

OPTIMASI PARAMETER ALGORITMA K-MEANS PADA DATA REKAM MEDIS KLINIK ANANDA MEDIKA RANCAH

Zahra Alfirda Sofyan¹, Rudi Kurniawan², Yudhistira Arie Wijaya³

¹⁾ Program Studi Teknik Informatika, STMIK IKMI Cirebon

²⁾ Program Studi Rekayasa Perangkat Lunak, STMIK IKMI Cirebon

³⁾ Program Studi Sistem Informasi, STMIK IKMI Cirebon

STMIK IKMI Cirebon, Jalan Perjuangan No. 10B Majasem, Kota Cirebon, Indonesia

zahasopyan18@gmail.com

ABSTRAK

Data rekam medis merupakan informasi medis terkait identifikasi pasien, penyakit dan riwayat pengobatan. Peningkatan jumlah data rekam medis dari waktu ke waktu menyulitkan identifikasi pola penting yang tersembunyi dalam dataset. Oleh karena itu, diperlukan teknik analisis data yang tepat untuk mengelola data rekam medis secara optimal. Penelitian ini bertujuan untuk menemukan parameter terbaik yang menghasilkan *cluster optimum* berdasarkan *Davies Bouldin Index* (DBI), semakin rendah nilai DBI menunjukkan *cluster* terbaik. Metode yang digunakan dalam penelitian ini adalah *clustering* dengan Algoritma K-Means. Algoritma K-Means diimplementasikan dengan tools Rapidminer versi 10.3. Parameter yang digunakan dalam melakukan eksperimen terhadap nilai DBI adalah parameter *Measure Types Numerical Measure*. Dengan tiga kali percobaan *Numerical Measure* yang berbeda yaitu *Euclidean Distance*, *Manhattan Distance*, dan *Dynamic Time Warping Distance*, serta proses iterasi menggunakan parameter *Max Optimization* dimulai dari 1 hingga 10. Dari hasil eksperimen menunjukkan parameter terbaik adalah *Measure Types Numerical Measure* dengan *Manhattan Distance* karena menghasilkan nilai DBI terkecil dibandingkan parameter lainnya. Proses iterasi algoritma K-Means berakhir pada iterasi ke-3 karena pada iterasi tersebut diperoleh keanggotaan *cluster* yang sama dengan iterasi sebelumnya. Hal ini menunjukkan bahwa algoritma telah mencapai nilai konfiden.

Kata kunci: K-Means, *Davies Bouldin Index* (DBI), *Numerical Measure*, *Manhattan Distance*.

1. PENDAHULUAN

Rekam medis merupakan informasi medis yang mencakup rekaman riwayat dari pasien, biasanya dicatat oleh pihak penyedia pelayanan kesehatan seperti rumah sakit atau klinik [1]. Pengelolaan data rekam medis yang terorganisir dengan baik dapat meningkatkan pelayanan kesehatan [2].

Pengelolaan data rekam medis sering menjadi permasalahan di rumah sakit atau klinik. Permasalahan yang sering terjadi adalah banyak nya data pasien dari waktu ke waktu. Semakin banyaknya data pasien yang terkumpul menyebabkan kompleksitas yang tinggi karena banyak variabel yang terlibat. Hal ini menyulitkan upaya identifikasi pola penting dari data rekam medis. Oleh karena itu, permasalahan analisis data rekam medis yang besar dan kompleks ini menjadi relevan dan penting untuk diselesaikan guna mendukung para tenaga medis. Untuk mengatasi permasalahan tersebut diperlukan teknik analisis data mining untuk menganalisis data rekam medis yang akurat dan efisien. Data mining bertujuan untuk menemukan serta menggali pengetahuan berharga dari data atau informasi [3]. Kemampuan data mining untuk *clustering* sangat bernilai dalam proses pengambilan keputusan klinis berbasis data.

Penelitian lainnya dilakukan oleh Fajri & Purnamasari yang berjudul “Klasterisasi Pola Penyebaran Penyakit Pasien Berdasarkan Usia Pasien Menggunakan K-Means Clustering”. Temuan penting peneliti adalah kelompok usia yang paling terpengaruh adalah dewasa dan anak-anak dan algoritma

pengelompokan K-Means efektif dalam mengelompokkan data pasien berdasarkan usia [4].

Tujuan utama dari penelitian ini adalah untuk menemukan parameter terbaik yang menghasilkan *cluster optimum* berdasarkan *Davies Bouldin Index* (DBI). Hasil dari penelitian dapat dimanfaatkan oleh pihak manajemen Klinik Ananda Medika Rancah dalam proses pengambilan keputusan yang lebih tepat sasaran terkait pelayanan kesehatan kepada pasien klinik.

Pada penelitian ini menggunakan pendekatan *Knowledge Discovery In Database* (KDD) yang meliputi beberapa tahapan yaitu seleksi data, pengolahan, transformasi, Data Mining, dan hasil evaluasi [5]. Pendekatan yang diterapkan ialah menggunakan algoritma K-Means. Algoritma ini memiliki kemampuan mengelompokkan data kedalam kelompok yang serupa berdasarkan kesamaan karakteristik, memungkinkan identifikasi pola dalam data secara sistematis dan terstruktur [6]. Dengan demikian, data yang memperlihatkan ciri-ciri serupa akan ditempatkan dalam satu klaster yang sama, sedangkan data yang memiliki perbedaan karakteristik akan ditempatkan dalam klaster yang berbeda [7]. Pada penelitian ini software Rapidminer digunakan sebagai alat implementasi algoritma K-Means. Pemodelan dievaluasi dengan memanfaatkan *Davies Bouldin Index* (DBI) untuk menilai kualitas *clustering*. Semakin rendah nilai DBI atau mendekati nol nilai DBI yang didapatkan, maka semakin optimal data tersebut dikelompokkan [8].

Oleh karena itu, peneliti mengangkat sebuah judul penelitian “Optimasi Parameter Algoritma K-Means Pada Data Rekam Medis Klinik Ananda Medika Rancah”. Penelitian ini memiliki implikasi penting dalam bidang informatika dan kesehatan karena dapat mengembangkan teknik analisis data yang lebih baik. Hasil penelitian berpotensi memberikan kontribusi dalam pengembangan sistem pendukung keputusan klinis berbasis data mining untuk meningkatkan kualitas layanan kesehatan.

2. TINJAUAN PUSTAKA

Peneliti memerlukan landasan teori. Bagian landasan teori ini bertujuan untuk menyajikan dasar teori yang menjadi acuan utama dalam penelitian ini. Pemahaman mendalam terhadap konsep-konsep kunci seperti Data Mining, *Clustering*, K-Means, RapidMiner, *Davies Bouldin Index (DBI)*, *Euclidean Distance*, *Manhattan Distance* dan *Dynamic Time Warping Distance*.

2.1. Data Mining

Data mining yaitu suatu proses untuk menemukan hubungan serta pola-pola menarik dengan meneliti sekumpulan data dalam jumlah besar [9]. Tujuannya untuk mengekstrak informasi dan pengetahuan yang akurat, bernilai dan berpotensi dari data tersebut, sehingga melalui data mining dapat dimengerti dan berguna bagi pengambilan keputusan [10].

2.2. Clustering

Clustering adalah salah satu metode dalam pembelajaran tanpa pengawasan (*unsupervised learning*), di mana sekumpulan data (dataset) akan dikelompokkan atau dipartisi ke dalam *cluster*/kelompok yang berbeda berdasarkan tingkat kesamaan atau kemiripan tertentu antar data [11].

2.3. K-Means

Algoritma K-Means bekerja dengan empat langkah utama. Pertama, k objek dipilih secara acak dari data sebagai *centroid* awal kluster. Kedua, setiap objek dialokasikan ke *cluster* terdekat berdasarkan jarak *euclidean* atau ukuran kesamaan lainnya terhadap *centroid cluster*. Ketiga, posisi *centroid* setiap *cluster* diperbarui berdasarkan rata-rata dari objek yang ada di dalam *cluster* tersebut. Keempat, langkah kedua dan ketiga diulang terus menerus (iterasi) hingga *cluster* yang terbentuk stabil dan *centroid* tidak berubah posisi lagi antar iterasi. Proses iterasi inilah yang memungkinkan K-Means mengelompokkan data secara efisien [12].

2.4. RapidMiner

RapidMiner dirancang agar dapat digunakan dengan mudah, memungkinkan pengguna untuk dengan mudah membuat dan menguji berbagai model, bahkan tanpa memiliki pengalaman pemrograman [13].

2.5. Davies Bouldin Index (DBI)

Davies Bouldin Index (DBI) diperkenalkan oleh David L. Davies dan Donald W. Bouldin sebagai metode untuk mengevaluasi hasil kluster (*cluster*). Evaluasi menggunakan DBI melihat pengukuran internal kluster agar dapat menilai baik atau tidaknya hasil pengklasteran dilihat dari kuantitas dan kedekatan antar data dalam kluster. Kohesi diartikan sebagai total kedekatan antara data terhadap titik *centroid* (pusat) kluster tersebut. Sebaliknya, separasi bergantung pada jarak antara titik pusat *centroid* (pusat) kluster [14].

2.6. Euclidean Distance

Euclidean Distance adalah salah satu metode penghitungan yang digunakan untuk menilai jarak antara dua titik dalam *Euclidean space*, termasuk dalam bidang *Euclidean* dua dimensi, tiga dimensi, atau bahkan lebih [15]. Jarak *Euclidean* dapat dinyatakan sebagai akar kuadrat dari total selisih kuadrat antara koordinat objek [16]. Kemudian, nilai *Euclidean Distance* yang diperoleh digunakan untuk menilai tingkat kesamaan guna menentukan apakah suatu objek termasuk dalam kelas tertentu berdasarkan nilai kedekatan yang diperoleh [17].

2.7. Manhattan Distance

Manhattan Distance juga dikenal sebagai “*City Block Distance*”, didefinisikan sebagai jumlah total jarak antara semua atribut [18]. Jarak *Manhattan*, mengukur total jarak antara dua titik dihitung dengan menambahkan selisih absolut dari setiap atribut pada kedua titik tersebut. Teknik ini umumnya digunakan untuk mengevaluasi kesamaan atau perbedaan antara dua objek dan bergantung pada perhitungan jarak vertikal dan horizontal antara titik-titik tersebut [19].

2.8. Dynamic Time Warping Distance

Dynamic Time Warping Distance adalah suatu metode yang memanfaatkan pendekatan pemrograman dinamis untuk menyesuaikan dua rangkaian waktu dengan menghitung jarak minimum antara titik data [20]. Metode ini menggunakan teknik pemrograman dinamis untuk menemukan semua kemungkinan jalur dan memilih satu jalur yang menghasilkan jarak minimum antara dua deret waktu. Pada setiap langkah, jarak kumulatif setiap elemen matriks dihitung berdasarkan nilai minimum dari tiga tetangga terdekatnya [21].

3. METODE PENELITIAN

3.1. Sumber Data

Sumber data berupa data sekunder. Data diperoleh dari rekam medis Klinik Ananda Medika Rancah, selain itu melakukan hasil wawancara bersama seorang dokter dan 2 orang karyawan di Klinik Ananda Medika Rancah yang mengenai informasi tentang data rekam medis di klinik tersebut. Kemudian data diolah oleh peneliti. Dataset yang

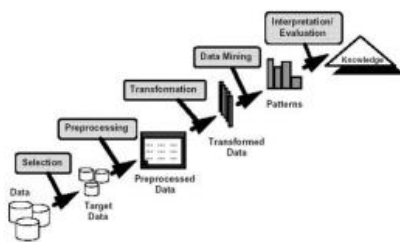
digunakan berjumlah 1.000 *record* data rekam medis Klinik Ananda Medika Rancah.

3.2. Teknik Pengumpulan Data

Metode pengumpulan data yang digunakan dalam penelitian ini adalah wawancara dan observasi. Wawancara dilakukan dengan mengajukan pertanyaan secara langsung kepada narasumber yang terlibat dalam pengelolaan data rekam medis di Klinik Ananda Medika Rancah, yaitu seorang dokter dan dua orang karyawan klinik. Tujuannya adalah untuk mendapatkan informasi mendalam terkait data rekam medis pasien. Sementara itu, observasi juga dilakukan dengan cara mengamati secara langsung proses pengelolaan data rekam medis di klinik tersebut. Hal ini dilakukan untuk mendapatkan gambaran yang lebih komprehensif mengenai objek penelitian yang dijadikan kasus studi.

3.3. Teknik Analisis Data

Teknik analisis data yang digunakan dalam penelitian adalah dengan proses *Knowledge Discovery In Database* (KDD). Berikut adalah tahapan-tahapan yang ada dalam proses KDD dapat dilihat pada Gambar 1. sebagai berikut:



Gambar 1. Tahapan KDD
sumber: Munawar & Purnamasari, 2023 [22]

Berikut ini adalah rangkaian proses penerapan metode KDD yang dilakukan dalam penelitian yaitu:

a. Data Selection

Di tahap ini dilakukan proses menganalisis data yang relevan karena sering ditemukan tidak semua data diperlukan dalam proses data mining. Data diambil dari Klinik Ananda Medika Rancah dari tanggal 29 -31 Oktober tahun 2023 sebanyak 1.000 record.

b. Preprocessing

Proses preprocessing merupakan tahap pembersihan data dan perbaikan data termasuk penghapusan duplikasi, verifikasi konsistensi, dan koreksi kesalahan data, termasuk kesalahan cetak. namun pada tahap ini hasil menunjukan pada dataset tidak mempunyai *Missing Value* atau kesalahan pada dataset.

c. Transformation

Pada tahap ini, transformasi data diperlukan untuk mengubah informasi tanpa entitas yang jelas menjadi bentuk data yang valid atau dapat diolah dalam proses Data Mining. Dalam penelitian ini, 13 atribut digunakan, di mana 12 atribut akan

diubah menjadi tipe data numerik menggunakan operator *Nominal To Numerical* dalam RapidMiner.

d. Data Mining

Pada tahap ini, dilakukan penerapan algoritma K-Means untuk mengekstrak pola informasi dari data. Evaluasi *Cluster Distance Performance* difokuskan pada penentuan nilai K optimum, diukur dengan metode validasi *internal Davies Bouldin Index* (DBI). Langkah-langkah melibatkan penentuan nilai K dari 2 hingga 20, menggunakan parameter *Measure Types Numerical Measure*. Eksperimen dilakukan menggunakan *Numerical Measure* yang berbeda yaitu *Euclidean Distance*, *Manhattan Distance*, dan *Dynamic Time Warping Distance*, untuk menghasilkan *cluster optimum* berdasarkan nilai DBI.

e. Interpretation/evaluation

Pada tahap *interpretation/evaluation*, peneliti menganalisis hasil klasterisasi untuk memahami relevansi temuan. Evaluasi pola dilakukan setelah klaster terbentuk untuk memastikan kualitasnya, memastikan konsistensi dan kesesuaian dengan fakta yang sudah diketahui sebelumnya.

4. HASIL DAN PEMBAHASAN

Implementasi algoritma K-Means *clustering* ini dirancang dengan menggunakan tools RapidMiner. Pada penelitian ini, digunakan tahapan *Knowledge Discovery in Database* (KDD) yang terdiri dari beberapa tahap, yaitu sebagai berikut:

4.1. Data Selection

Data dalam penelitian ini berupa file excel yang diperoleh dari dataset Klinik Ananda Medika Rancah dengan jumlah record sebanyak 1.000 record, terdiri dari 16 atribut. Dari 16 atribut 1 atribut dijadikan ID yaitu atribut register. Selengkapnya dapat dilihat pada Tabel 1 sebagai berikut:

Tabel 1. Atribut *Default* Dataset

No	Atribut	Type
1.	Tanggal	<i>Date-Time</i>
2.	Register	<i>Integer</i>
3.	Jenis Kelamin	<i>Nominal</i>
4.	Diagnosa	<i>Nominal</i>
5.	Admin	<i>Nominal</i>
6.	Jasa Pemeriksaan	<i>Nominal</i>
7.	Cek Kolesterol	<i>Nominal</i>
8.	Injeksi Dexametason	<i>Nominal</i>
9.	Injeksi Neurobion	<i>Nominal</i>
10.	Nebulizer Selling	<i>Nominal</i>
11.	Inj. Ranitidine	<i>Nominal</i>
12.	Injeksi Ketorolac	<i>Nominal</i>
13.	Cek Asam Urat	<i>Nominal</i>
14.	Debridement Luka Ringan	<i>Nominal</i>
15.	Home Visit	<i>Nominal</i>
16.	Cek Gula Darah	<i>Nominal</i>

Dataset rekam medis memiliki 16 atribut. Kemudian dilakukan filtering data sehingga hanya 13

atribut yang dipilih untuk pengolahan lebih lanjut. Filtering ini mengeluarkan 3 atribut dengan nilai korelasi 0 yang dianggap kurang berkontribusi bagi tujuan pemodelan. 13 Atribut terpilih tersebut antara lain: Registrasi, Jenis Kelamin, Diagnosa, Admin, Jasa Pemeriksaan, Cek Kolesterol, Injeksi Dexametason, Injeksi Neurobion, Nebulizer Selling, Inj. Ranitidine, Injeksi Ketorolac, Debridement Luka Ringan, dan Home Visit.

4.2. Preprocessing

Dalam tahapan *Knowledge Discovery in Database* (KDD) tahap *processing* bertujuan untuk menghilangkan *Missing Value* pada data yang akan digunakan sehingga pada saat proses tidak muncul kesalahan. *Missing Value* perlu ditangani karena dapat mengganggu proses *modelling* data mining selanjutnya. Dikarenakan pada dataset rekam medis Klinik Ananda Medika Rancah tidak ditemukan *Missing Value* atau data yang tidak memiliki nilai maka tahap *processing* data tidak dilakukan.

Name	Type	Missing	Statistics
REGISTRASI	Integer	0	221023001
JENIS KELAMIN	Normal	0	AKUT (AK) (435)
DIAGNOSA	Normal	0	AKUT (AK) (435)
ADMIN	Normal	0	AKUT (AK) (435)
JASA PEMERIKSAAN	Normal	0	AKUT (AK) (435)
CEK KOLESTEROL	Normal	0	AKUT (AK) (435)
INJEKSI DEXAMETASON	Normal	0	AKUT (AK) (435)

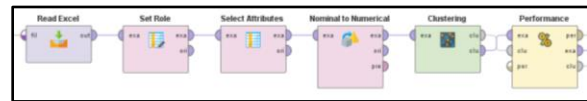
Gambar 2. Hasil Statistik

4.3. Transformation

Transformation data merupakan proses untuk mengubah tipe data atribut agar sesuai dengan analisis yang dibutuhkan dalam data mining. Pada penelitian ini, dilakukan *transformation* data karena terdapat beberapa atribut dengan tipe data non-numerik yang perlu diubah menjadi bentuk numerik agar dapat diproses lebih lanjut. Atribut non-numerik tersebut meliputi: Jenis Kelamin, Diagnosa, Admin, Jasa Pemeriksaan, Cek Kolesterol, Injeksi Dexametason, Injeksi Neurobion, Nebulizer Selling, Inj. Ranitidine, Injeksi Ketorolac, Debridement Luka Ringan, dan Home Visit. Ke-12 atribut non-numerik ini kemudian ditransformasikan ke dalam bentuk numerik.

4.4. Data Mining

Clustering dilakukan sebagai bagian dari tahap pemodelan data mining untuk menganalisis dan memperoleh informasi berharga dari data rekam medis. Proses *clustering* menghasilkan K=4 sebagai *cluster optimum*. Proses pemodelan tahap data mining menggunakan algoritma K-Means pada tools RapidMiner versi 10.3 dapat dilihat dalam Gambar 3 sebagai berikut:



Gambar 3. Pemodelan Data Mining dengan algoritma K-Means

Parameter yang digunakan dalam operator *clustering* menggunakan K-Means tampak pada Tabel 2 sebagai berikut:

Tabel 2. Operator K-Means

Parameter	Isi
K	4
Max Runs	10
Measure Types	Numerical Measure
Numerical Measure	Euclidean Distance, Manhattan Distance, dan Dynamic Time Warping Distance.
Max Optimization	1-10

4.5. Interpretation/Evaluation

Parameter yang digunakan dalam melakukan eksperimen terhadap nilai DBI adalah parameter *Measure Types Numerical Measure*. Dengan tiga kali percobaan *Numerical Measure* yang berbeda yaitu *Euclidean Distance*, *Manhattan Distance*, dan *Dynamic Time Warping Distance*, serta proses iterasi menggunakan parameter *Max Optimization* dimulai dari 1 hingga 10.

Berdasarkan hasil percobaan pengelompokkan data menggunakan algoritma K-Means pada nilai K=4 dari *Numerical Measure* dengan *Euclidean Distance* diperoleh hasil seperti pada Tabel 3 sebagai berikut:

Tabel 3. Hasil K=4 Numerical Measure dengan Euclidean Distance

K	Numerical Measure	Max Optimization	Davies Bouildin
4	Euclidean Distance	1	0.269
		2	0.270
		3	0.270
		4	0.270
		5	0.270
		6	0.270
		7	0.270
		8	0.270
		9	0.270
		10	0.270

Berdasarkan hasil percobaan pengelompokkan data menggunakan algoritma K-Means pada nilai K=4 dari *Numerical Measure* dengan *Manhattan Distance* diperoleh hasil seperti pada Tabel 4 sebagai berikut:

Tabel 4. Hasil K=4 *Numerical Measure* dengan *Manhattan Distance*

K	<i>Numerical Measure</i>	<i>Max Optimization</i>	<i>Davies Bouildin</i>
4	<i>Manhattan Distance</i>	1	0.267
		2	0.268
		3	0.268
		4	0.268
		5	0.268
		6	0.268
		7	0.268
		8	0.268
		9	0.268
		10	0.268

Berdasarkan hasil percobaan pengelompokan data menggunakan algoritma K-Means pada nilai K=4 dari *Numerical Measure* dengan *Dynamic Time Warping Distance* diperoleh hasil seperti pada Tabel 5 sebagai berikut:

Tabel 5. Hasil K=4 *Numerical Measure* dengan *Dynamic Time Warping Distance*

K	<i>Numerical Measure</i>	<i>Max Optimization</i>	<i>Davies Bouildin</i>
4	<i>Dynamic Time Warping Distance</i>	1	0.269
		2	0.270
		3	0.270
		4	0.270
		5	0.270
		6	0.270
		7	0.270
		8	0.270
		9	0.270
		10	0.270

5. KESIMPULAN DAN SARAN

Hasil pengujian data rekam medis Klinik Ananda Medika Rancah menggunakan algoritma K-Means *Clustering* pada RapidMiner menunjukkan bahwa parameter terbaik adalah *Measure Types Numerical Measure* dengan *Manhattan Distance* karena menghasilkan nilai DBI terkecil dibandingkan parameter lainnya. Dengan nilai DBI sebesar 0.268, serta proses iterasi algoritma K-Means berakhir pada iterasi ke-3 karena pada iterasi tersebut diperoleh keanggotaan *cluster* yang sama dengan iterasi sebelumnya. Meskipun demikian, disarankan untuk penelitian selanjutnya untuk menerapkan algoritma *clustering* lain, mengingat setiap algoritma memiliki karakteristik yang berbeda, sehingga dapat memperkaya analisis perbandingan kinerja dan hasil algoritma *clustering*.

DAFTAR PUSTAKA

- [1] W. S. Nanda, A. M. H. Pardede, and M. Simanjuntak, "Data Mining Analysis For Medical Record Data Clasterization Using K-Means Algorithm In Sylvani Binjai Hospital," *Indones. J. Educ. Comput. Sci.*, 2023, [Online]. Available: <https://jurnal.intekom.id/index.php/indotech/arti>
- [2] W. N. Purba, G. A. Sembiring, M. T. Turnip, and ..., "Penerapan Data Mining Untuk Pengelolaan Data Rekam Medis Menggunakan Metode K-Means Clustering Pada Rumah Sakit Royal Prima Medan," *J. Tekinkom*, 2023, [Online]. Available: <http://jurnal.murnisadar.ac.id/index.php/Tekinkom/article/view/857>
- [3] D. A. Manalu and G. Gunadi, "Implementasi Metode Data Mining K-Means Clustering Terhadap Data Pembayaran Transaksi Menggunakan Bahasa Pemrograman Python Pada Cv Digital Dimensi," *Infotech J. Technol. Inf.*, vol. 8, no. 1, pp. 43–54, 2022, doi: 10.37365/jti.v8i1.131.
- [4] M. B. Fajri and S. D. Purnamasari, "Klasterisasi Pola Penyebaran Penyakit Pasien Berdasarkan Usia Pasien Menggunakan K-Means Clustering," *J. Inf. echnology Ampera*, 2022, [Online]. Available: <https://journal-computing.org/index.php/journal-ita/article/view/315>
- [5] U. S, "Penerapan Data Mining Dengan Mengimplementasikan Algoritma K-Means Dalam Proses Clustering Untuk Pengelompokan Mahasiswa Calon Penerima Beasiswa KIP," *Build. Informatics, Technol. Sci.*, vol. 5, no. 1, 2023, doi: 10.47065/bits.v5i1.3411.
- [6] M. D. Permana, A. L. Hananto, E. Novalia, B. Huda, and ..., "Klasterisasi Data Jamaah Umrah pada Tanurmutmainah Tour Menggunakan Algoritma K-Means," *J. KomtekInfo*, 2023, [Online]. Available: <https://jkomtekinfo.org/ojs/index.php/komtekinf/article/view/332>
- [7] I. A. Misdiyanto, Yustina Suhandini T, "Identifikasi Jenis-Jenis Burung Lovebird Menggunakan Pengolahan Citra Digital Dengan Metode K-Means Clustering," *J. Sains Komput. Inform. (J-SAKTI)*, vol. 4, no. 2, pp. 445–456, 2020.
- [8] R. Sidik, N. Suarna, and A. R. Dikananda, "Analisa Data Set Peminatan Siswa Menggunakan Algoritma K- Means Dengan Optimize Parameter Di Sekolah Menengah Kejuruan (Studi Kasus : Smk Pui Gegesik)," vol. 7, no. 2, pp. 1197–1203, 2023.
- [9] M. Sholihul, I. G. Putu, W. Wadashwara, and F. Bimantoro, "Implementasi Algoritma Apriori Untuk Analisis Transaksi Penjualan Obat (Studi Kasus: Apotek Gilda Farma 2)," *J. Teknol. Informasi, Komput. dan Apl. Vol.*, vol. 4, no. 1, pp. 63–74, 2022.
- [10] L. Setiyani, M. Wahidin, D. Awaludin, and ..., "Analisis Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Data Mining Naïve Bayes: Systematic Review," *Fakt.*, 2020, [Online]. Available: <https://journal.lppmunindra.ac.id/index.php/Fakt>

- or_Exacta/article/view/5548
- [11] M. R. Putri, G. S. Nugraha, and R. DwiYansaputra, "Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Pendidikan Menggunakan Metode K-Means Clustering," *J. Comput. Sci. Informatics Eng.*, vol. 7, no. 1, pp. 1–8, 2023, doi: 10.29303/jcosine.v7i1.509.
 - [12] R. Astuti and K. Ukar, "Implementasi Data Mining dengan Metode Clustering Algoritma K-Means untuk Pengelompokan Data Tilang di Instansi Pemerintah," *Media Inform.*, vol. 20, no. 2, pp. 109–121, 2021, doi: 10.37595/mediainfo.v20i2.77.
 - [13] A. F. Khairani, A. Nazir, T. Darmizal, and ..., "Klasterisasi Peserta BPJS Berdasarkan Rekam Medis Menggunakan Algoritma K-Means," *J. Comput. ...*, 2023, [Online]. Available: <http://ejurnal.seminar-id.com/index.php/josyc/article/view/3442>
 - [14] S. Mutrofin, T. Wicaksono, and A. Murtadho, "Perbandingan Kinerja Algoritma Kmeans dengan Kmeans Median pada Deteksi Kanker Payudara," *J. Inf. dan Teknol.*, vol. 5, no. 1, pp. 88–91, 2023, doi: 10.37034/jidt.v5i1.274.
 - [15] N. Puspitasari, A. Ardin Maulana, and F. Alameka, "K-Means untuk Klasterisasi Daerah Rawan Penyakit Demam Berdarah," vol. 13, no. 1, pp. 40–52, 2023, [Online]. Available: <http://sisfotenika.stmikpontianak.ac.id/index.php/ST>
 - [16] G. P. I. Rani, A. Aziz, and M. P. T. Sulistyono, "Implementasi Euclidean Dan Chebyshev Distance Pada K-Medoids Clustering," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 6, no. 2, pp. 710–715, 2022, doi: 10.36040/jati.v6i2.5443.
 - [17] R. Nuraini, "Implementasi Euclidean Distance dan Segmentasi K-Means Clustering Pada Identifikasi Citra Jenis Ikan Nila," *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 3, no. 1, pp. 1–8, 2022.
 - [18] R. Hidayati, A. Zubair, A. H. Pratama, and L. Indana, "Analisis Silhouette Coefficient pada 6 Perhitungan Jarak K-Means Clustering Silhouette Coefficient Analysis in 6 Measuring Distances of K-Means Clustering," *Techno.COM*, vol. 20, no. 2, pp. 186–197, 2021.
 - [19] R. A. Shafira, Yahfizham, and A. Muliani, "Menentukan Jarak Terpendek Dalam Pengiriman Barang Dengan Perbandingan Euclidean Distance Dan Manhattan Distance," *J. Sci. Soc. Res.*, vol. 6, no. 3, pp. 31–37, 2023.
 - [20] I. Salsabila, S. Anwar, and R. Radhiah, "Perbandingan Kualitas Suara Smartphone Menggunakan Metode Dynamic Time Warping (DTW)," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 1, pp. 82–90, 2021, doi: 10.29207/resti.v5i1.2764.
 - [21] M. Irfan, T. Ammar, and J. Jesslyn, "K-Medoids Clustering dengan Jarak Dynamic Time Warping dalam Mengelompokkan Provinsi di Indonesia Berdasarkan Kasus Aktif Covid-19," vol. 4, pp. 685–692, 2021.
 - [22] K. K. Munawar and A. I. Purnamasari, "Implementasi Algoritma K-Means Clustering Pada Klasterisasi Kasus Hiv Di Jawa Barat," vol. 7, no. 2, 2023.