Institut Riset dan Publikasi Indonesia (IRPI)
MALCOM: Indonesian Journal of Machine Learning and Computer Science
Journal Homepage: https://journal.irpi.or.id/index.php/malcom
Vol. 5 Iss. 4 October 2025, pp: 1160-1171
ISSN(P): 2797-2313 | ISSN(E): 2775-8575

Segmentation of Promotion Target Areas Using the K-Means Clustering Algorithm at Universities

Segmentasi Wilayah Target Promosi Menggunakan Algortima *K-Means* Clustering pada Perguruan Tinggi

Zela Poiema Christy Mantohana Napitupulu[1*], Jong Jek Siang[2], Andhika Galuh Prabawati[3]

[1,2,3]Program Studi Sistem Informasi, Fakultas Teknologi Informasi,
Universitas Kristen Duta Wacana, Indonesia

E-Mail: [1]zela.poiema@si.ukdw.ac.id, [2]jjsiang@staff.ukdw.ac.id, [3]andhika.galuh@staff.ukdw.ac.id

Received Jun 20th 2025; Revised Aug 12th 2025; Accepted Sep 03rd 2025; Available Online Oct 30th 2025
Corresponding Author: Zela Poiema Christy Mantohana Napitupulu

Abstract

The admissions unit of a private university actively promotes to high schools in various forms. Target schools are selected based on experience and personnel availability, but have not used historical data analysis. In this research, a promotion area segmentation system is built using K-Means Clustering to group potential schools, cities, and provinces based on previous years' enrollment activities. The net data used is 7177 rows of data, which is taken from data on prospective student applicants for the period 2020 to 2024. The data includes accepted applicants, choice 1, choice 2, registration time, registration, and cancellation of registration. The data processing process includes the stages of cleaning, normalizing school names, standardizing school names using fuzzy string matching, weighting the status of applicants, and calculating scores. Clustering using K-Means is performed on the results of the normalized score calculation. The optimal number of clusters is determined using the Silhouette Coefficient method. The output is the centroid value and cluster label for each clustered entity. The system successfully clusters the potential schools, cities, and provinces of the promotion site. The output of the system is compared with the real location of promotions that have been carried out by the Admissions unit. The results show that of the 125 schools where promotion took place, only 48 schools matched the systems results. The system identified 109 potential and sufficiently potential schools, 44.04% of which had already been reached by promotion.

Keyword: Data Analys, K-Means Clustering, Promotion, Promotion Strategy, Region Segmentation

Abstrak

Unit admisi sebuah perguruan tinggi swasta aktif melakukan promosi ke sekolah menengah atas dengan berbagai bentuk. Target sekolah dipilih berdasarkan pengalaman dan ketersediaan personil, namun belum menggunakan analisis histori data. Pada penelitian ini dibangun sistem segmentasi wilayah promosi menggunakan *K-Means Clustering* untuk mengelompokkan potensi sekolah, kota, dan provinsi berdasarkan aktivitas pendaftaran tahun-tahun sebelumnya. Data bersih yang digunakan sejumlah 7177 baris data, yang diambil dari data calon mahasiswa pendaftar periode 2020 hingga 2024. Data meliputi pendaftar yang diterima, pilihan 1, pilihan 2, waktu pendaftaran, registrasi dan membatalkan registrasi. Proses pengolahan data meliputi tahapan pembersihan normalisasi nama sekolah, dan standarisasi nama sekolah menggunakan *fuzzy string matching*, pembobotan status pendaftar, dan perhitungan skor. Pengklasteran menggunakan *K-Means* dilakukan pada hasil perhitungan skor yang telah dinormalisasi. Jumlah klaster optimal ditentukan menggunakan metode *Silhouette Coefficient*. Luaran berupa nilai *centroid* dan label klaster untuk setiap entitas yang diklaster. Sistem berhasil mengelompokkan potensi sekolah, kota, dan provinsi tempat promosi. Luaran sistem dibandingkan dengan lokasi riil promosi yang telah dilakukan unit Admisi. Hasil menunjukkan bahwa dari 125 sekolah tempat promosi, hanya 48 sekolah yang sesuai dengan hasil sistem. Sistem mengidentifikasi 109 sekolah potensial dan cukup potensial, 44,04% yang sudah dijangkau promosi.

Kata Kunci: *K-Means Clustering*, Segmentasi Wilayah, *Silhouette Coefficient*, Strategi Promosi

1. PENDAHULUAN

Unit Admisi sebuah Perguruan Tinggi Swasta rutin mengelola proses penerimaan mahasiswa baru yang meliputi kegiatan promosi dengan berbagai bentuk ke sekolah-sekolah, mengelola proses pendaftaran, seleksi dan registrasi calon mahasiswa baru. Dalam promosi, penentuan lokasi dan sekolah-sekolah potensial menjadi aspek penting agar promosi menjadi efektif dan efisien mengingat keterbatasan sumber daya manusia dan anggaran [1]. Upaya promosi yang dilakukan mencakup berbagai strategi, seperti periklanan di berbagai *platform*, penyelenggaraan webinar, *workshop*, *expo* secara langsung maupun virtual, dan *campus visit*. Namun, dengan meningkatnya lokasi persebaran promosi, unit admisi menghadapi tantangan dalam menetapkan lokasi pemasaran yang potensial. Masalahnya, penentuan target sekolah promosi hanya dilakukan berdasarkan history tahun-tahun sebelumnya dan belum menggunakan data pendaftar. Apabila keputusan target penentuan target promosi tidak didasarkan pada analisis data tentang potensi pendaftaran dapat menyebabkan ketidaktepatan dalam menjangkau lokasi dan sekolah-sekolah yang memiliki potensi pendaftaran yang tinggi [2]. Analisis yang dilakukan secara manual membutuhkan waktu yang cukup lama karena calon mahasiswa berasal dari sekolah yang beragam, baik dari negeri maupun swasta yang tersebar di beragam wilayah [2]. Kunjungan promosi seringkali dilakukan secara *incidental* dan random karena tidak adanya pendataan sekolah yang menjadi prioritas [3][4]. Penetapan sekolah untuk promosi hanya didasarkan pada subjektivitas unit promosi seperti jumlah siswa di sekolah, tanpa mempertimbangkan aspek lainnya seperti jumlah siswa yang tertarik dan masuk menjadi mahasiswa di universitas tersebut [5]. Proses pengelompokkan data mahasiswa baru masih dilakukan dengan cara perhitungan manual sehingga berisiko menimbulkan kesalahan input dan membutuhkan waktu yang lama [6]. Pada penelitian ini dibuat segmentasi sekolah target promosi menggunakan metode Kmeans menggunakan data pendaftar secara holistik selama tahun 2020-2024 mulai dari pendaftaran, tes seleksi, hingga registrasi. Tujuannya agar diperoleh daftar sekolah yang secara historis potensial. Hasil penelitian penting bagi unit Admisi agar target area promosi terarah dan tidak dilakukan secara insidental berdasarkan subyektivitas pegawai Admisi.

Analisis model *Recency*, *Frequency*, *Monetary* (RFM) dan *K-Means clustering* digunakan untuk mengelompokkan tingkat potensial lokasi promosi berdasarkan *dataset* Penerimaan Mahasiswa Baru (PMB) [3]. Pengambilan keputusan akan sekolah potensial yang menjadi target promosi didapatkan dari klaterisasi atau pengelompokkan data calon mahasiswa pendaftar yang diklasifikasikan berdasarkan beberapa indikator seperti asal sekolah, jumlah lulusan, dan faktor lainnya menggunakan metode *Fuzzy C Means* (FCM) yang tidak memerlukan banyak parameter [5]. Penelitian lainnya membandingkan metode *clustering K-Means* dengan *K-Medoids* untuk mengelompokkan calon mahasiswa berdasarkan asal daerah, sekolah, dan jurusan ditemukan bahwa *K-Means* menunjukkan performa lebih baik dan lebih cocok digunakan untuk klasterisasi dengan distribusi data yang lebih seimbang [6]. Diantara metode *clustering* seperti *K-Means clustering*, *Fuzzy C Means*, dan *Hiearrichical clustering*, metode *K-Means* adalah algoritma paling sederhana yang menggunakan *unsupervised learning* untuk memecahkan masalah *clustering* [7][8]. Meskipun beberapa penelitian telah menjadi solusi terhadap permasalahan yang dihadapi namun terdapat beberapa kekurangan. Penelitian terbatas pada variabel atau indikator yang digunakan dalam proses klasterisasi belum mencakup faktor-faktor penting dari data calon mahasiswa [5].

Penelitian PMB [3] hanya menggunakan data siswa yang sudah melakukan registrasi dalam proses Kmeans. Kebaruan penelitian ini terletak pada penggunaan data calon mahasiswa secara holistik di semua level mulai pendaftaran, tes seleksi, hasil tes, hingga registrasi. Nilai positif atau negatif diberikan di tiap level. Siswa yang mengundurkan diri setelah melakukan registrasi menyebabkan nilai negatif pada sekolah asalnya.

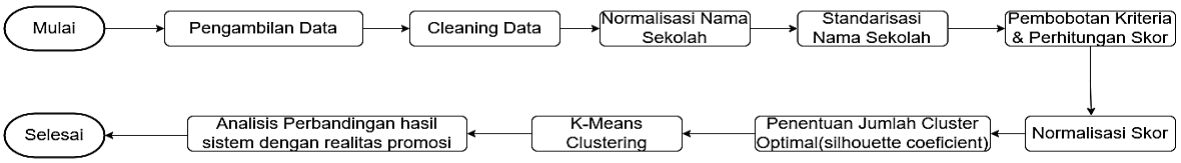
Penelitian ini bertujuan menghasilkan sistem berbasis web yang dapat melakukan segmentasi potensi sekolah menengah atas untuk promosi. Sistem dirancang untuk membantu unit Admisi dalam mengidentifikasi wilayah dan sekolah yang memiliki potensi tinggi sebagai target promosi. Berbeda dari penelitian sebelumnya yang umumnya hanya menggunakan sedikit parameter dan tanpa pembobotan, penelitian ini mengintegrasikan beberapa indikator penting, seperti jumlah pendaftar, jumlah yang lulus seleksi, waktu pendaftaran, jumlah registrasi, dan jumlah pembatalan registrasi. Masing-masing indikator diberikan bobot tertentu sesuai tingkat kontribusinya terhadap potensi sekolah.

2. METODOLOGI PENELITIAN

Metodologi penelitian merupakan pendekatan terstruktur yang digunakan untuk merancang, melaksanakan, dan menganalisis penelitian guna mencapai tujuan yang telah ditentukan [9]. Tahapan penelitian yang dilakukan dalam penelitian ini tampak pada Gambar 1.

Tahap awal dalam penelitian ini dilakukan pengambilan data. Terdapat 3 macam data yang digunakan, yaitu data calon mahasiswa baru dan data referensi program studi (ref_prodi), data realitas promosi, dan data sekolah seluruh Indonesia. *Dataset* calon mahasiswa baru yang digunakan adalah data periode 2020 hingga 2024 yang memiliki 22 *field* tetapi hanya 13 *field* yang digunakan yaitu no_daftar, tahun, tgl_daftar, pilihan1, pilihan2, kelamin, nama_sekolah, kota_sek, kab_sek, prop_sek, terima, nim, dan batal. *Dataset* referensi program studi (ref_prodi) memuat informasi mengenai kode program studi beserta nama program studi.

Dataset ini digunakan untuk memetakan kode program studi ke nama program studi yang sesuai. Data realitas promosi berupa dokumen yang memuat lokasi-lokasi promosi yang telah dilakukan oleh admisi pada tahun 2020 hingga 2024. Data sekolah seluruh Indonesia diperoleh melalui metode pengumpulan data sekunder dari repositori GitHub. Data ini yang akan digunakan sebagai referensi dalam proses standarisasi nama sekolah.



Gambar 1. Langkah Penelitian

2.1. Cleaning Data

Cleaning data merupakan serangkaian proses yang dilakukan untuk membersihkan, memodifikasi, atau menyiapkan data mentah sebelum digunakan dalam proses pemodelan atau analisis lanjutan [10]. Data akan diproses dalam tahap pembersihan seperti pengisian nilai kosong berdasarkan konteks, penghapusan kolom yang tidak relevan, serta penghapusan data duplikat. Hasil *cleaning* data dapat dilihat pada Tabel 1.

Tabel 1. *Cleaning* Data

Permasalahan Data	Jumlah	Jumlah Sesudah Dibersihkan
Jumlah Atribut (Kolom)	24	13
Nilai Kosong (*Missing Values*)	69	7177

2.2. Normalisasi Nama Sekolah

Dilakukan normalisasi nama sekolah untuk menyelaraskan variasi penulisan nama sekolah, mendefinisikan singkatan, agar seragam. Proses normalisasi bisa dilihat pada Tabel 2.

Tabel 2. Normalisasi Nama Sekolah

Sebelum	Sesudah
SMA NEGERI 1 WAMENA	SMAN 1 WAMENA
SMA PANGUDI LUHUR YK	SMA PANGUDI LUHUR YOGYAKARTA
SMA N 1 GIRSIP PARAPAT	SMAN 1 GIRSANG SIPANGAN PARAPAT
SMA NEGRI 2 LUWU	SMAN 2 LUWU

2.3. Standarisasi Nama Sekolah

Normalisasi hanya untuk menangani variasi kecil, sehingga masih diperlukan standarisasi variasi nama sekolah menggunakan *fuzzy string matching* terhadap referensi data sekolah seluruh Indonesia. Konsep utama dalam pencarian *fuzzy string matching* adalah menentukan apakah sebuah string yang dicari memiliki kemiripan dengan *string* yang ada di dalam kamus, meskipun tidak identik dalam susunan karakternya [11][12]. Proses ini berupa pencocokan wilayah bertingkat, pencocokan nama sekolah menggunakan token_set_ratio, partial_ratio, token_sort_ratio, fuzz.ratio dengan *threshold*, penyaringan kandidat berdasarkan jenis sekolah, angka, dan status sekolah, pemilihan kandidat terbaik berdasarkan skor *fuzzy* tertinggi dan jumlah kata yang paling mirip. Hasil standarisasi nama sekolah bisa dilihat pada Tabel 3.

Tabel 3. Standarisasi Nama Sekolah

Nama Sekolah Normalisasi	Nama Sekolah Referensi	Skor Kemiripan	Hasil Standarisasi
SMA BOPKRI 1 YOGYAKARTA	SMAS BOPKRI 1	98	SMAS BOPKRI 1
SMK SANTO FRANSISKUS ASISI	SMK ASISI	100	SMK ASISI
SMA PANGUDI LUHUR SANTO YOHANES	SMAS PANGUDI LUHUR SANTO YOHANES KETAPANG	99	SMAS PANGUDI LUHUR SANTO YOHANES KETAPANG

2.4. Pembobotan Kriteria dan Perhitungan Skor

Pada tahap ini diberikan bobot pada masing-masing kriteria yang sudah ditentukan untuk menilai potensi setiap sekolah serta potensi program studi yang ada pada masing-masing sekolah. Diberikan nilai bobot yang berbeda beda sesuai dengan tingkat kepentingannya dalam menentukan potensi suatu sekolah dan potensi program studi. Bisa dilihat kriteria dan bobot untuk potensi sekolah pada Tabel 4.

Pada K1 sampai K4, nilai bobotnya diatur linier terhadap potensi calon mahasiswa untuk kuliah. Sebaliknya, K5 diberi bobot negatif karena pembatalan registrasi.

Tabel 4. Kriteria dan Bobot untuk Potensi Sekolah

Kode Kriteria	Kriteria	Bobot
K1	Jumlah Pendaftar	0,1
K2	Jumlah Lulus Seleksi	0,2
K3	Waktu pendaftaran awal	0,3
K4	Jumlah Registrasi	0,3
K5	Jumlah Batal Registrasi	-0,3

Perhitungan dilakukan berdasarkan kriteria yang sudah di tentukan dan di kelompokkan per sekolahnya, kemudian dikalikan dengan bobotnya sehingga diperoleh nilai K1, K2, K3, K4, dan K5. Selanjutnya dilakukan penjumlahan pada nilai K1 hingga K5 untuk mendapatkan skor akhir sekolah, hasilnya tampak pada kolom paling kanan Tabel 5.

Tabel 5. Contoh Tabel Data Agregasi Sesuai Kriteria

Sekolah	Total Pendaftar	K1	Total Lulus	K2	Total Registrasi	K3	Total Batal	K4	Daftar Periode Awal	K5	Skor Akhir
A	176	17,6	155	31	129	38,7	5	-1,5	50	15	100,8
I	89	8,9	78	15,6	61	18,3	10	-3	35	10,5	50,3
D	71	7,1	63	12,6	50	15	2	-0,6	26	7,8	41,9

Skor akhir dari setiap sekolah di sebuah kota dijumlahkan hingga diperoleh skor akhir kota. Skor akhir tiap kota di sebuah provinsi dijumlahkan hingga diperoleh skor akhir provinsi. Tabel 6 adalah contoh perhitungan skor akhir kota dan Provinsi D.I. Yogyakarta.

Tabel 6. Contoh Tabel Data Agregasi Sesuai Kriteria

Provinsi	**Kota**	**Skor Akhir Kota**	**Skor Akhir Provinsi**
D.I. Yogyakarta	Yogyakarta	458,2	664,4
	SLEMAN	125,9	
	Bantul	66,5	
	Gunung Kidul	8,9	
	Kulon Progo	4,9	

Pada tingkat program studi, dibuat kriteria dan bobot yang berbeda untuk menentukan potensi program studi berdasarkan data pendaftar. Bisa dilihat pada Tabel 7 untuk kriteria dan bobot potensi program studi.

Tabel 7. Kriteria dan Bobot untuk Potensi Program Studi

Kode Kriteria	**Kriteria**	**Bobot**
K1	Jumlah Pilihan 1	0,3
K2	Jumlah Pilihan 2	-0,1
K3	Jumlah lulus seleksi	0,2
K4	Jumlah Registrasi	0,3

Pada K1, K3, dan K4, nilai bobtnya diatur linear terhadap potensi program studi dalam data pendaftar. Sebaliknya, K2 diberi bobot negatif karena penurunan prioritas di level program studi.

Selanjutnya dilakukan perhitungan berdasarkan kriteria program studi potensial. Tabel 8 adalah contoh perhitungan skor akhir beberapa sekolah di program studi sistem informasi berdasarkan nilai kriteria dan bobot Tabel 7.

Tabel 8. Contoh Tabel Data Agregasi Sesuai Kriteria Program Studi Sistem Informasi

Sekolah	Pilihan 1	K1	Pilihan 2	K2	Lulus Seleksi	K3	Jumlah Registrasi	K4	Skor Akhir
A	25	7,5	25	-2,5	25	5,0	21	6,3	16,3
T	13	3,9	15	-1,5	13	2,6	13	3,9	8,9
U	8	2,4	7	-0,7	9	1,8	5	1,5	5,0

Skor akhir dari setiap sekolah berdasarkan program studi dijumlahkan pada level kota untuk memperoleh skor akhir kota dan skor akhir provinsi. Tabel 9 adalah contoh perhitungan skor akhir kota dan provinsi di program studi sistem informasi di Provinsi D.I. Yogyakarta.

Tabel 9. Skor Akhir Kota dan Provinsi Berdasarkan Program Studi

Prodi	Provinsi	Kota	Skor Akhir Kota	Skor Akhir Provinsi
Sistem Informasi	D.I. Yogyakarta	Yogyakarta	61,2	87,1
		Sleman	18,2	
		Bantul	8	
		Gunung Kidul	0,1	
		Kulon Progo	-0,4	

Setelah ditemukan skor akhir dari sekolah, kota, dan provinsi, proses selanjutnya untuk mengklaster program studi di masing masing sekolah potensial atau tidak itu sama prosesnya dengan mengelompokkan sekolah, bedanya hanya pada kriteria dan pembobotan ini.

2.5. Normalisasi Skor

Dilakukan normalisasi pada skor akhir menggunakan *StandardScaler* agar nilai skor akhir lebih sebanding. *Standard Scaler* adalah teknik yang menggunakan normalisasi *Z-score* untuk menstandarkan data [13]. *StandardScaler* akan mengubah data agar memiliki rata-rata 0 dan standar deviasi 1 [14]. Proses ini penting karena skor akhir memiliki skala yang berbeda, dan normalisasi memastikan bahwa skor akhir memiliki kontribusi yang setara dalam proses *clustering* [15]. Persamaan untuk normalisasi skor menggunakan *StandardScaler* bisa dilihat pada Persamaan 1.

$$x_i' = \frac{x_i - \bar{x}}{s} \tag{1}$$

Dengan x_i' adalah nilai hasil normalisasi, $\bar{x}$ adalah rata-rata seluruh nilai dalam atribut tersebut, s adalah standar deviasi dari atribut tersebut, x_i adalah nilai data asli suatu atribut sebelum dinormalisasi [13].

Tabel 10 menunjukkan hasil perhitungan normalisasi skor akhir yang dihasilkan dari tahap perhitungan dan pembobotan menggunakan *StandardScaler* menggunakan rumus *StandardScaler* persamaan (1) pada provinsi DIY.

Tabel 10. Normalisasi Skor Akhir Sekolah, Kota, dan Provinsi

Provinsi	Kota	Sekolah	Skor Normalisasi Sekolah	Skor Normalisasi Kota	Skor Normalisasi Provinsi
D.I. Yogyakarta	Yogyakarta	A	24,19	1,93	3,92
		D	9,82		
	Sleman	I	11,87	-0,041	
		J	3,38		
	Bantul	N	3,43	-0,39	
		O	1,24		
	Gunung Kidul	P	-0,003	-0,73	
	Kulon Progo	R	0,02	-0,75	

2.6. Penentuan Jumlah Klaster Optimal dengan Silhouette Coefficient

Pada tahap ini dilakukan analisis untuk menentukan jumlah klaster yang optimal dengan menggunakan metode *Silhouette Coefficient*. *Silhouette Coefficient* merupakan metode yang menggabungan dua metode yaitu metode *cohesion* dan *separation* [16][17]. *Cohesion* bertujuan mengukur seberapa dekat hubungan antara satu objek dengan objek-objek lain dalam klaster yang sama, dan metode *separation* bertujuan mengukur seberapa jauh suatu klaster terpisah dari klaster lain [16]. Nilai *Silhouette Coefficient* yang mendekati +1 menunjukkan bahwa data telah dikelompokkan dengan baik ke dalam klaster, sebaliknya, nilai yang mendekati -1 menunjukkan bahwa pengelompokkan datanya memburuk [18]. Analisis ini dilakukan pada tiga level yaitu level sekolah per kota, level kota per provinsi, dan level pada seluruh provinsi. Penentuan jumlah klaster optimal pada tiap level menggunakan skor normalisasi. Persamaan untuk *silhouette coefficient* ada beberapa tahap. Menghitung rata-rata jarak antara suatu data i dengan semua data lain yang berada dalam klaster yang sama dengan cara memisahkan data i itu sendiri, bisa dilihat pada Persamaan 2.

$$a(i) = \frac{1}{|A|-1} \sum_{j \epsilon A, j \neq i} d(i,j) \tag{2}$$

Dengan $a(i)$ adalah rata-rata jarak antara data i dengan semua data lain pada klaster A, $d(i,j)$ adalah jarak antara data i dengan data j [16].

Rata-rata jarak antara data i dan semua data dalam klaster lain dihitung menggunakan Persamaan 3.

$$d(i,C) = \frac{1}{|C|}\sum_{j\epsilon c} d(i,j) \quad (3)$$

Dengan $d(i,C)$ adalah rata-rata jarak antara data i dan semua data dalam klaster C dan C adalah klaster lain yang berbeda dari klaster A [16].

Menghitung d(i,C) untuk semua C, maka diambil nilai terkecil dengan Persamaan 4:

$$b(i) = min_{c\neq A}\, d(i,C) \quad (4)$$

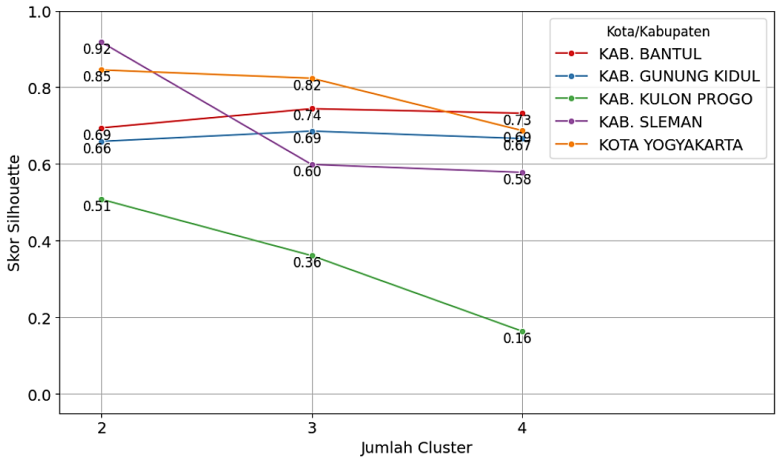
Dengan $b(i)$ adalah nilai jarak rata-rata terkecil antara data i dan klaster lain, Klaster B yang mencapai minimum yaitu (d(i,B) =b(i)) disebut tetangga dari objek (i) [16].

Nilai *Silhouette Coefficient* untuk setiap data i dihitung menggunakan Persamaan 5:

$$s(i) = \frac{(bi-ai)}{\max(ai,bi)} \quad (5)$$

Dengan $s(i)$ adalah Nilai *Silhouette Coefficient* [16].

Jumlah klaster optimal yang didapatkan menggunakan metode *silhouette coefficient* tampak pada Gambar 2 dan Tabel 12. Tetapi dalam sistem jumlah klaster optimal ditampilkan secara langsung sebagai informasi berbasis teks. Dalam analisis ini, dilakukan pengujian dengan jumlah klaster berbeda (2, 3, dan 4) untuk setiap tingkat analisis. Skor *Silhouette* dari masing-masing jumlah klaster dihitung, dan nilai tertinggi dipilih sebagai representasi jumlah klaster terbaik.



Gambar 2. Visualisasi Jumlah Klaster Optimal pada Sekolah di Prov. D.I. Yogyakarta

Gambar 2 merupakan visualisasi penggunaan *silhouette coefficient* untuk mendapatkan hasil optimal jumlah klaster pada sekolah di Provinsi Papua.

Tabel 11. Sampel Jumlah Klaster Optimal pada Sekolah Per-Kota di Prov. Papua

Kota Sekolah	Jumlah Klaster Optimal Sekolah	Skor Silhouette
Biak Numfor	2	0,80
Jayapura	3	0,70
Merauke	2	0,84
Mimika	2	0,80
Nabire	2	0,76
Jayapura	3	0,67

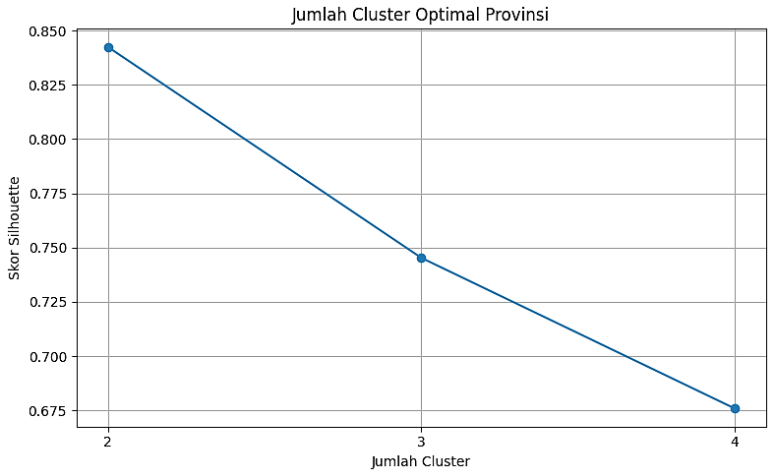
Tabel 11 menunjukkan hasil analisis jumlah klaster optimal dengan skor *silhouette* tertinggi pada sekolah-sekolah dalam tiap kota. Sebagai contoh, untuk sekolah-sekolah di Kota Biak Numfor, sistem merekomendasikan 2 klaster dengan skor *sillhouette* 0,80. Pada hasil penentuan menggunakan *silhouette coefficient* yang dilakukan, jumlah klaster optimal pada sekolah-sekolah dalam tiap kota bervariasi. Hal ini menunjukkan bahwa pola pengelompokan sekolah di masing-masing wilayah memiliki karakteristik yang

berbeda-beda dalam hal distribusi dan kemiripan. Provinsi dengan jumlah asal sekolah pendaftar yang lebih banyak dan tersebar cenderung menghasilkan jumlah klaster optimal yang lebih banyak.

Tabel 12. Sampel Jumlah Klaster Optimal Kota Per-Provinsi

Provinsi	Jumlah Klaster Optimal Kota	Skor Silhouette
D.I. Yogyakarta	4	0,99
Kalimantan Barat	3	0,71
Jawa Timur	2	0,81

Tabel 12 merupakan sampel hasil analisis jumlah klaster optimal pada tingkat kota dalam provinsi. Sebagai contoh, untuk kota-kota di Prov. D.I. Yogyakarta, jumlah klaster optimal nya adalah 4 dengan skor 0,99 yang mendekati 1 menunjukkan pemisahan yang cukup baik antar kelompok.



Gambar 3. Visualisasi Jumlah Klaster Optimal pada Provinsi

Gambar 3 merupakan hasil analisis jumlah klaster optimal pada tingkat seluruh provinsi. Tampak bahwa jumlah klaster sekolah tingkat provinsi yang optimal sebanyak 2 klaster. Pada tingkat provinsi, sekolah dibagi menjadi 2 klaster, yaitu klaster provinsi yang sekolahnya potensial ditindaklanjuti dan klaster provinsi yang sekolahnya tidak potensial.

2.7. *K-Means* Clustering

Metode *K-Means* digunakan untuk mengelompokkan data ke dalam klaster dengan cara meminimalkan variasi atau jarak antar data [19]. Pada sistem, setelah pengguna memilih jumlah klaster optimal, sistem melakukan pengelompokan wilayah dan sekolah target PMB menggunakan algoritma *K-Means Clustering* pada tiga level yaitu sekolah, kota, dan provinsi. *Clustering* pada level sekolah dilakukan menggunakan skor normalisasi sekolah yang dilakukan per kota untuk memastikan hasil yang lebih adil dan representatif. *Clustering* pada level kota menggunakan skor normalisasi kota yang dikelompokkan dalam lingkup provinsi. *Clustering* pada level provinsi, berdasarkan skor normalisasi provinsi. *K-Means* bekerja secara iteratif untuk mencapai konvergensi, dimulai dengan menginisialisasi *centroid* awal dan mengelompokkan data berdasarkan jarak *Euclidean* terdekat [20]. Setelah itu, posisi *centroid* dihitung ulang hingga perubahan antar iterasi sangat kecil atau mencapai batas maksimum (100 iterasi). Pada penelitian ini, tabel yang disajikan pada masing masing *clustering* menggunakan posisi *centroid* yang sudah stabil sebagai hasil akhir dari proses iterasi dari kmeans.fit. Formula jarak *Euclidean* bisa dilihat pada Persamaan 6.

$$dij = \sqrt{(Xi - Xj)^2} \tag{6}$$

Dengan dij adalah jarak *Euclidean* antara titik data ke-i dan titik/*centroid* ke-j, Xi adalah nilai data ke-i (misalnya nilai skor normalisasi sekolah ke-1), Xj adalah nilai *centroid* klaster ke-j (misalnya pusat klaster ke-0 hasil dari *K-Means*).

2.8. Analisis Perbandingan Hasil Sistem dengan Realitas Promosi

Pada tahap akhir, dilakukan analisis perbandingan hasil *clustering* yang dihasilkan oleh sistem dengan data realitas lokasi promosi yang telah dilakukan oleh unit Admisi pada tahun-tahun sebelumnya. Analisis

perbandingan ini dilakukan secara manual diluar sistem. Tujuan dari langkah ini untuk menilai apakah penentuan lokasi promosi yang sudah dilakukan admisi di tahun-tahun sebelumnya itu tepat sasaran.

3. HASIL DAN PEMBAHASAN

3.1. Clustering Level Sekolah

Berikut ini hasil *clustering* pada level sekolah dalam kota dengan memilih yang memiliki jumlah klaster optimalnya 2. Bisa dilihat pada Tabel 13 *centroid* hasil akhir sekolah per kota dari proses iterasi yang sudah stabil pada Kota di Provinsi Yogyakarta.

Tabel 13. Contoh Hasil Akhir *Centroid* Sekolah Per-Kota yang Sudah Stabil pada Kota di Provinsi Yogyakarta

Kota	Nilai *Centroid* Data	
	C0	C1
Yogyakarta	2,64	66,46
Sleman	0,82	59,38
Kulon Progo	-1,19	-0,14

Setelah mengetahui nilai *centroid* kemudian dilakukan perhitungan jarak antar skor normalisasi dengan pusat klaster memakai jarak *Euclidean* dengan formula yang ada pada persamaan (6), hasilnya tampak pada Tabel 14.

Tabel 14. Hasil Perhitungan Jarak Pusat Klaster Sekolah

Nama Sekolah	Skor Normalisasi	Jarak Setiap Data dengan Pusat		Jarak Minimum
		C0	C1	
A	120,99	118,35	54,52	54,52
B	71,10	68,45	4,63	4,63

Berdasarkan hasil jarak minimun yang diperoleh dari Tabel 14, data dikelompokkan ke dalam klaster dan diberikan label potensi pada setiap klaster yang terbentuk. Penentuan label dilakukan berdasarkan nilai *centroid* dari masing-masing klaster. Sistem mebandingkan nilai *centroid* dari klaster 0 dengan klaster 1. Klaster dengan *centroid* tertinggi dianggap mewakili kelompok sekolah potensial. Bisa dilihat pada Tabel 15.

Tabel 15. Pengelompokkan Klaster dan Pemberian Label Sekolah

Kota	Nama Sekolah	Skor Normalisasi	Nilai *Centroid*	Klaster	Label
Yogyakarta	E	24,62	16,81	Klaster 0	Tidak Potensial
	F	9,007			
	A	120,99	96,04	Klaster 1	Potensial
	B	71,10			

3.2. Clustering Level Kota

Clustering pada level kota menggunakan skor normalisasi kota. Pada contoh Tabel 16, *centroid* kota dalam provinsi menggunakan yang jumlah klaster optimalnya 4 yang ada di Provinsi D.I. Yogyakarta dan Papua.

Tabel 16. Contoh Hasil Akhir *Centroid* Kota Per-Provinsi di D.I. Yogyakarta dan Papua yang Sudah Stabil

Provinsi	Nilai *Centroid* Data			
	C0	C1	C2	C3
D.I. Yogyakarta	-0,74	1,93	-0,39	-0,04
Papua	-0,55	2,54	0,22	0,93

Setelah mengetahui nilai k dan pusat klaster kota kemudian dilakukan perhitungan jarak antara skor normalisasi kota dengan nilai *centroid* data memakai jarak *Euclidean* dengan formula yang ada pada persamaan (6), hasilnya tampak pada Tabel 17.

Setelah mengetahui jarak minimum pada Tabel 17, data di kelompokkan ke dalam klasternya dan diberikan label untuk tiap klasternya berdasarkan nilai rata rata data pada tiap klasternya. Disini ada 4 klaster sehingga terbentuk 4 label yaitu paling potensial, potensial, cukup potensial, dan tidak potensial. Bisa dilihat pada Tabel 18.

Dari skor normalisasi pada Tabel 18, skor negatif menandakan bahwa kota itu memiliki nilai dibawah rata-rata, yang berarti tingkat potensi terhadap indikator lebih rendah. Dari kota daerah pendaftar di Provinsi Papua, ada 6 kota yang berpotensi untuk dilakukan promosi, karena skor normalisasinya positif.

Tabel 17. Hasil Perhitungan Jarak Pusat Klaster Kota

Kota	Skor Normalisasi	Jarak Setiap Data dengan Pusat				Jarak Minimum
		C0	C1	C2	C3	
Yogyakarta	1,93	2,67	0	2,32	1,97	0
Sleman	-0,04	0,70	1,97	0,35	0	0
Bantul	-0,39	0,35	2,32	0	0,35	0
Gunung Kidul	-0,73	0,01	2,66	0,34	0,69	0,01
Kulon Progo	-0,75	0,01	2,68	0,36	0,71	0,01
Mimika	2,53	3,09	0,01	2,31	1,59	0,01
Jayapura	0,93	1,49	1,61	0,71	0	0
Merauke	0,39	0,95	2,15	0,17	0,53	0,17
Biak Numfor	0,20	0,76	2,34	0,01	0,72	0,01
Nabire	-0,24	0,31	2,79	0,46	1,17	0,31

Tabel 18. Pengelompokkan Klaster dan Pemberian Label Kota

Provinsi	Kota	Skor Normalisasi	Nilai *Centroid*	Klaster	Label
D.I. YOGYAKARTA	Gunung Kidul	-0,73	-0,74	Klaster 0	Tidak Potensial
	Kulon Progo	-0,75			
	Yogyakarta	1,93	1,93	Klaster 1	Potensial
	Bantul	-0,39	-0,39	Klaster 2	Tidak Potensial
	Sleman	-0,04	-0,04	Klaster 3	Tidak Potensial
PAPUA	Nabire	-0,24	-0,24	Klaster 0	Tidak Potensial
	Jayapura	2,56	2,54	Klaster 1	Potensial
	Mimika	2,53			
	Merauke	0,39	0,29	Klaster 2	Kurang Potensial
	Biak Numfor	0,20			
	Jayapura	0,93	0,93	Klaster 3	Cukup Potensial

3.3. Clustering Level Provinsi

Clustering pada level provinsi, berdasarkan skor normalisasi provinsi. Berikut nilai *centroid* dari data yang didapat dengan ketentuan menyesuaikan jumlah klaster optimal provinsi yaitu 2. Bisa dilihat hasil nilai *centroid* provinsi yang sudah stabil pada Tabel 19.

Tabel 19. Contoh Hasil Akhir *Centroid* Provinsi yang Sudah Stabil

Centroid	Nilai *Centroid*
C0	-0,21
C1	3,61

Dilakukan perhitungan jarak antar skor normalisasi provinsi dengan pusat klaster yang ada pada tabel 19 memakai jarak *Euclidean* dengan formula yang ada pada persamaan (6), hasilnya tampak pada Tabel 20.

Tabel 20. Sampel Hasil Perhitungan Jarak Pusat Klaster Provinsi

Provinsi	Skor Normalisasi	Jarak Setiap Data dengan Pusat		Jarak Minimum
		C0	C1	
D.I. Yogyakarta	3,923	4,143	0,308	0,308
Jawa Tengah	3,306	3,525	0,308	0,308
Papua	1,123	1,342	2,491	1,342

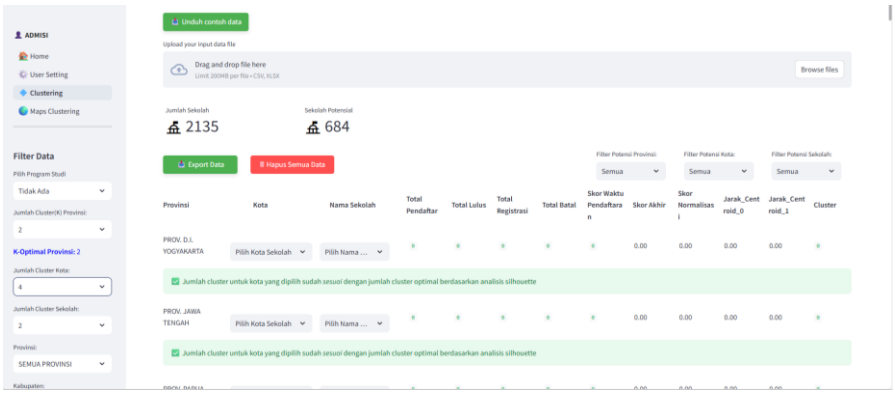
Berdasarkan hasil jarak minimun yang diperoleh dari Tabel 20, data dapat dikelompokkan ke dalam klaster dan diberikan label untuk tiap klasternya, bisa dilihat pada Tabel 21 untuk klaster dan label klaster.

Tabel 21. Pengelompokkan Klaster dan Pemberian Label pada Klaster Provinsi

Klaster	Provinsi	Skor Normalisasi	Nilai Rata-rata (*centroid*)	Label
Klaster 0	Papua	1,123	-0,205	Tidak Potensial
Klaster 1	D.I. Yogyakarta	3,923	3,6145	Potensial
	Jawa Tengah	3,306		

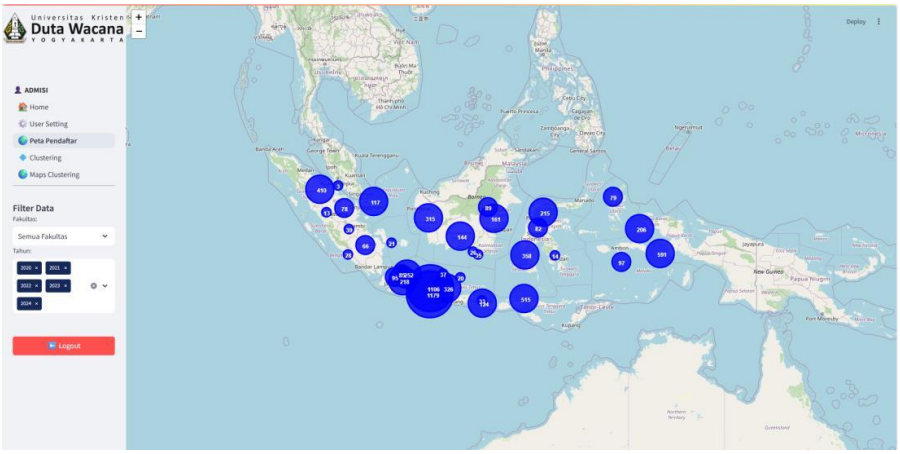
3.4. Sistem Klaster

Gambar 4 adalah hasil implementasi dari sistem klaster potensi dari masing masing sekolah, kota, dan provinsi pendaftar. Unit admisi maupun pimpinan dapat melihat sekolah, daerah yang potensial menjadi target promosi, lengkap dengan level potensialnya. Sistem dapat menampilkan daftar sekolah potensial di sebuah Kota/ Kabupaten maupun dalam bentuk scatter diagram diatas peta daerah tersebut untuk mempermudah visualisasi. Dengan adanya sistem ini, unit admisi dapat dengan mudah menentapkan prioritas sekolah mana saja yang perlu dikunjungi untuk menghemat waktu dan biaya promosi.



Gambar 4. Tampilan *Clustering* Data Pendaftar

Sistem juga memvisualkan hasil *clustering* dalam bentuk peta geografis. Tampilan dapat dilihat pada Gambar 5. Ukuran lingkaran berbanding lurus dengan jumlah pendaftar sehingga bagian admisi dapat dengan mudah memutuskan target promosi potensialnya.



Gambar 5. Tampilan Maps *Clustering* Data Pendaftar

3.5. Diskusi

Pada penelitian ini tingkat potensial sebuah sekolah dihitung dari kumulatif nilai potensial semua siswa yang mendaftar di universitas tersebut selama 2020-2024. Tingkat potensial sebuah kota/kabupaten dihitung dari kumulatif tingkat potensial sekolah di kota/kabupaten tersebut. Nilai potensial sebuah sekolah tidak hanya dihitung dari jumlah siswa pendaftar saja, namun juga mempertimbangkan jumlah siswa yang mengikuti tes seleksi, lulus tes seleksi, melakukan registrasi, serta membatalkan registrasi. Masing-masing komponen tersebut diberi nilai positif dan negatif sesuai kepentingannya. Poin tertinggi diberikan pada siswa yang mendaftar hingga melakukan registrasi. Pengurangan poin diberikan apabila ada siswa yang membatalkan registrasi atau mendaftar tapi tidak mengikuti tes. Metode *K-Means* sangat tepat untuk membuat segmentasi sekolah potensial di universitas yang akan menjadi fokus target promosi. Hasil ini sejalan dengan yang diperoleh [3], maupun [5] di universitas lain. Koefisien *Silhouette* yang diterapkan memberikan jumlah klaster optimal sebanyak 2-4 klaster di tingkat kota/kabupaten, dan 2 klaster di tingkat provinsi.

Pada penelitian [3], data yang disegmentasi difokuskan pada siswa yang melakukan registrasi, tanpa memperhitungkan proses lain. Kebaruan penelitian ini terletak pada proses penentuan nilai potensial sekolah yang memperhitungkan semua proses, mulai pendaftaran hingga registrasi, serta memberikan penalti pada siswa yang melakukan tindakan negatif seperti pembatalan registrasi atau tidak mengikuti tes.

Hasil akhir sistem berupa provinsi, kota/kabupaten dan sekolah potensial promosi. Selain berupa tabel daftar sekolah potensial, sistem juga memberikan visualisasi peta scatter diagram yang ukuran titiknya linier terhadap potensial daerah/sekolah. Hasil luaran sistem selanjutnya dibandingkan dengan data riil tujuan promosi yang telah dilakukan. Dari 125 sekolah yang telah dijangkau oleh kegiatan promosi, hanya 48 sekolah yang sesuai dengan hasil sistem, yaitu sekolah yang dikategorikan sebagai potensial dan cukup potensial. Sistem merekomendasikan total 109 sekolah potensial dan cukup potensial, namun hanya 44,04% dari jumlah tersebut sudah dijangkau. Hasil ini menunjukkan bahwa promosi yang selama ini dilakukan belumlah tepat karena unit admisi tidak menggunakan data history sebagai dasar penentuan sekolah target promosi. Hasil penelitian dapat dikembangkan menjadi sebuah sistem informasi yang terintegrasi dengan sistem penerimaan mahasiswa baru yang sudah ada sehingga data, hasil clustering maupun grafik dapat terupdate secara real time

4. KESIMPULAN

Penerapan sistem memakai algoritma *K-Means clustering* telah berhasil mengelompokkan wilayah, sekolah, dan program studi pada tiap sekolah berdasarkan tingkat potensinya terhadap pendaftaran mahasiswa baru. Sistem ini secara dinamis dapat menentukan jumlah klaster berdasarkan evaluasi kualitas klaster menggunakan *silhouette coefficient*, sehingga dapat merekomendasikan jumlah klaster optimal. Dari total 34 provinsi yang dianalisis, terdapat 8 provinsi potensial di luar D.I. Yogyakarta dan Jawa Tengah. Meskipun beberapa wilayah secara keseluruhan terklasifikasi sebagai "tidak potensial", sistem mampu mengidentifikasi sekolah-sekolah tertentu yang memiliki potensi tinggi di dalam wilayahnya. Hasil analisis perbandingan keluaran sistem dengan realitas promosi menunjukkan bahwa 48 sekolah yang dijangkau promosi sesuai dengan kategori potensial dan cukup potensial menurut sistem. Dari 109 sekolah yang direkomendasikan sebagai potensial dan cukup potensial, 44,04% yang menjadi target promosi. Pengembangan penelitian dengan membandingkan metode Kmeans dengan metode lain sangatlah dimungkinkan agar diperoleh hasil yang lebih *reliable*

REFERENSI

[1] R. Toro and dan Sri Lestari, "Perbandingan Algoritma Klasifikasi Untuk Penentuan Lokasi Promosi Penerimaan Mahasiswa Baru Pada IIB Darmajaya Lampung Comparison of Classification Algorithm to Determine the Location of New Student Admission Promotion on IIB Darmajaya Lampung," *Februari*, vol. 22, no. 1, pp. 223–234, 2023.

[2] 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," 2022.

[3] H. Hairani, D. Susilowati, I. Puji Lestari, K. Marzuki, and L. Z. A. Mardedi, "Segmentasi Lokasi Promosi Penerimaan Mahasiswa Baru Menggunakan Metode RFM dan K-Means Clustering," *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 21, no. 2, pp. 275–282, Mar. 2022, doi: 10.30812/matrik.v21i2.1542.

[4] I. Fauzi, "Penerapan Data Mining Untuk Menentukan Lokasi Promosi Sekolah Dengan Metode K-Means Clustering (Studi Kasus: Smp Islam Al Syukro Universal)," 2025.

[5] D. Vernanda, N. N. Purnawan, and T. H. Apandi, "Penerapan Fuzzy C Means Untuk Menentukan Target Promosi Penerimaan Mahasiswa Baru," *Jurnal Ilmiah Ilmu dan Teknologi Rekayasa*, vol. 2, no. 2, Feb. 2020, doi: 10.31962/jiitr.v2i2.63.

[6] D. M. Putri, A. S. Ilmananda, and N. Prisanta, "Penggunaan Algoritma K-Means dan K-Medoids untuk Pengembangan Strategi Promosi Penerimaan Mahasiswa Baru," *SMATIKA JURNAL*, vol. 14, no. 02, pp. 388–398, Dec. 2024, doi: 10.32664/smatika.v14i02.1474.

[7] I. Ariati, R. Nugraha Norsa, L. Akhsan, and J. Heikal, "Segmentasi Pelanggan Menggunakan K-Means Clustering Studi Kasus Pelanggan Uht Milk Greenfield," *Jurnal Ilmiah Indonesia*, vol. 2023, no. 7, pp. 629–643, 2023, doi: 10.36418/cerdika.xxx.

[8] R. Nur Aulia, R. Taufiqillah, and P. Dewi, "Analysis Customer Segmentation on Individual Life Insurance in Kalimantan Province Using K-Means Clustering with SPSS," *Puspita Dewi INNOVATIVE: Journal Of Social Science Research*, vol. 4, pp. 4462–4476, 2024.

[9] N. Trezandy Lapatta, R. Ardiansyah, and D. Shinta Angreni, "Donor Segmentation Analysis Using the RFM Model and K-Means Clustering to Optimize Fundraising Strategies," 2024. [Online]. Available: http://jurnal.polibatam.ac.id/index.php/JAIC

[10] Y. Elda, S. Defit, Y. Yunus, and R. Syaljumairi, "Klasterisasi Penempatan Siswa yang Optimal untuk Meningkatkan Nilai Rata-Rata Kelas Menggunakan K-Means," *Jurnal Informasi dan Teknologi*, pp. 103–108, Sep. 2021, doi: 10.37034/jidt.v3i3.130.

[11] H. Rumapea, "Deteksi Kemiripan Artikel Melalui Keywords Dengan Metode Fuzzy String Matching Dalam Natural Language Processing," *METHOMIKA Jurnal Manajemen Informatika dan Komputerisasi Akuntansi*, vol. 5, no. 1, pp. 60–66, Apr. 2021, doi: 10.46880/jmika.Vol5No1.pp60-66.

[12] H. Nur Hanani, H. Jayadianti, H. Cahya Rustamaji, and U. Pembangunan Nasional Veteran Yogyakarta, "Fuzzy String Matching for Semi-Automatication of Words with Jaro Winkler Distance Algorithm on Microsoft Word Documents Fuzzy String Matching untuk Semi-Otomatisasi Pencocokan Kata dengan Algoritma Jaro Winkler Distance pada Dokumen Microsoft Word," pp. 145–160, 2021, [Online]. Available: www.myvocabulary.com

[13] L. B. V. de Amorim, G. D. C. Cavalcanti, and R. M. O. Cruz, "The choice of scaling technique matters for classification performance," Dec. 2022, doi: 10.1016/j.asoc.2022.109924.

[14] M. Arif, maruf Setiawan, A. Dwi Hartono, M. Arif Ma, and ruf Setiawan, "Menggunakan Metode Machine Learning Untuk Memprediksi Nilai Mahasiswa Dengan Model Prediksi Multiclass," *Jurnal Informatika: Jurnal pengembangan IT*, vol. 10, no. 1, pp. 190–204, 2025, doi: 10.30591/jpit.v9ix.xxx.

[15] F. D. Wahyuningtyas, A. Arafat, A. Stiawan, and D. Rolliawati, "Komparasi Algoritma Hierarchical, K-Means, dan DBSCAN pada Analisis Data Penjualan Melalui Facebook," *Explore: Jurnal Sistem Informasi dan Telematika*, vol. 14, no. 1, p. 7, Jun. 2023, doi: 10.36448/jsit.v14i1.2931.

[16] S. Paembonan, H. Abduh, and K. Kunci, "Penerapan Metode Silhouette Coeficient Untuk Evaluasi Clutering Obat Clustering; K-means; Silhouette coeficient," 2021. [Online]. Available: https://ojs.unanda.ac.id/index.php/jiit/index

[17] Y. I. Kurniawan, P. R. Anugrah, R. M. Sugihono, F. A. Abimanyu, and L. Afuan, "Pengelompokan Prioritas Negara Yang Membutuhkan Bantuan Menggunakan Clustering K-Means dengan Elbow dan Silhouette," *Jurnal Pendidikan dan Teknologi Indonesia (JPTI)*, vol. 3, no. 10, pp. 455–463, 2023, doi: 10.52436/1.jpti.343.

[18] "5. Pengelompokkan Data Kemiskinan Provinsi Jawa Barat Menggunakan Algoritma K-Means dengan Silhouette Coefficient 29-35," pp. 29–35, 2022, doi: 10.38204/tematik.v9i1.921.

[19] Febby Arisca Zurfani, Sawaluddin, Mardiningsih, and Muhammad Romi Syahputra, "Analisis Metode Clustering K-Means pada Zonasi Daerah Terdampak Banjir di Kota Medan dengan Evaluasi Silhouette Coefficient," *Algoritma : Jurnal Matematika, Ilmu pengetahuan Alam, Kebumian dan Angkasa*, vol. 2, no. 6, pp. 170–181, Nov. 2024, doi: 10.62383/algoritma.v2i6.270.

[20] W. N. Purba and R. Hartanto, "Perbandingan Penerapan Algoritma K-Means Dan Fuzzy C-Means Dalam Analisis Clustering Terhadap Pergerakan Harga Historis Saham Bank Rakyat Indonesia," *Jurnal Teknik Informasi dan Komputer (Tekinkom)*, vol. 7, no. 2, p. 865, Dec. 2024, doi: 10.37600/tekinkom.v7i2.1214.