

Strategi Promosi Penerapan Data Mining Mahasiswa Baru Dengan Metode K-Means Clustering

Agneresa^a, April Lia Hananto^{b*}, Shofa Shofiah Hilabi^c, Agustia Hananto^d, Tukino^e
^{a,b,c,d,e} Universitas Buana Perjuangan Karawang, Jl. Ronggo Waluyo, Telukjambe Timur, Kabupaten Karawang 41361,
Indonesia
^baprilia@ubpkarawang.ac.id

Abstract

which are held every new academic year. The amount of data generated through this process can become a new source of information for universities that can be used to determine an efficient and specific promotion strategy. In this study, Clustering and the K-Means algorithm are used in the Data Mining. The data in this study used data new student admissions for the 2020/2021 academic year, which totaled 2,479 applicants with attributes such as gender, school origin, origin of major, and chosen study program. This research uses RapidMiner software for implementation. Davies Bouldin Index is used to find the optimal number of Clusters and then produces 2 Clusters. Cluster 1 as a high interest Cluster totals 1945 data with a percentage of 78.82%. Cluster 1 characteristics are dominated by students from 52% public high schools and 21% private vocational schools. The most chosen study programs are Management with 22%. Cluster 2 as a low interest Cluster totaling 525 data with a percentage of 21.18% has characteristics dominated by students from 48% Public Vocational Schools and 45% Private Vocational Schools, the study programs that dominate the Cluster are Management Study Program 25% and Industrial Engineering 24%. This research produces an interesting pattern as a source of new information to determine a more effective and targeted promotion strategy in promoting each study program as to increase the number of prospective new student applicants in the following year.

Keywords: Clustering; Data Mining; DBI; K-Means.

Abstrak

Kegiatan penerimaan mahasiswa sudah menjadi kebiasaan tahunan bagi perguruan tinggi seperti Universitas Buana Perjuangan yang diadakan setiap tahun ajaran baru. Banyaknya data yang dihasilkan melalui proses ini dapat menjadi sumber informasi baru bagi perguruan tinggi yang dapat dimanfaatkan untuk menentukan strategi promosi yang efisien dan spesifik. Pada penelitian ini, teknik *Clustering* dan algoritma *K-Means* digunakan dalam proses *Data Mining*. Data dalam penelitian ini menggunakan data penerimaan mahasiswa baru tahun ajaran 2020/2021 yang berjumlah 2479 pendaftar dengan atribut seperti jenis kelamin, asal sekolah, asal jurusan, dan program studi yang dipilih. Penelitian ini menggunakan perangkat lunak RapidMiner untuk implementasi. *Davies Bouldin Index* digunakan untuk mencari jumlah *Cluster* yang optimal kemudian menghasilkan 2 *Cluster*. *Cluster 1* sebagai cluster minat tinggi berjumlah 1945 data dengan persentase sebesar 78,82%. karakteristik cluster 1 didominasi oleh mahasiswa yang berasal dari SMA Negeri 52% dan SMK Swasta 21% Prodi yang paling banyak dipilih yaitu Manajemen sebanyak 22%. *Cluster 2* sebagai cluster minat rendah berjumlah 525 data dengan persentase sebesar 21,18% mempunyai karakteristik yang didominasi oleh mahasiswa berasal dari SMK Negeri 48% dan SMK Swasta 45% program studi yang paling mendominasi pada cluster yaitu Prodi Manajemen 25% dan Teknik Industri 24%. Penelitian ini menghasilkan pola yang menarik sebagai sumber informasi baru untuk menentukan strategi promosi yang lebih efektif dan tepat sasaran dalam mempromosikan setiap program studi sehingga dapat meningkatkan jumlah calon pendaftar mahasiswa baru di tahun berikutnya.

Kata Kunci: Clustering; Data Mining; DBI; K-Means.

1. Pendahuluan

Pertumbuhan teknologi diiringi dengan merambatnya perkembangan prasarana teknologi informasi telah mengubah cara organisasi, industri, dan bahkan pemerintah bekerja, berpikir, dan bertindak[1]. Teknologi dan informasi yang semakin pesat membuat informasi data menjadi tidak terhitung jumlahnya dalam kehidupan manusia, terlihat jelas di bidang pengolahan data skala besar dalam penyimpanan data. Ini menjadi daya tarik yang besar bagi perusahaan dan organisasi, baik negeri maupun swasta[2]. Jumlah data yang dikumpulkan oleh institusi pendidikan seperti universitas semakin meningkat setiap tahunnya. Seperti pada data pendaftaran calon mahasiswa baru[3] pengolahan data yang cepat, tepat, dan tepat menghasilkan informasi yang diperlukan merupakan keunggulan utama dari kemajuan teknologi informasi. Hal ini memungkinkan data yang sebelumnya dianggap tidak berharga menjadi sumber informasi setelah data tersebut diolah[4]. KDD memasukkan data mining sebagai salah satu komponennya (*Knowledge Discovery in Databases*), melibatkan sejumlah langkah seperti, Pemilihan data, *preprocessing*, transformasi, *mining*, dan penilaian hasil. *Clustering* adalah teknik yang digunakan dalam penambangan data. Metode analisis data yang disebut pengelompokan dapat dianggap sebagai salah satu yang mengelompokkan data ke dalam area dengan sifat yang serupa[5].

Universitas Buana Perjuangan Karawang merupakan salah satu Universitas swasta di Kota Karawang yang berdiri sejak tahun 2014. Universitas Buana Perjuangan Karawang mempunyai 7 fakultas dan 13 program studi. Sistem Informasi, Teknik Industri, Teknik Informatika, Hukum, Manajemen, Pendidikan Pancasila dan Kewarganegaraan, Pendidikan Agama Islam, Pendidikan Guru SD, dan Teknik Mesin adalah beberapa pilihan studi yang ditawarkan di UBP Karawang. Dalam proses penerimaan mahasiswa baru di Universitas Buana Perjuangan Karawang membutuhkan strategi promosi yang ditujukan untuk menjaring lebih banyak calon mahasiswa baru setiap tahunnya. Sejumlah besar data dihasilkan dalam bentuk profil siswa yang terdaftar selama proses penerimaan mahasiswa baru. Jika data diperoleh dengan prosedur yang benar, maka dapat mengungkapkan informasi yang sebelumnya tersembunyi dalam data profil mahasiswa dan membantu perguruan tinggi dalam mengembangkan rencana pemasaran baru berdasarkan pengelompokan sekolah mahasiswa[6].

Beberapa penelitian tentang topik ini telah dilakukan seperti: Penelitian terdahulu Fitri Yunita, Penerapan *Data Mining* Menggunakan Algoritma *K-Means Clustering* dengan implementasinya menggunakan *software* RapidMiner 5.3, menghasilkan hingga 3 *Cluster*[7]. Oki Oktaviarna Tesano dan kawan-kawan menggunakan *software* RapidMiner 9.10 untuk mengimplementasikan teknik CRIPS DM dalam penelitiannya, "Analisis Data Mining Dengan Algoritma *K-Means Clustering* Untuk Menentukan Strategi Promosi Mahasiswa Baru di STMIK Primakara" yang menghasilkan tiga *Cluster*[8] Penelitian Rivot Pormes menggunakan alat Matlab dan algoritma pengelompokan *K-Means* untuk mengidentifikasi lokasi pasokan siswa dan menghasilkan tiga kelompok: kelompok minat tinggi, sedang, dan rendah [9]. Rivot Pormes menggunakan alat Matlab dalam penelitiannya untuk membuat tiga klaster, klaster minat tinggi, sedang, dan rendah, untuk menganalisis area pasokan siswa[10]. Imam Mahmudi dkk menggunakan algoritma *Clustering* sebagai strategi promosi penerimaan mahasiswa baru di Universitas Hasyim Asy'ari, dan Jombang mengembangkan 5 *Cluster* sebagai hasil penelitiannya.

Dengan menggunakan pendekatan *Clustering*, data mahasiswa baru tahun ajaran 2020–2021 dikumpulkan untuk penelitian ini, mengubah data tidak terstruktur menjadi data terstruktur dan menciptakan wawasan baru yang dapat diterapkan di Universitas Buana Perjuangan. Teknik *K-Means Clustering* akan digunakan untuk data penerimaan mahasiswa baru tahun akademik 2020–2021 dalam penelitian ini.

2. Tinjauan Pustaka

2.1 *Data Mining*

Istilah "penambangan data" mengacu pada proses menemukan pengetahuan dalam basis data., tetapi ada juga pendapat bahwa data mining adalah teknik untuk belajar dan menganalisis pengetahuan[11] database yang melakukan tugas-tugas seperti kategorisasi, asosiasi, pengelompokan, estimasi, dan prediksi. Penambangan data digunakan untuk menemukan informasi dan pengetahuan. Salah satu cara untuk menjelaskan bagaimana penambangan data digunakan untuk mengekstraksi informasi dari kumpulan data yang sangat besar[12]. Penambangan data sering dianggap sebagai komponen penemuan pengetahuan dalam basis data (KDD), yang merupakan metode untuk memperoleh pengetahuan berharga dari data. Garis besar proses KDD adalah sebagai berikut.

a. *Selection and Addition*

Pada tahapan ini pemilihan atribut diterapkan pada data yang diperoleh, dengan hasil yang digabungkan menjadi dataset. Proses pembuatan dataset sangat penting karena berfungsi sebagai dasar untuk mempelajari data mining dan menemukan pola baru.

b. *Preprocessing and Data Cleaning*

Pembersihan dilakukan untuk meningkatkan ketergantungan data. Dengan menangani bidang kosong, menangani baris data yang tidak berguna, dan menghilangkan derau, data dapat dibersihkan. Tahap persiapan awal ini mungkin memerlukan teknik statistik canggih atau alat penambangan data tertentu.

c. *Transformation*

Untuk menyiapkan data dengan lebih baik dan menyiapkannya untuk model penambangan data, transformasi data digunakan.

d. *Data Mining*

Tiga tahapan data mining meliputi memilih model, memilih metode, dan benar-benar menggunakan data. Pola atau aturan baru dapat ditemukan dengan menerapkan data mining.

e. *Evaluation and Interpretation*

menggunakan penilaian dan interpretasi untuk menciptakan pengetahuan dalam bentuk pola. Selain itu, hal itu dilakukan untuk mendokumentasikan pengetahuan yang ditemukan untuk digunakan nanti.

f. *Discovered Knowledge*

Sistem sekarang dapat diperbarui dengan informasi yang baru ditemukan. Langkah ini menentukan seberapa aktif proses secara keseluruhan [13].

2.3 *Clustering*

Saat item dikelompokkan bersama dengan cara yang membuatnya terhubung satu sama lain dibandingkan dengan pengelompokan penempatan lainnya, proses ini dikenal sebagai *Clustering*. Agar item-item dalam sebuah *Cluster* menjadi sangat mirip satu sama lain, objek-objek tersebut diatur ke dalam *Cluster*. Keanekaragaman data dalam suatu kelompok dapat diminimalisir dengan mengelompokkan item atau data[5]. Metode k-means yang mencoba mengelompokkan item berdasarkan kualitas ke dalam k divisi merupakan salah satu teknik *Clustering* yang efektif dan cepat yang dapat digunakan[14].

2.4 *K-Means*

Metode pengelompokan data non-hierarkis yang disebut *K-Means* membagi data menjadi satu atau lebih *Cluster* atau pengelompokan. Bersama dengan *Cluster* lain, data dengan berbagai properti diintegrasikan[15]. Ide dasar metode ini adalah mengatur partisi k buah, pusatnya (*centroid*), dan rata-ratanya (*mean*). Pendekatan *K-Means* dimulai dengan pembentukan partisi *Cluster* dan secara iteratif memperbaiki partisi ini hingga tidak ada perubahan yang terdeteksi[16]. Secara umum, algoritma K-Means dapat digunakan untuk mengatasi masalah dengan meringkas sejumlah besar item, memahami perilaku, dan mencari prospek baru. Akan lebih mudah untuk membahas sifat atau sifat masing-masing kelompok[17]. Langkah-langkah algoritma K-Means adalah sebagai berikut:[18]

1. Pilih k sebagai jumlah *Cluster* yang diinginkan untuk dibangun. Tentukan pusat cluster.
2. Centroid awal ditentukan secara acak sesuai jumlah cluster yang ditentukan.
3. Dengan menggunakan persamaan Euclidean, tentukan jarak setiap titik data dari pusat *Cluster*.

$$d_{ij} = \sqrt{(x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2 + \dots + (x_{ki} - x_{kj})^2} \quad (1)$$

D_{ij} adalah jarak antara I pusat data cluster j, X_{ki} adalah data I pada atribut data k, X_{kj} adalah titik pusat j ke atribut k

4. Hitung nilai rata-rata pada setiap cluster untuk memperbaharui *centroid*
5. Ulangi tahap 3 jika terdapat data yang masih berpindah *cluster*.

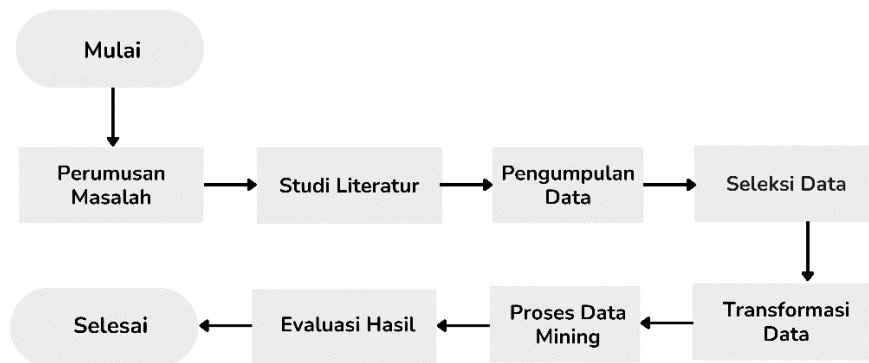
2.5 *Index Davies Bouldin*

David L. Davies dan Donald W. Bouldin menemukan DBI pada tahun 1979. Dalam pendekatan pengelompokan dimana cohesivities Salah satu teknik yang digunakan untuk mengevaluasi reliabilitas atau kuantitas kluster adalah penjumlahan jarak antara data dengan pusat kluster dari kluster yang diikuti. yang paling efektif adalah Davies-Bouldin Index. Kuantitas dan kedekatan data hasil *Cluster* dapat digunakan untuk menentukan apakah hasil *Cluster* baik atau buruk saat menggunakan Indeks Davies Bouldin sebagai ukuran evaluasi[19]. Pengukuran Indeks Davies-Bouldin meningkatkan jarak antara kluster C_i dan C_j

sekaligus berupaya mengurangi jarak antar titik di dalam kluster. Kesamaan sifat antar *Cluster* minimal ketika jarak antara mereka paling tinggi, membuat perbedaan di antara mereka terlihat lebih jelas. Jika setiap objek dalam *Cluster* memiliki tingkat karakter yang tinggi.[20]

3. Metode

Ada serangkaian kerangka kerja yang harus diikuti ketika menggunakan metodologi penelitian ini. Gambar berikut mengilustrasikan bagaimana penelitian dilakukan



Gambar 1. Alur Penelitian

- Rumusan Masalah, Masalah dalam penelitian ini adalah bagaimana menganalisis data mahasiswa baru yang jumlahnya terus bertambah. Ini akan memungkinkan universitas untuk menggunakan dan memperoleh pengetahuan baru dalam mengatur strategi promosi yang tepat untuk Universitas Buana Perjuangan Karawang.
- Studi Literatur, Pada tahap ini peneliti melakukan kajian literatur terhadap materi yang berkaitan dengan tujuan dan judul penelitian. Artikel jurnal, penelitian sebelumnya tentang *K-Means Clustering Algorithm*, *Data Mining*, dan bahan literatur lain yang mendukung tujuan penelitian digunakan sebagai sumber literatur.
- Pengumpulan Data, Untuk mendapatkan data mengenai data Penerimaan Mahasiswa Baru tahun 2020/2021 maka penulis melakukan survey ke bagian Pusat Data dan Informasi (PUSDATIN) UBP Karawang dengan mengajukan surat penelitian.
- Seleksi Data, tidak semua data dapat digunakan dalam penelitian ini, hanya data yang sesuai untuk dilakukan analisis sebagai strategi promosi saja yang akan di pilih, data yang diambil. Data yang digunakan yaitu data jenis kelamin, asal sekolah, asal jurusan dan Prodi yang dipilih.
- Transformasi Data, seluruh data yang di dapat seperti data jenis kelamin, asal sekolah, asal jurusan dan Prodi yang dipilih merupakan data dengan tipe nominal, sehingga harus dilakukan transformasi dari data nominal ke dalam data numerik agar dapat dilakukan proses data mining.
- Proses *Data Mining*, algoritma k-means clustering digunakan pada penelitian ini dengan menggunakan perangkat lunak Rapidminer.
- Evaluasi Hasil, Pada titik ini, ditentukan apakah studi yang ada telah mencapai tujuan yang diinginkan dengan menghasilkan pola tipikal dan model prediktif.

4. Hasil dan Pembahasan

4.1 Data Awal

Saat item dikelompokkan bersama dengan cara yang membuatnya terhubung satu sama lain dibandingkan dengan pengelompokan penempatan lainnya, proses ini dikenal sebagai *Clustering*. Agar item-item dalam sebuah *Cluster* menjadi sangat mirip satu sama lain, objek-objek tersebut diatur ke dalam *Cluster*. Keanekaragaman data dalam suatu kelompok dapat diminimalisir dengan mengelompokkan item atau data[5]. Metode k-means yang mencoba mengelompokkan item berdasarkan kualitas ke dalam k divisi merupakan salah satu teknik *Clustering* yang efektif dan cepat yang dapat digunakan[14].

Tabel 1. Data Awal

No	No.Pend	Jk	Asal Sekolah	Asal Jurusan	Jurusan Dipilih	Alasan Masuk Ubp
1	20200001	P	Sma 4 karawang	Ips	Pendidikan guru sekolah dasar	Says ingin belajar dengan benar dan ingin mencapai cita-cita

2	20200003	L	Sman 1 jatisari	Ipa	Ilmu hukum	Ingin menjadi pengacara
3	20200004	P	Smk kesehatan bhakti kencana subang	Farmasi	Farmasi	Melanjutkan studi
4	20200006	L	Sman 3 karawang	Ips	Sistem informasi	Ingin melanjutkan pendidikan
5	20200007	P	Smk negeri 1 karawang	Rekayasa perangkat lunak	Akuntansi	Karena saya yakin kampus ini adalah yang terbaik
6	20200008	L	Smk bina karya 1 karawang	Teknik mesin	Psikologi	Deket , sesuai jurusan
7	20200009	L	Sman 1 cigombong	Ipa	Manajemen	Untuk memperluas ilmu, memperdalam bidang yang saya inginkan
8	20200010	L	Smk 1 adiwarna	Listrik	Teknik industri	Ingin menambah wawasan dan lingkaran pertemanan serta mewujudkan cita-cita
9	20200011	L	Smkn 3 kuningan	Teknik kendaraan ringan	Teknik mesin	Fasilitasnya lengkap
....
2484	20209011	P	Smks rosma	Otomatisasi dan tata kelo	Pend. Pancasila & kewarganegaraan	Untuk menambah wawasan dan jaraknya dekat dari rumah

4.2 Seleksi Data

Pengolahan data sangat penting karena informasi yang terkumpul masih berupa data mentah yang perlu diolah. Pembersihan melibatkan penghapusan data kosong, data yang tidak konsisten atau tidak perlu, dan data kosong atau tidak jelas. Ini juga termasuk mengisi informasi yang hilang dan memperbaiki kesalahan pemrosesan. Tabel 2 mencantumkan karakteristik yang dipertimbangkan dalam penyelidikan ini dan mencakup hal-hal berikut:

4.2 Transformasi Data

Karena data penerimaan mahasiswa baru dikumpulkan secara acak, maka setiap informasi yang terekam harus diinisialisasi atau diubah ke dalam bentuk angka. Ini diperlukan untuk menangani data menggunakan metode pengelompokan k-means. Selama proses transformasi, angka 1, 2, 3, dan seterusnya digunakan sebagai inisial untuk data dengan frekuensi terendah hingga tertinggi.

1) Transformasi Data Jenis Kelamin

Tabel 3. Transformasi Data Jenis Kelamin

Jenis kelamin	Keterangan	Frekuensi	Inisialisasi
L	Laki-laki	1256	1
P	Perempuan	1223	2

2) Transformasi Data Asal Sekolah

Table 4. Transformasi Data Asal Sekolah

Asal sekolah	Frekuensi	Inisialisasi
Sma negeri	1032	1
Smk swasta	647	2
Smk negeri	501	3
Lainnya	156	4
Sma swasta	142	5

3) Transformasi Data Jurusan Asal

Table 5. Transformasi Data Asal Jurusan

Jurusan asal	Keterangan	Frekuensi	Inisialisasi
Ipa	Ilmu pengetahuan alam	864	1
Ips	Ilmu pengetahuan sosial	457	2
Tkj	Teknik komputer jaringan	190	3
Tm	Teknik mesin	145	4
Tkr	Teknik kendaraan ringan	115	5
Ak	Akuntansi	104	6
Fm	Farmasi	88	7

AP	Administrasi Perkantoran	80	8
Tbsm	Teknik bisnis sepeda motor	73	9
Ti	Teknik industri	73	10
Ot	Otomotif	60	11
Rpl	Rekayasa perangkat lunak	54	12
Te	Teknik elektro	36	13
Tav	Teknik audio video	29	14
Tgb	Teknik gambar bangunan	21	15
Ka	Kimia analisis	18	16
Argb p	Agribisnis pertanian	18	17
Tataboga	Tataboga	17	18
Bahasa	Bahasa	15	19
Perhotelan	Perhotelan	12	20
Agama	Agama	10	21

4) Transformasi Data Jurusan yang dipilih

Tabel 6. Transformasi data prodi ubp

Prodi di UBP	Frekuensi	Inisialisasi
Manajemen	570	1
Teknik industri	351	2
Psikologi	311	3
Teknik informatika	262	4
Ilmu hukum	226	5
Akuntansi	164	6
Pendidikan guru sekolah dasar	164	7
Farmasi	140	8
Teknik mesin	119	9
Pend. Pancasila & kewarganegaraan	91	10
Sistem informasi	81	11

Hasil keseluruhan data setelah dilakukan *pre-processing* data dapat dilihat pada table 7.

Table 7. Hasil Transformasi Data

No	Prodi UBP	Sekolah asal	Jurusan asal	Jenis kelamin
1	5	1	1	1
2	8	3	7	2
3	11	1	1	1
4	6	3	12	2
5	3	2	4	1
6	1	1	1	1
7	2	2	13	1
8	9	3	5	1
....
2479	10	2	8	2

4.4 Pengolahan Data Mining

K-Means *Clustering* dan RapidMiner Studio versi 9.10.00 digunakan untuk mengolah data penerimaan mahasiswa baru. Pada penelitian ini untuk menentukan *Cluster* yang paling optimal menggunakan Nilai *Davies Bouldin* terkecil dengan menguji sebanyak 9 *Cluster* untuk mendapatkan Optimasi jumlah *Cluster* terbaik, berikut jumlah *Cluster* dan nilai DBI pada Tabel 8.

Tabel 8. Hasil Nilai DBI

Cluster	Hasil DBI
C=2	0.788
C=3	0.888
C=4	0.935
C=5	0.820
C=6	0.892
C=7	0.914

C=8	0.928
C=9	0.915
C=10	0.903

Berdasarkan hasil dari tabel 8. Nilai optimalisasi *Cluster* yang optimal adalah jumlah C2, dengan nilai DBI terkecil yang mendekati 0 sebesar 0,788. Oleh karena itu, penelitian ini menggunakan dua *Cluster*. Pada Gambar 2, setiap anggota *Cluster* terlihat.

Cluster Model

```
Cluster 0: 1954 items
Cluster 1: 525 items
Total number of items: 2479
```

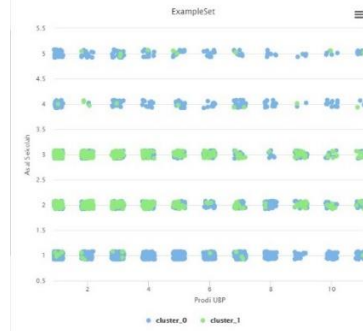
Gambar 2. Model *Cluster*

Pada Gambar 3. Merupakan titik pusat atau centroid dari masing-masing *Cluster*

Attribute	cluster_0	cluster_1
Prodi UBP	4.381	3.709
Asal Sekolah	1.963	2.533
Asal Jurusan	2.395	11.851
Jenis Kelamin	1.538	1.389

Gambar 3. Tabel *Centroid*

Ada dua kelompok data yang ditunjukkan pada Gambar 3 di atas. Dilihat dari titik distribusi/ titik centroidnya. *Cluster* pertama terlihat pada warna biru, sedangkan *Cluster* kedua terlihat dari titik sebaran warna hijau. Gambar 4 di bawah mengilustrasikan hal ini.



Gambar 4. Pemodelan *K-Means Clustering* RapidMiner Studio

Kemudian melakukan analisa dari hasil *Cluster* yang telah dilakukan. Tabel 9 dan 10 menunjukkan hasil analisis *Cluster*.

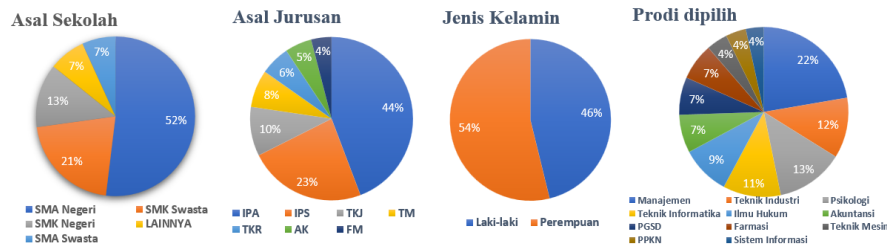
Tabel 9. Hasil *Cluster* Pertama

Hasil <i>Cluster</i> satu			
<i>Cluster</i> ke-1 berjumlah 1954 Mahasiswa			
Prodi dipilih	Asal Sekolah	Asal Jurusan	Jenis Kelamin
Manajemen = 436	SMA Negeri = 1015	IPA = 864	Laki-laki = 902
Teknik Industri = 225	SMK SWASTA = 408	IPS = 457	Perempuan = 1052
Psikologi = 253	SMK Negeri = 252	TKJ = 190	
Teknik Informatika = 216	LAINNYA = 146	TM = 145	
Ilmu Hukum = 183	SMA SWASTA = 133	TKR = 115	
Akuntansi = 142		AK = 104	
PGSD = 140		FM = 79	
Farmasi = 137			
Teknik Mesin = 79			
PPKN = 76			
Sistem Informasi = 67			

Tabel 10. Hasil *Cluster* Kedua

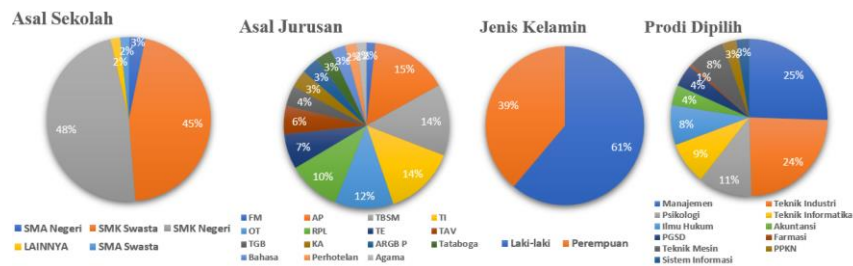
Hasil <i>Cluster</i> Dua			
Cluster ke-2 berjumlah 525 Mahasiswa			
Prodi dipilih	Asal Sekolah	Asal Jurusan	Jenis Kelamin
Manajemen = 134	SMA Negeri = 17	AP = 80	Laki-laki = 321
Teknik industri = 126	Smk swasta = 239	Tbsm = 73	Perempuan = 204
Psikologi = 58	SMK Negeri = 250	TI = 73	
Teknik informatika = 46	Lainnya = 10	Ot = 60	
Ilmu hukum = 43	Sma swasta = 9	Rpl = 54	
Akuntansi = 22		TE = 36	
Pgsd = 24		Tav = 29	
Farmasi = 3		TGB = 21	
Teknik mesin = 40		Ka = 18	
Ppkn = 15		Argb p = 18	
Sistem informasi = 14		Tataboga = 17	
		Bahasa = 15	
		Perhotelan = 12	
		Agama = 10	
		FM = 9	

4.5 Evaluasi Hasil



Gambar 5. Pie Diagram Hasil *Cluster* 1

Pada Gambar 5. merupakan visualisasi dari hasil perhitungan dari *Cluster* pertama, *Cluster* dengan tingkat peminat banyak dengan jumlah data 1945 data didominasi oleh mahasiswa yang berasal dari SMA Negeri sebesar 52% SMK Swasta sebanyak 21%. Sedangkan asal jurusan didominasi oleh jurusan IPA sebesar 44%, IPS 23%. Dengan Prodi yang dipilih didominasi oleh Prodi Manajemen dengan persentase sebesar 22%. *Cluster* pertama ini didominasi oleh mahasiswa perempuan sebesar 54%.



Gambar 6. Pie Diagram Hasil *Cluster* 2

Pada Gambar 6. Merupakan visualisasi dari hasil perhitungan *Cluster* kedua, *Cluster* dengan tingkat peminat rendah jumlah data 525 data yang didominasi oleh mahasiswa yang berasal dari SMK Negeri dengan persentase sebesar 48% dan SMK Swasta dengan persentase sebesar 45%. Sedangkan berdasarkan asal jurusan didominasi oleh jurusan Administrasi Perkantoran (AP) dengan persentase sebesar 15% dan Teknik Bisnis Sepeda Motor (TBSM) sebesar 14% dan Teknik Industri (TI) Sebesar 14%. Prodi yang dipilih didominasi oleh Prodi manajemen dengan persentase sebesar 25% dan Teknik Industri dengan persentase sebesar 24%. Pada *Cluster* ini didominasi oleh mahasiswa berjenis kelamin laki-laki dengan persentase sebesar 61%.

Mengacu pada hasil pengolahan data yang dilakukan dengan menggabungkan pendekatan *Clustering* dengan algoritma k-means seperti yang telah dilakukan di atas maka terbentuk pola-pola yang dapat di pelajari mengenai strategi promosi yang harus dilakukan ke depannya, yaitu dengan cara melakukan promosi untuk masing-masing program studi pada setiap sekolah yang paling banyak diminati dari setiap *Cluster* sehingga promosi dapat berjalan dengan tepat sasaran.

5. Kesimpulan

Berlandaskan dari hasil penelitian yang telah dilakukan pengolahan data mahasiswa baru Universitas Buana Perjuangan Karawang tahun 2020/2021 menggunakan teknik *K-Means Clustering* dengan total record data 2479. Pengolahan data kemudian menggunakan perangkat lunak RapidMiner. Maka dapat ditarik kesimpulan:

1. Berdasarkan hasil dari penentuan jumlah *Cluster* dilakukan dengan Davies Bouldin Index (DBI) maka *Cluster* yang optimal berjumlah 2 *Cluster* dengan nilai DBI yang hampir mendekati 0 yaitu sebesar 0,788.
2. *Cluster* 1 berjumlah 1954 data persentase sebesar 78.82% yang merupakan *Cluster* tingkat tinggi dan *Cluster* 2 berjumlah 525 data persentase sebesar 21.18% yang merupakan *Cluster* tingkat rendah.
3. Strategi promosi yang bisa dilakukan nantinya untuk penerimaan mahasiswa baru yaitu dengan cara mengikuti *Cluster* yang telah terbentuk berdasarkan asal sekolah dan program studi yang paling banyak diminati di setiap asal sekolah mahasiswa. Promosi juga dapat dilakukan guna mengoptimalkan program studi yang memiliki sedikit peminat. Langkah ini memungkinkan memiliki peluang besar dalam upaya meningkatkan jumlah pendaftar baru bagi Universitas Buana Perjuangan Karawang.

References

- S. S. Hilabi, U. Buana, P. Karawang, and B. Huda, "TechnoXplore Jurnal Ilmu Komputer & Teknologi Informasi," 2019.
- A. Lia Hananto et al., "Analysis of Drug Data Mining with Clustering Technique Using K-Means Algorithm," *J Phys Conf Ser*, vol. 1908, no. 1, 2021, doi: 10.1088/1742-6596/1908/1/012024.
- M. Tonggiroh and M. Taher Jufri, "DATA MINING STRATEGI PROMOSI PADA UNIVERSITAS YAPIS PAPUA MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING," 2018.
- D. D. Pertiwi and R. Taufiq, "ANALISIS DAN DESAIN SISTEM INFORMASI PENGOLAHAN NILAI SISWA DI SMK AVICENA RAJEG," pp. 29–35, 2020.
- C. Nas, "Data Mining Pengelompokan Bidang Keahlian Mahasiswa Menggunakan Algoritma K-Means (Studi Kasus : Universitas Cic Cirebon)," 2020.
- N. Ayu Rahmalinda and A. Jananto, "Penerapan Metode K-Means Clustering Dalam Menentukan Strategi Promosi Berdasarkan Data Penerimaan Mahasiswa Baru."
- F. Yunita, "PENERAPAN DATA MINING MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING PADA PENERIMAAN MAHASISWA BARU (STUDI KASUS : UNIVERSITAS ISLAM INDRAGIRI)," 2018.
- O. Oktaviarna Tensao et al., "INFORMASI (Jurnal Informatika dan Sistem Informasi) Analisa Data Mining dengan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Mahasiswa Baru Pada STMIK Primakara."
- R. Pormes and D. H. F. Manongga, "Penggunaan Algoritma Clustering K-means Untuk Melihat Daerah-Daerah Penyuplai Mahasiswa Di Biro Promosi UKSW," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 5, pp. 2443–2229, 2019, doi: 10.28932/jutisi.v5i2.1968.
- M. Sobri Sungkar, "ANALISIS MINAT MAHASISWA MEMASUKI PROGRAM STUDI TEKNIK DENGAN MENGGUNAKAN METODE K-MEANS CLUSTERING DI POLITEKNIK HARAPAN BERSAMA," vol. 5, no. 5, 2020.
- A. Lia Hananto, S. Sofiah Hilabi, and D. Noviani, "Design of Customer Satisfaction Application at BCA Kcp Rengasdengklok Using C.45 Algorithm Method," *Buana Information Technology and Computer Sciences (BIT and CS)*, vol. 3, no. 1, pp. 11–16, 2022, doi: 10.36805/bit-cs.v3i1.2048.
- Y. Yuliani, "Algoritma Random Forest Untuk Prediksi Kelangsungan Hidup Pasien Gagal Jantung Menggunakan Seleksi Fitur Bestfirst," *Infotek : Jurnal Informatika dan Teknologi*, vol. 5, no. 2, pp. 298–306, Jul. 2022, doi: 10.29408/jit.v5i2.5896.
- M. R. Muttaqin and M. Defriani, "Algoritma K-Means untuk Pengelompokan Topik Skripsi Mahasiswa," *ILKOM Jurnal Ilmiah*, vol. 12, no. 2, pp. 121–129, 2020, doi: 10.33096/ilkom.v12i2.542.121-129.
- D. P. Sari, "Implementasi Algoritma K-Means Dalam Menentukan Tingkat Penyebaran Pandemi Covid-19 Di Sumatera Barat," *Computer Based Information System Journal*, vol. 9, no. 1, pp. 50–56, 2021, doi: 10.33884/cbis.v9i1.3646.
- R. Julianti Hablum, A. Khairan, J. Jati Metro, and K. Ternate Selatan, "JIKO (Jurnal Informatika dan Komputer) Ternate CLUSTERING HASIL TANGKAP IKAN DI PELABUHAN PERIKANAN NUSANTARA (PPN) TERNATE MENGGUNAKAN ALGORITMA K-MEANS," 2019.
- D. Marlina, N. Lina, A. Fernando, and A. Ramadhan, "Implementasi Algoritma K-Medoids dan K-Means untuk Pengelompokan Wilayah Sebaran Cacat pada Anak," *Jurnal CoreIT: Jurnal Hasil Penelitian Ilmu Komputer dan Teknologi Informasi*, vol. 4, no. 2, p. 64, 2018, doi: 10.24014/coreit.v4i2.4498.
- Tukino and Baenil Huda, "TechnoXplore Jurnal Ilmu Komputer & Teknologi Informasi," 2019.
- K. D. R. Sianipar, S. W. Siahaan, M. Siregar, F. I. R.H Zer, and D. Hartama, "PENERAPAN ALGORITMA K-MEANS DALAM MENENTUKAN TINGKAT KEPUASAN PEMBELAJARAN ONLINE PADA MASA PANDEMI COVID-19," *Jurnal Teknologi Informasi*, vol. 4, no. 1, 2020, doi: 10.36294/jurti.v4i1.1258.

- E. Muningsih, I. Maryani, and V. R. Handayani, "Penerapan Metode K-Means dan Optimasi Jumlah Cluster dengan Index Davies Bouldin untuk Clustering Propinsi Berdasarkan Potensi Desa," *Jurnal Sains dan Manajemen*, vol. 9, no. 1, pp. 95–100, 2021.
- M. Herviany, S. P. Delima, T. Nurhidayah, and Kasini, "Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Daerah Rawan Tanah Longsor di Provinsi Jawa Barat," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 1, no. 1, pp. 34–40, 2021.