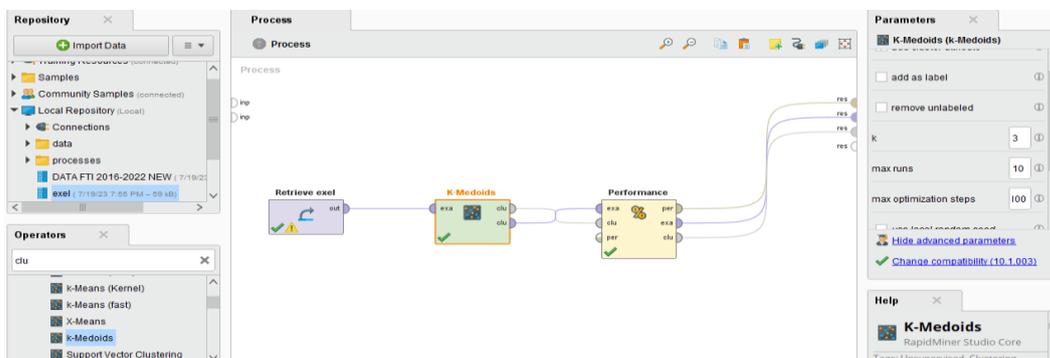


Gambar 3. Proses Impor Dataset ke RapidMiner

Pada Gambar 3, proses impor data ke RapidMiner dilakukan dengan menggunakan operator *Retrieve Excel* dengan fitur *configuration wizard*. Fitur tersebut digunakan untuk memasukkan file data serta memilih tipe data, atribut, dan role. Selanjutnya dilakukan pemodelan desain *clustering* menggunakan algoritma K-Means dan K-Medoids seperti yang ditunjukkan pada Gambar 4 dan Gambar 5.

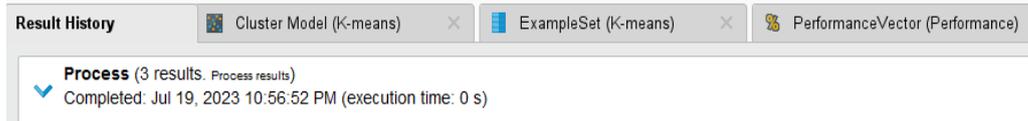


Gambar 4. Desain Input Proses dari Algoritma K-Means

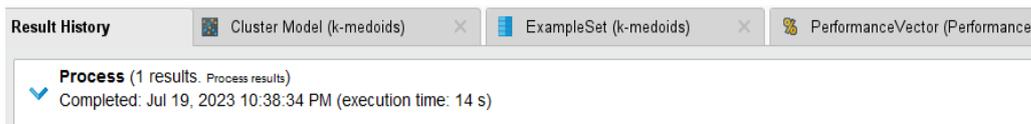


Gambar 5. Desain Input Proses dari Algoritma K-Medoids

Proses pengujian *clustering* dimulai dengan operator *Retrieve Excel* dengan fitur *import configuration wizard*. Fitur ini digunakan untuk memasukan *file dataset* yaitu data mahasiswa FTI Unmer Malang dari tahun 2016-2022 ke dalam RapidMiner. Pada *dataset* ini kemudian dilakukan klasterisasi dan pengujian untuk algoritma K-Means dan K-Medoids. Caranya yaitu dengan dihubungkan ke operator *Performance* untuk menghasilkan *output* dari algoritma *clustering*, dan terakhir dihubungkan ke *Result*. Setelah pengujian berjalan, maka didapatkan informasi lama waktu pemrosesan seperti yang ditunjukkan pada Gambar 6 dan Gambar 7.

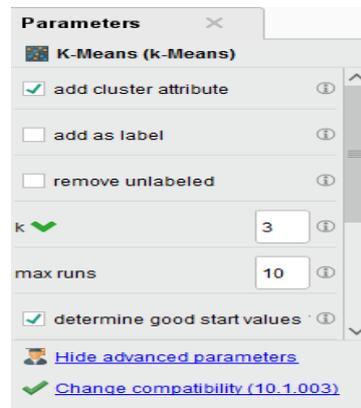


Gambar 6. Result History menggunakan Algoritma K-Means

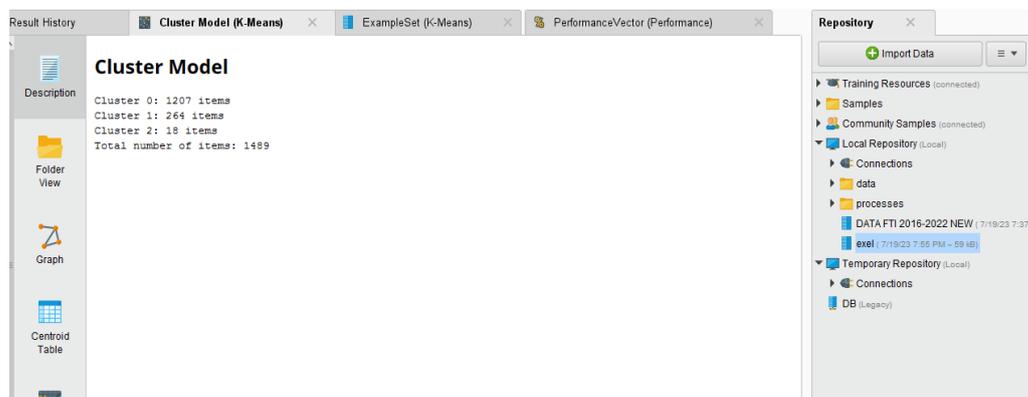


Gambar 7. Result History menggunakan Algoritma K-Medoids

Pada penelitian ini, jumlah cluster yang ditentukan adalah 3, seperti yang diperlihatkan pada Gambar 8. Model *cluster* diperlihatkan pada Gambar 9 untuk K-Means dan Gambar 10 untuk K-Medoids. Pada gambar ditunjukkan bahwa *cluster* yang dihasilkan berjumlah 3, dimulai dari *cluster 0* dan diakhiri dengan *cluster 3*. Total *dataset* di dalam *cluster* berjumlah 1489 dengan jumlah *item* yang berbeda pada tiap-tiap *cluster*.

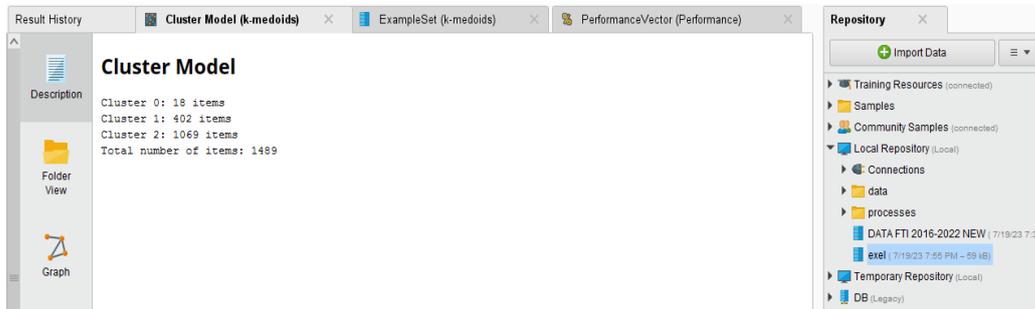


Gambar 8. Penentuan Jumlah Cluster



Gambar 9. Hasil Cluster Algoritma K-Means

Gambar 9 menunjukkan hasil *clustering* menggunakan algoritma K-Means dengan *cluster* 0 berjumlah 1.207 *item*, *cluster* 1 berjumlah 264 *item*, dan *cluster* 2 berjumlah 18 *item*. Gambar 10 menunjukkan hasil *clustering* menggunakan algoritma K-Medoids dengan *cluster* 0 berjumlah 18 *item*, *cluster* 1 berjumlah 402 *item*, dan *cluster* 2 berjumlah 1069 *item*.



Gambar 10. Hasil *Cluster* Algoritma K-Medoids

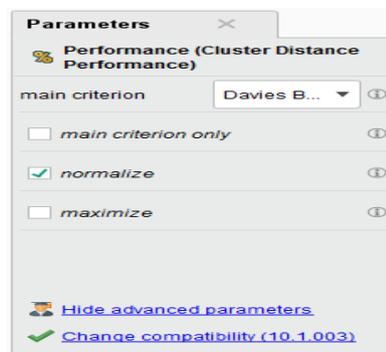
Gambar 11 dan Gambar 12 menunjukkan data *centroid* baru yang terbentuk setelah proses *clustering* data menggunakan RapidMiner untuk Algoritma K-Means dan K-Medoids. *Output* ini memberikan informasi tentang titik *centroid* pada setiap *cluster*, berdasarkan hasil pemrosesan data yang telah dilakukan. Untuk mengevaluasi kualitas dari hasil *clustering*, digunakan metrik DBI dengan parameter seperti yang diperlihatkan pada Gambar 13.

Attribute	cluster_0	cluster_1	cluster_2
PROVINSI	2.129	14.333	9.148
ASAL SEKOLAH	3.383	4.167	4.189
kota	16.178	1007.389	125.398
jurusan	1	1	1

Gambar 11. Tampilan *Centroid Table* untuk Algoritma K-Means

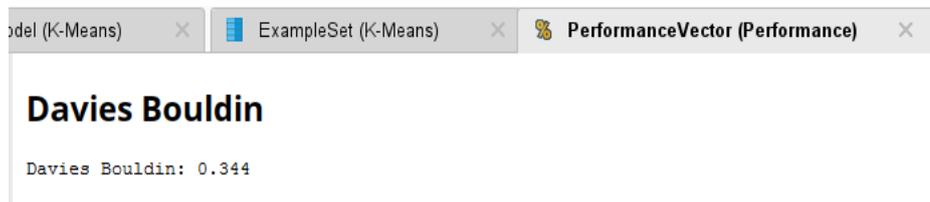
Attribute	cluster_0	cluster_1	cluster_2
PROVINSI	2	1	1
ASAL SEKOLAH	1	6	2
kota	1015	101	1
jurusan	1	1	1

Gambar 12. Tampilan *Centroid Table* untuk Algoritma K-Medoids

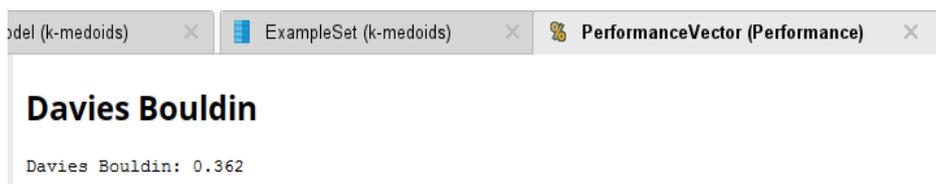


Gambar 13. Parameter Kinerja DBI pada Algoritma *Clustering*

Hasil dari perhitungan DBI dan deskripsi *Performance* menunjukkan *centroid distance* pada setiap *cluster* di dalam *Performance Vector*. Hasil perhitungan DBI untuk K-Means dan K-Medoids masing-masing ditunjukkan pada Gambar 14 dan Gambar 15. Semakin rendah nilai DBI, maka semakin baik kualitas *cluster* tersebut.

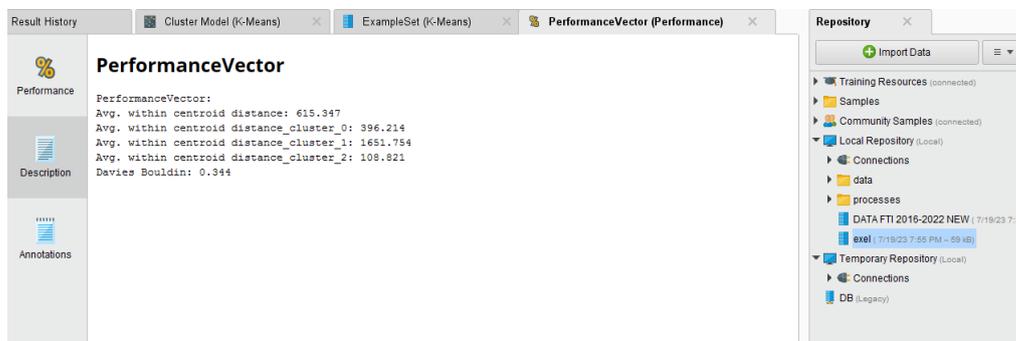


Gambar 14. Hasil Pengukuran DBI untuk Algoritma K-Means

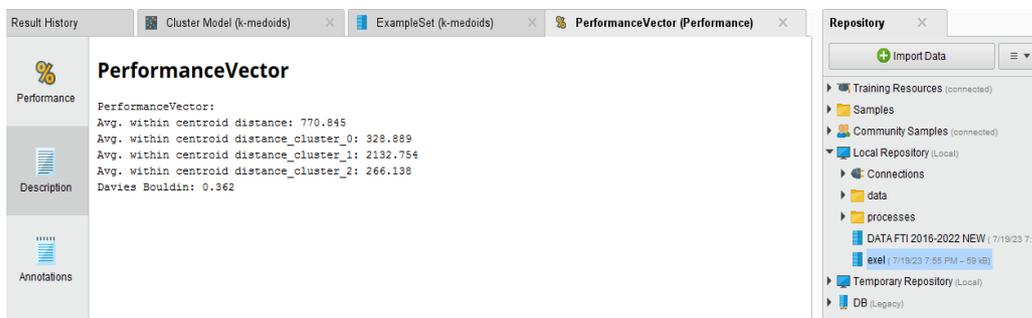


Gambar 15. Hasil Pengukuran DBI untuk Algoritma K-Medoids

Berdasarkan Gambar 14 dan Gambar 15 diperoleh nilai DBI K-Means yaitu 0,344 dan K-Medoids yaitu 0,362. Dari hasil tersebut, ditunjukkan bahwa K-Means memiliki nilai DBI yang lebih kecil, yang berarti bahwa hasil klaster memiliki kemiripan yang tinggi antar objek di dalam kelompoknya. Deskripsi *Performance* yang menghasilkan *centroid distance* pada setiap *cluster* ditunjukkan pada Gambar 16 untuk algoritma K-Means dan Gambar 17 untuk algoritma K-Medoids.



Gambar 16. Hasil DBI dan *Centroid Distance Performance Vector* pada Algoritma K-Means



Gambar 17. Hasil DBI dan *Centroid Distance Performance Vector* pada Algoritma K-Medoids

Berdasarkan hasil klasterisasi menggunakan algoritma K-Means, ditunjukkan *cluster 0* memiliki 1.207 data, yang dapat diartikan bahwa kelompok pertama adalah kategori wilayah tinggi peminat. *Cluster 1* memiliki 264 data, dimana kelompok kedua adalah kategori wilayah sedang peminat.

Cluster 2 memiliki 18 data, sehingga kelompok ketiga merupakan kategori wilayah rendah peminat. Dari hasil klasterisasi dengan algoritma K-Medoids, diperoleh *cluster 0* dengan 18 data, yang dapat diartikan bahwa kelompok pertama adalah wilayah paling rendah peminat. *Cluster 1* memiliki 402 data, dimana kelompok kedua merupakan wilayah sedang peminat. *Cluster 2* memiliki 1069 data, sehingga diartikan bahwa kelompok ketiga adalah wilayah paling tinggi peminat.

Melalui hasil *clustering*, diperoleh informasi mengenai kategori kelompok wilayah berdasarkan jumlah peminat. Hal ini dapat dimanfaatkan untuk menentukan wilayah mana saja yang menjadi prioritas dalam kegiatan promosi. Wilayah dengan kategori tinggi peminat perlu mendapatkan perhatian lebih dengan mengalokasikan sumber daya lebih besar mulai dari waktu, tenaga kerja, hingga dana. Hal ini diperlukan agar investasi yang dikeluarkan dapat seimbang dengan hasil yang didapatkan.

Agar promosi dapat dilakukan dengan efektif dan tepat sasaran, maka kampanye pemasaran perlu ditingkatkan. Salah satu strategi yang perlu dipertimbangkan adalah penggunaan media penyebaran informasi berdasarkan segmentasi kelompok mahasiswa hasil klasterisasi. Untuk daerah perkotaan, maka sumber informasi yang dapat digunakan adalah media sosial dan *website*. Informasi yang disediakan harus lengkap dan terstruktur, termasuk informasi mengenai beasiswa, kerja sama, dan pendaftaran *online*. Sedangkan untuk daerah pelosok, maka informasi juga perlu disebar dengan cara tradisional yaitu melalui media cetak dan iklan.

Hasil penelitian ini diharapkan dapat membantu FTI Unmer Malang dalam pengambilan keputusan untuk menentukan strategi promosi terbaik. Dengan demikian, diharapkan dapat meningkatkan jumlah mahasiswa baru pada periode selanjutnya dan mencapai target penerimaan yang lebih optimal.

5. Penutup

Berdasarkan hasil perbandingan antara dua metode *clustering* yaitu K-Means dan K-Medoids, algoritma K-Means menunjukkan performa yang lebih baik dengan indeks DBI sebesar 0,344 dibandingkan dengan algoritma K-Medoids dengan indeks DBI sebesar 0,362. Sehingga dapat ditarik kesimpulan bahwa algoritma K-Means lebih baik dan direkomendasikan untuk digunakan dalam membantu menentukan strategi promosi mahasiswa baru di perguruan tinggi. Selain itu, dapat disimpulkan bahwa algoritma K-Means lebih baik digunakan untuk klasterisasi dengan distribusi data yang lebih seimbang.

Penentuan strategi promosi dapat dilakukan berdasarkan jumlah peminat di setiap kategori wilayah dari masing-masing *cluster* yang terbentuk. Dengan hasil *clustering* yang lebih baik, memungkinkan perguruan tinggi untuk menargetkan promosi secara lebih efisien kepada kelompok mahasiswa tertentu. Hal ini dapat dilakukan dengan memfokuskan kegiatan promosi pada daerah-daerah tertentu yang lebih prospektif. Daerah dengan distribusi peminat lebih tinggi perlu diprioritaskan dalam kaitannya dengan sumber daya yang perlu dipersiapkan. Strategi komunikasi dan promosi dapat dirancang dengan lebih efektif sesuai dengan preferensi dari kelompok-kelompok mahasiswa.

Penelitian selanjutnya dapat dikembangkan dengan memperdalam analisis strategi pemasaran untuk penerimaan mahasiswa baru di perguruan tinggi. Selain itu, penerapan algoritma *clustering* lainnya dapat digunakan untuk memberikan perspektif baru ataupun memberikan hasil *clustering* yang lebih akurat.

Referensi

- [1] Nina Dewi Lashwaty, Ina Sholihah Widiati, and Hadis Turmudi, "Peran Penggunaan Media Online Dalam Penerimaan Mahasiswa Baru," *Jurnal STIE Semarang*, vol. 12, no. 3, 2020.
- [2] Iqbal Fauzy and Himawan Pramaditya, "Peramalan Penerimaan Mahasiswa Baru Dengan Perbandingan Metode Regresi Linier Sederhana Dan Single Eksponensial Smoothing," *Jurnal Riset Mahasiswa Bidang Teknologi Informasi*, vol. 6, no. 2, 2024.

- [3] Muhammad Rizal and Lanihayati Sandiana, "Aplikasi Pemasaran Perumahan Berbasis Teknologi Augmented Reality," *Inspiration: Jurnal Teknologi Informasi dan Komunikasi*, vol. 6, no. 2, 2016.
- [4] E. F. L. Awalina and W. I. Rahayu, "Optimalisasi Strategi Pemasaran dengan Segmentasi Pelanggan Menggunakan Penerapan K-Means Clustering pada Transaksi Online Retail," *Jurnal Teknologi dan Informasi*, vol. 13, no. 2, pp. 122–137, Aug. 2023, doi: 10.34010/jati.v13i2.10090.
- [5] R. Mythily, A. Banu, and S. Raghunathan, "Clustering Models for Data Stream Mining," *Procedia Comput Sci*, vol. 46, pp. 619–626, 2015, doi: 10.1016/j.procs.2015.02.107.
- [6] A. S. Ritonga and I. Muhandhis, "Clustering Data Tweet E-Commerce Menggunakan Metode K-Means (Studi Kasus Akun Twitter Bibli Indonesia)," *SMATIKA JURNAL*, vol. 12, no. 01, pp. 75–84, Jun. 2022, doi: 10.32664/smatika.v12i01.665.
- [7] G. Dwilestari, M. Mulyawan, and I. Ali, "Analisis Clustering menggunakan K-Medoid pada Data Penduduk Miskin Indonesia," *JURSIMA J. Sist. Inf. dan Manaj*, vol. 9, no. 3, pp. 282–290, 2021.
- [8] T. H. Sardar and Z. Ansari, "An analysis of MapReduce efficiency in document clustering using parallel K-means algorithm," *Future Computing and Informatics Journal*, vol. 3, no. 2, pp. 200–209, Dec. 2018, doi: 10.1016/j.fcij.2018.03.003.
- [9] Eko Prasetyo and Aldo Sahala, *Data mining : mengolah data menjadi informasi menggunakan MATLAB*. Yogyakarta: ANDI OFFSET, 2014.
- [10] Gde Agung Brahmama Suryanegara, Adiwijaya, and Mahendra Dwifabri Purbolaksono, "Peningkatan Hasil Klasifikasi pada Algoritma Random Forest untuk Deteksi Pasien Penderita Diabetes Menggunakan Metode Normalisasi," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 1, pp. 114–122, Feb. 2021, doi: 10.29207/resti.v5i1.2880.
- [11] L. Gayatri and H. Hendry, "Pemetaan Penyebaran Covid-19 Pada Tingkat Kabupaten/Kota Di Pulau Jawa Menggunakan Algoritma K-Means Clustering," *Sebatik*, vol. 25, no. 2, pp. 493–499, Dec. 2021, doi: 10.46984/sebatik.v25i2.1307.
- [12] Ardi Ramdani, Christian Dwi Sofyan, Fauzi Ramdani, Muhamad Fauzi Arya Tama, and Muhammad Angga Rachmatsyah, "Algoritma Klasifikasi Data Mining Untuk Memprediksi Masyarakat Dalam Menerima Bantuan Sosial," *Jurnal Ilmiah Sistem Informasi*, vol. 1, no. 2, pp. 39–47, Jul. 2022, doi: 10.51903/juisi.v1i2.363.