

Pemetaan Penyebaran Anggota Muhammadiyah Berdasarkan Tingkat Kepadatan Menggunakan Metode K-Means Clustering

Egi Jati Ramadhan¹, Rudiman²

^{1,2}Program Studi Teknik Informatika

Universitas Muhammadiyah Kalimantan Timur

Jl. Ir. H. Juanda No.15, Sidodadi, Kec. Samarinda Ulu, Kota Samarinda, Kalimantan Timur

e-mail: ¹egijatiramadhan@gmail.com, ²rudiman@gmail.com

Abstrak

Berdasarkan jumlah penduduk beragama Islam yang tergabung dalam organisasi Muhammadiyah khususnya pada provinsi Kalimantan Timur Kabupaten Penajam Paser Utara yang belum diketahui penyebaran organisasi Muhammadiyah, maka perlunya dilakukan pemetaan. Penelitian ini menggunakan metode K-Means dengan informasi geografis untuk melakukan pemetaan anggota organisasi Muhammadiyah yang bertujuan untuk mengetahui luas penyebaran anggota Muhammadiyah secara visual yang berada di kabupaten Penajam Paser Utara. Setelah dilakukan tahapan observasi, interview dan penyebaran questionnaire melalui perijinan pada tingkat PWM di Samarinda maka diperoleh jumlah data keanggotaan sebesar 205 penduduk. Pada tahap penelitian dilakukan proses data cleaning dari dataset yang telah dikumpulkan, dan kemudian transformasi data dalam algoritma K-Means Clustering serta visualisasi data kedalam Quantum GIS. Pada setiap cluster akan divisualisasikan dan dibagi menjadi tiga cluster serta dipisah berdasarkan keanggotaan masing-masing cluster. Pengelompokan keanggotaan yang telah terbentuk dibagi dalam tingkat penyebaran anggota Muhammadiyah menjadi 3 cluster yaitu cluster 0, cluster 1, dan cluster 2 dengan tambahan atribut data. Dari hasil penelitian ini diperoleh bahwa Kabupaten Penajam Paser Utara memiliki total 168 anggota Muhammadiyah yang tersebar di beberapa kelurahan. Dengan Cluster 0 terdiri dari 21 wilayah kelurahan, Cluster 1 terdiri dari 3 wilayah kelurahan, dan Cluster 2 terdiri dari 1 wilayah kelurahan.

Kata kunci— Clustering, Pemetaan, Big Data Muhammadiyah

1. PENDAHULUAN

Mayoritas penduduk Indonesia memeluk agama Islam sesuai dengan data Badan Pusat Statistika (BPS) tahun 2021 menunjukkan sekitar 236,53 juta atau 86,88 % penduduk Indonesia merupakan pemeluk agama Islam. Tentu angka tersebut terbagi kedalam beberapa kelompok atau organisasi besar yaitu Nahdlatul Ulama (NU), Muhammadiyah dan organisasi Islam lainnya. Muhammadiyah adalah sebuah organisasi Islam yang besar di Indonesia. Bahkan merupakan gerakan kemanusiaan terbesar di dunia di luar gerakan kemanusiaan yang dilaksanakan oleh gereja, sebagaimana disinyalir oleh seorang James L. Peacock [1]. Luasnya penyebaran dan banyaknya anggota Muhammadiyah di Indonesia serta kurangnya data terkait setiap anggota Muhammadiyah membuat sulitnya melihat perkembangan dan penyebarannya organisasi di daerah-daerah tertentu. Bahkan pada provinsi Kalimantan Timur khususnya kabupaten Penajam Paser Utara, pada tahun 2019 pemeluk agama Islam yang tercatat pada BPS di kabupaten Penajam Paser Utara berjumlah 164,9 ribu jiwa. Namun dari data tersebut belum dapat dipastikan berapa jumlah pemeluk agama Islam yang telah bergabung dalam organisasi Muhammadiyah.

Berdasarkan permasalahan tersebut maka diperlukan adanya pemetaan anggota organisasi Muhammadiyah dengan sistem informasi geografis yang bertujuan untuk mengetahui

luas penyebaran anggota Muhammadiyah secara visual yang berada di kabupaten Penajam Paser Utara. Dengan cara mengumpulkan suatu data terkait yang selanjutnya dikelompokkan menjadi tiga kelompok berdasarkan tingkat kepadatannya. Pengelompokan data anggota Muhammadiyah dapat dilakukan dengan menerapkan algoritma pengelompokan yang terdapat dalam data mining salah satunya adalah K-Means Clustering sehingga menjadi informasi yang bermanfaat guna memudahkan dalam melihat perkembangan dan penyebaran anggota organisasi Muhammadiyah secara visual.

Metode K-Means termasuk algoritma klasterisasi yang banyak digunakan pada penelitian sebelumnya dikarenakan prosesnya yang sederhana untuk diimplementasikan dan dijalankan [2]. Metode tersebut termasuk dalam metode berbasis partisi dengan cara kerja membagi sekumpulan data kedalam beberapa kelompok. Dalam Sistem Informasi Geografis yang digabungkan dengan teknik pengelompokan maka akan menampilkan posisi penyebaran data pada suatu kondisi yang sesungguhnya dan menemukan solusi dari permasalahan yang ada dengan membuat suatu sistem informasi berbasis komputer [3].

2. METODE PENELITIAN

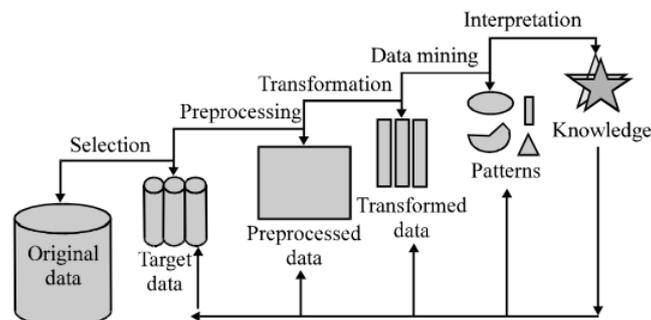
2.1 Data Mining

Data mining atau yang dikenal sebagai *pattern recognition* ialah suatu metode yang berguna untuk mengolah data yang memiliki tujuan untuk menemukan pola yang belum terlihat dari kumpulan data tersebut. Hasil yang diperoleh dari pengolahan data dengan metode ini dapat digunakan untuk mengambil keputusan di masa depan karena *data mining* merupakan teknik pemrosesan data dalam skala besar sehingga dapat menjadi peran penting dalam berbagai bidang seperti keuangan, Industri, ilmu teknologi dan cuaca. *Data mining* umumnya menggambarkan suatu metode dengan pengelompokan, klasifikasi, regresi, serta pemilihan variabel dan market basket analisis [4]. Dalam hal ini menambang big data Muhammadiyah merupakan suatu cara untuk mengeksplorasi data dalam amal usaha Muhammadiyah (AUM) sehingga pengelompokan warga Muhammadiyah berdasarkan tingkat kepadatannya di suatu wilayah dapat dieksploitasi dengan tujuan untuk memajukan suatu AUM di daerah tersebut.

2.2 Data preparasi

Langkah pertama yang dilakukan dalam proses *data mining* adalah pengumpulan dan pemilihan data dari sumber data yaitu penyebaran *questionnaire*. Pada tahap kedua adalah *preprocessing* yang ditujukan guna memperbaiki kualitas dari suatu data, transformasi dan tahap interpretasi serta evaluasi yang dapat menghasilkan keluaran tentang suatu pengetahuan baru yang dapat memberikan suatu kontribusi yang lebih baik [5].

Data baris mungkin memiliki catatan yang tidak lengkap, nilai *noise*, *outlier* dan data yang tidak konsisten sehingga perlu dilakukan pembersihan data. Pembersihan data adalah langkah pertama dalam teknik prapemrosesan data yang digunakan untuk menemukan nilai yang hilang, mengenali outlier dan, memperbaiki data yang tidak konsisten [6].



Gambar 1. Langkah-langkah Penemuan Pengetahuan

Data kotor ini akan mempengaruhi prosedur mining dan menyebabkan output yang tidak dapat diandalkan dan buruk. Oleh karena itu, penting bagi beberapa rutinitas pembersihan data untuk digunakan.

2.3 K-Means

K-Means adalah suatu cara mengelompokkan data nonhierarki atau sekatan sehingga dapat mempartisi suatu data ke dalam bentuk kelompok bisa dua kelompok atau lebih. Dengan menggunakan Metode ini maka data akan dipisahkan ke dalam suatu kelompok dan data yang mempunyai kesamaan akan dimasukkan dalam kelompok yang serupa [7].

Ide dasar algoritma K-Means yang cukup sederhana yaitu dapat memperkecil Sum of Squared Error (SSE) diantara objek-objek data dengan sejumlah K *centroid* yang sudah ditetapkan di awal. Prinsip utama dari suatu algoritma yaitu menyusun K buah partisi/pusat (*centroid*)/ rata-rata (*mean*) dari sekumpulan data. Penyelesaiannya akan menghasilkan titik *centroid* yang dijadikan tujuan dari algoritma K-Means [8]. Setelah iterasi K-Means berhenti, setiap objek dalam dataset menjadi anggota dari suatu cluster. Melalui K-means clustering setiap data akan dikelompokkan berdasarkan jarak *centroid* terdekat [9]. Adapun langkah-langkah algoritma K-Means mencakup beberapa point yaitu :

1. Tahap awal ialah menentukan jumlah Cluster(k) pada dataset. Pada dataset kasus ini jumlah *cluster* ditentukan menjadi 3 *cluster*.
2. Menentukan secara random titik pusat *cluster* atau *centroid* sebanyak data 3 *centroid*.
3. Rumus Eudclidean Distance digunakan untuk menghitung jarak yang paling dekat dengan titik pusat atau *centroid* yaitu :

$$D = \sqrt{(x_i + s_i) + (y_i + t_i)} \tag{1}$$

Keterangan :

- D = Euclidean Distance.
- i = Banyak objek.
- x, y = titik Kordinat objek.
- s, t = titik Kordinat *centroid*.

4. Perhitungan anggota *cluster* yang sekarang dengan pusat *cluster*. Pusat *cluster* yaitu rata-rata dalam suatu *cluster*. Adapun rumus yang dapat digunakan yaitu :

$$v_{ij} = \frac{1}{n} \sum_{K=0}^{N_i} x_{kj} \tag{2}$$

Keterangan :

- v_{ij} = *Centroid* rata-rata *cluster* ke-i untuk variabel ke-j.
- N_i = Jumlah anggota *cluster* ke-i.
- i, K = Index dari *cluster*.
- j = Index dari variabel.
- x_{kj} = Nilai data ke-k variabel ke-j untuk *cluster*.

5. Kemudian lakukan langkah ke-3 dan langkah ke-4 kembali dan iterasi mencapai *centroid* bernilai optimal.

2.4 Davies Boildin Index

Index Davies Bouldin (DBI) adalah suatu metode yang dapat digunakan dalam prosedur *clustering* untuk mengukur suatu validitas atau jumlah *cluster* yang optimal. Adapun metode dari pengelompokan yang dimana kohesi dapat diartikan sebagai jumlah kedekatan data dengan pusat *cluster* dari *cluster* yang diikuti. Pada evaluasi ini menggunakan DBI yang mempunyai skema penilaian *cluster* internal yang mengidentifikasi *cluster* yang baik atau buruk berdasarkan jumlah dan kedekatan antar data dari skor *cluster* [10].

3. HASIL DAN PEMBAHASAN

Pada penelitian ini data yang digunakan merupakan hasil pengumpulan kuisisioner dan juga data langsung yang didapatkan dari Pimpinan Muhammadiyah Kalimantan Timur (PWM). Kedua data tersebut kemudian digabungkan sehingga menjadi kumpulan data yang utuh dan lengkap. Dari hasil proses tersebut diperoleh data sebanyak 168 data. Namun dari data tersebut belum terdapat atribut titik kordinat yang diperlukan pada proses perancangan Sistem Informasi Geografis sehingga dibutuhkan penambahan atribut tersebut pada data dengan mencari titik kordinat setiap anggota dengan bantuan Google Maps berdasarkan alamat yang tertera pada data tersebut.

Tabel 1. Data Anggota Muhammadiyah Penajam Paser Utara

No	Nama	Cabang	Alamat	...	Kelurahan
1	Anggota 1	Waru	Waru RT 06 RW 03	...	Waru
2	Anggota 2	Babulu	Babulu Darat Rt 11	...	Babulu Darat
3	Anggota 3	Penajam	Jl Propinsi Km 3	...	Penajam
⋮	⋮	⋮	⋮	⋮	⋮
167	Anggota 4	Waru	Perum SMA 2 N PPU Jl.Negara Km 25	...	Waru
168	Anggota 5	Babulu	Babulu Darat Rt 11	...	Babulu Darat

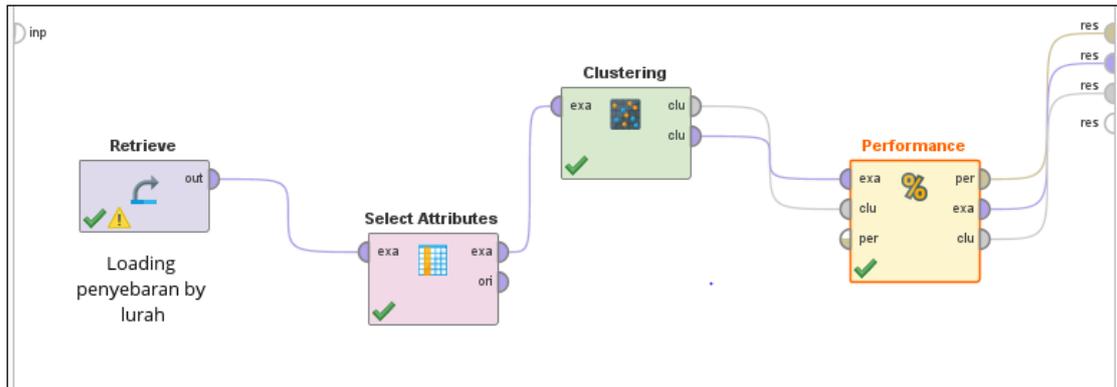
3.1 Data Transformasi

Tahapan ini dilakukan untuk mengubah isi data atau tipe data sebelum memasuki tahapan *clustering*. Atribut yang akan ditransformasi adalah kelurahan. Penambahan atribut baru yaitu jumlah anggota yang dibagi dua yaitu laki-laki dan perempuan. Nilai atribut jumlah anggota didapatkan dari membagi setiap baris data yang dikategorikan berdasarkan masing-masing kelurahan. Data yang telah melewati proses data transformasi yang tersaji pada Tabel 2.

Tabel 2. Data Hasil Tranformasi

Desa	Jumlah anggota	
	Laki laki	Perempuan
Penajam	10	11
Babulu Darat	14	19
Waru	6	4
Gunung Makmur	4	15
Rawa Mulia	2	3
Sri Raharja	2	5
Sebakung Jaya	1	2
Gunung Intan	1	0
Rintik	4	1
Labangka	6	9
Api Api	1	2
Petung	2	3
Girimukti	3	2
Saloloang	2	2
Lawe Lawe	2	0
Nipah Nipah	1	1
Sungai Parit	3	0
Nenang	0	2
Gunung Seteleng	5	4
Maridan	2	0
Bumi Harapan	3	3
Bukit Raya	0	1
Semai Dua	1	0
Argo Mulyo	2	1
Tengin Baru	1	1

Pada penelitian ini permodelan atau pengelompokan data menggunakan bantuan *tool* RapidMiner untuk mengimplementasikan algoritma K-Means Clustering. Tahapan proses permodelan yang dilakukan dalam aplikasi Rapidminer ditampilkan pada Gambar 2.



Gambar 2. Design Process Pengelompokkan Data Anggota Muhammadiyah

Adapun proses pada Gambar 2 merupakan gambaran keseluruhan proses *clustering*, terdapat 4 tahapan yang terdapat dalamnya yaitu :

1. Tahap pertama ialah melakukan *import* dalam ke dalam repository Rapid Miner menggunakan operator *import* data, jumlah data yang dimasukkan sebanyak 25 data.
2. Setelah melakukan import data, langkah selanjutnya adalah memilih atribut apa saja yang akan dikelompokkan atau diproses menggunakan algoritma K-Means Clustering. Dari tiga atribut yang ada pada dataset, dua atribut yang akan diproses lebih lanjut yaitu atribut laki-laki dan perempuan.
3. Tahap ketiga merupakan proses utama, yaitu segmentasi kelompok pada data keanggotaan Muhammadiyah. Berikut adalah penerapan algoritma K-Means menggunakan asumsi parameter input yaitu jumlah dataset sebanyak n data dengan jumlah inisialisasi *centroid* k yang telah ditentukan secara acak. Data yang digunakan berjumlah 25 data. Percobaan dilakukan menggunakan parameter berikut:
 $K = 3$
 $Data = 25$
 $Atribut = 2$
4. Cluster Distance Performance ditujukan untuk mengambil suatu model *cluster centroid* dan *clustered set* menjadi input dalam mengevaluasi dari kinerja model sesuai dengan *centroid cluster*. Dua ukuran kinerja didukung rata-rata dalam jarak *cluster* dan Davies Bouldin indeks. Dalam model ini akan dicari nilai dari DBI untuk melihat presentase kesamaan antar data dalam suatu *cluster*.

Setelah semua proses dan skema dijalankan, maka hasil klasterisasi ditampilkan pada Tabel 3.

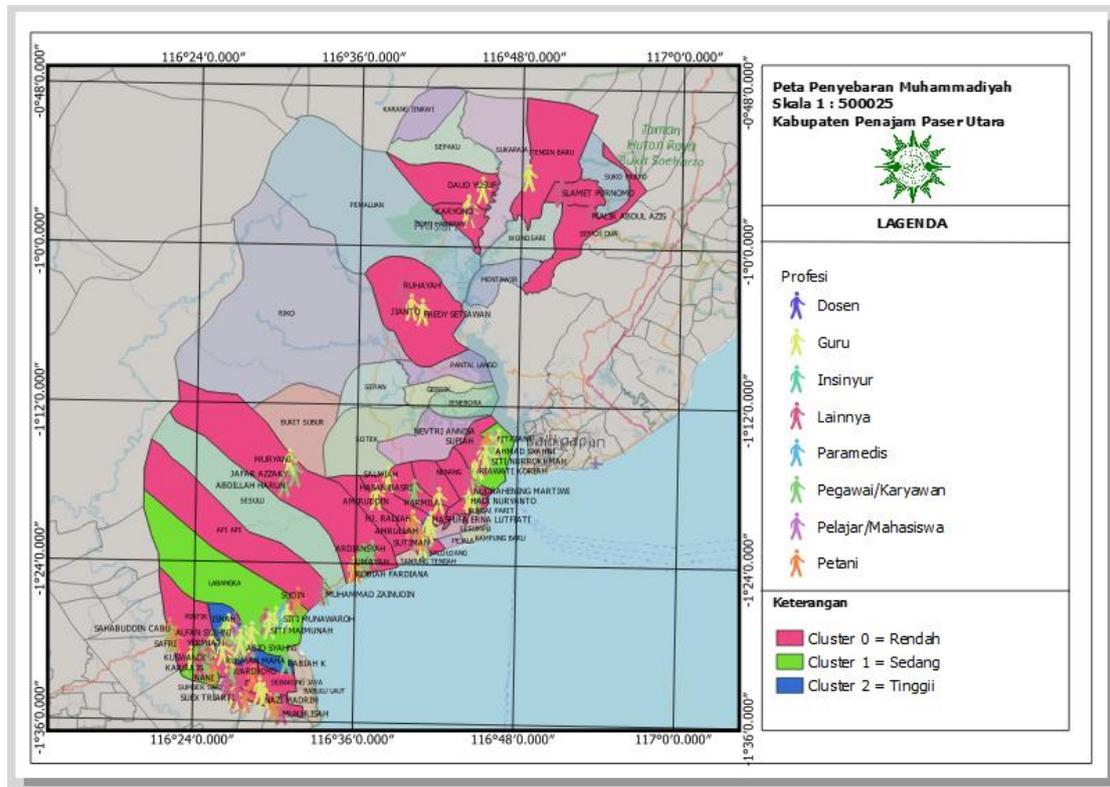
Tabel 3. Hasil Cluster

Cluster	Anggota
Cluster 0	Waru, Rawa Mulia, Sri Raharja, Sebakung Jaya, Gn Intan, Rintik, Api-Api, Petung, Girmukti, Saloloang, Lawe-Lawe, Nipah, Sungai Parit, Nenang, Gn Seteleng, Maridan, Bumi Harapan, Bukit Raya, Semoi 2, Argo Mulyo, Tengin Baru.
Cluster 1	Penajam, Gn Makmur, Labangka
Cluster 2	Babulu Darat



Gambar 3. Nilai DBI

Hasil nilai DBI dengan input data model *cluster centroid* ini dan *clustered set* menjadi input dalam menilai kinerja model berdasarkan kedekatan titik *centroid cluster* menghasilkan angka 0,218. Hasil dari tahapan ini digambarkan pada peta Kabupaten Penajam Paser Utara pada Gambar 4.



Gambar 4. Penyebaran Anggota Muhammadiyah

Peta administrasi kabupaten Penajam Paser Utara terdapat titik-titik atau plot penyebaran anggota Muhammadiyah. Pengkategorian setiap titik berdasarkan pekerjaan masing-masing anggota yang disimbolkan dengan warna dan dijelaskan pada bagian kanan peta beserta pembagian setiap kelurahan yang telah melalui labeling cluster yang terdapat didalamnya.

4. KESIMPULAN

Pada penelitian ini diperoleh kesimpulan yaitu hasil pengelompokan Kabupaten/Kota di Kabupaten Penajam Paser Utara berdasarkan jumlah anggota yang tersebar di Penajam Paser Utara menggunakan metode K-Means Clustering sebagai berikut:

1. Pemetaan anggota Muhammadiyah di Kabupaten Penajam Paser Utara menggunakan algoritma K-Means Clustering berdasarkan tingkat penyebaran di setiap kelurahan menghasilkan tiga *cluster* yaitu kelurahan dengan tingkat penyebaran anggota Muhammadiyah yang tinggi, penyebaran dengan tingkat sebaran sedang, dan penyebaran dengan tingkat sebaran yang rendah. Adanya hasil penelitian ini maka diharapkan menjadi informasi bagi Pimpinan Muhammadiyah ataupun pihak yang bersangkutan dalam pengambilan kebijakan mengenai berbagai hal.
2. Total keseluruhan anggota Muhammadiyah yang ada pada kabupaten Paser Utara ialah 168 anggota yang tersebar di 25 kelurahan yang ada di kabupaten tersebut dan peta digital yang dihasilkan memberi informasi titik atau plot tiap-tiap anggota yang bermukim di wilayah kelurahan pada kabupaten Penajam Paser Utara.

5. SARAN

Penelitian ini hanya mengkaji mengenai metode non hirarki, yaitu pada metode K-Means Clustering dalam pengaplikasiannya. Pada peneliti selanjutnya diharapkan dapat melakukan *cluster* menggunakan metode *clustering* yang lain mengingat cukup banyak metode analisis *cluster* yang dapat dikembangkan dan mengaplikasikan pada bidang ilmu lain.

DAFTAR PUSTAKA

- [1] A. Hairit, “Dinamika Pendidikan Islam Multikultural Perspektif Lembaga Pendidikan Muhammadiyah,” *Journal of Islamic Education Policy*, vol. 5, no. 1, pp. 43–58, 2020.
 - [2] A. Muhidin and I. Baragigiratri, “Pemetaan Penduduk Calon Penerima Bantuan Renovasi Rumah Desa Pesangkalan Menggunakan Algoritma Clustering K-Means,” *Jurnal SIGMA*, vol. 8, no. 2, pp. 147–152, 2018.
 - [3] G. Y. Hilman, B. Sasmito, and A. P. Wijaya, “Pemetaan Daerah Rawan Kriminalitas Di Wilayah Hukum Poltabes Semarang Tahun 2013 Dengan Menggunakan Metode Clustering,” *Jurnal Geodesi Undip*, vol. 4, no. 1, pp. 32–42, 2015.
 - [4] R. T. Vlandari, “Pengelompokan Tingkat Keamanan Wilayah Jawa Tengah Berdasarkan Indeks Kejahatan Dan Jumlah Pos Keamanan Dengan Metode Klastering K-Means,” *Jurnal Ilmiah SINUS*, vol. Vol 14, No, no. ISSN :1693-1173, pp. 59–72, 2016.
 - [5] K. Fatmawati and A. P. Windarto, “Data Mining: Penerapan Rapidminer Dengan K-Means Cluster Pada Daerah Terjangkit Demam Berdarah Dengue (Dbd) Berdasarkan Provinsi,” *Computer Engineering, Science and System Journal*, vol. 3, no. 2, p. 173, 2018, doi: 10.24114/cess.v3i2.9661.
 - [6] Suad A. Alasadi and Wesam S. Bhaya, “Review of Data Preprocessing Techniques in Data Mining,” *Journal of Engineering and Applied Sciences*, vol. 12, no. 16, pp. 4102–4107, 2017.
 - [7] G. Gustientiedina, M. H. Adiya, and Y. Desnelita, “Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan,” *Jurnal Nasional Teknologi Dan Sistem Informasi*, vol. 5, no. 1, pp. 17–24, 2019.
 - [8] J. Sumarlin, “Identifikasi Profil Dosen Berdasarkan Nilai Keandalan Perkuliahan Dengan Metode Klasterisasi K-Means,” in *Proceeding SINTAK 2019*, 2019, pp. 204–213.
 - [9] M. Wati, W. H. Rahmah, N. Novirasari, Havaluddin, E. Budiman, and Islamiyah, “Analysis K-Means Clustering to Predicting Student Graduation,” in *Journal of Physics: Conference Series*, 2021, vol. 1844, no. 1. doi: 10.1088/1742-6596/1844/1/012028.
 - [10] R. K. Dinata, H. Novriando, N. Hasdyna, and S. Retno, “Reduksi Atribut Menggunakan Information Gain Untuk Optimasi Cluster Algoritma K-Means,” *J. Edukasi dan Penelit. Inform.*, vol. 6, no. 1, pp. 48–53, 2020.
-