

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

IMPLEMENTASI K-MEANS CLUSTERING PADA DATA PENGELOMPOKAN PENDAFTARAN MAHASISWA BARU (STUDI KASUS UNIVERSITAS ABDURRAB)

TUGAS AKHIR

Disusun Sebagai Salah Satu Syarat
Untuk Memperoleh Gelar Sarjana Teknik
Pada Jurusan Teknik Informatika

Oleh

MUAMMAD HANIF ABDURROHMAN

NIM. 11950111707



UIN SUSKA RIAU

FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SULTAN SYARIF KASIM RIAU

PEKANBARU

2024



LEMBAR PERSETUJUAN

IMPLEMENTASI K-MEANS CLUSTERING PADA DATA PENGELOMPOKAN PENDAFTARAN MAHASISWA BARU (STUDI KASUS UNIVERSITAS ABDURRAB)

TUGAS AKHIR

Oleh

MUHAMMAD HANIF ABDURROHMAN

NIM. 11950111707

Telah diperiksa dan disetujui sebagai Laporan Tugas Akhir
di Pekanbaru, pada tanggal 15 Januari 2023

Pembimbing I,

Dr. Elin Haerani, S.T , M.kom

NIP. 19810523 200710 12 003

- Hak Cipta Dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



LEMBAR PENGESAHAN

IMPLEMENTASI K-MEANS CLUSTERING PADA DATA PENGELOMPOKAN PENDAFTARAN MAHASISWA BARU (STUDI KASUS UNIVERSITAS ABDURRAB)

Oleh

MUHAMMAD HANIF ABDURROHMAN

NIM. 11950111707

Telah dipertahankan di depan sidang dewan penguji
sebagai salah satu syarat untuk memperoleh gelar Sarjana Teknik
pada Universitas Islam Negeri Sultan Syarif Kasim Riau

Pekanbaru, 15 Januari 2024

Mengesahkan,

Ketua Jurusan,

Dekan,

Dr. Hartono, M.Pd

NIP. 19640301 199203 1 003

Iwan Iskandar, M.T

NIP. 19821216 201503 1 003

DEWAN PENGUJI

Ketua : Iwan Iskandar, M.T
 Pembimbing I : Dr. Elin Haerani, S.T., M.Kom
 Pembimbing II : -
 Penguji I : Fadhilah Syafrica, S.T., M.Kom
 Penguji II : Lola Oktavia, S.S.T., M.T.I

Hak Cipta Dilindungi Undang-Undang
 1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

© Hak cipta milik UIN Suska Riau



SURAT PERNYATAAN

Saya yang bertandatangan di bawah ini:

Nama : Muhammad Hanif Abdurrohman
 NIM : 11950111707
 Tempat, Tgl. Lahir : Depok, 24 Desember 2000
 Fakultas : Sains dan Teknologi
 Prodi : Teknik Informatika
 Judul Jurnal :

IMPLEMENTASI K-MEANS CLUSTERING PADA DATA PENGELOMPOKAN PENDAFTARAN MAHASISWA BARU (STUDI KASUS UNIVERSITAS ABDURRAB)

Menyatakan dengan sebenar-benarnya bahwa:

1. Penulisan jurnal dengan judul sebagaimana tersebut di atas adalah hasil pemikiran dan penelitian saya sendiri.
2. Semua kutipan pada karya tulis saya ini sudah disebutkan sumbernya.
3. Oleh karena itu jurnal saya ini, saya nyatakan bebas dari plagiat.
4. Apabila di kemudian hari terbukti terdapat plagiat dalam penulisan jurnal saya tersebut, maka saya bersedia menerima sanksi sesuai peraturan perundang-undangan.

Demikianlah Surat Pernyataan ini saya buat dengan penuh kesadaran dan tanpa paksaan dari pihak manapun juga.

Pekanbaru, 16 Januari 2024
 Yang membuat pernyataan



M. Hanif Abdurrohman
 NIM. 11950111707

Hak Cipta Dilindungi Undang-Undang
 1. Dilarang mengutip, salin, atau seluruh karya tulis ini tanpa mengacu ke sumbernya.
 a. Penelitian hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

© Hak cipta milik UIN Suska Riau
 State Islamic University of Sultan Syarif Kasim Riau

IMPLEMENTASI K-MEANS CLUSTERING PADA DATA PENGELOMPOKAN PENDAFTARAN MAHASISWA BARU (STUDI KASUS UNIVERSITAS ABDURRAB)

Muhammad Hanif Abdurrohan¹⁾, Elin Haerani²⁾, Fadhilah Syafria³⁾, Lola Oktavia⁴⁾

^{1,2,3)} Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sultan Syarif Kasim Riau
^{1,2,3)} Jl. Soebrantas N0 155 Pekanbaru – Riau – Indonesia

E-mail: 11950111707@students.uin-suska.ac.id, elin.haerani@uin-suska.ac.id, fadhilah.syafria@uin-suska.ac.id,
lola.oktavia@uin-suska.ac.id

ABSTRAK

Menghadapi dinamika kompleks pendaftaran mahasiswa baru, metode k-means clustering diperkenalkan sebagai pendekatan utama. Fokusnya adalah pada Universitas Abdurrab, di mana diselidiki berbagai atribut calon mahasiswa, termasuk jenis kelamin, pendidikan orangtua, penghasilan orangtua, kota/kabupaten asal, provinsi, usia, dan pilihan program studi. Dengan algoritma k-means clustering, tujuan penelitian adalah mengungkap pola yang mendasari preferensi dan karakteristik kelompok mahasiswa baru. Hasil dari penelitian ini memberikan wawasan mendalam mengenai faktor-faktor yang mempengaruhi keputusan penerimaan mahasiswa baru di lingkungan kampus Universitas Abdurrab. Pada penelitian ini Davies-Bouldin Index (DBI) digunakan sebagai metode untuk menentukan jumlah cluster optimal, nilai DBI terendah adalah 1,5 yang terjadi pada 8 cluster. Hal ini menunjukkan bahwa 8 cluster adalah jumlah cluster optimal untuk data yang telah ditransformasi dan siap dilakukan k-means clustering. Setelah melaksanakan proses clustering dengan metode K-Means yang melibatkan pembentukan 8 cluster, untuk menemukan pola dan wawasan dari hasil clustering, ada dua cara yang digunakan pada penelitian ini, pertama membuat heatmap korelasi fitur yang ditampilkan, dapat diperoleh informasi mengenai hubungan antar variabel. Nilai korelasi berkisar antara -0.4 hingga 1.0 dimana nilai positif menunjukkan korelasi positif dan nilai negatif menunjukkan korelasi negatif. Korelasi positif berarti bahwa jika nilai salah satu variabel meningkat, maka nilai variabel lainnya juga cenderung meningkat. Sebaliknya, korelasi negatif berarti bahwa jika nilai salah satu variabel meningkat, maka nilai variabel lainnya cenderung menurun.

Kata Kunci: : K-Means Clustering, Pendaftaran Mahasiswa Baru, Universitas Abdurrab, Seleksi Mahasiswa, Analisis Data

ABSTRACT

Facing the complex dynamics of freshman enrollment, the k-means clustering method was introduced as the main approach. The focus is on Abdurrab University, where various attributes of prospective students are investigated, including gender, parental education, parental income, hometown, province, age, and choice of study program. With the k-means clustering algorithm, the purpose of the study is to uncover the underlying patterns of preferences and characteristics of new student groups. The results of this study provide in-depth insight into the factors that influence the decision to admit new students in the campus environment of Abdurrab University. In this study Davies-Bouldin Index (DBI) was used as a method to determine the optimal number of clusters, the lowest DBI value was 1.5 which occurred in 8 clusters. This shows that 8 clusters is the optimal number of clusters for data that has been transformed and is ready for k-means clustering. After carrying out the clustering process with the K-Means method which involves the formation of 8 clusters, to show patterns and insights from the clustering results, there are two ways used in this study, first make a heatmap of the correlation of features displayed, information can be obtained about the relationship between variables. The correlation value ranges from -0.4 to 1.0 where positive values indicate a positive correlation and negative values indicate a negative correlation. A positive correlation means that if the value of one variable increases, then the value of the other variable also tends to increase. Conversely, negative correlation means that if the value of one variable increases, then the value of the other variable tends to decrease.

Keyword: K-Means Clustering, New Student Admission, Abdurrab University, Student Selection, Data Analysis

PENDAHULUAN

Peningkatan jumlah pendaftar mahasiswa baru di era globalisasi menuntut pengelolaan data penerimaan mahasiswa yang lebih efisien dan terstruktur. Dalam konteks ini, penggunaan metode *clustering* seperti *K-Means* menjadi relevan untuk mengelompokkan data pendaftaran mahasiswa baru. Penelitian ini bertujuan untuk menginvestigasi implementasi algoritma *K-Means clustering* dalam proses pengelompokan data pendaftaran mahasiswa baru di Universitas Abdurrab. Penerapan teknik ini diharapkan dapat memberikan wawasan yang mendalam terkait pola-pola tersembunyi dalam data pendaftaran, sehingga meningkatkan pemahaman institusi terhadap karakteristik calon mahasiswa. Dengan fokus pada studi kasus di Universitas Abdurrab, penelitian ini tidak hanya berkontribusi pada pengembangan metode pengelompokan mahasiswa baru, tetapi juga memberikan pandangan kontekstual terhadap dinamika penerimaan mahasiswa di lingkungan perguruan tinggi tersebut. Dengan demikian, implementasi *K-Means clustering* diharapkan dapat menjadi solusi yang efektif dan efisien dalam mengelola data pendaftaran mahasiswa baru untuk mendukung keputusan strategis di tingkat universitas. Data Mining adalah suatu proses menganalisis pola data yang tersembunyi menurut berbagai perspektif untuk kategorisasi menjadi informasi yang berguna, yang dikumpulkan di area umum, data warehouse untuk analisis yang efisien, algoritma Data Mining, memfasilitasi pengambilan keputusan bisnis, dan informasi lainnya[1]. Dari beberapa definisi tadi dapat disimpulkan bahwa Data mining adalah proses pengumpulan informasi dalam jumlah besar, kemudian dijadikan informasi yang dapat dipergunakan. Metode dari Data Mining yang akan digunakan pada

penelitian ini adalah *clustering*. *Clustering* adalah proses partisi satu set objek data ke dalam himpunan bagian yang disebut dengan *cluster*[2]. *Clustering* merupakan salah satu konsep dalam unsupervised learning. *Clustering* berkaitan dengan menemukan struktur atau pola dalam kumpulan data yang tidak dikategorikan. Algoritma *clustering* akan memproses data dan menemukan *cluster* atau grup alami jika ada dalam data[3]. Kita dapat mengubah berapa banyak *cluster* yang harus diidentifikasi oleh algoritma yang digunakan. Hal Ini memungkinkan kita untuk menyesuaikan rincian banyaknya kelompok data.

Penelitian yang sudah pernah dilakukan dan berhubungan dengan penelitian ini antara lain : Penelitian[4] dengan topik sama perihal Pendaftaran mahasiswa baru. Menggunakan metode regresi linear sederhana. Mendapatkan hasil prediksi Pendaftaran mahasiswa baru dari prodi manajemen dengan hasil pengujian tingkat error sebesar 3,444% atau tingkat akurasi 96,556%. Kemudian penelitian[5] dengan topik sama Penerapan Enterprise Architecture (EA) pada Universitas X Palembang menggunakan TOGAF sebagai panduan telah dievaluasi melalui pengukuran pada empat domain utama: Architecture Business, Architecture Data, Architecture Application, dan Architecture Technology. Evaluasi tersebut melibatkan penilaian berdasarkan pertanyaan dengan skala nilai. Hasilnya menunjukkan bahwa penerapan EA berada pada tingkat baik (skala 2) dengan rentang nilai yang diperoleh. Kesimpulan ini menunjukkan bahwa Universitas X Palembang telah berhasil menerapkan Enterprise Architecture sesuai pedoman TOGAF..

Penelitian ini berfokus pada algoritma *K-Means Clustering*. Beberapa penelitian yang

menggunakan metode *K-Means* adalah [6] Teknik pengelompokan pada data balita di Rokan Hulu menghasilkan dua kelompok, yaitu kelompok dengan status gizi baik dan kelompok dengan status gizi buruk [7] Berdasarkan hasil yang diperoleh dari implementasi metode *k-means clustering*, ditemukan bahwa satu provinsi termasuk dalam kelompok dengan populasi tinggi, yaitu Jawa Timur. Sementara itu, terdapat tiga provinsi yang masuk dalam kelompok dengan populasi sedang, antara lain Sumatera Utara, Jawa Barat, dan Jawa Tengah. Untuk kelompok dengan populasi rendah, terdapat 30 provinsi, termasuk di antaranya Aceh, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kepulauan Bangka Belitung, Kepulauan Riau, DKI Jakarta, DI Yogyakarta, Banten, Bali, NTB, NTT, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat, dan Papua. Beberapa Teknik yang akan digunakan pada penelitian kali ini akan dijabarkan. Data mining adalah proses ekstraksi pola atau informasi yang berguna dari sejumlah besar data. Tujuan dari data mining adalah menemukan hubungan tersembunyi, tren, pola, atau pengetahuan yang dapat membantu pengambilan keputusan yang lebih baik dalam suatu organisasi[8]. Penjelasan lainnya juga mendefinisikan wacana Data Mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning buat mengidentifikasi informasi yang berguna dan bermanfaat dari database besar[9]. Tahapan Data Mining adalah serangkaian langkah atau proses yang dilakukan untuk mendapatkan pengetahuan

yang berharga dari kumpulan data. Tahapan-tahapan tersebut biasanya melibatkan pemahaman tentang data, persiapan data, pemodelan, dan evaluasi. Biasa disebut dengan KDD[10]. *Clustering* adalah teknik dalam Data Mining yang digunakan untuk mengelompokkan objek-objek data yang serupa ke dalam kelompok-kelompok atau klaster-klaster. Tujuan utama dari *clustering* adalah menemukan pola atau struktur yang tersembunyi dalam data tanpa adanya label atau klasifikasi sebelumnya [11]. Dalam proses *clustering*, objek-objek data dikelompokkan berdasarkan kesamaan atau kedekatan antara mereka. Kedekatan antar objek dapat diukur menggunakan metrik atau jarak tertentu, seperti jarak Euclidean atau jarak Manhattan. Objek-objek yang memiliki jarak yang lebih dekat dianggap lebih serupa daripada objek-objek yang memiliki jarak yang lebih jauh. Menurut [12] Algoritma *K-Means Clustering* adalah metode pengelompokan data yang membagi data ke dalam K kelompok berdasarkan kesamaan atribut. Langkah-langkahnya adalah:

Inisialisasi: Pilih K titik awal sebagai pusat kelompok.

Pengelompokan: Atribusikan setiap data ke kelompok terdekat berdasarkan jarak Euclidean atau Manhattan.

Perbarui Pusat Kelompok: Hitung ulang pusat kelompok dengan mengambil rata-rata dari data dalam kelompok.

Iterasi: Ulangi langkah 2 dan 3 hingga tidak ada perubahan pusat kelompok atau mencapai batas iterasi.

Algoritma ini berusaha meminimalkan varians dalam kelompok dan berhenti ketika jarak total tidak berubah secara signifikan. Hasilnya adalah data terbagi menjadi K kelompok yang saling berdekatan. Penelitian memakai Davies-Bouldin Index pernah

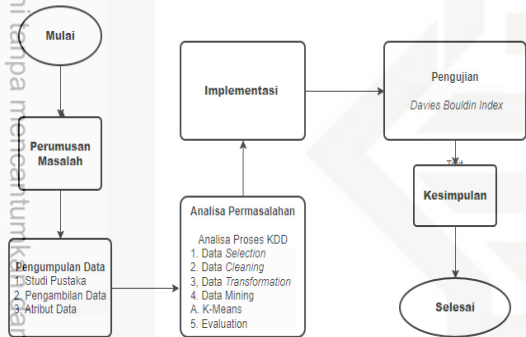
dilakukan oleh [13] Dari analisis data yang telah dilakukan, disimpulkan bahwa penerapan model *K-Means* dengan penilaian menggunakan indeks Davies-Bouldin (DBI) dapat mengoptimalkan jumlah *cluster*.

METODE

Penelitian ini bertujuan untuk menemukan pola-pola yang menunjukkan preferensi dan karakteristik data pendaftaran mahasiswa baru.

Kerangka Kerja Penelitian

Untuk mempermudah dalam pengerjaan penelitian ini, maka dibuatkanlah kerangka kerja penelitian seperti pada gambar 3.1 berikut ini.



Gambar 1. Kerangka Kerja Penelitian

Perumusan Masalah

Perumusan masalah adalah langkah awal dalam merencanakan penelitian. Pada tahap ini melibatkan identifikasi, pemilihan, serta perumusan yang dapat mengahasikan solusi untuk penelitian yang ingin diselesaikan. Rumusan masalahnya adalah bagaimana menerapkan metode *K-Means* untuk mengelompokkan Pendaftaran mahasiswa baru di Universitas Abdurrab .

Pengumpulan Data

Pengumpulan data ini dapat dikategorikan menggunakan data sekunder. Data ini didapat

dari Puskom (pusat komunikasi) Universitas Abdurrab berupa excel dimana berisi 824 data penerimaan mahasiswa tahun 2023. Data tersebut masih mentah dan mesti dilakukan proses data mining. Atribut yang dipergunakan pada penelitian ini adalah atribut data pengelompokan pendaftaran mahasiswa baru Universitas Abdurrab yang meliputi seperti Jenis kelamin, Pendidikan ayah & ibu, Penghasilan ayah & ibu, Kota/ Kabupaten asal, provinsi, usia dan prodi yang akan dipilih.

Tabel 1. Atribut Data

No	Atribut	Keterangan
1	Program Studi	Nama program studi pilihan
2	Jenis Kelamin	Jenis kelamin mahasiswa
3	Pendidikan ayah & ibu	Tingkat pendidikan yang telah dicapai oleh orang tua
4	kota_kab	Nama kota asal
5	provinsi	Nama provinsi asal
6	Pendapatan ayah & ibu	Pendapatan perbulan orang tua
7	Usia	Usia mahasiswa

Analisa Proses KDD

a. Data Cleaning

Pembersihan terhadap data-data artinya langkah awal buat data yang kita gunakan, di tahapan pembersihan data kita akan melakukan proses mendeteksi, memperbaiki dan menghapus data yang tidak diharapkan pada database. Pembersihan data pada data Pendaftaran mahasiswa yang akan di teliti dikarenakan data mentah dari suatu tempat yang tidak dapat untuk langsung di analisis, data kosong (*missing value*), pengecekan data duplikat, inkonsisten data, dan ketidaksesuaian data (*outlier*). Dengan menggunakan pembersihan dapat mempengaruhi peforma dari teknik Data Mining karena data akan berkurang setelah

2. Dilarang mengummumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

diproses. Data yang awalnya 824 menjadi 823 dikarenakan terdapat satu baris data yang *missing value* yakni datanya kosong.

b. Data Selection

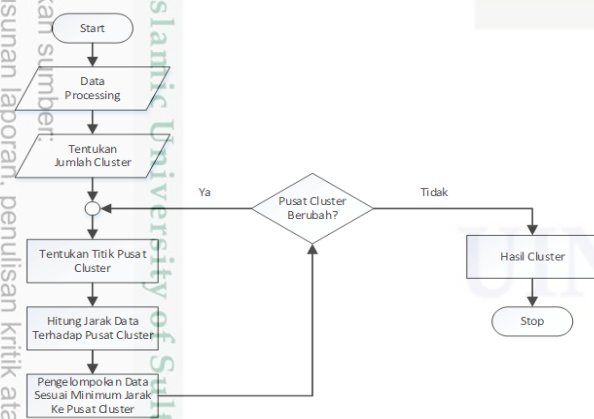
Pada tahap ini dilakukan tahapan seleksi data. Atribut yang digunakan dari data ini hanyalah Jenis kelamin, Pendidikan ayah & ibu, Penghasilan ayah & ibu, Kota/ Kabupaten asal, provinsi, usia dan prodi yang akan dipilih.

c. Transformation

Karena metode *K-Means Clustering* merupakan metode yang mampu digunakan apabila data berupa angka, proses tranformasi data merupakan mengubah data atribut yang selain angka ke dalam nilai angka agar data dapat diolah menggunakan algoritma *K-Means Clustering*, yaitu Kota/ Kabupaten asal, provinsi, dan prodi yang akan dipilih. Data tersebut diubah dari *string* menjadi *numeric* agar dapat diolah oleh algoritma *K-Means Clustering*

d. Data Mining

Setelah data melalui tahap pre-processing dan tahap tranformation maka data telah bisa diolah menggunakan metode *K-Means Clustering*. Metode *K-Means Clustering* merupakan proses pengelompokan data ke dalam sebuah *cluster* menggunakan titik pusat berbeda setiap *cluster*. Tahap algoritma *K-Means Clustering* dapat di lihat pada gambar



Gambar 2. Tahap Algoritma *K-Means*

e. Interpretation atau Evaluation

Proses pencarian pola atau informasi baru yang menarik dan bermanfaat di suatu kumpulan data yang terpilih dari algoritma *K-Means*, berupa visualisasi *heatmap* dan *characteristics* dari hasil *cluster*

Implementasi

Pada [14] Algoritma *K-Means* merupakan algoritma yang membutuhkan parameter input sebanyak *k* dan membagi sekumpulan *n* objek kedalam *k cluster* sehingga tingkat kemiripan antar anggota dalam suatu *cluster* tinggi sedangkan tingkat kemiripan dengan anggota pada *cluster* lain sangat rendah.

Rumus jarak yang umum digunakan dalam algoritma *k-means* adalah jarak Euclidean. Misalkan *x* adalah titik data, *c* adalah centroid, dan *d* adalah jarak antara *x* dan *c*, rumus jarak Euclidean adalah:

$$d(x,c) = \sqrt{((x_1 - c_1)^2 + (x_2 - c_2)^2 + \dots + (x_n - c_n)^2)}$$

Persamaan 1 Rumus Jarak Euclidean

Di sini, (*x*₁, *x*₂, ..., *x*_{*n*}) dan (*c*₁, *c*₂, ..., *c*_{*n*}) adalah koordinat dari titik data dan centroid dalam ruang *n*-dimensi. Jarak yang terpendek antara centroid dengan dokumen menentukan posisi *cluster* suatu dokumen. Adapun rumus iterasi lainnya didefinisikan sebagai berikut :

$$(i) = x_1 + x_2 + x_3 + \dots + x_n \sum x$$

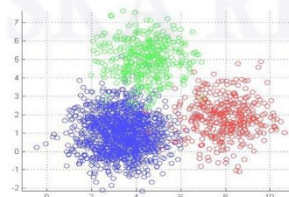
Persamaan 2 Rumus iterasi

Keterangan :

*X*₁ = Nilai data record ke-1

*X*₂ = Nilai data record ke-2

$\sum x$ = jumlah data record



Gambar 3. *K-means clustering*

2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



Pengujian

Tahap pengujian ini merupakan langkah untuk melakukan evaluasi terhadap software yang telah selesai dikembangkan pada tahap implementasi sebelumnya dan menguji hasilnya.

Menurut [15] Davies-Bouldin Index (DBI) adalah metrik evaluasi yang digunakan untuk mengukur kualitas *clustering* dalam algoritma *K-Means*. Nilai DBI yang lebih rendah menunjukkan *clustering* yang lebih baik, dengan *cluster* yang lebih terpisah dan kompak. DBI menggabungkan intra-*cluster* similarity dan inter-*cluster* dissimilarity. Untuk menghitung DBI, langkah-langkahnya adalah:

Lakukan *K-Means clustering* dan tentukan pusat *cluster* untuk setiap *cluster*.

Hitung jarak antara pusat *cluster* dan titik data dalam *cluster*.

Hitung rata-rata jarak untuk setiap *cluster*, menunjukkan seberapa kompaknya *cluster*.

Hitung dissimilarity antara pusat *cluster* menggunakan metrik jarak.

Temukan *cluster* lain yang memiliki dissimilarity minimum dengan setiap *cluster*.

Hitung DBI dengan menggunakan rumus:

$$DBI = (1 / k) * \Sigma(\text{maximum Rij})$$

Berdasarkan rumus diatas k adalah jumlah *cluster* dan Rij adalah rasio dissimilarity antara *cluster*. Semakin kecil nilai DBI, semakin baik *clustering*. DBI sering digunakan bersama metrik lain untuk evaluasi dan pemilihan jumlah *cluster* optimal.

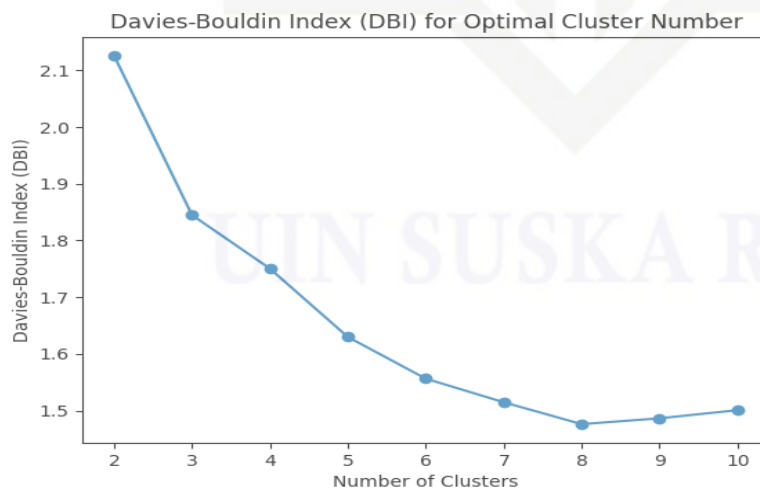
Kesimpulan

melibatkan rekapitulasi temuan, interpretasi hasil, kaitan dengan tujuan penelitian dan literatur, implikasi praktis dan teoretis, serta kesimpulan akhir dengan dorongan untuk diskusi lanjutan

HASIL

Davies-Bouldin Index (DBI).

Pada penelitian ini Davies-Bouldin Index (DBI) digunakan sebagai metode untuk menentukan jumlah cluster optimal.



Gambar 4. Davies-Bouldin Index (DBI)

2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

a. Pengutipan harus disertai dengan informasi yang benar-benar, akurat, dan dapat dipertanggungjawabkan, penulisan kritik atau tinjauan suatu masalah.

c. Pengutipan harus disertai dengan informasi yang benar-benar, akurat, dan dapat dipertanggungjawabkan, penulisan kritik atau tinjauan suatu masalah.

d. Pengutipan harus disertai dengan informasi yang benar-benar, akurat, dan dapat dipertanggungjawabkan, penulisan kritik atau tinjauan suatu masalah.

e. Pengutipan harus disertai dengan informasi yang benar-benar, akurat, dan dapat dipertanggungjawabkan, penulisan kritik atau tinjauan suatu masalah.

f. Pengutipan harus disertai dengan informasi yang benar-benar, akurat, dan dapat dipertanggungjawabkan, penulisan kritik atau tinjauan suatu masalah.

g. Pengutipan harus disertai dengan informasi yang benar-benar, akurat, dan dapat dipertanggungjawabkan, penulisan kritik atau tinjauan suatu masalah.

h. Pengutipan harus disertai dengan informasi yang benar-benar, akurat, dan dapat dipertanggungjawabkan, penulisan kritik atau tinjauan suatu masalah.

i. Pengutipan harus disertai dengan informasi yang benar-benar, akurat, dan dapat dipertanggungjawabkan, penulisan kritik atau tinjauan suatu masalah.

j. Pengutipan harus disertai dengan informasi yang benar-benar, akurat, dan dapat dipertanggungjawabkan, penulisan kritik atau tinjauan suatu masalah.

k. Pengutipan harus disertai dengan informasi yang benar-benar, akurat, dan dapat dipertanggungjawabkan, penulisan kritik atau tinjauan suatu masalah.

l. Pengutipan harus disertai dengan informasi yang benar-benar, akurat, dan dapat dipertanggungjawabkan, penulisan kritik atau tinjauan suatu masalah.

m. Pengutipan harus disertai dengan informasi yang benar-benar, akurat, dan dapat dipertanggungjawabkan, penulisan kritik atau tinjauan suatu masalah.

n. Pengutipan harus disertai dengan informasi yang benar-benar, akurat, dan dapat dipertanggungjawabkan, penulisan kritik atau tinjauan suatu masalah.

o. Pengutipan harus disertai dengan informasi yang benar-benar, akurat, dan dapat dipertanggungjawabkan, penulisan kritik atau tinjauan suatu masalah.

p. Pengutipan harus disertai dengan informasi yang benar-benar, akurat, dan dapat dipertanggungjawabkan, penulisan kritik atau tinjauan suatu masalah.

q. Pengutipan harus disertai dengan informasi yang benar-benar, akurat, dan dapat dipertanggungjawabkan, penulisan kritik atau tinjauan suatu masalah.

r. Pengutipan harus disertai dengan informasi yang benar-benar, akurat, dan dapat dipertanggungjawabkan, penulisan kritik atau tinjauan suatu masalah.

s. Pengutipan harus disertai dengan informasi yang benar-benar, akurat, dan dapat dipertanggungjawabkan, penulisan kritik atau tinjauan suatu masalah.

t. Pengutipan harus disertai dengan informasi yang benar-benar, akurat, dan dapat dipertanggungjawabkan, penulisan kritik atau tinjauan suatu masalah.

u. Pengutipan harus disertai dengan informasi yang benar-benar, akurat, dan dapat dipertanggungjawabkan, penulisan kritik atau tinjauan suatu masalah.

v. Pengutipan harus disertai dengan informasi yang benar-benar, akurat, dan dapat dipertanggungjawabkan, penulisan kritik atau tinjauan suatu masalah.

w. Pengutipan harus disertai dengan informasi yang benar-benar, akurat, dan dapat dipertanggungjawabkan, penulisan kritik atau tinjauan suatu masalah.

x. Pengutipan harus disertai dengan informasi yang benar-benar, akurat, dan dapat dipertanggungjawabkan, penulisan kritik atau tinjauan suatu masalah.

y. Pengutipan harus disertai dengan informasi yang benar-benar, akurat, dan dapat dipertanggungjawabkan, penulisan kritik atau tinjauan suatu masalah.

z. Pengutipan harus disertai dengan informasi yang benar-benar, akurat, dan dapat dipertanggungjawabkan, penulisan kritik atau tinjauan suatu masalah.

aa. Pengutipan harus disertai dengan informasi yang benar-benar, akurat, dan dapat dipertanggungjawabkan, penulisan kritik atau tinjauan suatu masalah.

ab. Pengutipan harus disertai dengan informasi yang benar-benar, akurat, dan dapat dipertanggungjawabkan, penulisan kritik atau tinjauan suatu masalah.

ac. Pengutipan harus disertai dengan informasi yang benar-benar, akurat, dan dapat dipertanggungjawabkan, penulisan kritik atau tinjauan suatu masalah.

ad. Pengutipan harus disertai dengan informasi yang benar-benar, akurat, dan dapat dipertanggungjawabkan, penulisan kritik atau tinjauan suatu masalah.

ae. Pengutipan harus disertai dengan informasi yang benar-benar, akurat, dan dapat dipertanggungjawabkan, penulisan kritik atau tinjauan suatu masalah.

af. Pengutipan harus disertai dengan informasi yang benar-benar, akurat, dan dapat dipertanggungjawabkan, penulisan kritik atau tinjauan suatu masalah.

ag. Pengutipan harus disertai dengan informasi yang benar-benar, akurat, dan dapat dipertanggungjawabkan, penulisan kritik atau tinjauan suatu masalah.

ah. Pengutipan harus disertai dengan informasi yang benar-benar, akurat, dan dapat dipertanggungjawabkan, penulisan kritik atau tinjauan suatu masalah.

ai. Pengutipan harus disertai dengan informasi yang benar-benar, akurat, dan dapat dipertanggungjawabkan, penulisan kritik atau tinjauan suatu masalah.

aj. Pengutipan harus disertai dengan informasi yang benar-benar, akurat, dan dapat dipertanggungjawabkan, penulisan kritik atau tinjauan suatu masalah.

ak. Pengutipan harus disertai dengan informasi yang benar-benar, akurat, dan dapat dipertanggungjawabkan, penulisan kritik atau tinjauan suatu masalah.

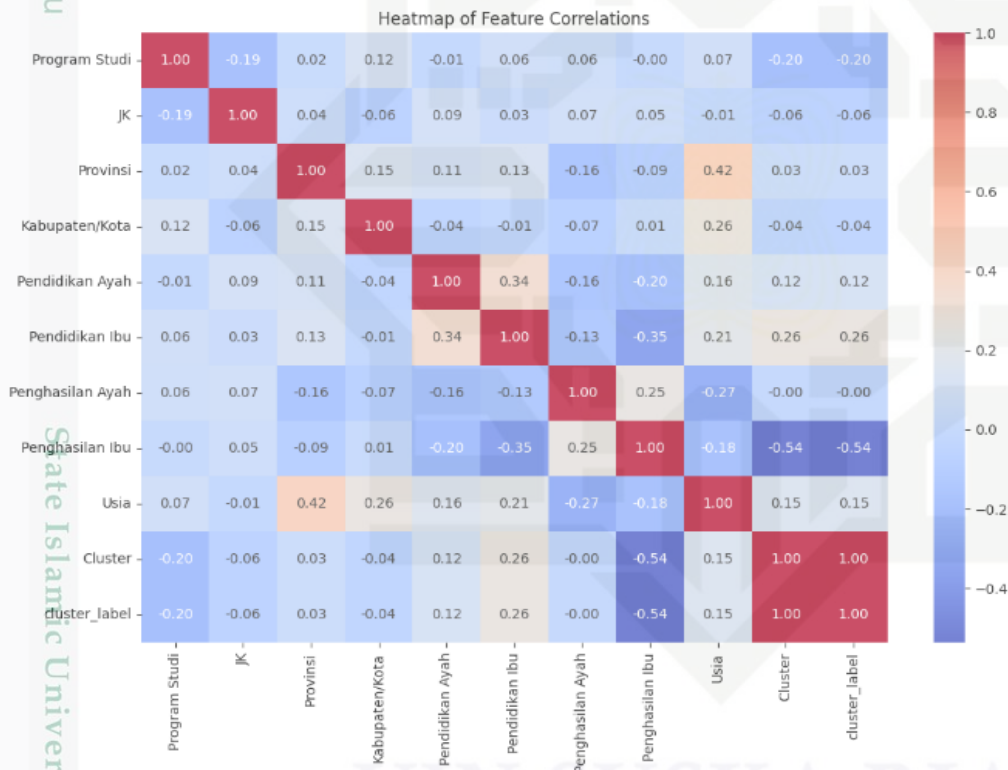
al. Pengutipan harus disertai dengan informasi yang benar-benar, akurat, dan dapat dipertanggungjawabkan, penulisan kritik atau tinjauan suatu masalah.

Berdasarkan Gambar 4, nilai DBI terendah adalah 1,5 yang terjadi pada 8 cluster. Hal ini menunjukkan bahwa 8 cluster adalah jumlah cluster optimal untuk data yang telah ditransformasi dan siap dilakukan k-means clustering.

Heatmap of Correlations

Setelah melaksanakan proses clustering dengan metode K-Means yang melibatkan pembentukan 8 cluster, untuk menemukan pola dan wawasan dari hasil clustering, ada dua cara yang digunakan pada penelitian ini, pertama membuat heatmap korelasi

fitur yang ditampilkan, dapat diperoleh informasi mengenai hubungan antar variabel. Nilai korelasi berkisar antara -0.4 hingga 1.0 dimana nilai positif menunjukkan korelasi positif dan nilai negatif menunjukkan korelasi negatif. Korelasi positif berarti bahwa jika nilai salah satu variabel meningkat, maka nilai variabel lainnya juga cenderung meningkat. Sebaliknya, korelasi negatif berarti bahwa jika nilai salah satu variabel meningkat, maka nilai variabel lainnya cenderung menurun.



Gambar 5. Heatmap of Correlations

Beberapa temuan penting dari heatmap ini adalah:

- a. Terdapat korelasi positif yang kuat antara Penghasilan Ayah dan Penghasilan Ibu (0.86). Hal ini menunjukkan bahwa orang tua yang

memiliki penghasilan tinggi cenderung memiliki pasangan yang juga memiliki penghasilan tinggi.

- b. Usia memiliki korelasi positif dengan Jenis Kelamin (0.28) dan Provinsi (0.25) namun

2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

memiliki korelasi negatif dengan Cluster (-0.23). Hal ini menunjukkan bahwa usia berpengaruh terhadap jenis kelamin dan provinsi asal responden, serta kelompok atau segmen yang mereka masuki.

c. Pendidikan Ayah memiliki korelasi negatif dengan Jenis Kelamin (-0.21). Hal ini menunjukkan bahwa ayah yang memiliki pendidikan tinggi cenderung memiliki anak perempuan.

d. Program Studi memiliki korelasi positif

dengan Kabupaten/Kota (0.22) dan Pendidikan Ibu (0.21). Hal ini menunjukkan bahwa program studi yang dipilih responden dipengaruhi oleh lokasi dan tingkat pendidikan ibu mereka.

e. Jenis Kelamin memiliki korelasi negatif dengan Pendidikan Ibu (-0.18) dan Penghasilan Ibu (-0.17). Hal ini menunjukkan bahwa ibu yang memiliki pendidikan dan penghasilan tinggi cenderung memiliki anak laki-laki.

Cluster Characteristics

Tabel 2 Cluster Characteristics

Program Studi	Jenis Kelamin	Provinsi	Kabupaten/Kota	Pendidikan Ayah	Pendidikan Ibu	Penghasilan Ayah	Penghasilan Ibu	Usia	Cluster
0,490473	0,408257	0,599083	0,325786	0,688613	0,597477	0,766972	0,806422	0,047655	0
0,236048	0,446078	0,607843	0,821861	0,731257	0,653799	0,498039	0,472549	0,064542	1
0,762139	0,411348	0,597872	0,344651	0,780142	0,768617	0,717021	0,256028	0,051024	2
0,769231	0,25974	0,579221	0,429124	0,803667	0,806818	0,192208	0,241558	0,056277	3
0,187257	0,383838	0,59596	0,298732	0,78669	0,761995	0,223232	0,229293	0,054714	4
0,780543	0,22549	0,589216	0,881727	0,692618	0,713235	0,666667	0,368627	0,068627	5
0,619658	0,444444	0,9	0,960993	0,98366	0,982639	0,061111	0,055556	1	6
0,2386	0,422156	0,591617	0,312269	0,77633	0,782934	0,734132	0,224551	0,05173	7

Kedua adalah *characteristic* setiap *cluster* yang dapat menghasilkan informasi menarik yaitu :

a. *Cluster 0* memiliki nilai rata-rata tertinggi untuk Penghasilan Ayah dan Penghasilan Ibu, yaitu 0.766972 dan 0.806422. Hal ini menunjukkan bahwa *cluster 0* terdiri dari responden yang memiliki orang tua dengan penghasilan yang tinggi. *Cluster 0* juga memiliki nilai rata-rata terendah untuk Usia, yaitu 0.047655. Hal ini mengindikasikan bahwa *cluster 0* terdiri dari responden yang berusia muda.

b. *Cluster 1* memiliki nilai rata-rata tertinggi

untuk Kabupaten/Kota, yaitu 0.821861. Hal ini menunjukkan bahwa *cluster 1* terdiri dari responden yang berasal dari kabupaten/kota yang berbeda-beda. *Cluster 1* juga memiliki nilai rata-rata terendah untuk Program Studi, yaitu 0.236048. Hal ini mengindikasikan bahwa *cluster 1* terdiri dari responden yang memiliki program studi yang beragam.

c. *Cluster 2* memiliki nilai rata-rata tertinggi untuk Program Studi, yaitu 0.762139. Hal ini menunjukkan bahwa *cluster 2* terdiri dari responden yang memiliki program studi yang sama atau sejenis. *Cluster 2* juga memiliki nilai

rata-rata terendah untuk Penghasilan Ibu, yaitu 0.256028. Hal ini mengindikasikan bahwa *cluster* 2 terdiri dari responden yang memiliki ibu dengan penghasilan yang rendah.

d. *Cluster* 3 memiliki nilai rata-rata tertinggi untuk Pendidikan Ayah dan Pendidikan Ibu, yaitu 0.803667 dan 0.806818. Hal ini menunjukkan bahwa *cluster* 3 terdiri dari responden yang memiliki orang tua dengan tingkat pendidikan yang tinggi. *Cluster* 3 juga memiliki nilai rata-rata terendah untuk JK, yaitu 0.259740. Hal ini mengindikasikan bahwa *cluster* 3 terdiri dari responden yang mayoritas berjenis kelamin laki-laki.

e. *Cluster* 4 memiliki nilai rata-rata tertinggi untuk Pendidikan Ayah dan Pendidikan Ibu, yaitu 0.786690 dan 0.761995. Hal ini menunjukkan bahwa *cluster* 4 terdiri dari responden yang memiliki orang tua dengan tingkat pendidikan yang tinggi. *Cluster* 4 juga memiliki nilai rata-rata terendah untuk Penghasilan Ayah dan Penghasilan Ibu, yaitu 0.223232 dan 0.229293. Hal ini mengindikasikan bahwa *cluster* 4 terdiri dari responden yang memiliki orang tua dengan penghasilan yang rendah.

f. *Cluster* 5 memiliki nilai rata-rata tertinggi untuk Kabupaten/Kota, yaitu 0.881727. Hal ini menunjukkan bahwa *cluster* 5 terdiri dari responden yang berasal dari kabupaten/kota yang berbeda-beda. *Cluster* 5 juga memiliki nilai rata-rata terendah untuk JK, yaitu 0.225490. Hal ini mengindikasikan bahwa *cluster* 5 terdiri dari responden yang mayoritas berjenis kelamin laki-laki.

g. *Cluster* 6 memiliki nilai rata-rata tertinggi untuk Provinsi, Kabupaten/Kota, Pendidikan Ayah, dan Pendidikan Ibu, yaitu 0.900000, 0.960993, 0.983660, dan 0.982639. Hal ini menunjukkan bahwa *cluster* 6 terdiri dari responden yang berasal dari provinsi dan kabupaten/kota yang sama, dan memiliki orang tua dengan tingkat pendidikan yang sangat

tinggi. *Cluster* 6 juga memiliki nilai rata-rata tertinggi untuk Usia, yaitu 1.000000. Hal ini mengindikasikan bahwa *cluster* 6 terdiri dari responden yang berusia tua.

h. *Cluster* 7 memiliki nilai rata-rata tertinggi untuk Penghasilan Ayah, yaitu 0.734132. Hal ini menunjukkan bahwa *cluster* 7 terdiri dari responden yang memiliki ayah dengan penghasilan yang tinggi. *Cluster* 7 juga memiliki nilai rata-rata terendah untuk Penghasilan Ibu, yaitu 0.224551. Hal ini mengindikasikan bahwa *cluster* 7 terdiri dari responden yang memiliki ibu dengan penghasilan yang rendah.

KESIMPULAN

Dari hasil analisis *clustering* yang telah dilakukan pada data penerimaan mahasiswa baru, didapat kesimpulan sebagai berikut :

- Penggunaan *K-Means clustering* dapat efektif dalam mengelompokkan responden berdasarkan karakteristik demografis dan sosial ekonomi mereka. Visualisasi hasil *clustering* dengan *heatmap* dapat menggambarkan dengan jelas korelasi antar variabel. Analisis korelasi ini memberikan wawasan penting yang dapat membimbing penelitian lebih lanjut terkait karakteristik dan pola yang mungkin terdapat dalam data.
- Characteristics* tiap *cluster* dapat digunakan untuk menggambarkan profil dari responden yang tergabung dalam *cluster* tersebut. Profil responden dapat memberikan wawasan penting bagi penelitian lebih lanjut mengenai preferensi, kebutuhan, dan perilaku dari responden. *Characteristics* tiap *cluster* juga dapat dijadikan dasar untuk melakukan segmentasi pasar, strategi pemasaran, dan rekomendasi produk atau layanan yang sesuai dengan *cluster* tersebut.

DAFTAR PUSTAKA

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 2. Dilarang mengutip sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.
 3. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 4. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
- [1] M. Arhami and M. Nasir, *Data Mining - Algoritma dan Implementasi*. Penerbit Andi, 2020. [Online]. Available: https://books.google.co.id/books?hl=id&lr=&id=AtcCEAAQBAJ&oi=fnd&pg=PP1&dq=ARHAMI+NASIR&ots=hDhlOM5Wr8&sig=jqLWHIAwU9vMbVwAEHH9hbxJKj8&redir_esc=y#v=onepage&q=ARHAMINASIR&f=false
 - [2] R. K. Dinata, S. Safwandi, N. Hasdyna, and N. Azizah, "Analisis K-Means Clustering pada Data Sepeda Motor," *INFORMAL Informatics J.*, vol. 5, no. 1, p. 10, 2020, doi: 10.19184/isj.v5i1.17071.
 - [3] S. Aulia, "Klasterisasi Pola Penjualan Pestisida Menggunakan Metode K-Means Clustering (Studi Kasus Di Toko Juanda Tani Kecamatan Hutabayu Raja)," *Djtechno J. Teknol. Inf.*, vol. 1, no. 1, pp. 1–5, 2021, doi: 10.46576/djtechno.v1i1.964.
 - [4] A. Muhartini, O. Sahroni, S. Rahmawati, Dwi, T. Febrianti, and I. Mahuda, "Analisis Peramalan Jumlah Penerimaan Mahasiswa Baru Dengan Menggunakan Metode Regresi Linear Sederhana Ajeng," *J. Bayesian J. Ilm. Stat. dan Ekon.*, vol. 1, 2021, doi: 10.30604/jika.v7i2.1507.
 - [5] L. Lathifah, "Penerapan Enterprise Architecture pada Penerimaan Mahasiswa Baru menggunakan TOGAF di Universitas X Palembang," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 7, no. 3, pp. 647–655, 2020, doi: 10.35957/jatisi.v7i3.565.
 - [6] D. Dona and M. Rifqi, "Penerapan Metode K-Means Clustering Untuk Menentukan Status Gizi Baik Dan Gizi Buruk Pada Balita (Studi Kasus Kabupaten Rokan Hulu)," *Rabit J. Teknol. dan Sist. Inf. Univrab*, vol. 7, no. 2, pp. 179–191, 2022, doi: 10.36341/rabit.v7i2.2171.
 - [7] E. Ramadanti and M. Muslih, "Penerapan Data Mining Algoritma K-Means Clustering Pada Populasi Ayam Petelur Di Indonesia," 2022 doi: 10.36341/rabit.v7i1.2155.
 - [8] N. Sunanto and G. Falah, "Penerapan Algoritma C4.5 Untuk Membuat Model Prediksi Pasien Yang Mengidap Penyakit Diabetes," *Rabit J. Teknol. dan Sist. Inf. Univrab*, vol. 7, no. 2, pp. 208–216, 2022, doi: 10.36341/rabit.v7i2.2435.
 - [9] "Undang Undang Republik Indonesia Nomor 20 Tahun 2003 Tentang Sistem Pendidikan Nasional." 2003.
 - [10] E. Buulolo, *Data Mining Untuk Perguruan Tinggi*. Sleman: CV BUDI UTAMA, 2020. [Online]. Available: https://books.google.co.id/books?hl=id&lr=&id=K_SDwAAQBAJ&oi=fnd&pg=PP1&dq=Buulolo+2020&ots=Kezq5Rx3Qk&sig=O70NsVBI38e-bhW-3r7uQB-0EHU&redir_esc=y#v=onepage&q=Buulolo2020&f=false
 - [11] I. Nuryani and D. Darwis, "Analisis Clustering Pada Pengguna Brand Hp Menggunakan Metode K-Means," *Pros. Semin. Nas. Ilmu Komput.*, vol. 1, no. 1, p. 2021, 2021.
 - [12] A. Sulistiyawati and E. Supriyanto, "Implementasi Algoritma K-means Clustering dalam Penentuan Siswa Kelas Unggulan," *J. Tekno Kompak*, vol. 15, no. 2, p. 25, 2021, doi: 10.33365/jtk.v15i2.1162.
 - [13] E. Muningsih, I. Maryani, and V. R. Handayani, "Penerapan Metode K-Means dan Optimasi Jumlah Cluster dengan Index Davies Bouldin untuk Clustering Propinsi Berdasarkan Potensi Desa," *J. Sains dan Manaj.*, vol. 9, no. 1, p. 96, 2021, [Online]. Available: www.bps.go.id
 - [14] F. Yunita, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Pada Penerimaan Mahasiswa Baru," *Sistemasi*, vol. 7, no. 3, p. 238, 2018, doi: 10.32520/stmsi.v7i3.388.
 - [15] W. Gie and D. Jollyta, "Perbandingan Euclidean dan Manhattan Untuk Optimasi Cluster Menggunakan Davies Bouldin Index: Status Covid-19 Wilayah Riau," *Pros. Semin. Nas. Ris. Dan Inf. Sci.*, vol. 2, no. April, pp. 187–191, 2020.