

PERBANDINGAN ALGORITME K-MEANS CLUSTERING DAN FUZZY C-MEANS UNTUK PENENTUAN STRATEGI PROMOSI UNIVERSITAS KATOLIK DARMA CENDIKA

Christopher Kevin Herijanto, Yosefina Finsesia Riti
Teknik Informatika, Universitas Katolik Darma Cendika
Jl. Dr. Ir. H. Soekarno No.201, Surabaya
Christopher.herijanto@student.ukdc.ac.id

ABSTRAK

Universitas Katolik Darma Cendika, sebagai salah satu Perguruan Tinggi Swasta (PTS) di Kota Surabaya, secara rutin melibatkan diri dalam kegiatan promosi setiap tahunnya. Proses promosi ini melibatkan investasi biaya yang signifikan, maka dari itu hasil yang optimal diharapkan dapat dicapai dari upaya tersebut. Keberhasilan promosi dapat dinilai dari jumlah mahasiswa yang mendaftar pada awal tahun ajaran baru. Penentuan strategi promosi bukanlah tugas yang sederhana, mengingat perlu mempertimbangkan beberapa aspek tertentu. Untuk membantu dalam menentukan strategi promosi di Universitas Katolik Darma Cendika, pendekatan data mining dapat digunakan. Probabilitas statistika memungkinkan penemuan pola baru dari kumpulan data yang tersimpan. Salah satu teknik dalam probabilitas statistika adalah clustering, dan algoritme yang digunakan di sini adalah K-Means dan Fuzzy C-Means. Setiap algoritme memiliki kinerja dan akurasi yang berbeda tergantung pada karakteristik data yang digunakan. Dengan melakukan perbandingan antara 2 algoritme tersebut, penelitian ini dapat memberikan hasil algoritme yang lebih baik untuk digunakan sebagai strategi promosi dengan pengujian Silhouette Coefficient dan pengujian jumlah iterasi. Pengujian *Silhouette Coefficient* menunjukkan bahwa algoritme *K-Means* lebih unggul dibandingkan *Fuzzy C-Means*. Nilai *Silhouette Coefficient* rata-rata algoritme *K-Means* sebesar 0,2833. Sedangkan nilai *Silhouette Coefficient* rata-rata algoritme *Fuzzy C-Means* sebesar -0.1456.

Kata kunci : *K-Means, Fuzzy C-Means, clustering, probabilitas statistika, strategi promosi*

1. PENDAHULUAN

Universitas Katolik Darma Cendika merupakan salah satu perguruan tinggi swasta (PTS) terkemuka di Kota Surabaya. Setiap tahunnya Universitas Katolik Darma Cendika aktif mengikuti kegiatan promosi untuk menarik minat calon mahasiswa baru. Selama ini, hasil promosi universitas di Universitas Katolik Darma Cendika masih kurang maksimal dengan lokasi dan target promosi yang kurang efektif. Ini dapat dibuktikan dengan data mahasiswa baru yang mulai menurun pada tahun 2023. Maka dari itu, Strategi promosi harus dipertimbangkan dengan cermat, terutama ketika menentukan lokasi dan target promosi supaya tidak salah sasaran. Pertimbangan diberikan kepada daerah asal dan latar belakang akademik mahasiswa yang terdaftar.

Penerapan data mining dan probabilitas statistika merupakan pendekatan yang sangat relevan untuk menentukan strategi promosi Universitas Katolik Darma Cendika. Data mining memungkinkan pengolahan data yang ada khususnya data mahasiswa Universitas Katolik Darma Cendika. Data mining mengekstraksi informasi yang berguna, yang sebelumnya tidak diketahui, namun dapat dipahami dari sejumlah besar data dengan tujuan membuat keputusan bisnis yang penting[1]. Teknik probabilitas statistika digunakan untuk mengidentifikasi pola tersembunyi dan memprediksi tren masa depan. Data mining adalah proses yang menggunakan teknik probabilitas statistika, probabilitas memberikan dasar matematis yang kuat untuk analisis data mining[1].

Diantara berbagai teknik dalam probabilitas statistika, metode clustering adalah salah satu pendekatan, dan algoritme K-Means, Hierarchical clustering, Naive Bayes, Fuzzy C-Means adalah bagian dari metode clustering ini. Pengolahan data yang dilakukan pada 148 record data dari masing - masing nilai data pretest siswa-siswa SMA Islam Ta'alumul HudaBumiayu tahun ajaran 2018/2019 menggunakan *algoritme* K-means dan *algoritme* Naive Bayes didapatkan akurasi sebesar 80% dengan predikat good classification[2]. Berdasarkan hasil dari penelitian yang telah dilakukan maka dapat ditarik kesimpulan algoritme Naive Bayes maupun algoritme Fuzzy C-means dapat membantu pengelompokan data kemiskinan dengan akurasi algoritme Naive Bayes sebesar 74% sedangkan akurasi algoritme Fuzzy C-means sebesar 67%[3]. Implementasi *algoritme* k-means clustering telah memudahkan clusterisasi desa tervaksinasi Covid-19 pada Kecamatan Ujung Padang. Penerapan Fuzzy C-Means dalam Mengelompokkan Kabupaten/Kota Berdasarkan Fasilitas Pelayanan Kesehatan Di Jawa Timur yang memiliki nilai Silhouette Coefficient pada hasil cluster adalah 0,695, sehingga cluster dikatakan sudah baik[4]. Berdasarkan analisis dan pembahasan yang telah dilakukan pada perbandingan metode klasifikasi di kota Tegal untuk naive bayes dan regresi logistik pada status pengguna kb tahun 2014, Pengklasifikasian menggunakan metode Regresi Logistik Biner lebih baik dibandingkan dengan metode Naive Bayes[5]. Metode K-Means dan Fuzzy C-Means adalah dua teknik yang umum digunakan

dalam analisis data dan pengelompokan (clustering). Keduanya memiliki hasil yang berbeda tergantung pada karakteristik data dan tujuan analisis, maka dari itu penelitian ini bertujuan untuk menganalisis algoritme yang lebih baik untuk melakukan clustering data mahasiswa dan menggunakan hasil dari clustering untuk menentukan strategi promosi yang optimal.

Algoritme pengelompokan K-means clustering didasarkan pada jarak terpendek dan membagi data Anda menjadi beberapa kelompok berdasarkan kriteria numerik tertentu[6], [7]. Dalam konsep Fuzzy C-Means, status keanggotaan suatu objek atau data tidak dinyatakan secara pasti dengan nilai 1 (sebagai anggota) atau nilai 0 (bukan anggota), melainkan ditentukan dengan suatu nilai derajat keanggotaan yang memiliki rentang nilai antara 0 hingga 1[8]. Melakukan analisis perbandingan terhadap algoritme K-Means dan Fuzzy C-Means memiliki tujuan untuk Universitas Katolik Darma Cendika untuk memiliki kebermanfaatan, terkhususnya dalam mengidentifikasi algoritme yang lebih unggul untuk melakukan proses clustering data mahasiswa dan menentukan strategi promosi. Informasi ini dapat digunakan untuk mengembangkan strategi promosi yang lebih tepat sasaran dan efektif, mengoptimalkan alokasi sumber daya promosi, dan meningkatkan jumlah mahasiswa baru. Dengan adanya penelitian ini, Pihak PMB Universitas Katolik Darma Cendika dapat terbantu dalam menentukan strategi promosi yang lebih tepat.

2. TINJAUAN PUSTAKA

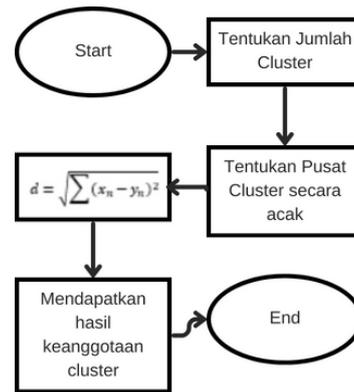
2.1. Data yang Digunakan dan Tools

Tools yang digunakan berupa Google Colab dengan Google Compute Engine dan Python v3. Pengujian ini juga menggunakan beberapa library python berupa numpy, matplotlib, dan scikit-learn untuk menunjang hasil yang lebih maksimal.

Data yang digunakan berupa Dataset yang sudah preprocessed bersumber dari rekapitulasi data penerimaan MABA (Mahasiswa Baru) tahun akademik 2017 - 2023 Universitas Katolik Darma Cendika, dimana mahasiswa yang di terima menjadi mahasiswa baru pada Universitas ini sebanyak 1469 data MABA.

2.2. K-Means Clustering

K-Means Clustering merupakan algoritme pengelompokan non-hierarkis yang membagi data menjadi satu atau lebih kelompok dengan kesamaan karakteristik[9]. Algoritme ini melakukan partisi data sehingga objek dengan karakteristik serupa dikelompokkan bersama, sedangkan objek dengan karakteristik yang berbeda dikelompokkan secara terpisah[10]. Algoritme K-Means termasuk dalam kategori fungsi clustering, yang merujuk pada proses pengelompokan data, observasi, atau kasus berdasarkan kemiripan objek yang sedang diinvestigasi. Langkah-langkah melakukan clustering dengan metode K-Means adalah sebagai berikut[11]:



Gambar 1. Proses Algoritme K-Means

- a. Tentukan jumlah cluster k.
- b. Inisialisasikan pusat-pusat cluster k, yang sering dilakukan secara acak. Pusat-pusat cluster diberikan nilai awal dengan menggunakan angka-angka acak.
- c. Tetapkan semua data atau objek ke cluster terdekat. Kedekatan antara dua benda diukur dengan jarak antara keduanya. Kedekatan data dengan suatu cluster ditentukan oleh jarak antara data dan pusat cluster. Pada tahap ini, kita perlu menggunakan teori jarak Euclidean untuk menghitung jarak dari setiap datum ke setiap pusat cluster dengan rumus sebagai berikut: (1).

$$D(i, j) = \sqrt{(X1i - Y1j)^2 + (X2i - Y2j)^2 + \dots + (Xki - Ykj)^2} \tag{1}$$

Dimana:

D (i,j) = Jarak data ke i ke pusat cluster j

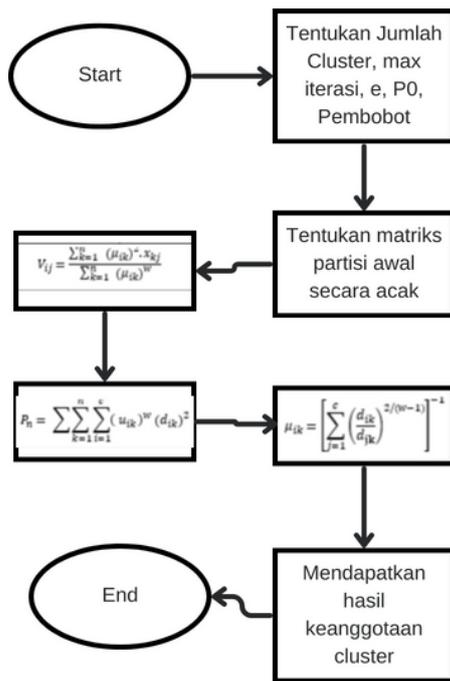
Xki = Data ke i pada atribut data ke k

Xkj = Titik pusat ke j pada atribut ke k

- d. Menghitung ulang pusat cluster dengan mempertimbangkan keanggotaan cluster saat ini. Pusat cluster dapat dihitung sebagai rata-rata seluruh data dalam cluster, atau sebagai median cluster jika diinginkan. Rata-rata bukanlah satu-satunya ukuran yang dapat Anda gunakan.
- e. Tetapkan setiap objek ke cluster dengan pusat terdekat. Jika pusat cluster tidak berubah mak proses clustering dianggap selesai. Jika tidak, kembali ke langkah 3 hingga tidak ada perubahan yang terjadi pada pusat klaster atau jumlah iterasi maksimum yang dikonfigurasi tercapai.

2.3. Fuzzy C-Means

Jim Bezdek memperkenalkan Fuzzy C-Means untuk pertama kalinya pada tahun 1981. Pendekatan ini merupakan perkembangan dari algoritme K-Means dengan mengintegrasikan prinsip fuzzy ke dalamnya (Luthfi, 2007). Berbeda dengan K-Means, Fuzzy C-Means memungkinkan data yang di-cluster menjadi anggota dari setiap cluster yang ada. Hubungan antara data dan cluster ditentukan oleh nilai keanggotaannya yang bervariasi dalam rentang 0 hingga 1[12], [13]. Proses alur dari cara bekerja Fuzzy C-Means dapat dilihat pada Gambar 2.



Gambar 2. Proses *Algoritme Fuzzy C-Means*

2.4. Teknik Evaluasi

Salah satu teknik evaluasi untuk menilai kualitas suatu kluster adalah menggunakan *Silhouette Coefficient*, seperti yang dijelaskan oleh Furqon pada tahun 2016. Metode ini memanfaatkan perhitungan jarak *Euclidean* dalam evaluasinya[12]. Secara ringkas, rumus *Silhouette Coefficient* dapat dinyatakan dalam persamaan (2).

$$Si = \frac{(bi-ai)}{\max(bi,ai)} \tag{2}$$

Dimana:

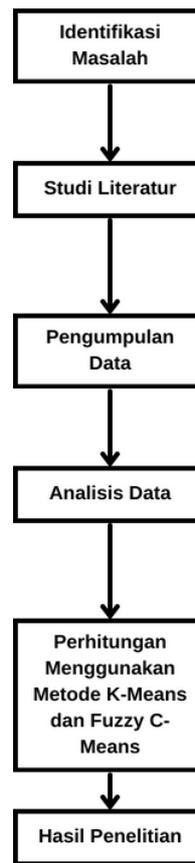
- Si = *Silhouette Coefficient*
- ai = jarak rata-rata objek i ke seluruh objek dalam cluster yang sama
- bi = nilai minimum rata-rata jarak objek i ke objek lain dalam cluster berbeda

Nilai *Silhouette Coefficient* adalah rentang (-1) sampai 1. Semakin tinggi nilainya, semakin tinggi kualitasnya.

3. METODE PENELITIAN

Tahapan dalam menyelesaikan penelitian perbandingan algoritme K-Means dan Fuzzy C-Means untuk penentuan strategi promosi universitas katolik darma cendika, dapat dilihat pada Gambar 3.

a. Identifikasi Permasalahan
 Mengidentifikasi masalah merupakan langkah yang esensial untuk memahami akar permasalahan, menentukan penyebabnya, dan mencari solusi yang tepat. Dalam penelitian ini, fokus pada identifikasi lokasi promosi penerimaan siswa baru di Universitas Katolik Darma Cendika.



Gambar 3. Tahapan Penelitian

- b. Studi Pustaka
 Pada tahap ini, dilakukan pencarian literatur dari berbagai sumber, seperti buku dan internet, guna memperkaya konsep dan teori yang mendukung penelitian. Hal ini bertujuan untuk membangun dasar pengetahuan yang solid dan relevan.
- c. Identifikasi Permasalahan
 Mengidentifikasi masalah merupakan langkah yang esensial untuk memahami akar permasalahan, menentukan penyebabnya, dan mencari solusi yang tepat. Dalam penelitian ini, fokus pada identifikasi lokasi promosi penerimaan siswa baru di Universitas Katolik Darma Cendika.
- d. Studi Pustaka
 Pada tahap ini, dilakukan pencarian literatur dari berbagai sumber, seperti buku dan internet, guna memperkaya konsep dan teori yang mendukung penelitian. Hal ini bertujuan untuk membangun dasar pengetahuan yang solid dan relevan.
- e. Pengumpulan Data
 Langkah ini melibatkan proses pengumpulan data mahasiswa di Universitas Katolik Darma Cendika yang dilakukan dengan wawancara dengan pihak PMB Universitas Katolik Darma Cendika dan didapatkan 1469 data mahasiswa.
- f. Analisis Data
 Dalam melakukan penelitian, analisis data menjadi kunci untuk memahami informasi dengan lebih baik. Proses ini membantu dalam menemukan solusi untuk permasalahan penelitian, khususnya

terkait keputusan pegawai dalam menentukan lokasi promosi.

- g. Penerapan Metode K-Means dan Fuzzy C-Means Tahap ini menjadi krusial karena melibatkan penggunaan *algoritme* K-Means dan Fuzzy C-Means untuk mengelompokkan data berdasarkan jarak terdekat dengan centroid. Metode ini digunakan untuk mendapatkan perbandingan hasil *algoritme* terkait penentuan lokasi promosi penerimaan siswa baru yang belum sesuai dengan cluster yang tepat.

4. HASIL DAN PEMBAHASAN

Dalam penelitian ini, evaluasi dilakukan melalui pengujian hasil jumlah iterasi dan penilaian Silhouette Coefficient. Parameter input yang digunakan terdiri dari:

- K = 6;
- Maks iterasi = 100;
- P0 = 0;
- W = 2;
- $\Theta = 0.01$;

Ket:

- a. K adalah jumlah cluster yang ingin dibentuk dalam *algoritme* K-Means dan Fuzzy C-Means.
- b. Maks Iterasi dalam konteks ini adalah proses perbaikan pusat cluster (centroid) yang dilakukan berulang kali hingga *algoritme* mencapai konvergensi atau mencapai batas iterasi maksimum.
- c. P0 merujuk ke inisialisasi atau kondisi awal dari pusat cluster dalam *algoritme*.
- d. W adalah bobot atau faktor pembobotan dalam *algoritme* Fuzzy C-Means. Dalam Fuzzy C-Means, nilai W (sering disebut sebagai parameter fuzziness atau parameter weighting exponent) menentukan derajat keanggotaan data terhadap berbagai cluster.
- e. Θ adalah parameter yang menunjukkan nilai threshold atau ambang batas konvergensi. Dalam konteks *algoritme* clustering, ini biasanya mengacu pada perubahan maksimum yang diizinkan antara iterasi pusat cluster.

Pengujian dilaksanakan sebanyak 10 kali dengan menggunakan 1469 data. Pengujian ini menggunakan program python dengan memanfaatkan fitur pandas, numpy, sklearn.metrics, sklearn.preprocessing, dan skfuzzy. Pendekatan ini diambil karena setiap *algoritme* memiliki unsur acak yang dapat memengaruhi hasil, sehingga pengulangan pengujian diperlukan untuk mendapatkan gambaran yang lebih konsisten. Pada penelitian ini menggunakan data sampling asal daerah mahasiswa yang dapat dilihat pada Tabel 1, asal sekolah mahasiswa yang dapat dilihat pada Tabel 2, dan jurusan mahasiswa yang dapat dilihat pada Tabel 3.

Tabel 1. Asal Daerah Mahasiswa

No	Asal Daerah	Jumlah Mahasiswa	Inisialisasi
1	Surabaya	953	1
2	Sidoarjo	372	2
3	Manggarai	40	3
4	Mojokerto	33	4
5	Gresik	29	5
6	Manggarai Barat	15	6
7	Kediri	15	7
8	Manggarai Timur	12	8

Tabel 2. Asal Sekolah Mahasiswa

No	Asal Sekolah	Jumlah Mahasiswa	Inisialisasi
1	SMA	776	1
2	SMK	297	2
3	SMAN	175	3
4	JURUSAN LAINNYA	159	4
5	SMKN	62	5

Tabel 3. Jurusan Mahasiswa

No	Jurusan Mahasiswa	Jumlah Mahasiswa	Inisialisasi
1	Akuntansi	408	1
2	Man Pemasaran	387	2
3	Hukum	176	3
4	Man Perhotelan	173	4
5	Industri	104	5
6	Informatika	98	6
7	Arsitektur	67	7
8	Akupunktur	56	8

4.1. Analisa Hasil K-Means

Langkah pertama dari proses K-Means Clustering merupakan penentuan jumlah cluster untuk pengujian ini yang berjumlah 3 cluster. Setelah itu, pada iterasi – 1 dilakukan penentuan nilai Centeroid / titik pusat awal yang digunakan pada penelitian diambil secara random dari data sampling.

Tabel 4. Centroid / Titik Pusat Awal

No	Sampling	Daerah	Sekolah	Jurusan	Cluster
1	3	1	1	3	C1
2	7	1	1	7	C2
3	9	4	3	8	C3

Adapun hasil perhitungan jarak data ke – 1 sampai data ke – 10 terhadap centeroid / titik pusat awal C1, C2 dan C3 dapat dilihat pada Tabel 5.

Tabel 5. Hasil Perhitungan Jarak Data ke 1 sampai 10

No	Sampling	C1	C2	C3	Hasil
1	S1	2	6	7.874007874	C1
2	S2	3	4.123105626	4.123105626	C1
3	S3	0	4	6.164414003	C1
4	S4	6	6	3.741657387	C3
5	S5	7.874007874	9.695359715	8.124038405	C1
6	S6	6.782329983	6.164414003	3.741657387	C3
7	S7	2.236067977	6.08276253	7.549834435	C1
8	S8	4	0	3.741657387	C2
9	S9	6.164414003	3.741657387	0	C3
10	S10	5.916079783	7.681145748	6.403124237	C1

Pada iterasi 2, seluruh data ditempatkan pada cluster terdekat. Menghitung ulang pusat cluster menggunakan keanggotaan cluster saat ini. Pusat cluster adalah rata-rata seluruh data/objek dalam suatu cluster tertentu. Anda juga dapat menggunakan median cluster. Oleh karena itu, mean (rata-rata) bukanlah satu-satunya ukuran yang dapat digunakan. Hitung pusat cluster baru.

- Untuk cluster 1 terdapat 6 data
- Untuk cluster 2 terdapat 1 data.
- Untuk cluster 3 terdapat 3 data.

Tabel 6. Centroid / Titik Pusat Awal Iterasi - 2

No	Cluster	K1	K2	K3
1	C1	3.504359272	6.263728936	6.706420763
2	C2	4	0	3.741657387
3	C3	6.315581329	5.302023797	2.494438258

Berdasarkan Gambar 4, didapatkan data hasil clustering yang dilakukan dengan algoritme K-Means berupa cluster 1 dengan 5 data, cluster 2 dengan 2 data dan cluster 3 dengan 3 data. Dapat terlihat juga untuk hasil rata-rata pusat cluster dari setiap 3 cluster yang sudah ditentukan. Jumlah iterasi yang dilakukan juga tertera yaitu 4 kali. Skor Silhouette Coefficient K-Means dari percobaan tersebut juga ditampilkan yaitu 0.35082.

```

Hasil K-Means Clustering:
No Sampling Cluster_KMeans
0 1 S1 Distance_C1
1 2 S2 Distance_C1
2 3 S3 Distance_C1
3 4 S4 Distance_C3
4 5 S5 Distance_C2
5 6 S6 Distance_C3
6 7 S7 Distance_C1
7 8 S8 Distance_C1
8 9 S9 Distance_C3
9 10 S10 Distance_C2

Pusat Cluster K-Means:
['C1': array([1.6, 1.4, 3.2]), 'C2': array([7. , 4. , 1.5]), 'C3': array([5.33333333, 3.33333333, 6.33333333])]
Jumlah Iterasi K-Means: 4
Skor Silhouette K-Means: 0.3508178392744172
    
```

Gambar 4. Hasil Akhir K-Means

4.2. Analisa Hasil Fuzzy C-Means

Berdasarkan Gambar 5, didapatkan data hasil clustering yang dilakukan dengan algoritme Fuzzy C-Means berupa cluster 1 dengan 4 data dan cluster 3 dengan 6 data. Dapat terlihat juga untuk hasil pusat cluster berisi 10 data pusat cluster dari setiap 3 cluster yang sudah ditentukan. Jumlah iterasi yang dilakukan juga tertera yaitu 11 kali. Skor Silhouette Coefficient Fuzzy C-Means dari percobaan tersebut juga ditampilkan yaitu 0.18435.

Hasil Fuzzy C-Means Clustering:

```

No Sampling Cluster_Fuzzy
0 1 S1 1
1 2 S2 3
2 3 S3 1
3 4 S4 3
4 5 S5 3
5 6 S6 1
6 7 S7 1
7 8 S8 3
8 9 S9 3
9 10 S10 3
    
```

Pusat Cluster Fuzzy C-Means:

```

[[[0.95937234 0.25234286 0.94457164 0.10127599 0.04970107 0.12429172
0.94572491 0.3666713 0.03739964 0.00631709]
[0.02205205 0.5020105 0.03605792 0.33646374 0.06627992 0.64332272
0.02850375 0.51770529 0.0146521 0.00956341]
[0.01857561 0.24564664 0.01937043 0.56226027 0.884019 0.23238556
0.02577134 0.11562341 0.04794826 0.9841195 ]]
Jumlah iterasi Fuzzy C-Means: 11
Skor Silhouette Fuzzy C-Means: 0.18434966382627668
    
```

Gambar 5. Hasil Akhir Fuzzy C-Means

4.3. Evaluasi Hasil K-Means dan Fuzzy C-Means

Setelah melakukan pengujian, kedua algoritme K-Means dan Fuzzy C-Means menghasilkan 3 cluster dengan hasil yang berbeda.

Pengujian Silhouette Coefficient dilaksanakan untuk mengevaluasi mutu suatu kelompok. Silhouette Coefficient diperoleh dengan membandingkan jarak antar data dalam satu kelompok dengan jarak data ke kelompok lainnya. Perbandingan nilai Silhouette Coefficient dapat ditemukan pada Tabel 7.

Tabel 7. Hasil Pengujian Silhouette Coefficient

Pengujian Silhouette		
Percobaan	K-Means	FCM
1	0.3508	-0.09979
2	0.3288	-0.2253
3	0.3495	-0.0139
4	0.3547	-0.205
5	0.3571	-0.151
6	-0.004	-0.1507
7	0.2205	-0.1419
8	0.3508	-0.1404
9	0.3036	-0.1491
10	0.2212	-0.1792
Average	0.2833	-0.1456

Pengujian Silhouette Coefficient menunjukkan bahwa algoritme K-Means lebih unggul dibandingkan Fuzzy C-Means. Nilai Silhouette Coefficient rata-rata algoritme K-Means sebesar 0,2833. Sedangkan nilai Silhouette Coefficient rata-rata algoritme Fuzzy C-

Means sebesar -0.1456. Perbandingan hasil jumlah iterasi dapat ditemukan pada Tabel 8.

Tabel 8. Hasil Jumlah Iterasi

Jumlah Iterasi		
Percobaan	K-Means	FCM
1	7	10
2	95	12
3	100	13
4	100	11
5	74	8
6	100	19
7	100	10
8	66	13
9	100	11
10	100	10
Average	84.2	11.7

Hasil jumlah iterasi dari setiap algoritme menunjukkan bahwa algoritme *Fuzzy C-Means* lebih unggul dibandingkan *K-Means*. Jumlah iterasi rata-rata algoritme *Fuzzy C-Means* sebesar 11.7. Sedangkan jumlah iterasi rata-rata algoritme *K-Means* sebesar 84.2 yang menunjukkan bahwa *K-Means* membutuhkan waktu lebih lama untuk mendapatkan hasil *clustering* dibandingkan *Fuzzy C-Means*.

Berdasarkan pengujian nilai *Silhouette Coefficient* yang lebih baik, yang mengartikan bahwa hasil dari algoritme tersebut lebih baik. Hasil dari Algoritme *K-Means* lebih unggul untuk dijadikan strategi promosi untuk Universitas Katolik Darma Cendika dalam mendapatkan hasil yang lebih valid dibandingkan dengan Algoritme *Fuzzy C-Means*. Oleh karena itu, strategi promosi yang berdasarkan hasil *K-Means* mungkin memberikan segmentasi target yang lebih jelas dan valid. Penggunaan *K-Means* dapat memberikan keuntungan dalam memahami karakteristik mahasiswa dan menyusun strategi promosi yang lebih terarah.

Berdasarkan hasil proses *clustering* menggunakan *K-Means* maka didapatkan sebuah informasi untuk penentuan promosi sebagai berikut:

a. *Cluster 1* (C1)

Cluster 1 didominasi asal daerah dari Kota Surabaya, Untuk latar belakang pendidikan sekolah berasal SMA Swasta dan jurusan yang diminati adalah akuntansi.

b. *Cluster 2* (C2)

Cluster 2 asal daerah dari Gresik, latar belakang pendidikan sekolah berasal dari SMA Swasta dan jurusan yang diminati adalah informatika.

c. *Cluster 3* (C3)

Cluster 3 didominasi asal daerah dari Manggarai Timur, latar belakang pendidikan sekolah berasal dari Jurusan lainnya dan jurusan diminati adalah akuntansi.

5. KESIMPULAN DAN SARAN

Berdasarkan penelitian yang telah dilakukan, beberapa simpulan dapat diambil terkait penggunaan Algoritme *K-Means* dan *Fuzzy C-Means* dalam

melakukan pengelompokan data dan menentukan strategi promosi di Universitas Katolik Darma Cendika. Penelitian ini menunjukkan bahwa *K-Means* memiliki kualitas hasil yang lebih baik dibandingkan *Fuzzy C-Means*, dengan nilai rata-rata *Silhouette Coefficient* sebesar 0,2833. Meskipun demikian, *Fuzzy C-Means* menunjukkan proses *clustering* yang lebih cepat berdasarkan jumlah iterasi, yaitu 11.7. Berdasarkan analisis *Silhouette Coefficient*, *K-Means* lebih unggul dalam melakukan proses *clustering* data mahasiswa dan lebih valid untuk menentukan target promosi Universitas Katolik Darma Cendika dibandingkan *Fuzzy C-Means*. Selain itu, penerapan kedua algoritme ini menghasilkan tiga kelompok utama, yaitu asal daerah mahasiswa, asal sekolah mahasiswa, dan jurusan mahasiswa. Hasil dari *K-Means Clustering* menunjukkan bahwa strategi promosi sebaiknya difokuskan pada daerah asal mahasiswa, khususnya Kota Surabaya, dengan penekanan pada siswa dari sekolah SMA Swasta dan jurusan akuntansi sebagai fokus utama.

DAFTAR PUSTAKA

[1] R. Setiawan, "Penerapan Data Mining Menggunakan Algoritme *K-Means Clustering* Untuk Menentukan Strategi Promosi Mahasiswa Baru (Studi Kasus : Politeknik Lp3i Jakarta)," *J. Lentera Ict*, vol. 3, no. 1, pp. 76–92, 2016.

[2] M. Khozi Khozali, N. Mega Saraswati, and R. Cipta Sigitta Hariyono, "Penerapan Algoritme *K-Means* Dan Algoritme *Naïve Bayes* Dalam Pemilihan Konsentrasi Jurusan Siswa Di Sma Islam Ta'Alumul Huda Bumiayu," *Ijir*, vol. 2, no. 2, pp. 19–25, 2021.

[3] P. S. Saputra, "Perbandingan Algoritme *Fuzzy C-Means* Dan Algoritme *Naive Bayes* Dalam Menentukan Keluarga Penerima Manfaat (Kpm) Berdasarkan Status Sosial Ekonomi (Sse) Terendah," *JST (Jurnal Sains dan Teknol.)*, vol. 10, no. 1, pp. 1–8, 2021, doi: 10.23887/jstundiksha.v10i1.23340.

[4] Y. R. Prayogi and S. N. Budiman, "Color Grading Systems to Classify Ripeness of Apple Mango Fruit," *Inf. J. Ilm. Bid. Teknol. Inf. dan Komun.*, vol. 3, no. 2, pp. 57–61, 2018, doi: 10.25139/inform.v3i2.1010.

[5] N. Rajagukguk, D. Ispriyanti, and Y. Wilandari, "Perbandingan Metode Klasifikasi Regresi Logistik Biner Dan *Naive Bayes* Pada Status Pengguna Kb Di Kota Tegal Tahun 2014," *Concept Commun.*, vol. null, no. 23, pp. 301–316, 2019.

[6] A. Asroni and R. Adrian, "Penerapan Metode *K-Means* Untuk *Clustering* Mahasiswa Berdasarkan Nilai Akademik Dengan Weka Interface Studi Kasus Pada Jurusan Teknik Informatika UMM Magelang," *Semesta Tek.*, vol. 18, no. 1, pp. 76–82, 2016, doi: 10.18196/st.v18i1.708.

[7] M. R. Alhapizi, M. Nasir, and I. Effendy,

- “Penerapan Data Mining Menggunakan Algoritme K-Means Clustering Untuk Menentukan Strategi Promosi Mahasiswa Baru Universitas Bina Darma Palembang,” *J. Softw. Eng. Ampera*, vol. 1, no. 1, pp. 1–14, 2020, doi: 10.51519/journalsea.v1i1.10.
- [8] D. L. Rahakbauw, V. Y. I. Ilwaru, and M. H. Hahury, “Implementasi Fuzzy C-Means Clustering Dalam Implementation Of Fuzzy C-Means Clustering In,” *J. Ilmu Mat. dan Terap.*, vol. 11, pp. 1–12, 2019, [Online]. Available: <https://media.neliti.com/media/publications/277582-implementasi-fuzzy-c-means-clustering-da-3afa5ba1.pdf>
- [9] N. Azmi, F. Helmiah, and S. Sudarmin, “Implementasi Metode K-Means Sebagai Upaya Penentuan Lokasi Promosi Penerimaan Siswa Baru,” *Build. Informatics, Technol. Sci.*, vol. 3, no. 4, pp. 649–660, 2022, doi: 10.47065/bits.v3i4.1456.
- [10] M. R. Marajari, K. Nasution, and T. Haramaini, “Analisis Data Mahasiswa Untuk Mendukung Strategi Promosi Pada Universitas Medan Area Menggunakan Algoritme K-Means Clustering,” *Semnastek*, pp. 100–107, 2021.
- [11] N. Silalahi, “Penentuan Strategi Promosi Universitas Budi Darma Menggunakan Algoritme K-Means Clustering,” *TIN Terap. Inform. Nusant.*, vol. 1, no. 1, pp. 40–46, 2020.
- [12] R. Syarif, M. T. Furqon, and S. Adinugroho, “Perbandingan Algoritme K-Means dengan Algoritme Fuzzy C Means (FCM) dalam Clustering Moda Transportasi Berbasis GPS,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 10, pp. 4107–4115, 2018, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [13] D. T. Kusuma and N. Agani, “Prototipe Komparasi Model Clustering Menggunakan Metode K-Means Dan FCM untuk Menentukan Strategi Promosi : Study Kasus Sekolah Tinggi Teknik-PLN Jakarta,” *J. TICOM*, vol. 3, no. 3, pp. 1–10, 2015, doi: 10.13140/RG.2.2.35612.08326.