

Analisa Data Mining Menggunakan Algoritma Agglomerative Hierarchical Clustering untuk Menentukan Strategi Promosi Calon Mahasiswa Baru

Fakhrian Fadlia Adiwijaya^{a1}, Chrismikha Hardyanto^{a2}, Riani Lubis^{a3}

^aProgram Studi Teknik Informatika, Universitas Komputer Indonesia
Jl. Dipatiukur No.102-116 Bandung 40132

¹fakhrian@email.unikom.ac.id

²chrismikha@email.unikom.ac.id

³riani.lubis@email.unikom.ac.id

Abstrak

Bertambahnya jumlah perguruan tinggi swasta ataupun negeri di Indonesia menambah kompetisi dalam menarik perhatian calon mahasiswa, peningkatan ini mengharuskan institusi mencari strategi promosi yang lebih efektif. Data penerimaan mahasiswa baru pada periode tahun 2021 sampai 2023 menunjukkan, dari 25 program studi yang ada, 6 program studi mengalami penurunan jumlah pendaftaran, dan 3 lainnya hanya mengalami peningkatan yang tidak signifikan. Pada penelitian ini akan menggunakan data penerimaan mahasiswa baru dan media promosi untuk merancang strategi promosi yang efektif melalui data mining, menggunakan algoritma *Agglomerative Hierarchical Clustering* (AHC). Hasil penelitian menunjukkan bahwa dengan membagi konten promosi menjadi tiga fokus utama berdasarkan preferensi calon mahasiswa, perguruan tinggi dapat meningkatkan efektivitas promosi. Konten pertama menargetkan fakultas manajemen, ilmu budaya, dan desain dengan proporsi konten untuk manajemen (40.65%), desain komunikasi visual (35.10%), akuntansi (9.63%), sastra Jepang (8.16%), dan desain interior (6.45%), menggunakan media promosi website dan Instagram dengan penekanan pada website. Konten kedua menargetkan fakultas teknik dan ilmu komputer dengan proporsi konten untuk teknik informatika (57.84%), sistem informasi (23.01%), teknik arsitektur (8.35%), sistem komputer (7.47%), dan teknik sipil (3.33%), juga menggunakan website dan Instagram dengan fokus pada website. Konten ketiga berfokus pada fakultas ilmu sosial dan ilmu politik, dengan proporsi konten untuk ilmu komunikasi (74.13%), hubungan internasional (21.29%), dan ilmu pemerintahan (4.57%), dengan penekanan pada Instagram. Pendekatan ini memungkinkan perguruan tinggi untuk lebih memahami preferensi dan kebutuhan calon mahasiswa, serta menciptakan pesan promosi yang lebih personal dan relevan, meningkatkan daya tarik dan keberhasilan dalam rekrutmen mahasiswa baru.

Kata kunci: Data Mining, Clustering, Agglomerative Hierarchical Clustering (AHC), Promosi, Segmentasi Calon Mahasiswa

Data Mining Analysis Using Agglomerative Hierarchical Clustering Algorithm to Determine Promotion Strategies for Prospective New Students

Abstract

The increase in the number of private and public universities in Indonesia has intensified the competition to attract prospective students. This growth compels institutions to seek more effective promotional strategies. Data on new student admissions from 2021 to 2023 indicates that out of 25 study programs, 6 experienced a decline in the number of applicants, while 3 showed only insignificant increases. This study utilizes new student admission data and promotional media to design effective promotional strategies through data mining, using the Agglomerative Hierarchical Clustering (AHC) algorithm. The results show that by dividing promotional content into three main focuses based on prospective students' preferences, universities can enhance the effectiveness of their promotions. The first content targets the faculties of management, cultural sciences, and design, with content

proportions for management (40.65%), visual communication design (35.10%), accounting (9.63%), Japanese literature (8.16%), and interior design (6.45%), using website and Instagram as the primary promotional media, with an emphasis on the website. The second content targets the faculties of engineering and computer science, with content proportions for informatics engineering (57.84%), information systems (23.01%), architecture engineering (8.35%), computer systems (7.47%), and civil engineering (3.33%), also using website and Instagram, with a focus on the website. The third content focuses on the faculties of social and political sciences, with content proportions for communication studies (74.13%), international relations (21.29%), and government studies (4.57%), with an emphasis on Instagram. This approach allows universities to better understand the preferences and needs of prospective students, creating more personal and relevant promotional messages, ultimately increasing the attractiveness and success of new student recruitment.

Keywords: Data Mining, Clustering, Agglomerative Hierarchical Clustering (AHC), Promotion, Prospective Student Segmentation

I. PENDAHULUAN

Pada tahun 2022, terdapat 4004 perguruan tinggi baik negeri maupun swasta yang dikelola oleh Kementerian Pendidikan, Kebudayaan, Riset dan Teknologi serta Kementerian Agama, tren ini menunjukkan peningkatan sebesar 0,73% dibandingkan dengan tahun 2021 (sumber : www.bps.go.id). Kondisi ini mengakibatkan peningkatan persaingan antar perguruan tinggi dalam menarik calon mahasiswa baru. Setiap institusi kini terdorong untuk mencari strategi yang lebih efektif dalam mempromosikan program studi atau fakultas mereka kepada calon mahasiswa [1].

Promosi yang efektif sangat penting karena preferensi calon mahasiswa terhadap perguruan tinggi, jenis program studi, lokasi, metode pembelajaran, dan faktor lainnya dapat berubah seiring waktu [1]. Perkembangan teknologi dan media sosial telah mengubah cara calon mahasiswa mengakses informasi tentang lembaga pendidikan. Oleh karena itu, perguruan tinggi perlu memanfaatkan platform digital dan media sosial secara optimal untuk mencapai audiens mereka dengan lebih efektif [2,3].

Memahami dengan baik siapa target audiens lembaga pendidikan atau lulusan sekolah tertentu merupakan kunci dalam mengembangkan strategi promosi yang efektif. Dengan mempertimbangkan faktor-faktor tersebut dan memahami konteks permasalahan, lembaga pendidikan dapat merancang strategi promosi yang tepat untuk menarik dan mempertahankan mahasiswa baru. Hal ini akan membantu perguruan tinggi untuk terus berkembang dan sukses menghadapi dinamika dunia pendidikan [4].

Pada penelitian ini dilakukan analisis terhadap data penerimaan mahasiswa baru pada periode waktu tahun 2021 sampai 2023, berdasarkan data tersebut diketahui dari 25 program studi yang ada, 6 di antaranya mengalami penurunan jumlah pendaftar, sementara 3 lainnya mengalami peningkatan yang tidak signifikan. Jika kondisi ini dibiarkan, penurunan tersebut dapat berlanjut di tahun-tahun mendatang. Oleh karena itu, diperlukan tindakan yang konkret untuk mengatasi masalah ini.

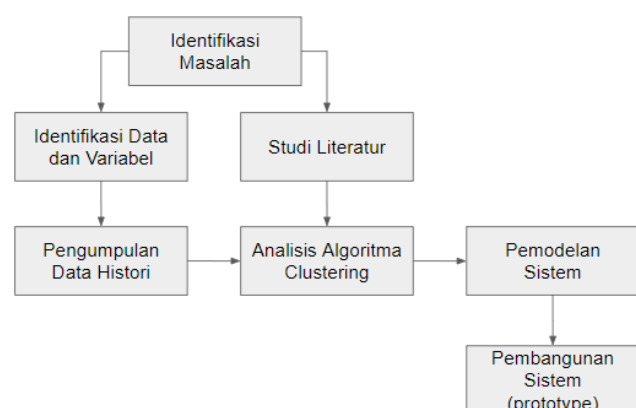
Setiap awal tahun ajaran baru, perguruan tinggi melakukan proses penerimaan mahasiswa baru yang menghasilkan data dalam jumlah besar. Sayangnya, data tersebut belum sepenuhnya dimanfaatkan untuk menentukan strategi promosi. Data yang ada dapat diubah menjadi pengetahuan yang berguna bagi perguruan tinggi untuk melakukan evaluasi dan prediksi. [5,6,7]. Pada

penelitian ini pihak perguruan tinggi yang diteliti masih kesulitan dalam menentukan konten promosi yang baik dan efektif untuk menarik calon mahasiswa berdasarkan preferensi calon mahasiswa.

Dengan demikian, diperlukan sebuah sistem informasi prediksi strategi promosi mahasiswa baru melalui penerapan data mining menggunakan metode *clustering*. Algoritma *clustering* dapat digunakan untuk mengelompokkan calon mahasiswa berdasarkan karakteristik seperti minat akademik, latar belakang geografis, demografi, atau preferensi program studi [8,9]. Hal ini memungkinkan perguruan tinggi untuk mengidentifikasi segmen pasar yang berbeda dan merancang strategi promosi yang lebih tepat sasaran untuk setiap kelompok [10,11]. Dengan pemahaman yang lebih mendalam tentang preferensi dan kebutuhan calon mahasiswa di setiap kelompok, perguruan tinggi dapat menciptakan pesan promosi yang lebih personal dan relevan. Hal ini dapat meningkatkan efektivitas promosi, karena calon mahasiswa cenderung lebih tertarik dengan pesan yang sesuai dengan minat dan kebutuhan mereka [9].

II. METODOLOGI

Tahapan penelitian yang dilakukan oleh peneliti dapat dilihat pada Gambar 1. Adapun metode analisis algoritma clustering akan menggunakan tahapan *Knowledge Discovery in Databases* (KDD) yang dapat dilihat pada Gambar 2



Gambar 1. Tahapan Penelitian

Adapun langkah-langkah yang dilaksanakan pada penelitian ini adalah sebagai berikut :

1. Identifikasi masalah; pada tahap ini dilakukan identifikasi masalah berdasarkan fakta yang ada.
2. Identifikasi data dan variabel; pada tahap ini dilakukan identifikasi data yang digunakan yaitu data penerimaan mahasiswa baru untuk mendapatkan variabel asal sekolah, asal provinsi, preferensi program studi dan data media promosi untuk mendapatkan variabel media promosi yang digunakan oleh perguruan tinggi seperti televisi, radio, website, facebook, twitter, instagram, koran, brosur, billboard, youtube, dan tiktok.
3. Studi literatur; pada tahap ini dilakukan studi pustaka yang berkaitan dengan metode clustering serta algoritma *Agglomerative Hierarchical Clustering* (AHC) yang digunakan untuk membentuk kelompok promosi.
4. Pengumpulan data histori; pada tahap ini dilakukan pengumpulan data histori, yaitu :
 - a. Data penerimaan mahasiswa baru yang mendaftar pada periode waktu 2021 s/d 2022
 - b. Data media promosi yang digunakan pada periode waktu 2021 s/d 2022
5. Analisis algoritma clustering; pada tahap ini akan digunakan algoritma AHC untuk pembentukan kelompok. Algoritma AHC dipilih karena algoritma ini lebih fleksibel dalam menentukan jumlah cluster dan mampu menggabungkan cluster yang paling mirip terlebih dahulu, sehingga menghasilkan segmentasi yang lebih intuitif [12]. Selain itu, AHC lebih efektif dalam menangani data dengan skala yang berbeda dan memberikan visualisasi yang lebih jelas dalam proses clustering [13,14,15]. Untuk tahapan analisis akan menggunakan *Knowledge Discovery in Databases* (KDD).
6. Pemodelan sistem; pada tahap ini dilakukan analisis dan perancangan sistem yang akan dibangun menggunakan pendekatan *Object Oriented Design* (OOD).
7. Pembangunan sistem; pada tahap ini dilakukan pembangunan sistem berbasis web menggunakan bahasa pemrograman python.

Tahapan yang digunakan pada penelitian ini menggunakan *Knowledge Discovery in Databases* (KDD). KDD adalah metode dalam data mining untuk mengekstraksi apa yang dianggap sebagai pengetahuan sesuai spesifikasi ukuran dan batasan menggunakan database dengan melakukan terlebih dahulu tahapan preprocessing, sub sampling dan transformasi database yang diperlukan [12].

A. *Agglomerative Hierarchical Clustering*

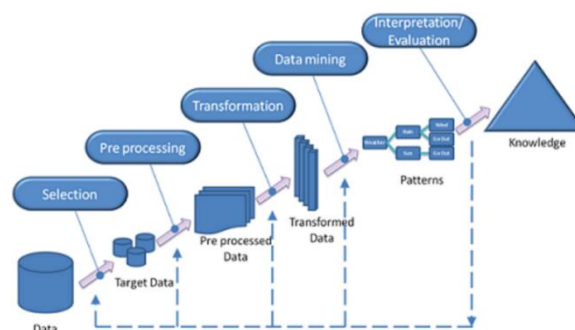
Clustering adalah teknik dalam analisis data yang digunakan untuk mengelompokkan objek-objek data ke dalam kelompok-kelompok yang serupa berdasarkan karakteristik tertentu. Salah satu metode clustering yang umum digunakan adalah *Hierarchical Agglomerative Clustering* (AHC). Algoritma ini dipilih karena pembentukan kelompok dilakukan dengan

membandingkan karakteristik berdasarkan kedekatan satu sama lain[17]. Tahapan pada algoritma AHC adalah sebagai berikut [18] :

1. Inisialisasi : setiap data akan dianggap sebagai satu kluster besar.
2. Perhitungan Matriks Jarak : Matriks jarak antara setiap pasang data dihitung. Matriks jarak dapat menggunakan metode Euclidean, Manhattan, dan sebagainya.
3. Pengelompokan Data : Pasangan data yang paling dekat diidentifikasi dan digabungkan menjadi satu kluster baru. Proses ini dilakukan secara iteratif hingga setiap data sama dengan setiap cluster, dan matriks jarak diperbaharui setiap kali kluster baru dibentuk.
4. Pembentukan Dendogram : Dendogram dibangun untuk merepresentasikan hierarki kluster. Dendogram ini menunjukkan bagaimana kluster-kluster terbentuk dan bagaimana mereka saling terkait satu sama lain.
5. Penentuan Jumlah Kluster : Tahap ini, dapat dipilih jumlah kluster yang optimal berdasarkan struktur dendogram atau berdasarkan tujuan analisisnya.

B. *Knowledge Discovery in Databases*

KDD adalah metode dalam data mining untuk mengekstraksi apa yang dianggap sebagai pengetahuan sesuai spesifikasi ukuran dan batasan menggunakan database dengan melakukan terlebih dahulu tahapan *preprocessing*, *sub sampling* dan *transformation* yang diperlukan [18]. Adapun tahapan proses KDD dapat dilihat pada Gambar 2.



Gambar 2. Tahapan Analisis Algoritma Clustering

Tahapan awal pada KDD adalah seleksi, pada tahap ini, data yang relevan dipilih dari dataset dan mengambil data yang penting untuk dilakukan analisis lebih lanjut. Sejalan dengan tujuan proses penemuan pengetahuan dari data, tahapan selanjutnya adalah *preprocessing*, tahapan ini perlu dilakukan karena data mentah seringkali mengandung *noise*, inkonsistensi, dan nilai yang hilang. Pada tahapan ini, proses pembersihan data dan integrasi dilakukan, hal ini termasuk penanganan nilai yang hilang, dan koreksi kesalahan. Tahapan ketiga adalah *transformation*, pada tahapan ini melibatkan perubahan data yang telah dilakukan *preprocessing* menjadi format yang sesuai untuk analisis dan mining. Proses ini dilakukan untuk meningkatkan kualitas data. Tahapan keempat adalah data mining, pada tahapan ini akan melibatkan berbagai teknik data mining pada data yang telah diubah. Algoritma data mining digunakan untuk menemukan pola

tersembunyi, tren, hubungan, dan asosiasi dalam dataset. Tahapan terakhir adalah *interpretation / evaluation*, tahapan ini terdiri dari representasi pengetahuan berdasarkan pola-pola yang ditemukan pada tahapan data mining, kemudian diinterpretasikan dalam konteks domain masalah. Ini melibatkan implikasi pengetahuan yang ditemukan dan mengevaluasi bagaimana pengetahuan tersebut dapat diterapkan untuk memecahkan masalah di dunia nyata.

III. HASIL DAN PEMBAHASAN

Penelitian ini akan memanfaatkan data penerimaan mahasiswa baru dan data media promosi yang memiliki beberapa atribut untuk proses data mining, guna merancang strategi promosi perguruan tinggi. Langkah-langkah data mining yang akan dilakukan untuk menentukan strategi promosi adalah sebagai berikut :

A. Selection

Untuk data yang akan diambil adalah data PMB periode tahun 2021 sampai 2022 yang berisi data penerimaan mahasiswa yang masuk ke suatu perguruan tinggi. Kemudian data media promosi yang berisi data media promosi yang digunakan calon mahasiswa untuk mengetahui perguruan tinggi.

Data penerimaan mahasiswa baru, berisi laporan pendaftaran mahasiswa yang telah melakukan pendaftaran ke perguruan tinggi yang terdiri dari 8 atribut dan 5,839 baris data. Data media, berisi laporan calon pendaftar dan media yang digunakan untuk mengetahui informasi mengenai pendaftaran perguruan tinggi, yang terdiri dari 15 atribut dan 7,592 baris data. Contoh data untuk masing-masing data penerimaan mahasiswa baru dan data media dapat dilihat pada Tabel I dan Tabel II

Tabel I Contoh Data Penerimaan Mahasiswa Baru

NIM	Nama	Asal Sekolah	Kota	Provinsi	Program Studi	Tahun Masuk	Jalur Masuk
10121001	HISYAM MUZAKI SUGANDHI	MA IBNU SINA	Bandung	Jawa Barat	TEKNIK INFORMATIKA-S1	2021	PMDK
10121002	MUHAMAD FAKHMY NUGRAHA	SMK NEGERI 8 BANDUNG	Bandung	Jawa Barat	TEKNIK INFORMATIKA-S1	2021	USM
10121003	GILANG SETIAWAN PUTRA	SMK KARYA PEMBANGUNAN BALEENDAH	Bandung	Jawa Barat	TEKNIK INFORMATIKA-S1	2021	USM
...
63822055	SYARIFAH NUR AISAH	MAS AL-QUR'AN AL-AMANAH	Bandung	Jawa Barat	SASTRA JEPANG-S1	2022	USM
63822701	MUHAMMAD FERDI FIRMANSYAH	SMK Teknologi Industri Pembangunan CIMAHI	Cimahi	Jawa Barat	SASTRA JEPANG-S1	2022	Konversi

Tabel II Contoh Data Media Promosi

A	B	C	D	E	F	G	H	I	J	K	L	M	N
HISYAM MUZAKI SUGANDHI	22122010	TEKNIK INFORMATIKA-S1	-	-	Ya	-	-	-	-	-	-	-	-
MUHAMAD FAKHMY NUGRAHA	22122000	TEKNIK INFORMATIKA-S1	-	-	Ya	-	-	Ya	-	-	-	-	-
GILANG SETIAWAN PUTRA	22122080	TEKNIK INFORMATIKA-S1	-	-	-	-	-	Ya	-	Ya	-	-	-
...
SYARIFAH NUR AISAH	22132687	SASTRA JEPANG-S1	-	-	-	-	-	-	-	-	-	-	-
MUHAMMAD FERDI FIRMANSYAH	22132400	SASTRA JEPANG-S1	-	-	Ya	-	-	-	-	-	-	-	-

Keterangan :

- A : Nama Calon Mahasiswa
- B : Nomor Pendaftar
- C : Program Studi
- D : Atribut Media Televisi
- E : Atribut Media Radio

- F : Atribut Media Website
- G : Atribut Media Facebook
- H : Atribut Media Twitter
- I : Atribut Media Instagram
- J : Atribut Media Koran

- K : Atribut Media Brosur
- L : Atribut Media Billboard
- M : Atribut Media Youtube
- N : Atribut Media TikTok

B. *Preprocessing*

Pada tahapan *preprocessing* akan dilakukan proses pembersihan data berdasarkan hasil dari tahapan *selection*. Tujuan dari tahapan ini adalah agar data yang digunakan pada tahapan data mining dapat menghasilkan kelompok yang baik. Pada tahapan ini hal pertama yang dilakukan adalah penggabungan data antara data penerimaan mahasiswa baru dengan data media promosi dengan melakukan join berdasarkan atribut nama dan program studi pada data penerimaan mahasiswa baru dengan atribut nama pendaftar dan program studi pada data media promosi. Setelah data digabungkan, tahapan selanjutnya adalah verifikasi kualitas data, dari 5.839 data PMB yang diperoleh terdapat 153 nilai null pada atribut asal sekolah, dan 31 nilai null pada atribut provinsi. Sementara untuk data media promosi terdapat 203 nilai null pada semua atribut media yang disebabkan tidak adanya kecocokan antara data penerimaan mahasiswa baru dengan data media promosi. atribut-atribut yang bernilai null tersebut akan dihapus dari basis data. Langkah selanjutnya adalah menambahkan atribut tipe sekolah untuk me-*minimalisir* keberagaman data sekolah, penambahan atribut tipe sekolah ini berdasarkan atribut asal sekolah, kemudian penambahan atribut fakultas berdasarkan karakter pertama dari atribut NIM. Setelah tahapan *preprocessing* selesai dilakukan, maka masuk ke tahapan berikutnya, yaitu tahapan *transformation*.

C. *Transformation*

Tahapan *transformation* dilakukan untuk membentuk data yang sudah dipilih dan diperbaiki untuk disesuaikan dengan kebutuhan pada tahapan data mining. Tahapan ini akan berbeda-beda tergantung dari jenis atau pola informasi pengetahuan yang ingin didapatkan. Tahapan pertama adalah penyesuaian pada atribut program studi disesuaikan dengan kode program studi di perguruan tinggi tempat penelitian, untuk penyesuaian atribut program studi dapat dilihat pada Tabel III.

Tabel III. Penyesuaian Atribut Program Studi

Nilai Atribut Program Studi	Format Nilai
Teknik Informatika - S1	1
Sistem Komputer - S1	2
Teknik Industri - S1	3
Teknik Arsitektur - S1	4
Sistem Informasi - S1	5
Perencanaan Wilayah dan Kota - S1	6
Teknik Komputer - D3	8
Manajemen Informatika - D3	9
Komputerisasi Akuntansi - D3	10
Teknik Sipil - S1	30
Teknik Elektro - S1	31
Akuntansi - S1	11
Manajemen - S1	12
Akuntansi - D3	13
Manajemen Pemasaran - D3	14
Kuangan dan Perbankan - D3	15

Nilai Atribut Program Studi	Format Nilai
Ilmu Hukum - S1	16
Ilmu Pemerintahan - S1	17
Ilmu Komunikasi - S1	18
Hubungan Internasional - S1	43
Desain Komunikasi Visual - S1	19
Desain Interior - S1	20
Desain Grafis - D3	21
Sastra Inggris - S1	37
Sastra Jepang - S1	38

Berikutnya adalah penyesuaian atribut tipe sekolah hasil dari penambahan atribut pada proses *pre processing*. Penyesuaian atribut tipe sekolah dilakukan berdasarkan urutan atribut tipe sekolah sesuai dengan jumlah kemunculan pada data hasil penggabungan. Untuk penyesuaian atribut asal sekolah dapat dilihat pada Tabel IV

Tabel IV. Penyesuaian Atribut Asal Sekolah

Nilai Atribut Tipe Sekolah	Format Nilai
SMA	1
SMK	2
MA	3
Pesantren	4
Home Schooling	5
PKBM	6
Universitas / Politeknik	7
Paket C	8

Berikutnya penyesuaian atribut provinsi, atribut provinsi akan diubah format datanya kedalam numerik. Perubahan ini berdasarkan urutan provinsi yang ada di Indonesia. Untuk penyesuaian atribut provinsi dapat dilihat pada Tabel V.

Tabel V. Penyesuaian Atribut Provinsi

Nilai Atribut Provinsi	Format Nilai
Nanggroe Aceh Darussalam (NAD)	1
Sumatera Utara	2
Sumatera Selatan	3
Sumatera Barat	4
Bengkulu	5
...	...
Papua Pegunungan	37
Papua Barat Daya	38

Penyesuaian untuk atribut media dilakukan dengan cara merubah kedalam format numerik, ketentuan yang diberlakukan adalah merubah menjadi nilai 1 untuk nilai "Ya" dan 0 untuk nilai "Tidak".

Setelah tahapan *transformation* selesai dilakukan, tahapan selanjutnya adalah pemilihan atribut sesuai kebutuhan dari data mining. Berdasarkan hasil analisis dipilihlah 15 atribut yang terdiri dari Provinsi, Program Studi, Televisi, Radio, Website, Facebook, Twitter, Instagram, Koran, Brosur, Billboard, Youtube, TikTok,

Tipe Sekolah, . Untuk baris data yang digunakan pada mining sebanyak 3671 baris data

D. Data Mining

Tahapan ini merupakan tahapan yang dilakukan untuk mencari informasi atau pengetahuan dalam metodologi *Knowledge Discovery in Databases*. Tahapan ini akan mengolah data yang sudah dipersiapkan pada tahapan sebelumnya. Algoritma dan teknik yang dilakukan pada tahapan ini dapat bervariasi, tergantung pada tujuan penggalian informasi yang akan dilakukan.

Metode yang digunakan untuk membentuk kelompok promosi adalah dengan menggunakan *clustering*. Sedangkan algoritma yang akan digunakan adalah *Agglomerative Hierarchical Clustering* dengan pendekatan *Ward* untuk menentukan penggabungan antara dua kelompok berdasarkan nilai variasi terkecil. Mekanisme data mining dari proses *clustering* menggunakan *Agglomerative Hierarchical Clustering* adalah sebagai berikut [19] :

- 1) Hitung matrik jarak antar data menggunakan *Euclidean Distance* menggunakan rumus :

$$dist(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \tag{1}$$

- 2) Gabungkan kelompok terdekat menjadi satu kelompok berdasarkan nilai variasi terkecil yang didapat berdasarkan perhitungan Ward menggunakan rumus :

$$V_{Ward}(A \cup B) = \frac{n_A \cdot n_B}{n_A + n_B} \cdot d(A, B)^2 \tag{2}$$

- 3) Hitung kembali jarak antar kelompok yang baru terbentuk dengan titik data lainnya.
- 4) Ulangi langkah 2 sampai 3 hingga terbentuk 1 kelompok untuk seluruh data.
- 5) Bentuk dendrogram untuk memvisualisasikan kelompok yang terbentuk untuk menentukan jumlah kelompok yang dapat digunakan.

Pada tahapan data mining ini akan digunakan dataset hasil dari tahapan *transformation* yang telah dilakukan. Data tersebut terdiri dari 15 atribut dan 3671 baris data. Contoh dataset yang digunakan dapat dilihat pada Tabel VI

Tabel VI. Contoh Dataset untuk Perhitungan Data Mining

Dmain	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
D1	18	1	0	0	1	0	0	0	0	0	0	0	0	3	1
D2	18	1	0	0	1	0	0	1	0	0	0	0	0	2	1
D3	18	1	0	0	0	0	0	1	0	1	0	0	0	2	1
D4	18	1	0	0	1	0	0	1	0	0	0	0	0	2	1
D5	18	1	0	0	1	0	0	1	0	0	0	0	0	2	1
D6	18	1	0	0	1	0	0	0	0	1	0	0	0	2	1
D7	18	1	0	0	0	0	0	1	0	0	0	0	0	2	1
D8	18	1	0	0	1	0	0	0	0	0	0	0	0	1	1
D9	18	1	0	0	0	0	0	0	0	0	0	0	0	2	1
D10	18	1	0	0	0	0	0	0	0	1	0	0	0	1	1
...
D3670	18	38	0	0	0	0	0	0	0	0	0	0	0	3	6
D3671	18	38	0	0	1	0	0	0	0	0	0	0	0	2	6

Keterangan :

- A : Atribut Provinsi
- B : Atribut Program Studi
- C : Atribut Media Televisi
- D : Atribut Media Radio
- E : Atribut Media Website
- F : Atribut Media Facebook
- G : Atribut Media Twitter
- H : Atribut Media Instagram
- I : Atribut Media Koran
- J : Atribut Media Brosur
- K : Atribut Media Billboard
- L : Atribut Media Youtube
- M : Atribut Media TikTok
- N : Atribut Tipe Sekolah
- O : Atribut Fakultas

Dari contoh data pada Tabel VI, lakukan perhitungan jarak menggunakan rumus *euclidean distance* untuk masing-masing atribut, sebagai contoh perhitungan jarak dilakukan untuk data ke-1 sampai ke-3

D1, D1 =

$$\sqrt{(18 - 18)^2 + (1 - 1)^2 + (0 - 0)^2 + (0 - 0)^2 + (1 - 1)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2} = 0$$

D1, D2 =

$$\sqrt{(18 - 18)^2 + (1 - 1)^2 + (0 - 0)^2 + (0 - 0)^2 + (1 - 0)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 1)^2 + (0 - 0)^2 + (0 - 0)^2} = 1.41$$

D1, D3 =

$$\sqrt{(18 - 18)^2 + (1 - 1)^2 + (0 - 0)^2 + (0 - 0)^2 + (1 - 1)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 1)^2 + (0 - 0)^2 + (0 - 1)^2 + (0 - 0)^2 + (0 - 0)^2 + (3 - 2)^2 + (1 - 1)^2} = 2$$

Setelah dilakukan perhitungan jarak antar data yang ada, langkah selanjutnya dilakukan perhitungan nilai variasi menggunakan perhitungan *ward*.

$$Ward(D1, D1) = \frac{1 \times 1}{1 + 1} \times 0 = 0$$

$$Ward(D1, D2) = \frac{1 \times 1}{1 + 1} \times 1.41 = 0.71$$

$$Ward(D1, D3) = \frac{1 \times 1}{1 + 1} \times 2 = 1$$

Perhitungan nilai *variance* antara gabungan masing-masing data dapat dilihat pada Tabel VII

Tabel VII Hasil Perhitungan Nilai Variance antar Gabungan Data pada Iterasi 1

Dmain	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	...	D3670	D3671
D1	0.00	0.71	1.00	0.71	0.71	0.71	0.87	1.00	0.71	1.22	...	18.67	18.67
D2	0.71	0.00	0.71	0.00	0.00	0.71	0.50	0.71	0.71	1.00	...	18.69	18.67
D3	1.00	0.71	0.00	0.71	0.71	0.71	0.50	1.00	0.71	0.71	...	18.69	18.69
D4	0.71	0.00	0.71	0.00	0.00	0.71	0.50	0.71	0.71	1.00	...	18.69	18.67
D5	0.71	0.00	0.71	0.00	0.00	0.71	0.50	0.71	0.71	1.00	...	18.69	18.67
D6	0.71	0.71	0.71	0.71	0.71	0.00	0.87	0.71	0.71	0.71	...	18.69	18.67
D7	0.87	0.50	0.50	0.50	0.50	0.87	0.00	0.87	0.50	0.87	...	18.68	18.68
D8	1.00	0.71	1.00	0.71	0.71	0.71	0.87	0.00	0.71	0.71	...	18.70	18.67
D9	0.71	0.71	0.71	0.71	0.71	0.71	0.50	0.71	0.00	0.71	...	18.67	18.67
D10	1.22	1.00	0.71	1.00	1.00	0.71	0.87	0.71	0.71	0.00	...	18.70	18.69
...
D3670	18.67	18.69	18.69	18.69	18.69	18.69	18.68	18.70	18.67	18.70	...	0.00	0.71
D3671	18.67	18.67	18.69	18.67	18.67	18.67	18.68	18.67	18.67	18.69	...	0.71	0.00

Berdasarkan hasil perhitungan nilai *variance*, didapatkan nilai terkecil selain hubungan dengan dirinya sendiri terletak pada D4 dengan D2. Maka dari itu, D4 dan D2 akan digabungkan menjadi satu cluster. Hasil penggabungan tersebut akan digunakan pada iterasi berikutnya sampai iterasi sebanyak data pada tabel minus 1, yang artinya akan dilakukan perhitungan jarak euclidean distance dan penggabungan data menggunakan ward sebanyak 3670 kali.

E. Interpretation / Evaluation

Tahapan interpretation / evaluation merupakan tahapan terakhir yang dilakukan untuk mencari informasi atau pengetahuan dalam metodologi *Knowledge Discovery in Databases*. Tahapan ini akan mengevaluasi pola informasi yang telah terbentuk dari proses data mining. Proses evaluasi yang dilakukan digunakan untuk menilai apakah pola informasi yang terbentuk telah sesuai berdasarkan hipotesa atau bertentangan.

Untuk melakukan evaluasi pada hasil clustering yang telah dilakukan, akan digunakan perhitungan DBI pada setiap kelompok data yang terbentuk. Untuk menghitung DBI, langkah-langkah yang harus dilakukan adalah sebagai berikut :

- 1) Tentukan nilai centroid dari setiap cluster yang sudah terbentuk, nilai ini dapat diperoleh dari rata-rata data dari setiap clusternya.
- 2) Hitung jarak setiap data dengan centroid di setiap cluster menggunakan rumus *euclidean distance*.
- 3) Hitung nilai *Sum of Square Within Cluster (SSW)* sebagai metrik kohesi dalam sebuah cluster.

$$SSW_i = \frac{1}{m_i} \sum_{j=0}^{m_i} d(x_j, c_i) \tag{3}$$

- 4) Hitung nilai *Sum of Square Between Cluster (SSB)* untuk separasi antara dua cluster

$$SSB_{i,j} = d(c_i, c_j) \tag{4}$$

- 5) Hitung rasio seberapa baik nilai perbandingan antara cluster ke-I dan cluster ke-j. nilainya didapat dari komponen kohesi dan separasi

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_{i,j}} \tag{5}$$

- 6) Cari nilai rasio terbesar (Rmaks), kemudian hitung DBI menggunakan rumus

$$DBI = \frac{1}{K} \sum_{i=1}^K Max(R_{i,j}) \tag{6}$$

Pada kasus pada penelitian ini, akan dilakukan pembentukan 3 kelompok, berdasarkan hasil yang diperoleh dari diagram dendrogram diketahui kelompok yang terbentuk untuk C1 sebanyak 2000 anggota, untuk C2 sebanyak 375 anggota dan C3 sebanyak 1296 anggota.

Langkah pertama yang dilakukan adalah mencari nilai centroid untuk masing-masing cluster dengan mencari nilai rata-rata. Titik centroid untuk C1 diketahui adalah (1.508, 16.742, 3.200, 14.590, 0.019, 0.011, 0.452, 0.026, 0.042, 0.436, 0.006, 0.143, 0.035, 0.005, 0.118) untuk C2 diketahui adalah (1.421, 17.424, 4.368, 38.227, 0.019, 0.013, 0.488, 0.027, 0.061, 0.387, 0.013, 0.149, 0.029, 0.003, 0.211) dan untuk C3 diketahui adalah (1.569, 17.963, 1.000, 2.305, 0.018, 0.014, 0.522, 0.025, 0.041, 0.418, 0.005, 0.173, 0.025, 0.028, 0.035).

Lakukan perhitungan jarak antara data dengan centroid pada masing-masing cluster, sebagai contoh dilakukan perhitungan jarak antara data ke-1 dengan centroid C1 sebagai berikut :

$$\sqrt{\begin{matrix} (1 - 1.508)^2 + (7 - 16.742)^2 + (1 - 3.2)^2 + (1 - 14.59)^2 \\ + (0 - 0.019)^2 + (0 - 0.011)^2 + (1 - 0.452)^2 + (1 - 0.026)^2 \\ + (0 - 0.042)^2 + (1 - 0.436)^2 + (0 - 0.006)^2 + (-0.143)^2 \\ + (0 - 0.035)^2 + (0 - 0.035)^2 + (0 - 0.005)^2 + (0 - 0.118)^2 \end{matrix}} = 16.949$$

Lakukan perhitungan *euclidean distance* untuk setiap data dalam suatu cluster dengan centroid clusternya. Kemudian lakukan perhitungan SSW untuk setiap cluster.

$$SSW_1 = \frac{1}{2000} \left(\begin{matrix} 16.949 + 17.852 + 14.618 \\ + 19.495 + 21.364 + 20.931 \\ \dots \\ + 6.898 + 6.908 + 4.636 \end{matrix} \right) = 6.271$$

$$SSW_2 = \frac{1}{375} \left(\begin{matrix} 4.950 + 5.803 + 9.175 \\ + 9.071 + 17.135 + 9.027 \\ \dots \\ + 17.994 + 9.067 + 9.101 \end{matrix} \right) = 5.803$$

$$SSW_3 = \frac{1}{1296} \left(\begin{matrix} 3.023 + 3.015 + 4.883 \\ + 4.173 + 3.020 + 4.879 \\ \dots \\ + 3.020 + 3.050 + 3.436 \end{matrix} \right) = 2.770$$

Diketahui berdasarkan perhitungan yang telah dilakukan nilai SSW untuk masing-masing cluster yaitu, C1 sebesar 6.271, C2 sebesar 5.803, dan C3 sebesar 2.770.

Selanjutnya lakukan perhitungan SSB untuk menghitung jarak antar cluster dan rasio untuk C1 dengan C2, C1 dengan C3 dan C2 dengan C3. Sebagai contoh dilakukan perhitungan nilai rasio untuk C1 dengan C2 sebagai berikut :

$$SSB_{1,2} = \sqrt{\begin{matrix} (1.508 - 1.421)^2 + (16.742 - 17.424)^2 + (3.2 - 4.368)^2 \\ + (14.590 - 38.227)^2 + (0.019 - 0.019)^2 + (0.011 - 0.013)^2 \\ + (0.452 - 0.488)^2 + (0.026 - 0.027)^2 + (0.042 - 0.061)^2 \\ + (0.436 - 0.387)^2 + (0.006 - 0.013)^2 + (0.143 - 0.149)^2 \\ + (0.035 - 0.029)^2 + (0.005 - 0.003)^2 + (0.118 - 0.221)^2 \end{matrix}} = 23.677$$

$$R_{1,2} = \frac{6.271 + 5.803}{23.677} = 0.510$$

Berdasarkan perhitungan yang telah dilakukan diketahui hasil untuk rasio C1 dengan C2 sebesar 0.510, untuk C1 dengan C3 sebesar 0.721 dan C2 dengan C3 sebesar 0.238. dari sini kita buat matriks seperti pada Tabel VIII untuk mencari Max Rasio.

Tabel VIII. Matrik Rasio

Rasio	C1	C2	C3	Max Rasio
C1	0	0.510	0.721	0.721
C2	0.510	0	0.238	0.510
C3	0.721	0.238	0	0.721

Dari Tabel VIII, diketahui nilai Max Rasio untuk masing-masing cluster. Selanjutnya lakukan perhitungan DBI.

$$DBI = \frac{1}{3} (0.721 + 0.510 + 0.721) = 0.651$$

Berdasarkan hasil perhitungan, diketahui nilai DBI yang dihasilkan untuk pembentukan 3 cluster adalah 0.651, nilai ini cukup baik, yang menandakan karakteristik anggota pada setiap kelompok sudah sangat baik. Untuk melihat karakteristik yang terbentuk antar kelompok disesuaikan dengan kebutuhan dari perguruan tinggi, sebagai contoh untuk interpretasi karakteristik kelompok yang telah dibentuk dapat dilihat pada Tabel IX

Tabel IX. Karakteristik Anggota Setiap Cluster

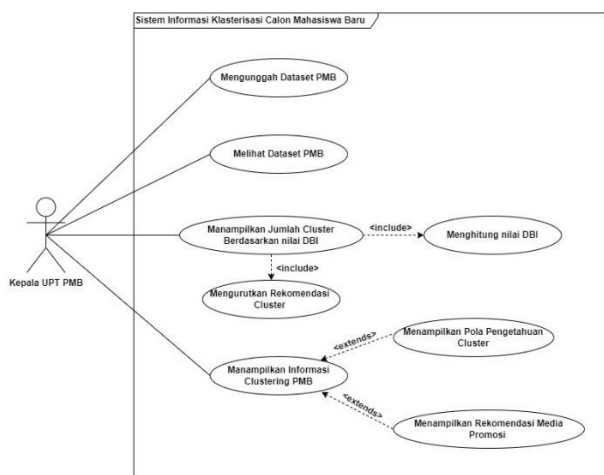
Cluster	Jumlah Anggota	Provinsi	Fakultas	Program Studi	Media Promosi
1	1432	Jawa Barat, Banten, Sumatera Barat, Bangka Belitung, DKI Jakarta	Manajemen, Ilmu Budaya, Desain	Manajemen (40.65%), Desain Komunikasi Visual (35.10%), Akuntansi (9.63%), Sastra Jepang (8.16%), Desain Interior (6.45%)	Website (39.73%), Instagram (35.89%), Brosur (12.37%), Tiktok (8.84%), Twitter (3.17%)
2	1605	Jawa Barat, Banten, DKI Jakarta, Jawa Tengah, Riau	Teknik dan Ilmu Komputer	Teknik Informatika (57.84%), Sistem Informasi (23.01%), Teknik Arsitektur (8.35%), Sistem Komputer (7.47%), Teknik Sipil (3.33%)	Website (43.76%), Instagram (34.59%), Brosur (14.45%), Tiktok (3.68%), Twitter (3.52%)
3	634	Jawa Barat, Banten, DKI Jakarta, Bangka Belitung, Sumatera Selatan	Ilmu Sosial dan Ilmu Politik	Ilmu Komunikasi (74.13%), Hubungan Internasional (21.29%), Ilmu Pemerintahan (4.57%)	Instagram (36.17%), Website (31.86%), Tiktok (17.13%), Brosur (10.06%), Twitter (4.79%)

Berdasarkan karakteristik yang dihasilkan untuk setiap cluster yang terbentuk, sebagai contoh apabila perguruan tinggi ingin membuat 3 jenis konten promosi yang fokus

pada preferensi calon mahasiswa, maka untuk konten pertama akan lebih berfokus pada fakultas manajemen, ilmu budaya, dan desain dengan persorsi dari 100% konten

promosi yang digunakan untuk memperkenalkan program studi manajemen sebesar 40.65%, desain komunikasi visual sebesar 35.10%, akuntansi sebesar 9.63%, sastra jepang sebesar 8.16% dan desain interior sebesar 6.45%. proporsi tersebut berdasarkan karakteristik dari kelompok pertama yang terbentuk, untuk media promosi yang digunakan dapat lebih berfokus pada 2 media yang terbanyak digunakan yaitu website dan instagram dengan fokus terbesar pada media website. untuk konten kedua akan lebih berfokus pada fakultas teknik dan ilmu komputer dengan proporsi dari 100% konten promosi yang digunakan untuk memperkenalkan program studi teknik informatika sebesar 57.84%, sistem informasi sebesar 23.01%, teknik arsitektur sebesar 8.35%, sistem komputer sebesar 7.47% dan teknik sipil sebesar 3.33%. proporsi tersebut berdasarkan karakteristik dari kelompok kedua yang terbentuk, untuk media promosi yang digunakan dapat lebih berfokus pada 2 media yang terbanyak digunakan yaitu website dan instagram dengan fokus terbesar pada media website. Untuk konten ketiga akan lebih berfokus pada fakultas ilmu sosial dan ilmu politik dengan proporsi dari 100% konten promosi yang digunakan untuk memperkenalkan program studi ilmu komunikasi sebesar 74.13%, hubungan internasional sebesar 21.29%, dan ilmu pemerintahan sebesar 4.57%. proporsi tersebut berdasarkan karakteristik dari kelompok ketiga yang terbentuk, untuk media promosi yang digunakan dapat lebih berfokus pada 2 media yang terbanyak digunakan yaitu website dan instagram dengan fokus terbesar pada media instagram.

Setelah tahapan analisis selesai dilakukan, dan mendapatkan pengetahuan yang dapat berguna bagi pemangku keputusan dalam hal ini Kepala UPT PMB, langkah selanjutnya adalah merancang dan membangun perangkat lunak, untuk perancangan perangkat lunak menggunakan use case yang dapat dilihat pada Gambar 3



Gambar 3. Use Case Diagram

IV. KESIMPULAN

Kesimpulan yang dapat diambil dari hasil penelitian yang telah dilakukan adalah :

1. Hasil analisis data penerimaan mahasiswa baru pada perguruan tinggi dapat memberikan informasi

karakteristik mahasiswa berdasarkan asal sekolah, asal provinsi, fakultas dan program studi yang dituju.

2. Membantu pemangku keputusan perguruan tinggi, dalam hal ini kepala UPT PMB untuk dapat menentukan konten promosi dan juga proporsi promosi pada prodi berdasarkan hasil kelompok yang merepresentasikan konten promosi yang dapat dibuat.
3. Membantu kepala UPT PMB dalam menentukan media promosi yang efektif untuk melakukan promosi berdasarkan karakteristik dari masing-masing kelompok yang merepresentasikan konten promosi yang dapat dibuat.

Adapun saran yang dapat dikembangkan dari penelitian ini adalah, data media promosi yang di dapat pada penelitian ini tidak mencantumkan tanggal pendaftaran mahasiswa. Atribut ini dapat disertakan pada proses analisis untuk dapat menghasilkan informasi waktu promosi yang tepat untuk setiap konten promosi yang telah dibuat.

DAFTAR PUSTAKA

- [1] A. Bellanov dan L. Nurhayati, "K-Means Clustering Analysis Untuk Menentukan Strategi Promosi Kampus," Jurnal Hasil Penelitian dan Karya Ilmiah Dalam Bidang Teknik Industri, vol. 9, no. 1, pp. 259-268, 2023.
- [2] C. Aditya Kusuma, A. Juanna, F. Atuna, and M. Ade Kahfiansyah, "Strategi Promosi dan Representasi Perguruan Tinggi dalam Website Institusi: Analisis Wacana Kritis," Value : Jurnal Manajemen dan Akuntansi, vol. 18, no. 3, pp. 1069-1080, Jan. 2024.
- [3] Mustakim, S. E. Rahmadani, and J. Adhiva, "Strategi Promosi Perguruan Tinggi Menggunakan Model Clustering dan Association," Jurnal Ilmiah Rekayasa dan Manajemen Sistem Informasi, 2020.
- [4] A. Bellanov dan L. Nurhayati, "K-Means Clustering Analysis Untuk Menentukan Strategi Promosi Kampus," Jurnal Hasil Penelitian dan Karya Ilmiah Dalam Bidang Teknik Industri, vol. 9, no. 1, pp. 259-268, 2023.
- [5] N. Rosyidah, Matin, and U. Rosyidi, "Internationalization in Higher Education: University's Effective Promotion Strategies in Building International Trust," European Journal of Educational Research, vol. 9, no. 1, pp. 351-361, 2020.
- [6] L. A. Setiyo dan B. Andoro, "Penerapan Algoritma K-Means Untuk Menentukan Strategi Promosi (Studi Kasus : Universitas Katolik Widya Mandala Kampus Kota Madiun)," dalam Seminar dan Konferensi Nasional IDEC 2021, 2021.
- [7] R. Andriani, R. F. Amanullah, D. Ninosari and K. , "Optimization Of Clustering Algorithm On Decision Support System Of Scholarship Recipients Using Analytical Hierarchy Process Method," in IOP conf. Series : Journal of Physics, 2018.
- [8] L. Brown, M. Johnson, and K. Lee, "Segmenting Customers with Data Mining Techniques," in Proceedings of the IEEE International Conference on Data Mining (ICDM), 2018, pp. 1123-1128.
- [9] J. Smith and R. Doe, "Sales Pattern Identification for Marketing Strategies Using Data Mining," American Institute of Physics Conference Proceedings, vol. 2677, no. 1, pp. 060003, 2020.
- [10] H. Zhang, Y. Liu, and X. Wang, "Design Mechanism of Precision Marketing Planning Platform Based on Data Mining Technology," IEEE Access, vol. 7, pp. 8765-8776, 2019.
- [11] M. Green and T. Hall, "Application of Naive Bayes Classifier Algorithm in Promotion Strategies," Journal of Information Systems and Informatics, vol. 22, no. 3, pp. 123-130, 2017.
- [12] Szymkowiak, A., Larsen, J. and Hnen, L.K. Hierarchical Clustering for Data Mining. Technical University of Denmark, Denmark. 2001.
- [13] M. Kaushik and B. Mathur, "Comparative Study of K-Means and Hierarchical Clustering Techniques," International Journal of Software & Hardware Research in Engineering, vol. 2, no. 6, pp. 93-98, 2014.

- [14] B. J. D. Sitompul, O. S. Sitompul and P. Sihombing, "Enhancement Clustering Evaluation Result of Davies-Bouldin Index with Determining Initial Centroid of K-Means Algorithm," in IOP conf. Series : Journal of Physics, 2019.
- [15] R. Andriani, R. F. Amanullah, D. Ninosari and K. , "Optimization Of Clustering Algorithm On Decision Support System Of Scholarship Recipients Using Analytical Hierarchy Process Method," in IOP conf. Series : Journal of Physics, 2018.
- [16] M. Khosrow-Pour, *Advanced Methodologies and Technologies in Network Architecture, Mobile Computing, and Data Analytics*. Hershey: IGI Global, 2018.
- [17] S. P. Borgatti, "How To Explain Hierarchical Clustering," *Connections*, vol. 17, no. 2, pp. 78-80, 1994.
- [18] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*, Morgan Kaufmann Publishers, 2000
- [19] F. F. Adiwijaya, A. Z. Fathurrahman, and C. Hardyanto, "Application of Agglomerative Hierarchical Clustering (AHC) for Grouping Prospective Scholarship Recipients," in *2023 International Conference on Informatics Engineering, Science & Technology (INCITEST)*, pp. 1-7, October 2023.