

Implementation of Data Clustering for Informatics Engineering Study Program Students at STMIK Widya Cipta Dharma Using the K-Means Method

Eksakta Ramang Kridho Haryono¹⁾, Siti Lailiyah²⁾, dan Muhammad Ibnu Sa'ad³⁾

^{1,2,3}Tenik Informatika, STMIK Widya Cipta Dharma

^{1,2,3}Jl. M.Yamin No. 25, Samarinda, 75123

E-mail: 2043055@wicida.ac.id¹⁾, lail.59a@wicida.ac.id²⁾, saad@wicida.ac.id³⁾

ABSTRACT

This research aims to cluster student data in the Informatics Engineering Study Program at STMIK Widya Cipta Dharma Samarinda using the K-Means Clustering method. The background of this study lies in the underutilization of the institution's large volume of academic and administrative data, which has not been effectively analyzed to support decision-making processes. This study applies the K-Means algorithm to group students based on similarities in academic and personal attributes. The dataset consists of records from 2019 to 2023 and includes variables such as Student ID Number, Semester, Gender, Religion, Marital Status, Employment Status, Occupation, Name of School of Origin, School Major, Province, City, District, Village, Postal Code, Father's Occupation, Mother's Occupation, Academic Year, GPA, Competition Name, Champion Rank, Registration Source, Reason for Choosing STMIK, Favorite Courses, Less Favorite Courses, and Hobbies. The optimal number of clusters is determined using the Elbow Method. The implementation is carried out using Jupyter Notebook, and the results are deployed into a web-based interactive platform. The results of this implementation indicate that the K-Means algorithm can effectively identify student groupings based on both academic performance and extracurricular involvement. These groupings provide valuable insights for developing tailored learning strategies, refining academic policies, supporting promotional efforts, and optimizing educational resources at STMIK Widya Cipta Dharma, ultimately contributing to the improvement of academic quality and institutional planning.

Keywords: Informatics Engineering, student clustering, K-Means

Implementasi Clustering Data Mahasiswa Program Studi Teknik Informatika STMIK Widya Cipta Dharma Menggunakan Metode K-Means

ABSTRAK

Penelitian ini bertujuan untuk mengelompokkan data mahasiswa pada Program Studi Teknik Informatika STMIK Widya Cipta Dharma Samarinda menggunakan metode K-Means Clustering. Latar belakang penelitian ini adalah belum optimalnya pemanfaatan data akademik dan administratif yang tersedia dalam jumlah besar, sehingga belum mampu memberikan dukungan yang maksimal dalam proses pengambilan keputusan. Penelitian ini menerapkan algoritma K-Means untuk mengelompokkan mahasiswa berdasarkan kesamaan atribut akademik dan non-akademik. Data yang digunakan mencakup tahun ajaran 2019 hingga 2023, dengan variabel seperti Nomor Induk Mahasiswa, Semester, Jenis Kelamin, Agama, Status Perkawinan, Status Bekerja, Pekerjaan, Nama Asal Sekolah, Jurusan Sekolah, Provinsi Asal, Kota/Kabupaten, Kecamatan, Kelurahan/Desa, Kode Pos, Pekerjaan Ayah, Pekerjaan Ibu, Tahun Ajaran, IPK, Nama Kompetisi, Peringkat Juara, Sumber Informasi Pendaftaran, Alasan Memilih STMIK, Mata Kuliah Favorit, Mata Kuliah Kurang Favorit, dan Hobi. Penentuan jumlah cluster optimal dilakukan menggunakan metode Elbow. Implementasi dilakukan menggunakan Jupyter Notebook, dan hasilnya diintegrasikan ke dalam platform web interaktif. Hasil implementasi menunjukkan bahwa algoritma K-Means mampu mengelompokkan mahasiswa secara efektif berdasarkan karakteristik akademik dan aktivitas ekstrakurikuler. Hasil pengelompokan ini memberikan wawasan yang bermanfaat untuk merancang strategi pembelajaran yang sesuai, menyusun kebijakan akademik, mendukung strategi promosi, dan mengoptimalkan sumber daya pendidikan di STMIK Widya Cipta Dharma, sehingga dapat meningkatkan kualitas akademik dan perencanaan institusi secara keseluruhan.

Kata Kunci: Teknik Informatika, Clustering Mahasiswa, K-Means

1. PENDAHULUAN

Sekolah Tinggi Manajemen Informatika dan Komputer Widya Cipta Dharma (STMIK WiCiDa) di

Kota Samarinda memiliki tiga program studi utama, yaitu Teknik Informatika, Sistem Informasi, dan Bisnis Digital. Sebagai institusi pendidikan tinggi berbasis

teknologi, STMIK WiCiDa menyimpan data akademik dan administrasi dalam jumlah yang cukup besar, mencakup 822 data mahasiswa aktif. Data tersebut terdiri atas data biodata mahasiswa, rencana studi, hasil studi berupa nilai dan indeks prestasi, serta informasi pendukung lainnya. Sayangnya, potensi besar dari data tersebut belum dimanfaatkan secara optimal untuk mendukung pengambilan keputusan strategis, evaluasi program studi, dan peningkatan kualitas proses pembelajaran.

Permasalahan utama yang dihadapi adalah belum adanya sistem analisis data yang mampu menyajikan informasi tersembunyi (*hidden pattern*) dari data mahasiswa yang ada, padahal informasi tersebut sangat penting bagi manajemen program studi untuk menyusun kebijakan berbasis data. Salah satu pendekatan yang dapat digunakan untuk menjawab tantangan tersebut adalah dengan menerapkan teknik *clustering* data mahasiswa guna memperoleh pemetaan karakteristik mahasiswa secara lebih sistematis.

Beberapa penelitian sebelumnya telah memanfaatkan metode *K-Means Clustering* dalam bidang pendidikan untuk mengelompokkan mahasiswa berdasarkan nilai, keaktifan, atau perilaku akademik. Namun, penerapan secara spesifik pada lingkungan Program Studi Teknik Informatika STMIK WiCiDa dengan menggunakan kombinasi atribut seperti mata kuliah wajib, mata kuliah pilihan, dan kemampuan algoritmik, masih sangat terbatas. Inilah yang menjadi celah kebaruan (*novelty*) dari penelitian ini.

Kontribusi utama dari penelitian ini adalah menyediakan informasi berbasis data yang dapat digunakan oleh pengelola program studi untuk merancang strategi pembelajaran yang lebih adaptif, serta mendukung evaluasi dan peningkatan kualitas mahasiswa secara menyeluruh. Sebagai solusi, penelitian ini mengimplementasikan metode *K-Means*, yaitu algoritma *clustering non-hierarki* yang membagi data ke dalam beberapa cluster berdasarkan kedekatan nilai atribut yang digunakan. Atribut yang digunakan dalam penelitian ini antara lain adalah nilai mata kuliah algoritma dan struktur data, mata kuliah wajib, dan mata kuliah pilihan, yang diharapkan dapat menghasilkan tiga kelompok utama, yaitu Cluster 1 (C1), Cluster 2 (C2), dan Cluster 3 (C3).

Melalui pendekatan ini, diharapkan Program Studi Teknik Informatika dapat memperoleh pemahaman yang lebih mendalam mengenai profil akademik mahasiswanya sehingga mampu meningkatkan mutu layanan pembelajaran serta mendukung perencanaan akademik yang lebih tepat sasaran.

2. RUANG LINGKUP

2.1 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka rumusan masalah penelitian ini ialah, Bagaimana hasil dari *Clustering* menggunakan metode *K-Means* pada data Mahasiswa Teknik Informatika STMIK Widya Cipta Dharma?

2.2 Batasan Masalah

Batasan Masalah pada penelitian ini ialah:

1. Teknik yang digunakan pada penelitian ini ialah Teknik *Clustering* dengan menggunakan metode *K-Means* dalam mengelompokkan data mahasiswa Program Studi Teknik Informatika STMIK Widya Cipta Dharma.
2. Data Mahasiswa yang menjadi objek penelitian adalah mahasiswa program studi Teknik Informatika STMIK Widya Cipta Dharma tahun 2019 sampai tahun 2023.
3. Variabel yang digunakan meliputi biodata diri dan akademik mahasiswa Program Studi Teknik Informatika yang meliputi ialah, Nama, NIM, jenis kelamin, semester, agama, Status Perkawinan, Status Bekerja, Pekerjaan, Nama Asal Sekolah, Jurusan Sekolah, Asal Provinsi, Asal Kota/Kabupaten, asal Kecamatan, Asal Kelurahan/Desa, No Kode Pos, Pekerjaan Ayah, Pekerjaan Ibu, Tahun ajaran masuk, IPK, Nama Komepetisi/lomba/kegiatan, Tingkat nasional yang diikuti, Peringkat juara, Informasi pendaftaran STMIK WiCiDa, Alasan kuliah di STMIK WiCiDa, Matakuliah yang disukai, Matakuliah yang kurang disukai, Hobby.
4. Data Mahasiswa yang digunakan merupakan data mahasiswa aktif yang ada pada tahun 2019 sampai 2023 pada program studi teknik informatika STMIK Widya Cipta Dharma.
5. Implementasi metode *K-Means* menggunakan *Jupyter Notebook*.
6. Hasil *Clustering* di *Deployment* dan berjalan di sisi server atau website.
7. Pengujian Metode *K-Means* menggunakan *Elbow Method*.

2.3 Tujuan Penelitian

Penelitian ini bertujuan untuk mengelompokkan data mahasiswa Program Studi Teknik Informatika menggunakan metode *K-Means Clustering* guna menghasilkan informasi yang berguna dalam mendukung pengambilan keputusan akademik. Hasil pengelompokan ini diharapkan dapat membantu dalam menyesuaikan metode pembelajaran, memantau kinerja mahasiswa, serta mengoptimalkan pemanfaatan sumber daya seperti dosen dan fasilitas. Selain itu, informasi seperti asal sekolah mahasiswa juga dapat dimanfaatkan sebagai strategi promosi yang lebih terarah bagi institusi.

2.4 Manfaat

1. Bagi Ketua Program Studi Teknik Informatika
Manfaat penelitian ini bagi Program Studi yaitu merupakan solusi yang tepat untuk membantu Program Studi Teknik Informatika STMIK Widya Cipta Dharma dalam mengetahui dan mengembangkan kualitas mahasiswa Program Studi Teknik Informatika. Dengan Menggunakan *Clustering* bermanfaat untuk mengelola data mahasiswa untuk meningkatkan prestasi akademik dan jumlah mahasiswa.
2. Bagi Mahasiswa
Manfaat penelitian ini yaitu untuk belajar untuk menerapkan secara langsung ilmu yang diperoleh selama dalam perguruan tinggi agar dapat meningkatkan kemampuan, memperluas wawasan, pengetahuan, pengalaman, memantapkan keterampilan, dapat mengetahui dan menganalisa serta membantu memberikan solusi, jawaban pada permasalahan yang ada sehingga dapat menyelesaikan masalah dengan baik.
3. Bagi Perguruan Tinggi
Manfaat penelitian ini bagi Program Studi yaitu merupakan solusi yang tepat untuk membantu Program Studi Teknik Informatika STMIK Widya Cipta Dharma dalam mengetahui dan mengembangkan kualitas mahasiswa Program Studi Teknik Informatika. Dengan Menggunakan *Clustering* bermanfaat untuk mengelola data mahasiswa untuk meningkatkan prestasi akademik dan jumlah mahasiswa.

3. BAHAN DAN METODE

3.1 Data Mining

Menurut Maulida (2018) *Data mining* adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan didalam *database*. *Data mining* adalah proses yang menggunakan teknik statistic, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar.

Menurut Mardi (2016) *Data mining* adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik-teknik, metode-metode, atau algoritma dalam *data mining* sangat bervariasi.

Menurut Syahril (2020) *Data Mining* adalah serangkaian proses untuk menambah serta mencari informasi yang selama ini tidak diketahui secara manual dari suatu basis data. Informasi yang dihasilkan diperoleh dengan cara mengekstraksi dan mengenali pola yang penting atau menarik dari data yang terdapat pada basis data. *Data mining* terutama digunakan untuk mencari pengetahuan yang terdapat dalam basis data yang besar sehingga sering disebut *Knowledge Discovery Databases (KDD)*.

3.2 Clustering

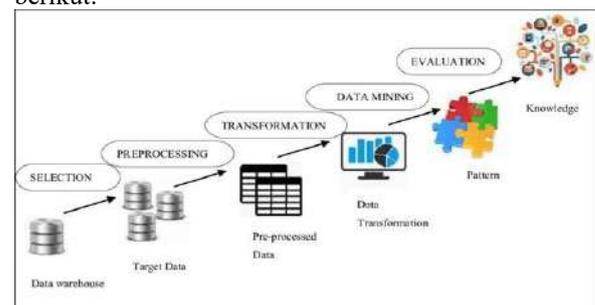
Menurut Abdurrahman (2018) *Clustering* didefinisikan sebagai suatu proses mengorganisir objek-objek menjadi satu kelompok yang anggotanya memiliki similaritas (kemiripan) tertentu. Similaritas dibedakan menjadi dua, yakni similaritas berdasarkan bentuk dan similaritas berdasarkan jarak.

Menurut Priyatman (2019) *Clustering* atau analisis *Cluster* adalah proses pembentukan kelompok data (*Cluster*) dari himpunan data yang tidak diketahui kelompok-kelompok atau kelas kelasnya dan proses menentukan data-data termasuk ke dalam *Cluster* yang mana. *Clustering* merupakan proses untuk mengetahui kelas-kelas taksonomi atau batryologi, atau analisis topologi dari data-data yang ada. Dilihat dari kacamata data *mining*, *Clustering* bukanlah proses klasifikasi. Karena dalam proses klasifikasi, data dikelompokkan ke dalam kelas-kelas yang telah diketahui sebelumnya.

Menurut Herlinda (2021) *Clustering* adalah proses pengelompokan titik data yang termasuk dalam kelompok yang sama lebih mirip satu sama lain daripada di kelompok yang berbeda, hanya berdasarkan informasi yang tersedia dengan titik data.

3.3 KDD (Knowledge Discovery in Database)

Menurut Saputra (2018) *Knowledge Discovery in Database (KDD)* adalah penerapan metode saintifik pada *Data Mining*. Dalam konteks ini *Data Mining* merupakan satu langkah dari proses KDD. KDD adalah kegiatan atau proses untuk mendapatkan informasi yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola dan hubungan dalam set data berukuran besar. Adapun tahapan atau rangkaian proses *Data Mining* dapat dibagi menjadi beberapa tahap. Tahap-tahap tersebut bersifat iterative. Terdapat 7 tahapan pada *Data Mining*, yaitu Tahapan-tahapan *Data Mining* seperti gambar 1 dapat dijelaskan sebagai berikut:



Gambar 1 Tahapan Data Mining

Figure 1 Data Mining Stages

Sumber : Binus University School of Information Systems (2021), Proses Data Mining KDD

1. *Data Cleaning*
Membersihkan data dari data yang tidak konsisten dan bersifat noise karena secara umum data yang tersedia pada *database* memiliki isian-isian yang

tidak sempurna karena hilang ataupun tidak valid. Selain itu, ada juga karakteristik data yang tidak relevan dengan kebutuhan untuk Data Mining.

2. **Data Integration**
Mengintegrasikan data dari berbagai sumber, dimana integrasi ini dilakukan pada data-data yang memiliki atribut unik. Pada proses integrasi perlu dilaksanakan dengan hati-hati karena kesalahan dapat mengakibatkan hasil yang menyimpang.
3. **Data Transformation**
Mentransformasikan data supaya sesuai untuk kebutuhan Data Mining. Aturan pemrograman untuk mengenali pola dari data untuk pelaksanaan transformasi ini sering disebut dengan algoritma. Beberapa algoritma Data Mining yang populer seperti: *Association Rule*, *K-Means Clustering*, *C.45*, *Naive Bayes Classifier*, *Apriori*, *Tree Structured*, dan lain-lain. Masing-masing algoritma ini akan menghasilkan output yang berbeda-beda.
4. **Proses Mining**
Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data.
5. **Pattern Evaluation**
Evaluasi atas pola yang ditemukan berguna untuk menemukan informasi emas (*knowledge*). Pada tahapan ini diharapkan adanya hasil berupa pola yang unik yang dapat digunakan sebagai model untuk memprediksi perilaku-perilaku bisnis yang sekiranya akan mempengaruhi sebuah keputusan manajemen.
6. **Knowledge Presentation**
Mempresentasikan *knowledge* yang sudah didapatkan dari user. Tahapan terakhir ini merangkum dan menerjemahkan analisa yang telah didapat dengan bahasa yang semudah mungkin karena melibatkan orang lain pada manajemen yang kemungkinan besar awam dengan Data Mining.

3.4 K-Means

Menurut Suryadi (2019) *K-Means* merupakan metode klasterisasi yang sering digunakan diberbagai bidang karena penggunaannya sederhana, mudah untuk di implementasikan, mampu untuk mengklaster data yang besar. Algoritma *K-Means* merupakan metode berbasis jarak yang membagi data kedalam sejumlah *Cluster* dan dalam setiap tahapan tertentu setiap objek harus masuk dalam kelompok, pada tahap selanjutnya objek dapat berpindah ke kelompok lain. Dalam algoritma ini pada dasarnya melakukan proses *Clustering* tetapi tergantung dari data yang didapat dan konklusi yang dicapai. Maka dari itu algoritma *K-Means* mempunyai aturan yaitu ada jumlah *Cluster* yang akan diinputkan dan hanya dapat memiliki atribut yang bertipe numerik.

Menurut Indraputra (2020) Algoritma *K-Means* merupakan salah satu algoritma dengan *partitional*, karena *K-Means* didasarkan pada penentuan jumlah

awal kelompok dengan mendefinisikan nilai *centroid* awalnya. Algoritma *K-Means* menggunakan proses secara berulang-ulang untuk mendapatkan basis data *Cluster*. Dibutuhkan jumlah *Cluster* awal yang diinginkan sebagai masukan dan menghasilkan titik *centroid* akhir sebagai output. Metode *K-Means* akan memilih pola k sebagai titik awal *centroid* secara acak atau random. Jumlah *iterasi* untuk mencapai *Cluster centroid* akan dipengaruhi oleh calon *Cluster centroid* awal secara random. Sehingga didapat cara dalam pengembangan algoritma dengan menentukan *centroid Cluster* yang dilihat dari kepadatan data awal yang tinggi agar mendapatkan kinerja yang lebih tinggi.

Menurut Sembiring (2022) Algoritma *K-Means* merupakan metode analisis kelompok yang mengarah pada pemartisian N objek pengamatan ke dalam kelompok (*Cluster*) dimana setiap objek pengamatan dimiliki oleh sebuah kelompok dengan mean (rata-rata) terdekat. Algoritma *K-Means* merupakan algoritma yang cukup mudah untuk diimplementasikan dan dijalankan, relatif cepat, mudah disesuaikan dan banyak digunakan. Algoritma *K-Means* data diklasifikasikan berdasarkan kedekatannya dengan *centroid*. Data dikelompokkan dengan memaksimalkan data yang sama dalam satu klaster dan meminimalkan data yang sama antar klaster. Langkah-langkah Metode *K-Means* antara lain sebagai berikut:

1. Tentukan K data sebagai *centroid*, K adalah jumlah *Cluster* yang diinginkan.
2. Jarak antara data dan pusat *Cluster* dihitung menggunakan *Euclidian Distance*. Untuk menghitung jarak semua data ke setiap titik pusat *Cluster* dapat menggunakan teori jarak *Euclidean* yang dirumuskan sebagai berikut :

$$d_{ij} = \sqrt{\sum_{k=1}^n \{x_{ik} - x_{jk}\}^2}$$

Keterangan :

d_{ij} = Jarak objek antara objek i dan j

n = Dimensi data

x_{ik} = Koordinat dari objek i pada dimensi k

x_{jk} = Koordinat dari objek j pada dimensi k

3. Data ditempatkan dalam *Cluster* yang terdekat, dihitung dari tengah *Cluster*.
4. Pusat *Cluster* baru akan ditentukan bila semua data telah ditetapkan dalam *Cluster* terdekat.
5. Proses penentuan pusat *Cluster* dan penempatan data dalam *Cluster* diulangi sampai nilai *centroid* tidak berubah lagi.

Dari definisi diatas dapat disimpulkan bahwa Algoritma *K-Means* adalah sebuah metode non hierarki yang pada awalnya mengambil sebagian banyaknya komponen populasi untuk dijadikan pusat *Cluster* awal.

Algoritma ini juga termasuk *partitioning Clustering* yang memisahkan data ke k daerah bagian yang terpisah.

3.5 centroid

Menurut Nasari (2016) *centroid* merupakan titik awal yang diperoleh secara acak, dan jumlah *centroid* sebanyak *Cluster* yang akan dibuat. *Centroid* awal merupakan titik pusat *Cluster* pertama atau awal pusat *Cluster*.

Menurut Retno (2019) *centroid* merupakan titik awal dimulainya sebuah pengelompokan didalam *Cluster* pada algoritma *K-Means*. Pengelompokan data dilakukan dengan menghitung jarak terdekat dengan titik pusat *Cluster* awal sebagai titik tengah dalam pembentukan kelompok atau *Cluster*.

Menurut Nabila (2021) *centroid* merupakan titik pusat *Cluster* pertama yang diperoleh secara acak dari data training. Pemilihan titik *centroid* awal akan sangat berpengaruh terhadap hasil *Cluster*.

3.6 Implementasi

Menurut Nugroho (2016) implementasi adalah suatu proses untuk melaksanakan kebijakan menjadi tindakan kebijakan dari politik ke dalam administrasi. Pengembangan kebijakan dalam rangka penyempurnaan suatu program.

Menurut Mustanir (2016) pengertian implementasi dirumuskan secara pendek, dimana "*to implementation*" (mengimplementasikan) berarti "*to provide means for carrying out* (menyediakan untuk melakukan sesuatu); *to give practical effect to*" (menimbulkan dampak/ akibat terhadap sesuatu). Beranjak dari rumusan implementasi tersebut dapat diperoleh gambaran bahwa "*to implementation* (mengimplementasikan) berkaitan dengan suatu aktifitas yang terlaksana melalui penyediaan sarana (misalnya: undang-undang, peraturan pemerintah, pedoman pelaksanaan, sumber daya dan lain-lain) sehingga dari aktifitas tersebut akan menimbulkan dampak/ akibat terhadap sesuatu.

Menurut Mamonto (2018) Implementasi adalah bermuara pada aktivitas, aksi, tindakan atau adanya mekanisme suatu sistem, implementasi bukan sekedar aktivitas, tapi suatu kegiatan yang terencana dan untuk mencapai tujuan kegiatan. Secara sederhana implementasi di artikan pelaksanaan atau penerapan.

3.7 Elbow Method

Menurut Pangestu (2022) Metode *Elbow* adalah metode untuk menghasilkan informasi dalam menentukan jumlah *Cluster* terbaik dengan cara melihat grafik perbandingan antara jumlah yang akan membentuk siku pada suatu titik.

Menurut Firdaus (2022) *Elbow Method* yaitu metode tertua untuk mengidentifikasi jumlah *Cluster* optimal pada dataset yang dianalisis, yang mana ide dasar dari metode tersebut yaitu menentukan $k = 2$ sebagai jumlah *Cluster* optimal awal k . Kemudian, jumlah *Cluster* optimal k tersebut akan terus bertambah dari langkah

pertama ke langkah yang paling maksimum untuk menentukan perkiraan jumlah *Cluster* optimal potensial dan akhirnya akan dibedakan jumlah *Cluster* optimal k sesuai dengan levelnya.

Menurut Febrianto (2024) Metode *Elbow* merupakan suatu metode yang digunakan untuk menghasilkan informasi dalam menentukan jumlah *Cluster* terbaik dengan cara melihat persentase hasil perbandingan antara jumlah *Cluster* yang akan membentuk siku pada suatu titik.

3.8 Black Box

Uminingsih (2022) *Black Box* Testing merupakan pengujian untuk menunjukkan kesalahan pada sistem aplikasi seperti kesalahan pada fungsi sistem aplikasi, serta menu aplikasi yang hilang. Jadi *Black Box* testing merupakan metode uji fungsionalitas sistem aplikasi. Dalam melakukan pengujian menggunakan masukan data acak dengan tujuan untuk mendapatkan hasil yang pasti. Dikatakan pasti artinya bila salah, maka ditolak oleh sistem informasi atau data input tersebut tidak dapat disimpan dalam *database*. Sedangkan bila data input benar maka dapat di terima/masuk di *database*, sedangkan bila data input benar maka dapat diterima/masuk di *database* sistem informasi.

4. PEMBAHASAN

4.1 Hasil Penelitian dan Pembahasan

Metode pengembangan sistem yang digunakan dalam penelitian adalah *Knowledge Discovery in Database* (KDD), dalam metode ini terdapat 5 tahap yang dilakukan setelah itu akan dilanjutkan dengan tahap implementasi dan tahap implementasi dan tahap pengujian

4.1.1 Tahapan Seleksi Data

Dalam penelitian ini data yang digunakan merupakan data mahasiswa yang masuk aktif dari program studi Teknik informatika STMIK Widya Cipta Dharma tahun 2019-2023. Data Mahasiswa didapatkan dari hasil google form yang telah dibagikan melalui grup kelas dan Angkatan.

4.1.2 Tahapan *Cleaning* Data

Data yang didapatkan adalah data yang diterapkan dari kusioner menggunakan Google form merupakan data mahasiswa program studi Teknik Informatika yang masih aktif dari Angkatan 2019-2023 dengan jumlah 100 data mahasiswa. Pada data mahasiswa terdapa 26 atribut terdiri dari Nama, NIM, jenis kelamin, semester, agama, Status Perkawinan, Status Bekerja, Pekerjaan, Nama Asal Sekolah, Jurusan Sekolah, Asal Provinsi, Asal Kota/Kabupaten, asal Kecamatan, Asal Kelurahan/Desa, No Kode Pos, Pekerjaan Ayah, Pekerjaan Ibu, Tahun ajaran masuk, IPK, Nama Kompetisi/lomba/kegiatan, Tingkat nasional yang

diikuti, Peringkat juara, Informasi pendaftaran STMIK WiCiDa, Alasan kuliah di STMIK WiCiDa,

. Atribut yang terdapat dalam data tersebut ialah:

1. Kolom “No” merupakan nomor sebagai pengurut data dalam tabel
2. Kolom “Nama” berisi nama mahasiswa.
3. Kolom “NIM” berisi nomor induk mahasiswa.
4. Kolom “Jenis Kelamin” berisi keterangan terkait jenis kelamin mahasiswa.
5. Kolom “Semester” berisi keterangan semester yang lagi ditempuh mahasiswa.
6. Kolom “Agama” berisi keterangan agama yang dianut mahasiswa.
7. Kolom “Status Perkawinan” berisi Status Perkawinan mahasiswa.
8. Kolom “Nama Pekerjaan” berisi nama profesi pekerjaan mahasiswa.
9. Kolom “Nama Asal Sekolah” berisi terkait asal sekolah yang pernah di tempuh mahasiswa.
10. Kolom “Jurusan Sekolah” berisi jurusan sekolah mahasiswa sewaktu masa sekolah.
11. Kolom “Asal Provinsi” berisi terkait asal provinsi tempat tinggal mahasiswa.
12. Kolom “Asal Kota/Kabupaten” berisi terkait asal kota/kabupaten tempat tinggal mahasiswa.
13. Kolom “Asal Kecamatan” berisi terkait asal kecamatan tempat tinggal mahasiswa.
14. Kolom “Asal Kelurahan/Desa” berisi terkait asal kelurahan tempat tinggal mahasiswa.
15. Kolom “No Kode Pos” berisi terkait no kode pos tempat tinggal mahasiswa.
16. Kolom”Pekerjaan Ayah” berisi terkait nama profesi pekerjaan ayah mahasiswa.
17. Kolom “Pekerjaan Ibu” berisi terkait nama profesi pekerjaan ibu mahasiswa.
18. Kolom “Tahun Ajaran Masuk” berisi terkait tahun ajaran masuk mahasiswa di STMIK WiCiDa.
19. Kolom “IPK” berisi terkait IPK mahasiswa selama kuliah STMIK WiCiDa.
20. Kolom”Nama Kompetisi/Lomba/Kegiatan, Tingkat Nasional” berisi terkait lomba yang pernah diikuti mahasiswa STMIK WiCiDa.
21. Kolom “Peringkat juara” berisi terkait peringkat juara lomba yang pernah diikuti mahasiswa.
22. Kolom “Informasi Pendaftaran STMIK WiCiDa” berisi Terkait mahasiswa mendapatkan informasi perndaftaran STMIK WiCiDa.
23. Kolom “Alasan Kuliah di STMIK WiCiDa” berisi terkait alasan mahasiswa masuk STMIK WiCiDa.
24. Kolom “Matakuliah yang di Sukai” berisi terkait matakulia yang disukai mahasiswa selama menempu perkuliahan.
25. Kolom “Matakuliah yang Kurang di Sukai” berisi terkait matakuliah yang kurang disukai mahasiswa selama menempu perkuliahan.
26. Kolom “Hobby” berisi terkait hobby mahasiswa.

Sebelum proses pemilihan data yang akan dianalisis lebih lanjut, sistem perlu mempersiapkan data terlebih dahulu. Tahap ini melibatkan pembacaan dan pemuatan data dari file CSV ke dalam sistem. Untuk keperluan ini, fungsi `pd.read_csv()` dari pustaka `pandas` pada bahasa pemrograman Python akan digunakan. Pustaka `pandas` sendiri merupakan sebuah pustaka yang sangat populer dalam bidang analisis data di Python, sehingga penggunaan fungsi `pd.read_csv()` akan memudahkan proses pembacaan data dari file CSV.

Proses pengisian data yang kosong proses ini sangat penting dilakukan untuk melakukan pemrosesan data seperti algoritma *k-means*. Algoritma *K-Means* tidak bisa berjalan dan akan menghasilkan error jika data dari sebuah atribut tidak terisi. Kemudian data yang telah di isi digabungkan menjadi satu `DataFrame`.

Sebelum data dianalisis menggunakan metode *clustering*, langkah penting yang perlu dilakukan adalah menghapus variabel-variabel yang tidak relevan. Proses ini bertujuan untuk menyederhanakan data, sehingga mempermudah proses *clustering* dan menghindari gangguan atau noise yang dapat mempengaruhi hasil analisis.

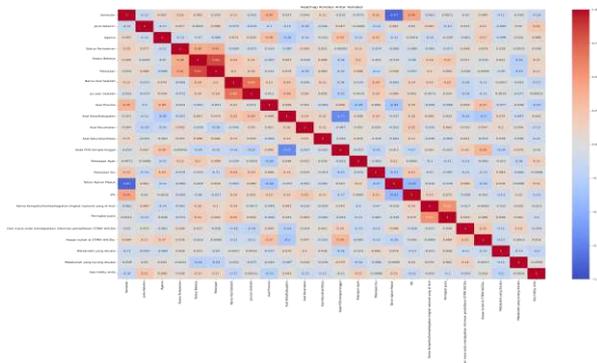
4.1.3 Tahapan Tranformasi Data

Pada proses ini dirubah sesuai dengan kebutuhan analisis yang dilakukan. Pada penelitian ini atribut yang digunakan ialah Semester, Jenis Kelamin, Agama, Status Perkawinan, Status Pekerjaan, Pekerjaan, Nama Asal Sekolah, Jurusan Sekolah, Asal Provinsi, Asal Kota/Kabupaten, Asal Kecamatan, Asal Kelurahan, No kode Pos, Pekerjaan Ayah, Pekerjaan Ibu, Tahun Ajaran Masuk, IPK, Nama Kompetisi/lomba/Kegiatan tingkat nasional yang diikuti peringkat juara'n, Darimana anda mendapatkan informasi pendaftaran STMIK WiCiDa, Alasan kuliah di STMIK WiCiDa, Matakuliah yang disukai, Matakuliah yang kurang disukai, Hobby. Proses perubahan data dalam variabel yang digunakan dalam proses pengolahan pada metode algoritma *kmeans*. proses ini dilakukan untuk mengubah data kategorial menjadi data numerik agar dapat dipahami oleh algoritma *k-means*. Data akan berubah menjadi angka yang bernilai berurutan setiap kolom.

Semester	Jenis kelamin	Agama	Status Perkawinan	Status Bekerja/n	Pekerjaan	Nama Asal sekolah	Jurusan sekolah	Asal Provinsi	Asal kota/kabupaten	...
0	0	0	0	0	0	58	1	3	15	...
1	3	0	0	0	0	25	11	2	15	...
2	3	0	0	0	1	48	18	2	14	...
3	1	0	0	0	0	2	7	2	8	...
4	0	0	0	0	0	55	15	2	14	...
...
95	0	0	0	0	0	44	18	6	14	...
96	0	0	2	0	0	54	12	6	2	...
97	0	0	2	0	0	13	5	6	6	...
98	0	0	0	0	0	19	7	6	12	...
99	0	0	2	0	0	23	5	6	12	...

Gambar 2 Hasil proses encoded
Figure 2 Results of the encoded process

Pada hasil heatmap, nilai korelasi tertinggi antara asal nama sekolah dengan jurusan sekolah,serta status bekerja dengan pekerjaan dan Nama lomba dan juara lomba.



Gambar 3 Tampilan Heatmap
Figure 3 Heatmap View

4.1.4 Tahapan Data Mining

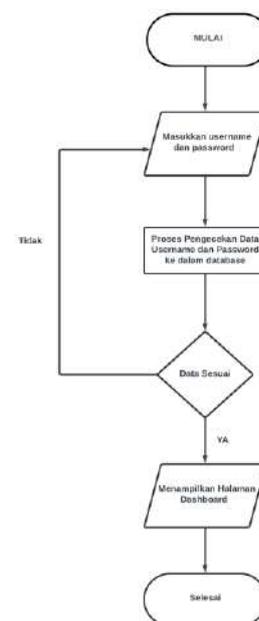
Pada tahap ini dilakukan pemilihan fitur yang akan digunakan dalam proses clustering, berdasarkan hasil analisis korelasi antar variabel. Fitur yang dipilih antara lain asal nama sekolah dengan jurusan sekolah, serta status bekerja dengan pekerjaan karena memiliki keterkaitan yang kuat. Pemilihan ini bertujuan untuk memastikan fitur yang digunakan benar-benar relevan. Selanjutnya, fitur terpilih disiapkan untuk dianalisis menggunakan metode clustering.

Proses *Clustering* telah dilakukan sebanyak 11 kali dengan berbagai kombinasi variabel untuk mendapatkan hasil yang paling optimal. Pada percobaan pertama, digunakan variabel Nama Asal Sekolah dan Jurusan Sekolah. Percobaan kedua menggunakan variabel Pekerjaan dan Status Pekerjaan, sedangkan percobaan ketiga memanfaatkan variabel Nama Lomba dan Peringkat Juara. Setiap percobaan dirancang berdasarkan analisis korelasi antar variabel untuk meningkatkan akurasi hasil *clustering*. Dari seluruh percobaan yang dilakukan, hasil terbaik diperoleh pada percobaan kedua yang terbaik dengan perolehan nilai korelasi 0.83. pada penelitian ini ditampilkan dua *cluster* tambahan dengan nilai korelasi 0,62 pada *cluster* pertama dan korelasi 0.55 pada *Cluster* ketiga. Pada percobaan pertama, variabel yang digunakan adalah Nama Asal Sekolah dan Jurusan Sekolah, sedangkan pada percobaan kedua menggunakan variabel Pekerjaan dan Status Pekerjaan serta percobaan ketiga menggunakan variabel Nama lomba dan Peringkat juara. Pemilihan variabel tersebut didasarkan pada nilai korelasi tertinggi dari hasil analisis sebelumnya.

4.1.5 Tahapan Evaluasi

Flowchart, atau bagan alur, adalah diagram yang menggambarkan urutan langkah, proses, dan keputusan dalam menjalankan suatu sistem atau program. Setiap elemen digambarkan menggunakan simbol standar

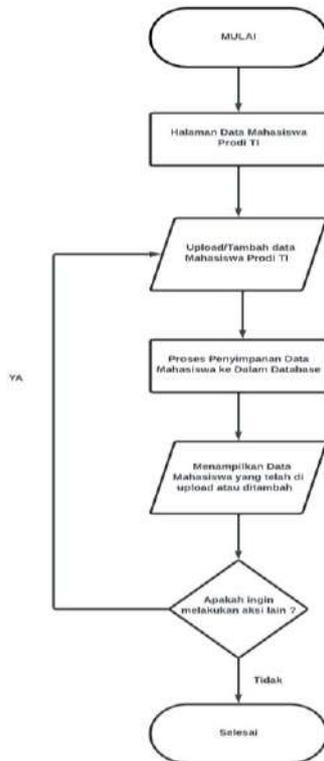
seperti oval untuk awal/akhir proses, persegi panjang untuk aktivitas, dan belah ketupat untuk keputusan. Simbol-simbol ini dihubungkan oleh panah yang menunjukkan alur logika antar proses. Flowchart sangat berguna dalam perancangan sistem karena memudahkan pemahaman alur kerja, identifikasi kesalahan logika, serta perbaikan ketidakefisienan proses. Dalam konteks sistem yang dikembangkan, flowchart digunakan untuk menggambarkan proses login pengguna sebelum mengakses aplikasi. Proses dimulai dari halaman login, di mana pengguna diminta memasukkan username dan password. Data tersebut kemudian diverifikasi dengan mencocokkannya terhadap informasi dalam database. Jika cocok, pengguna diberikan akses ke sistem; jika tidak, pengguna akan diarahkan kembali ke halaman login dan ditampilkan notifikasi bahwa data tidak valid. Proses ini penting untuk menjaga keamanan sistem dan memastikan hanya pengguna terdaftar yang dapat mengakses fitur aplikasi. Flowchart ini memberikan visualisasi yang jelas mengenai alur autentikasi, serta membantu dalam proses evaluasi dan pengembangan sistem lebih lanjut.



Gambar 4 Flowchart halaman login
Figure 4 Login page flowchart

Untuk memulai proses *clustering*, pengguna terlebih dahulu mengakses halaman data mahasiswa. Pada tahap ini, pengguna diminta untuk memasukkan data mahasiswa yang diperlukan. Setelah data diinputkan, sistem akan menampilkan informasi tersebut agar pengguna dapat memverifikasi bahwa semua data telah tercatat dengan benar. Di halaman ini, pengguna juga dapat melihat daftar lengkap data mahasiswa yang telah ditambahkan. Jika seluruh data sudah sesuai, pengguna

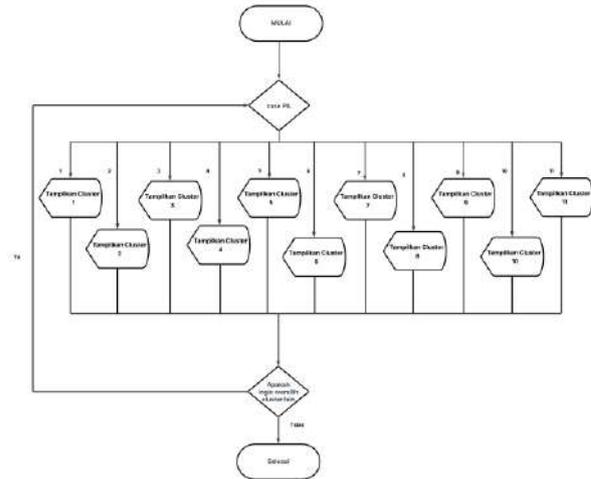
dapat melanjutkan ke tahap *clustering*, yaitu proses pengelompokan data mahasiswa berdasarkan kriteria tertentu. Tujuan dari proses ini adalah untuk mengidentifikasi pola dan kelompok dalam data mahasiswa sebagai bahan analisis lebih lanjut



Gambar 5 Flowchart Halaman Mahasiswa
Figure 5 Student Page Flowchart

Pada tahap awal, pengguna diarahkan ke halaman clustering yang dirancang secara interaktif untuk menampilkan hasil pengelompokan data mahasiswa menggunakan metode K-Means, salah satu algoritma populer dalam analisis data tidak terawasi (unsupervised learning) yang mengelompokkan data berdasarkan kemiripan karakteristik. Sistem akan menampilkan daftar cluster (case PIL) yang telah terbentuk berdasarkan atribut seperti asal sekolah, jurusan, hobi, mata kuliah favorit, dan faktor relevan lainnya. Setelah pengguna memilih salah satu cluster, sistem akan menampilkan hasil pengelompokan dalam bentuk tabel berisi daftar mahasiswa yang tergabung dalam cluster tersebut, serta visualisasi scatter plot dua dimensi. Setiap titik pada scatter plot mewakili satu mahasiswa dan diberi warna sesuai dengan cluster-nya, sementara simbol "X" berwarna merah menunjukkan posisi centroid, yaitu pusat kelompok berdasarkan rata-rata posisi anggota. Setelah hasil ditampilkan, pengguna dapat memilih untuk melihat cluster lain atau menyelesaikan proses.

Jikamemilih "Ya", pengguna diarahkan kembali ke daftar cluster; jika "Tidak", maka proses berakhir (Selesai). Visualisasi ini memberikan pemahaman intuitif terhadap pola pengelompokan, membantu mengidentifikasi distribusi anggota dalam tiap cluster, serta mendeteksi kemungkinan outlier. Penggunaan warna yang berbeda untuk setiap cluster memudahkan pengguna membedakan antar kelompok secara cepat dan akurat.

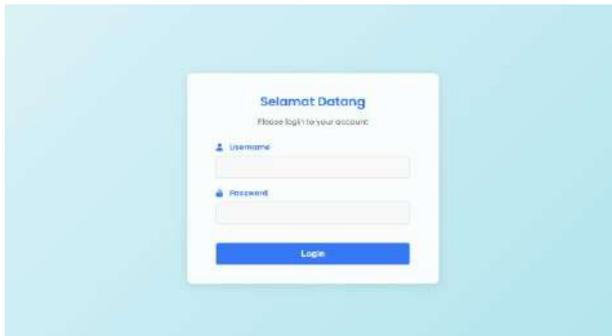


Gambar 6. Flowchart halaman cluster
Figure 6. Cluster page flowchart

4.1.5.1 Tampilan Halaman Aplikasi

1) Halaman Login

Halaman login merupakan tampilan awal yang muncul ketika sistem pertama kali dijalankan dan berfungsi sebagai gerbang utama untuk mengakses seluruh fitur yang tersedia dalam aplikasi. Pada halaman ini, pengguna diminta untuk memasukkan informasi autentikasi berupa username dan password sebagai langkah awal dalam proses verifikasi identitas. Setelah data diisi dengan benar, pengguna dapat menekan tombol "Login" untuk melanjutkan proses masuk ke dalam sistem. Sistem kemudian akan melakukan pencocokan data yang dimasukkan dengan informasi yang tersimpan di dalam database. Jika informasi tidak sesuai atau tidak terdaftar, maka proses login akan dinyatakan gagal, dan pengguna akan diarahkan kembali ke halaman login dengan peringatan untuk memeriksa kembali data yang telah dimasukkan. Sebaliknya, jika proses login berhasil dan data valid, pengguna akan secara otomatis diarahkan ke halaman dashboard sistem, yang menjadi pusat navigasi untuk mengakses fitur-fitur utama dalam aplikasi. Proses ini tidak hanya penting untuk memberikan pengalaman pengguna yang baik, tetapi juga krusial dalam menjaga keamanan dan membatasi akses hanya kepada pengguna yang memiliki otorisasi.



Gambar 7. Halaman Login
Figure 5. Login Page

2) Halaman Dashboard

Halaman utama dalam sistem ini menyajikan penjelasan lengkap mengenai dua bagian penting, yaitu Data Mahasiswa dan hasil *clustering*. Pada bagian Data Mahasiswa, pengguna dapat mengunggah, melihat, dan mengelola data mahasiswa yang tersimpan dalam sistem. Sementara itu, bagian Hasil *clustering* menampilkan hasil pengelompokan data yang telah diproses, disertai visualisasi yang memudahkan pengguna dalam memahami isi dan pola dari data tersebut. Tujuan dari halaman ini adalah agar pengguna dapat dengan mudah mengakses dan menganalisis informasi yang tersedia.



Gambar 8. Halaman Dashboard
Figure 8. Dashboard page

3) Halaman Data Mahasiswa

Halaman data mahasiswa merupakan salah satu komponen inti dalam sistem yang berfungsi untuk menampilkan informasi lengkap mengenai seluruh data mahasiswa yang telah diunggah oleh admin. Melalui halaman ini, admin memiliki akses penuh untuk melakukan berbagai pengelolaan data, seperti menambahkan data mahasiswa baru, mengedit data yang sudah ada, menghapus data yang tidak lagi relevan, serta melihat detail informasi setiap mahasiswa secara menyeluruh. Informasi yang tersedia meliputi berbagai atribut penting seperti nama, NIM, program studi, asal

Tampilan halaman ini disusun secara sistematis dalam bentuk tabel yang memudahkan admin dalam menavigasi dan memfilter data sesuai kebutuhan. Selain itu, halaman ini juga dilengkapi dengan fitur pencarian untuk mempercepat proses identifikasi data tertentu. Data mahasiswa yang ditampilkan pada halaman ini

tidak hanya berfungsi sebagai arsip digital, tetapi juga menjadi landasan utama dalam proses analisis dan pengelompokan data (*clustering*) yang dilakukan oleh sistem menggunakan algoritma K-Means.

No	Nama	NIM	Semester	Angkatan	Agama	Aksi
1	Rehanmahdi Ramadani	200101	1	Lah-Lah	Islam	
2	Herold Latipah	200102	1	Lah-Lah	Islam	
3	Fajar Herold	200103	1	Lah-Lah	Islam	
4	Latipah Herold	200104	1	Herold	Islam	
5	Arham Fajar Herold	200105	1	Lah-Lah	Islam	
6	Rehanmahdi	200106	1	Lah-Lah	Islam	

Gambar 9 Halaman Data Mahasiswa
Figure 9 Student Data Page

4) Halaman Clustering

Halaman ini menyajikan hasil *clustering* data mahasiswa dari program studi Teknik Informatika dalam bentuk beragam diagram yang menarik dan informatif. Tujuan dari halaman ini adalah untuk memberikan gambaran visual terhadap sebaran data mahasiswa berdasarkan variabel-variabel tertentu, sehingga memudahkan pengguna dalam membaca dan memahami informasi yang ditampilkan. Beberapa variabel yang divisualisasikan meliputi agama dan jenis kelamin. Pada variabel agama, diketahui bahwa sebanyak 81 mahasiswa beragama Islam, 15 mahasiswa beragama Kristen Protestan, dan 4 mahasiswa beragama Kristen Katolik. Sedangkan pada variabel jenis kelamin, jumlah mahasiswa laki-laki adalah 79 orang, sementara mahasiswa perempuan berjumlah 21 orang. Penyajian dalam bentuk diagram ini diharapkan dapat membantu pengguna dalam melakukan analisis lebih lanjut secara visual dan efisien.

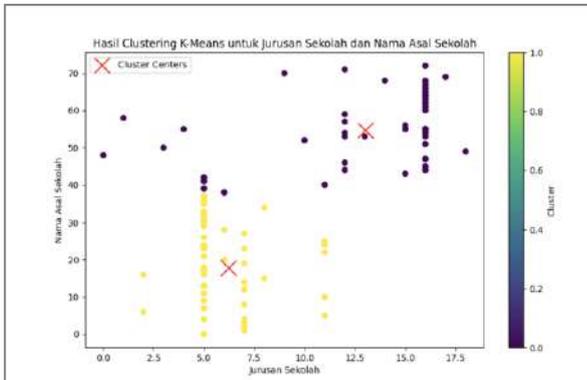


Gambar 10 Halaman *clustering* data mahasiswa
Figure 10 Student data clustering page

5) Halaman K-Means

Pada tampilan ini disajikan hasil *clustering* data mahasiswa yang diperoleh melalui implementasi metode K-Means, yang mengelompokkan data berdasarkan kemiripan karakteristik. Untuk mendukung proses

analisis, pengguna diberikan fleksibilitas untuk memilih dan menampilkan cluster spesifik sesuai kebutuhan. Fitur ini memungkinkan pengguna untuk mengamati lebih dalam mengenai distribusi data mahasiswa dalam setiap cluster, memahami pola yang terbentuk, serta mengevaluasi karakteristik unik dari masing-masing kelompok secara detail



Gambar 11 Tampilan hasil *clustering* menggunakan metode *k-means*

Figure 11 Display of clustering results using the *k-means* method

5. KESIMPULAN

Dengan adanya hasil Penelitian yang dilaksanakan dan berdasarkan uraian yang telah dibahas pada bab-bab sebelumnya, maka dapat disimpulkan bahwa:

- 1) Dari penelitian ini dapat disimpulkan bahwasannya telah dilakukannya *Clustering* menggunakan metode *K-Means* dimana dari hasil tersebut dari 11 kali percobaan. Dari 11 kali percobaan terdapat 3 *Cluster* terbaik dengan korelasi yang cukup tinggi yaitu *Cluster* pertama dengan nilai korelasi 0.68, *Cluster* kedua dengan nilai korelasi 0.83 dan *Cluster* ketiga dengan nilai korelasi 0.55.
- 2) *Cluster* pertama yang berisi variabel Nama asal sekolah dan jurusan sekolah, pada hasil penelitian ini *Cluster* pertama terdapat 2 titik *centroid* dimana pada pengelompokan pertama berisi data mahasiswa yang kebanyakan dari SMA dan MA sedangkan pengelompokan kedua itu berisi mahasiswa yang mayoritasnya dari SMK.
- 3) *Cluster* kedua yang berisi variabel pekerjaan dan Status pekerjaan, pada hasil ini penelitian ini *Cluster* kedua terdapat 2 titik *centroid* dimana pada pengelompokan pertama berisi data mahasiswa mayoritas belum bekerja sedangkan pengelompokan kedua itu berisi mahasiswa yang mayoritasnya sudah bekerja.
- 4) *Cluster* ketiga yang berisi variabel Nama lomba dan Peringkat juara, pada hasil penelitian ini *Cluster* ketiga terdapat 2 titik *centroid* dimana

pada pengelompokan pertama berisi data mayoritas mahasiswa yang pernah ikut lomba sedangkan pengelompokan ke dua itu berisi mahasiswa yang mayoritasnya tidak pernah ikut lomba.

- 5) Pengembangan sistem *Clustering* Data Mahasiswa Program Studi Teknik Informatika STMIK Widya Cipta Dharma berbasis website dibangun dengan menggunakan Bahasa pemrograman *Bootsrap*, PHP, HTML, Javascript dan MySQL untuk proses penyimpanan data ke dalam *database*.
- 6) Dapat membantu Program Studi Teknik Informatika STMIK Widya Cipta Dharma dalam melakukan pengelompokan data mahasiswa program studi Teknik Informatika.
- 7) Pemilihan variabel yang digunakan dalam *Clustering* dilakukan melalui analisis korelasi, sehingga fitur yang digunakan memiliki hubungan kuat dan relevan. Hal ini terbukti meningkatkan akurasi hasil *Clustering* dan kualitas pengelompokan data mahasiswa.
- 8) Hasil *Clustering* dapat digunakan untuk menyusun strategi peningkatan mutu akademik, penyesuaian metode pembelajaran, serta strategi promosi kampus berdasarkan sebaran asal sekolah atau wilayah mahasiswa. Informasi ini bermanfaat untuk menjangkau calon mahasiswa baru secara lebih tepat sasaran.

6. SARAN

Adapun saran-saran yang dapat dikemukakan berdasarkan kesimpulan diatas yaitu sebagai berikut:

- 1) Untuk penelitian selanjutnya peneliti menyarankan untuk memastikan bahwa proses pengumpulan data dilakukan dengan lebih konsisten dan lengkap. Data yang hilang atau tidak lengkap dapat mempengaruhi hasil analisis dan mengurangi validitas kesimpulan yang ditarik. Oleh karena itu, penggunaan metode pengisian nilai yang hilang yang lebih canggih mungkin diperlukan untuk memastikan bahwa data yang digunakan benar-benar representatif dan tidak bias.
- 2) Penting untuk melakukan validasi hasil *Clustering* dengan menggunakan metode evaluasi lain seperti Cohesion and Separation atau Davies-Bouldin Index. Metode evaluasi ini dapat membantu memastikan bahwa jumlah *Cluster* yang dipilih benar-benar optimal dan bahwa *Cluster* yang dihasilkan memiliki kualitas yang baik dalam hal pemisahan dan kekompakan. Validasi ini dapat memberikan keyakinan lebih pada hasil yang didapat dan membantu dalam membuat keputusan yang lebih tepat berdasarkan analisis.

- 3) Hasil *Clustering* dapat membantu memahami distribusi mahasiswa berdasarkan asal kota mereka, sehingga memberikan wawasan yang lebih mendalam tentang pola pendaftaran dan preferensi calon mahasiswa. Informasi ini dapat digunakan untuk menentukan daerah-daerah yang memerlukan strategi promosi atau sosialisasi pendaftaran yang lebih intensif, misalnya melalui kunjungan langsung ke sekolah-sekolah, penyelenggaraan seminar, atau kegiatan promosi digital yang disesuaikan dengan karakteristik daerah tersebut. Selain itu, hasil *Clustering* ini juga dapat digunakan untuk mengidentifikasi daerah yang memiliki potensi besar tetapi belum tergarap secara optimal, sehingga institusi dapat memfokuskan upayanya untuk menjangkau lebih banyak calon mahasiswa dari daerah tersebut. Dengan demikian, proses rekrutmen mahasiswa dapat menjadi lebih efektif, terarah, dan berbasis data.
- 4) Untuk penelitian selanjutnya, disarankan juga untuk menggunakan RapidMiner sebagai alat bantu dalam proses analisis data, khususnya dalam penerapan algoritma *Clustering* seperti *K-Means*. RapidMiner menyediakan antarmuka yang intuitif dan mendukung berbagai teknik pra-proses data, pemodelan, serta evaluasi secara visual dan efisien. Dengan menggunakan RapidMiner, peneliti dapat melakukan eksplorasi data yang lebih mendalam, membandingkan berbagai metode *Clustering*, dan menghasilkan visualisasi yang lebih informatif. Selain itu, integrasi fitur-fitur otomatisasi dan pelaporan dalam RapidMiner dapat meningkatkan akurasi serta efisiensi analisis, terutama bagi peneliti yang menginginkan proses data *mining* yang lebih sistematis dan terstruktur.

7. REFERENSI

- Abdurrahman, G. (2019). Clustering Data Kredit Bank Menggunakan Algoritma Agglomerative Hierarchical Clustering Average Linkage. *JUSTINDO (Jurnal Sistem & Teknologi Informasi Indonesia)*, Vol. 4, No. 1., 14.
- Febrianto, E. (2024). PENERAPAN ALGORITMA K-MEANS CLUSTERING UNTUK MEMETAKAN 4G BTS BEBAN TINGGI DI KABUPATEN BEKASI. *Jurnal Teknologi Pelita Bangsa*, 2.
- Firdaus, H. (2022). Analisa Cluster Menggunakan K-Means Dan Fuzzy C-Means Dalam Pengelompokan Provinsi Menurut Data Intesitas Bencana Alam Di Indonesia Tahun 2017-2021. *Jurnal Ilmiah Matematika*, 53.
- Herlinda, V., Darwis, D., & Dartono. (2021). ANALISIS CLUSTERING UNTUK RECREDESIALING FASILITAS KESEHATAN MENGGUNAKAN METODE FUZZY C-MEANS. *Jurnal Teknologi dan Sistem Informasi (JTSI) Vol 2, No. 2*, 95.
- Indraputra, R., & Fitriana, R. (2020). K-Means Clustering Data COVID-19. *Jurnal Teknik Industri Vol. 10 No. 3*, 277.
- Mamonto, N., Sumampouw, I., & Undap, G. (2018). IMPLEMENTASI PEMBANGUNAN INFRASTRUKTUR DESA DALAM PENGGUNAAN DANA DESA TAHUN 2017 (STUDI) DESA ONGKAW II KECAMATAN SINONSAYANG KABUPATEN MINAHASA SELATAN. *Jurnal Jurusan Ilmu Pemerintahan Volume 1 No. 1*, 3.
- Mardi, Y. (2016). Data Mining : Klasifikasi Menggunakan Algoritma C4.5. *Jurnal Edik Informatika*, 214.
- Maulida, L. (2018). PENERAPAN DATAMINING DALAM MENGELOMPOKKANKUNJUNGAN WISATAWAN KE OBJEK WISATA UNGGULAN DIPROV. DKI JAKARTA DENGAN K-MEANS. *Jurnal Informatika Sunan Kalijaga Vol. 2, No. 3*, 169.
- Mustanir, A., & Darmiah. (2016). IMPLEMENTASI KEBIJAKAN DANA DESA DAN PARTISIPASI MASYARAKAT DALAM PEMBANGUNAN DI DESA TETEJJI KECAMATAN TELLU LIMPOE KABUPATEN SIDENRENG RAPPANG. *Jurnal Politik Profetik Volume 04, No. 2*, 229.
- Nabila, Z., Isnain, A. R., Permata, & Abidin, Z. (2021). Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means. *Jurnal Teknologi Dan Sistem Informasi (JTSI)*, 103.
- Nasari, F., & Sianturi, C. J. (2016). Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Penyebaran Diare Di Kabupaten Langkat. *Cegito Smart Journal*, 113.
- Nugroho, A. H., Puspitasari, R., & Puspitasari, E. (2016). IMPLEMENTASI GEMAR MEMBACA MELALUI PROGRAM POJOK BACA DALAM MATA PELAJARAN IPS PADA SISWA KELAS VIII DI SMPN 2 SUMBER. 188.



- Pangestu, B. A., Kristiawan, N. A., & Sulistiyowati, N. (2022). Clustering Obat Untuk Menentukan Pola Pemasaran Efektif di Apotek Amarta Sehat. *Jurnal Ilmiah Wahana Pendidikan*, 119.
- Priyatman, H., Sajid, F., & Haldivany, D. (2019). Klasterisasi Menggunakan Algoritma K-Means. *JEPIN (Jurnal Edukasi dan Penelitian Informatika) Vol. 5 No. 1*, 63.
- Retno, S. (2019). *Peningkatan Akurasi Algoritma K-Means Dengan Clustering Purity Sebagai Titik Pusat Cluster Awal (Centroid)*. Medan: Universitas Sumatera Utara.
- Saputra, H. K. (2018). Analisis Data Mining Untuk Pemetaan Mahasiswa Yang Membutuhkan Bimbingan dan Konseling Menggunakan Algoritma Naive Bayes Classifier. *Jurnal Teknologi Informasi dan Pendidikan*, 15-16.
- Sembiring, S. N., Winata, H., & Kusnasari, S. (2022). Pengelompokan Prestasi Siswa Menggunakan Algoritma K-Means. *JURNAL SISTEM INFORMASI TGD Volume 1, Nomor 1*, 32-33.
- Suryadi, U. T., & Supriatna, Y. (2019). SISTEM CLUSTERING TINDAK KEJAHATAN PENCURIAN DI WILAYAH JAWA BARAT MENGGUNAKAN ALGORITMA K-MEANS. *Jurnal Teknologi Informasi dan Komunikasi* ISSN: 2252-4517 *STMIK Subang*, 17.
- Syahril, M., Erwansyah, K., & Yetri, M. (2020). Penerapan Data Mining Untuk Menentukan Pola Penjualan Peralatan Sekolah Pada Brand Wigglo Dengan Menggunakan Algoritma

Apriori. *Jurnal Teknologi Sistem Informasi Dan Sistem Komputer TGD*, 119.

- Uminingsih, Ichsanudin, M. N., Yusuf, M., & Suraya. (2022). PENGUJIAN FUNGSIONAL PERANGKAT LUNAK SISTEM INFORMASI PERPUSTAKAAN DENGAN METODE BLACK BOX TESTING BAGI PEMULA. *Jurnal Ilmiah Teknik dan Ilmu Komputer, Vol. 1No. 2, 3*.

UCAPAN TERIMA KASIH

Alhamdulillah, segala puji dan syukur penulis panjatkan ke hadirat Allah SWT atas segala rahmat, taufik, dan hidayah-Nya sehingga penulis dapat menyelesaikan penelitian dan penyusunan karya tulis ini dengan baik. Tanpa pertolongan-Nya, tugas akhir ini tidak akan dapat terselesaikan sebagaimana mestinya. Penulis juga menyampaikan rasa terima kasih yang sebesar-besarnya kepada dosen pembimbing, Ibu Siti Lailiyah S.Kom., M.Kom dan Bapak Muhammad Ibnu Sa'ad S.Kom., M.Kom, atas segala bimbingan, arahan, dan dukungan yang telah diberikan selama proses penelitian hingga selesainya penulisan tugas akhir ini. Penulis menyadari bahwa dalam penyusunan karya ini masih terdapat banyak kekurangan, baik dari segi isi maupun penyajian. Oleh karena itu, penulis sangat mengharapkan kritik dan saran yang membangun dari semua pihak demi perbaikan di masa mendatang. Penulis juga memohon maaf atas segala kekurangan dan kesalahan yang mungkin terjadi dalam penulisan ini. Akhir kata, penulis berharap semoga karya tulis ini dapat memberikan manfaat bagi para pembaca serta menjadi tambahan wawasan di bidang yang dibahas.