

KLASTERISASI PROFIL MAHASISWA BARU MENGGUNAKAN ALGORITME K-MEANS PADA FAKULTAS TEKNOLOGI INFORMASI UNIVERSITAS BUDI LUHUR

Vanessa Melati Al Aqsa¹, Deni Mahdiana^{2*}, Rusdah³, Dian Anubhakti⁴

^{1,2,3,4} Sistem Informasi, Fakultas Teknologi Informasi, Universitas Budi Luhur, DKI Jakarta, Indonesia

Email: ¹vanessamelati2@gmail.com, ^{2*}deni.mahdiana@budiluhur.ac.id, ³rusdah@budiluhur.ac.id, ⁴dian.anubhakti@budiluhur.ac.id

(* : corresponding author)

Abstrak-Fakultas Teknologi Informasi di Universitas Budi Luhur yang terdiri dari program studi yaitu Teknik Informatika, Sistem Informasi, dan Sistem Komputer. Universitas Budi Luhur ialah salah satu perguruan tinggi swasta di DKI Jakarta yang menjadi bagian dalam persaingan untuk mendapatkan mahasiswa baru setiap tahunnya. Permasalahan penelitian ini yaitu data penerimaan mahasiswa baru dalam 3 tahun terakhir yang melakukan daftar ulang calon mahasiswa baru mengalami penurunan, dikarenakan belum diketahuinya karakteristik atau profil calon mahasiswa baru yang masuk atau yang mendaftar di FTI UBL (Fakultas Teknologi Informasi Universitas Budi Luhur). Untuk menyelesaikan masalah tersebut pada penelitian ini diusulkan metode klasterisasi menggunakan Algoritma K-Means untuk mendapatkan karakteristik atau profil calon mahasiswa baru yang akan mendaftar FTI UBL sehingga dapat menjadi rekomendasi untuk tim penerimaan mahasiswa baru dalam menyusun strategi promosi. Penelitian ini menghasilkan 3 buah *cluster* yang terdiri dari *cluster* satu adalah *cluster* Teknik Informatika, asal sekolah SMK, dengan kota asal Kecamatan Larangan, Karang Tengah, Kebayoran Lama, dan Kembangan dengan jumlah 240 items, *cluster* dua *cluster* Teknik Informatika, asal sekolah SMK, dengan kota asal Kecamatan Pesanggrahan dan Pondok Aren adalah dengan jumlah 242 items, dan *cluster* tiga adalah *cluster* Teknik Informatika, asal sekolah SMK, dengan kota asal Ciledug dengan jumlah 205 items, dengan nilai Davies-Bouldin (DB) sebesar 0,470.

Kata Kunci: Klasterisasi, *K-Means*, Profil Mahasiswa Baru.

CLUSTERING OF NEW STUDENT PROFILES USING THE K-MEANS ALGORITHM AT THE FACULTY OF INFORMATION TECHNOLOGY, BUDI LUHUR UNIVERSITY.

Abstract- Faculty of Information Technology at Budi Luhur University which consists of study programs namely Informatics Engineering, Information Systems, and Computer Systems. Budi Luhur University is a private tertiary institution in DKI Jakarta which is part of the competition to get new students every year. The problem of this research is that data on new student admissions in the last 3 years who have re-registered prospective new students has decreased, due to the unknown characteristics or profiles of prospective new students who enter or register at FTI UBL (Faculty of Information Technology, Budi Luhur University). To solve this problem, this research proposes a clustering method using the K-Means Algorithm to obtain the characteristics or profiles of prospective new students who will register at FTI UBL so that they can become recommendations for the new student admissions team in developing promotion strategies. This study produced 3 clusters consisting of cluster one, namely the Informatics Engineering cluster, from SMK school, with cities from Larangan, Karang Tengah, Kebayoran Lama, and Kembangan sub-districts with a total of 240 items, clusters of two Informatics Engineering clusters, from SMK schools, with the city of origin of Pesanggrahan and Pondok Aren sub-districts is with a total of 242 items, and cluster three is the Informatics Engineering cluster, from a vocational school, with a city of origin in Ciledug with a total of 205 items, with a Davies-Bouldin (DB) value of 0.470.

Keywords: Clustering, *K-Means*, Profile of New Student.

1. PENDAHULUAN

Setiap tahun ajaran baru peminat dari penerimaan mahasiswa baru yang ingin masuk ke Fakultas Teknologi Informasi program studi yaitu Teknik Informatika, Sistem Informasi, dan Sistem Komputer di Universitas Budi Luhur sangat banyak. Universitas Budi Luhur ialah salah satu perguruan tinggi swasta di DKI Jakarta yang menjadi bagian dalam persaingan untuk mendapatkan mahasiswa baru setiap tahunnya. Mahasiswa baru merupakan seseorang yang sedang pada proses menimba ilmu buat menjalani pendidikan pada salah satu perguruan tinggi.

Pertumbuhan data penerimaan mahasiswa baru dari 3 tahun terakhir yang melakukan daftar ulang yaitu tahun ajaran 2019/2020 terdapat 360 mahasiswa baru kelas reguler dan 260 mahasiswa baru kelas sore, tahun ajaran 2020/2021 mahasiswa baru kelas reguler yaitu 351 mahasiswa dan 184 mahasiswa kelas sore, dan tahun ajaran 2021/2022 sebanyak 318 mahasiswa kelas reguler dan 107 mahasiswa kelas sore dari Fakultas Teknologi Informasi di Universitas Budi Luhur. Dari data 3 tahun terakhir di atas kecenderungan penerimaan mahasiswa baru menurun, dikarenakan belum diketahuinya karakteristik atau profil calon mahasiswa baru yang masuk atau yang mendaftar di FTI UBL (Fakultas Teknologi Informasi Universitas Budi Luhur). Salah satu metode yang bisa menangani permasalahan tersebut merupakan menggunakan penerapan data mining. Pemilihan metode yang akan digunakan pada data penerimaan mahasiswa baru ialah metode *clustering*.

Data mining adalah proses menemukan dan menganalisis data dalam jumlah besar dengan tujuan menemukan pola atau informasi yang menarik dari sejumlah besar data yang tersimpan dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma yang tepat sangat bergantung pada tujuan dan proses *knowledge discovery in database* (KDD) secara keseluruhan[1]. *Clustering* atau klasterisasi adalah metode pengelompokan data dalam proses partisi satu set objek data ke dalam himpunan bagian yang disebut dengan *cluster*. objek yang di dalam *cluster* memiliki karakteristik antar satu sama lainnya dan berbeda dengan *cluster* yang lain. Partisi tidak dilakukan secara manual melainkan dengan suatu algoritma *clustering*[2].

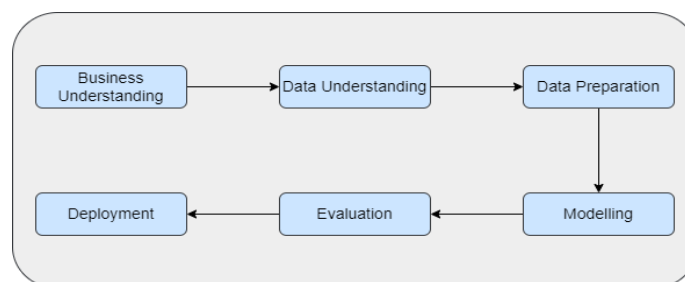
Terdapat metode-metode dari penelitian terdahulu yang sudah melakukan penentuan strategi promosi *clustering* terhadap data mahasiswa baru dengan menggunakan metode algoritma K-Means diantaranya adalah pengelompokan mahasiswa baru di Universitas Bina Darma Palembang, dimana *cluster* pertama adalah *cluster* yang terbanyak mendapatkan informasi dari brosur dan *cluster* 2 adalah *cluster* mendapatkan informasi dari teman dan kerabat [3]. Penelitian lainnya adalah pengelompokan mahasiswa baru di akademi maritim suaka bahari dengan mendapatkan jumlah *cluster* terbanyak ada pada *cluster* 16 adalah *cluster* mahasiswa yang berasal dari Jawa Barat, Jawa Tengah, Jawa Timur, DKI Jakarta, Jambi, Lampung, dan D.I. Yogyakarta [4]. Penelitian lainnya, pengelompokan mahasiswa baru dan menghasilkan *cluster* pertama adalah *cluster* SMK dan *cluster* kedua adalah SMA [5]. Selain itu, terdapat penelitian lain yang menggunakan algoritma berbeda yaitu algoritma Fuzzy C-Means dari penelitian pengelompokan data mahasiswa Universitas Islam Majapahit dimana daerah kecamatan, asal sekolah, dan program studi sebagai penentuan strategi promosi [6].

Selanjutnya terdapat peneliti lain yang menggunakan algoritma berbeda dari yang sebelumnya yaitu algoritma K-Medoids dari penelitian pengelompokan data mahasiswa di Stikers Perintis Padang, dimana *cluster* pertama adalah *cluster* dengan asal sekolah daerah Kerinci, *cluster* kedua dengan asal sekolah daerah Padang, *cluster* ketiga dengan asal sekolah daerah Agam, dan *cluster* keempat dengan asal sekolah daerah Bukittinggi [7]. Berdasarkan beberapa penelitian sebelumnya, maka pada penelitian ini mengusulkan untuk melakukan pengelompokan profil calon mahasiswa baru yang masuk atau mendaftar di FTI UBL (Fakultas Teknologi Informasi Universitas Budi Luhur) dengan metode K-Means. Sehingga tujuan penelitian ini adalah menerapkan algoritma K-Means untuk klasterisasi profil mahasiswa baru Fakultas Teknologi Informasi Universitas Budi Luhur sehingga dapat menjadi rekomendasi untuk tim penerimaan mahasiswa baru dalam menyusun strategi promosi.

2. METODE PENELITIAN

2.1 Metode CRISP-DM

CRISP-DM (*Cross Industry Standard Process Model for Data Mining*) merupakan metode yang banyak digunakan oleh para ahli dengan menggunakan proses pemodelan data di dalamnya. Tujuan dari metode CRISP-DM yaitu untuk menemukan pola yang menarik dan memiliki makna pada data yang digunakan. CRISP-DM memiliki tahapan dan kerangka kerja yang terstruktur sehingga pengguna metode ini akan lebih terarah dan mengetahui langkah yang harus dikerjakan dalam penelitian[8]. Pada tahapan ini menerapkan metode CRISP-DM (*Cross Industry Standard Process Model for Data Mining*) dengan beberapa tahapan di dalamnya. Model Proses CRISP-DM.



Gambar 1. Proses CRISP-DM

Berdasarkan Gambar 1, berikut adalah penjelasan tentang tahapan-tahapan yang dilakukan:

a. *Business Understanding*

Tahap pertama ialah *business understanding* untuk memahami tujuan dan kebutuhan yang ingin dicapai pada penelitian ini. Penelitian ini mempunyai maksud dan tujuan untuk klasterisasi profil mahasiswa baru Fakultas Teknologi Informasi Universitas Budi Luhur.

b. *Data Understanding*

Tahap ini dimulai pengumpulan data penerimaan mahasiswa baru pada tahun ajaran 2019/2020 yang didapat dari Direktur Teknologi Informasi Universitas Budi Luhur. Setelah itu dilanjutkan dengan proses untuk mendapatkan pemahaman yang mendalam tentang data tersebut, selain itu untuk mendeteksi apabila ada bagian yang menarik dari data yang akan digunakan buat hipotesa informasi yang tersembunyi.

c. *Data Preparation*

Pada tahap ini adalah tahapan dari semua kegiatan untuk membangun dataset akhir dari data mentah. Tahapan ini melakukan proses data *cleaning* dan data *transformation*. Setelah itu, data digabung ke dalam format yang sesuai untuk diproses dalam data mining.

d. *Modeling*

Pada tahap pemodelan dilakukan pemilihan dan penerapan teknik pemodelan serta beberapa parameternya yang akan disesuaikan untuk mendapatkan *cluster* yang optimal. Terdapat beberapa pemodelan untuk menyelesaikan masalah yang sama pada data mining. Dalam penelitian ini akan menerapkan algoritma K-Means dan menggunakan tool Rapid Miner.

K-Means adalah algoritma *clustering* yang sederhana untuk membangun kelompok data menjadi beberapa *cluster* (*k*). K-Means merupakan salah satu algoritma pelatihan *unsupervise*, awal mula algoritma ini dipublikasikan oleh Stuart Lloyd di tahun 1984 yang hingga saat ini merupakan algoritma yang banyak digunakan. Alasan algoritma ini banyak digunakan adalah karena K-Means cukup mudah untuk diimplementasikan, relatif cepat, dan mudah dimodifikasi [9].

e. *Evaluation*

Di tahap ini akan melakukan evaluasi memakai metode Davies Bouldin Index untuk memilih jumlah *cluster* terbaik dengan cara melihat persentase hasil perbandingan antara jumlah *cluster*.

Metode untuk mengevaluasi hasil *cluster* pada suatu metode *clustering* dengan menggunakan metode Davies-Bouldin Index. Landasan dari metode ini adalah nilai kohesi dan separasi. Proses *clustering*, kohesi merupakan jumlah dari kedekatan data terhadap *centroid* dari *cluster* yang diikuti. Sedangkan separasi didasarkan pada jarak antar *centroid* dari *cluster* tersebut [10].

f. *Deployment*

Tahapan ini dilakukan setelah menyelesaikan tahapan-tahapan di atas untuk menghasilkan laporan klasterisasi atau profil mahasiswa baru dalam menyusun strategi promosi yang tepat dengan karakteristik mahasiswa baru Fakultas Teknologi Informasi.

3. HASIL DAN PEMBAHASAN

3.1 *Business Understanding*

Penelitian ini dilakukan dengan menggunakan data primer yang didapatkan dari Direktur Teknologi Informasi Universitas Budi Luhur. Data yang diperoleh sebanyak 687 data. Data yang didapat yaitu data penerimaan mahasiswa baru tahun ajaran 2019/2020. Atribut dari data mahasiswa baru yang akan digunakan pada penelitian ini yaitu Asal Sekolah, Program Studi, Jurusan SLTA, Kota Asal, Status Kelas, Jenis Kelamin, dan Tahun Masuk.

3.2 *Data Preparation*

Di tahapan ini data yang didapat sebanyak 687 record mahasiswa baru tahun ajaran 2019/2020. Tahapan ini yang diterapkan di database buat menghapus noise, cleaning data, missing value, dan data yang tidak konsisten. Data Preprocessing pada penelitian ini dibagi menjadi beberapa langkah, yaitu cleaning data dan data

transformation. Langkah – langkah yang terdapat dalam tahapan ini tidak semuanya harus dilakukan. Berikut beberapa langkah – langkah yang dilakukan pada penelitian ini :

a. *Data Cleaning*

Pada tahapan ini data yang dikumpulkan memiliki beberapa bagian yang kosong dan hilang. Bagian tersebut dari atribut Jurusan SLTA dan Kota Asal. Salah satu cara untuk mengatasi *missing value* ialah dengan mengisi *missing value* tersebut. Pengisian *missing value* dalam penelitian ini dilakukan dengan cara yaitu mengisi manual *missing value* tersebut. Atribut Jurusan SLTA dengan *searching* jurusan yang ada di sekolah sesuai dengan data mahasiswa lalu dipilih Jurusan SLTA yang mendekati dengan pilihan program studi di Universitas Budi Luhur. Lalu, atribut kota asal mengisi secara manual dengan cara melihat asal sekolah dari mahasiswa. Setelah sudah tidak ada bagian yang kosong, lalu ke tahap selanjutnya yaitu data *transformation*.

b. *Data Transformation*

Pada tahapan ini untuk mengubah seluruh data yang terkumpul agar memudahkan ketika diolah dengan menggunakan metode K-Means *clustering*. Adapun variabel yang digunakan pada penelitian ini yaitu Asal Sekolah, Program Studi, Jurusan SLTA, Jenis Kelamin, Status Kelas, Kota Asal, dan Tahun Masuk.

Berikut ini merupakan hasil proses transformasi data menggunakan label encoding untuk Asal Sekolah seperti terlihat pada tabel 1, Program Studi pada tabel 2, Jurusan SLTA pada tabel 3, Jenis Kelamin pada tabel 4, Status Kelas pada tabel 5, Kota Asal pada tabel 6, dan Tahun Masuk pada tabel 7. Untuk pemilihan atribut Jurusan SLTA berdasarkan penelitian [3] :

1. Asal Sekolah

Tabel 1. Hasil Transformasi Data Asal Sekolah

Asal Sekolah	Kode
Sekolah Menengah Atas (SMA)	1
Sekolah Menengah Kejuruan (SMK)	2
Selain SMA dan SMK	3

2. Program Studi

Tabel 2. Hasil Transformasi Data Program Studi

Program Studi	Kode
Sistem Informasi	1
Teknik Informatika	2
Sistem Komputer	3

3. Jurusan SLTA

Tabel 3. Hasil Transformasi Data Jurusan SLTA

Jurusan SLTA	Kode
Teknik Komputer Jaringan	1
Multimedia	2
IPA	3
IPS	4
Akuntansi	5
Administrasi Perkantoran	6
Rekayasa Perangkat Lunak	7
Pariwisata	8
Teknik Otomatif	9
Teknik Informatika	10
...	...
Teknik Transmisi	26

4. Jenis Kelamin

Tabel 4. Hasil Transformasi Data Jenis Kelamin

Jenis Kelamin	Kode
Pria	1
Wanita	2

5. Status Kelas

Tabel 5. Hasil Transformasi Data Staus Kelas

Status Kelas	Kode
Karyawan	1
Reguler	2

6. Kota Asal

Tabel 6. Hasil Transformasi Data Kota Asal

Kota Asal	Kode
Kec. Angsana	1
Kec. Baros	2
Kec. Batu	3
Kec. Batauceper	4
Kec. Beji	5
Kec. Bekasi Utara	6
Kec. Benda	7
Kec. Biringkanaya	8
Kec. Bobotsari	9
Kec. Bojong	10
...	...
Kec. Warung Asem	117

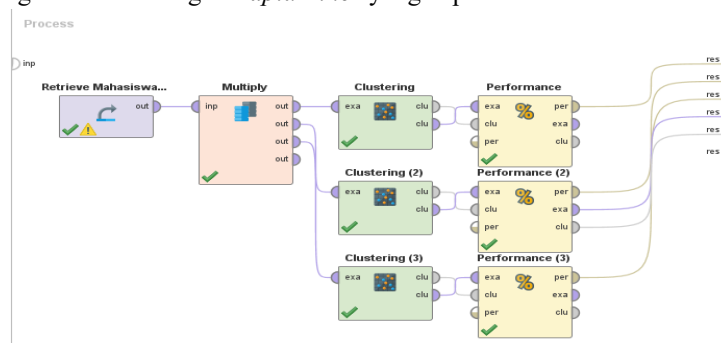
7. Tahun Masuk

Tabel 7. Hasil Transformasi Data Tahun Masuk

Tahun Masuk	Kode
2019	1

3.3 Modeling

Pada tahap ini melakukan tahap modeling untuk menentukan *cluster* yang optimal menggunakan k-means. Proses pada gambar 2 yang dilakukan dengan *RapidMiner* yang dapat dicermati.



Gambar 2. Proses dari Algoritma K-Means

Setelah melakukan proses rapid miner, maka dapat dilihat hasil pada tabel 8 nilai DBI paling kecil dengan penentuan jumlah *cluster* terbaik adalah berjumlah 3 *cluster* dengan nilai 0,470. Semakin kecil nilai DBI maka semakin optimal *cluster* yang dihasilkan.

Tabel 8. Hasil Davies Bouldin Index

Penentuan Jumlah <i>Cluster</i> Terbaik	Hasil Davies Bouldin Index
2 <i>Cluster</i>	0,510
3 <i>Cluster</i>	0,470
4 <i>Cluster</i>	0.494

Selanjutnya, gambar 3 merupakan hasil 3 *cluster* yang diperoleh :

Cluster Model

```
Cluster 0: 240 items
Cluster 1: 242 items
Cluster 2: 205 items
Total number of items: 687
```

Gambar 3. Hasil Jumlah *Cluster*

Dengan menggunakan pemodelan algoritma k-means *clustering* pada gambar di atas, menggunakan jumlah data 687 dan inialisasi jumlah *cluster* sebanyak 3, sesuai dengan pendefinisian nilai k dengan jumlah *cluster_0* : 240 items, *cluster_1* : 242 items, *cluster_2* : 205 items.

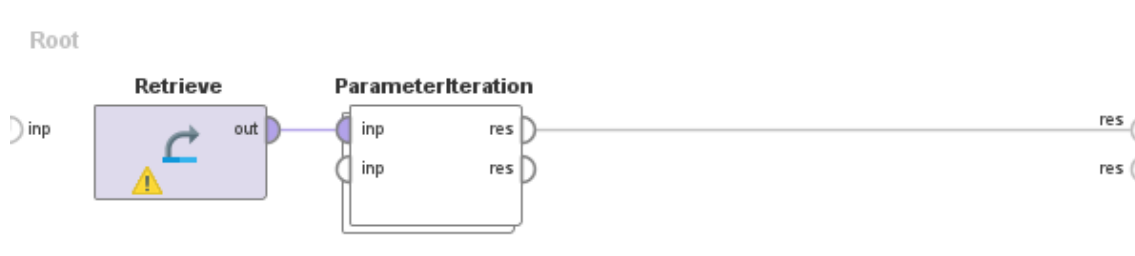
Dari data hasil penelitian *clustering* yang telah dilakukan pada gambar 3. Bisa disimpulkan bahwa *cluster* 1 program studi yang diminati adalah Teknik Informatika ,Sistem Informasi, Sistem Komputer tetapi untuk program studi yang dominan ialah Teknik Informatika, dengan dominan asal sekolah dari SMK, dan kota asal yaitu dominan dari Kec. Larangan, Kec. Karang Tengah, Kec. Kebayoran Lama, dan Kec. Kembangan.

Pada *cluster* 2 program yang diminati adalah Teknik Informatika, Sistem Informasi, Sistem Komputer tetapi untuk program studi yang dominan ialah Teknik Informatika, dengan dominan asal sekolah dari SMK, dan kota asal yaitu dominan dari Kec. Pesanggrahan dan Kec. Pondok Aren.

Selanjutnya *cluster* 3 program studi yang diminati adalah Teknik Informatika, Sistem Informasi, Sistem Komputer tetapi untuk program studi yang dominan ialah Teknik Informatika, dengan dominan asal sekolah dari SMK, dan kota asal yaitu dominan dari Kec. Ciledug.

3.4 Evaluation

Pada bagian ini disampaikan mekanisme pengukuran performa model terbaik. Metode uji dari penelitian ini menggunakan Davies-Bouldin Index untuk menguji jumlah *cluster* yang optimal. Proses gambar 4 dari pengujian yang menggunakan *tools* Rapidminer:



Gambar 4. Operator Pengujian

Proses di atas menggunakan operator *loop* parameters. Lalu dari hasil processLog menunjukkan nilai DB terkecil ialah 0,470 dengan jumlah 3 *cluster* yang optimal. Berikut pada gambar 5 yang menunjukkan hasil pengujian dengan nilai DB terkecil:

ProcessLog (11 rows, 3 columns)

k	DB	W
2	0.510	222.801
3	0.470	94.577
4	0.494	59.278
5	0.653	48.480
6	0.646	35.887
7	0.636	31.725
8	0.679	28.956
9	0.690	25.423
10	0.776	26.781
11	0.707	19.685
13	0.668	15.567

Gambar 5. Hasil Pengujian

4. KESIMPULAN

Berdasarkan hasil penelitian dapat diambil kesimpulan bahwa klusterisasi menggunakan algoritma K-Means mendapatkan karakteristik atau profil calon mahasiswa baru Fakultas Teknologi Informasi Universitas Budi Luhur sehingga dapat menjadi rekomendasi bagi tim pnb (penerimaan mahasiswa baru) untuk menyusun strategi promosi dengan menghasilkan 3 (tiga) *cluster* yang optimal yaitu *cluster* satu (*cluster* Teknik Informatika, asal sekolah SMK, dengan kota asal Kecamatan Larangan, Karang Tengah, Kebayoran Lama, dan Kembangan) dengan jumlah 240 items, *cluster* dua (*cluster* Teknik Informatika, asal sekolah SMK, dengan kota asal Kecamatan Pesanggrahan dan Pondok Aren) adalah dengan jumlah 242 items, dan *cluster* tiga (*cluster* Teknik Informatika, asal sekolah SMK, dengan kota asal Ciledug) dengan jumlah 205 items. Dengan nilai Davies Bouldin (DB) sebesar 0,470.

DAFTAR PUSTAKA

- [1] Y. Arie Wijaya, "Clustering Data Calon Siswa Baru Menggunakan Metode K-Means Di Sekolah Menengah Kejuruan Wahidin Kota Cirebon," J. Mhs. Tek. Inform., vol. 6, no. 2, hal. 552–559, 2022, [Daring]. Tersedia pada: <https://www.smkwahidincrb.sch.id/>.
- [2] Y. Mayona, R. Buaton, dan M. Simanjutak, "Data Mining Clustering Tingkat Kejahatan Dengan Metode Algoritma K-Means (Studi Kasus: Kejaksaan Negeri Binjai)," J. Inform. Kaputama, vol. 6, no. 3, hal. 2548–9739, 2022.
- [3] M. R. Alhapizi, M. Nasir, dan I. Effendy, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Mahasiswa Baru Universitas Bina Darma Palembang," J. Softw. Eng. Ampera, vol. 1, no. 1, hal. 1–14, 2020, doi: 10.51519/journalsea.v1i1.10.
- [4] T. Hartati, O. Nurdiawan, dan E. Wiyandi, "Analisis Dan Penerapan Algoritma K-Means Dalam Strategi Promosi Kampus Akademi Maritim Suaka Bahari," J. Sains Teknol. Transp. Marit., vol. 3, no. 1, hal. 1–7, 2021, doi: 10.51578/j.sitektransmar.v3i1.30.
- [5] N. Damanik dan M. Sigiroy, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Pada Penerimaan Mahasiswa Baru Sebagai Metode Promosi," J. Tek. Inform. Komput. Univers., vol. 4, no. 2, hal. 158, 2021, doi: 10.37600/tekinkom.v4i2.367.
- [6] Y. N. S. Abdul Malik, Yessy Diah Rosita, "Aplikasi Clustering Data Mahasiswa Universitas Islam Majapahit dengan Metode Fuzzy C-Means," Submit, vol. 2, no. 2, hal. 75–81, 2022.
- [7] E. Rahmah, E. Haerani, A. Nazir, dan S. Ramadhani, "Penerapan Algoritma K-Medoids Clustering Untuk Menentukan Strategi Promosi Pada Data Mahasiswa (Studi Kasus : Stikes Perintis Padang)," vol. 5, no. 3, hal. 556–564, 2022.
- [8] F. N. Dhewanyani, D. Amelia, D. N. Alifah, B. N. Sari, dan M. Jajuli, "Implementasi K-Means Clustering untuk Pengelompokan Daerah Rawan Bencana Kebakaran Menggunakan Model CRISP-DM," J. Teknol. dan Inf., vol. 12, no. 1, hal. 64–77, 2022, doi: 10.34010/jati.v12i1.6674.
- [9] A. A. Hussein, "Improve The Performance of K-means by using Genetic Algorithm for Classification Heart Attack," Int. J. Electr. Comput. Eng., vol. 8, no. 2, hal. 1256, 2018, doi: 10.11591/ijece.v8i2.pp1256-1261.
- [10] R. K. Dinata, H. Novriando, N. Hasdyna, dan S. Retno, "Reduksi Atribut Menggunakan Information Gain untuk Optimasi Cluster Algoritma K-Means," J. Edukasi dan Penelit. Inform., vol. 6, no. 1, hal. 48, 2020, doi: 10.26418/jp.v6i1.37606.