

PENERAPAN METODE K-MEANS DALAM MENENTUKAN KELOMPOK PENDALAMAN MATERI UJIAN NASIONAL

Yumi Novita Dewi¹, Harsih Rianto², Cahyani Budihartanti³,
Firstianty Wahyuhening Fibriany⁴

Program Studi Sistem Informasi^{1,2,3,4},

Fakultas Teknologi Informasi^{1,3}, Fakultas Teknik & Informatika^{2,4},

Universitas Nusa Mandiri^{1,3}, Universitas Bina Sarana Informatika^{2,4},

yumi.ymd@nusamandiri.ac.id¹, harsih.hhr@bsi.ac.id²,

cahyani.cbh@nusamandiri.ac.id³, firstianty.fbr@bsi.ac.id⁴

Received: November 30, 2022. **Revised:** December 16, 2021. **Accepted:**
December 21, 2021. **Issue Period:** Vol.6 No.1 (2022), Page 26-31

Abstrak: Pendalaman Materi (PM) merupakan kegiatan penting bagi siswa guna menghadapi ujian nasional. Ujian Nasional merupakan kegiatan pengukuran capaian kompetensi lulusan pada mata pelajaran tertentu secara nasional. Penelitian ini menggunakan 99 dataset berisikan data nilai rapor siswa yang diambil satu semester terakhir untuk menentukan kelompok (*cluster*) pendalaman materi terhadap mata pelajaran yang diujikan, dengan menggunakan metode *k-Means*. *k-Means* merupakan salah satu fungsi *clustering* yang melakukan proses pemodelan pengelompokan data dengan sistem partisi. Ada 4 mata pelajaran yang diujikan, yaitu; Bahasa Indonesia, Matematika, IPA, dan Bahasa Inggris. Dalam hal ini pengelompokan dibagi menjadi 2 *cluster* besar yaitu; kelas umum dan kelas khusus, dimana kelas khusus tersebut merupakan *variable* yang perlu penanganan khusus terhadap mata pelajaran yang diujikan. Perhitungan metode ini menggunakan *Rapid Miner 5.3*. dengan membagi menjadi 10 *cluster*, dimana diperoleh nilai terbaik pada *cluster* 2 dan 9, dengan nilai *centroid* 83,5 pada *cluster* 9 dimata pelajaran Bahasa Indonesia.

Kata kunci: *Cluster, Centroid, k-Means, Rapid Miner.*

Abstract: *The Study of Materials (PM) is an important activity for students to face the national exam. The National Examination is an activity measuring the achievement of graduate competencies in certain subjects nationally. This study uses 99 datasets containing data on student report cards taken in the last semester to determine the group (cluster) for deepening the material on the subjects tested, using the k-Means method. k-Means is one of the clustering functions that performs the process of modeling data grouping with a partition system. There are 4 subjects that are tested, namely; Indonesian, Mathematics, Science, and English. In this case the grouping is divided into 2 large clusters, namely; general class and special class, where the special class is a variable that needs special handling of the subjects being tested. Calculation of this method using Rapid Miner 5.3. by dividing into 10 clusters, where the best value is obtained in clusters 2 and 9, with a centroid value of 83.5 in cluster 9 in Indonesian subjects.*

Keywords: *Cluster; Centroid; k-Means; Rapid Miner.*



DOI: 10.52362/jisamar.v6i1.670

Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).

I. PENDAHULUAN

PERMENDIKBUD RI No. 4 Tahun 2018, Ujian Nasional adalah kegiatan untuk melakukan pengukuran terhadap capaian kompetensi lulusan peserta didik pada mata pelajaran tertentu secara nasional dengan mengacu pada Standar Kompetensi Lulusan [1]. Ditambahkan dalam permendikbud tersebut, kepentingan ujian nasional yang diatur dalam Peraturan Pemerintah No. 19 Tahun 2005 Bab 1 Pasal 1 Ayat 1, standar penilaian pendidikan adalah standar nasional pendidikan yang berkaitan dengan mekanisme, prosedur dan instrumen hasil belajar peserta didik. Pentingnya ujian nasional menjadikan sekolah memberikan perhatian khusus pada penilaian Ujian Nasional itu sendiri.

Pendalaman Materi (PM) merupakan rangkaian kegiatan penting yang dilakukakan oleh siswa, terutama siswa pada jenjang akhir dari satu tingkatan pembelajaran yang dimaksudkan untuk menyiapkan siswa dalam menghadapi ujian nasional. Dalam jurnal Implementasi Program Intensif Belajar yang ditulis oleh [2] untuk menghadapi Ujian Nasional di MAN 1 Jember program intensif belajar atau PM itu dapat meningkatkan nilai ujian nasional peserta didik hingga 68,73 %. Sekolah-sekolah menerapkan pendalaman pengayaan materi guna mencapai, menyesuaikan dan meningkatkan nilai standar dari lulusan peserta didik yang telah ditentukan negara.

Salah satu sekolah yang melakukan pendalaman materi adalah SMP swasta yang berada di Jakarta. Data yang didapat dari kepala sekolah dan wakil kepala sekolah bidang kurikulum. Diperoleh hasil dari wawancara yang dilakukan, bahwa di SMP tersebut seluruh siswa diwajibkan mengikuti PM terutama untuk siswa yang berada di kelas IX, dan dikelompokkan berdasarkan kelas. Namun metode pengelompokkan yang telah dibuat dan dilakukan selama ini dirasa kurang efektif, hal ini dikarenakan beberapa lulusannya tidak ada peningkatan nilai yang memuaskan. Maka, untuk tahun ajaran ini kedua wakepek itu akan menerapkan sistem baru dimana ada dua kelas PM, yang satu kelas umum dan yang satu kelas khusus, dimana pengelompokkan kelas khusus didasarkan pada nilai terendah nilai mata pelajaran ujian nasional yakni Matematika, IPA, Bahasa Indonesia dan Bahasa Inggris.

Berdasarkan kebutuhan pengelompokkan siswa tersebut maka pihak sekolah perlu melakukan pengolahan data (*data mining*) sebagai solusi pengambilan keputusan kelompok pendalaman materi tersebut. Pengolahan data, dalam hal ini *data mining* adalah proses dalam melakukan pendataan informasi-informasi penting dari data yang besar. Pengumpulan informasi yang penting tersebut kemudian dilakukan melalui sejumlah proses yang mencakup metode statistika, matematika, dan atau melalui teknologi artificial intelligence dengan menggunakan alat serta software yang menggunakan analisis statistik pada data dan menyaring serta menyimpan semula data tersebut [3][4]. Data mining tidak pernah lepas dari yang namanya dataset. Dalam pengolahan suatu data mining, dataset sangatlah dibutuhkan sebagai objek untuk mendapatkan hasil informasi. Nama lain dari objek yang sering digunakan adalah record, point, vector, pattern, event, observation, dan case. Sementara itu, baris yang menyatakan objek-objek data dan kolom disebut atribut. Atribut juga dapat disebut dengan variabel, field, fitur atau dimensi. Metode-metode yang digunakan dalam data mining berbentuk algoritma yang dapat digunakan mengekstrak pola keterhubungan antar data (asosiasi), polapola kedekatan data pada kelas-kelas tertentu (klasifikasi), dan juga untuk pengelompokan data menjadi beberapa klaster yang lebih homogen (klustering) [5]

Melalui penelitian terdahulu, dalam jurnal yang ditulis oleh [6] *Educational Data Mining* digunakan untuk memprediksi Hasil Belajar Siswa di SMAK Ora et Labora. Berdasarkan riset yang dilakukan, diperoleh kesimpulan bahwa data mining mampu membantu pihak sekolah guna meningkatkan hasil belajar siswa melalui beberapa metode-metodenya. Selain itu, dalam jurnal yang ditulis oleh [7], penggunaan data mining untuk klustering dilakukan untuk memperoleh kluster penanganan dan pelayanan kesehatan masyarakat di Kabupaten Blora.

Berdasarkan penelitian-penelitian sebelumnya, metode yang banyak dipilih adalah metode *Clustering*, *k-Means*. Menurut Celebi Kingravi (2013) dalam [8] algoritma *k-Means* merupakan salah satu algoritma populer yang digunakan dalam proses *clustering dataset* karena kesederhanaan algoritmanya. Senada dengan pendapat Jannah, [9], selain mampu menangani dataset dengan ukuran yang besar, algoritma ini juga memiliki tiga keunggulan yaitu kompleksitas waktu, kompleksitas ruang penyimpanan, dan pemrosesan tidak bergantung pada urutan *centroid* yang digunakan. Disamping itu, algoritma *k-Means* juga bersifat *versatile* yaitu mudah melakukan modifikasi disetiap tahapan dalam algoritma, misal dalam inisialisasi, fungsi penghitungan jarak, dan juga kriteria penghentian iterasi [10].



II. METODE DAN MATERI

2.1. Metode *k-Means*

Metode *k-Means* merupakan salah satu metode *Clustering Non-Hierarchical* yang berusaha mempartisi *dataset* ke dalam bentuk satu atau lebih kelompok (*cluster*). *Cluster* merupakan sekumpulan data yang serupa satu dengan yang lainnya dan berbeda dengan data di *cluster* lainnya. Penelitian ini menggunakan metode *k-Means*, dimana *k-Means* adalah metode yang efektif guna menentukan *cluster* dalam sebuah *dataset* dimana pada metode *k-Means* ini akan melakukan analisis kelompok yang mengacu pada permartisian *n* objek ke dalam *k* (*cluster*) berdasarkan nilai rata-rata (*means*) terdekat. Adapun tahapan metode *k-Means* ini adalah:

1. Menentukan jumlah *k* (*cluster*) dengan cara menentukan data yang menjadi *centroid* secara acak.
2. Menemukan *centroid* terdekat untuk setiap *record* dengan rumus *Euclidean* dengan cara menghitung nilai *centroid* masing-masing *cluster*.
3. Mengalokasikan masing-masing *dataset* ke *centroid* terdekat, dan memungkinkan untuk kembali ke tahap 2 apabila masih terdapat perpindahan data dari satu *cluster* ke *cluster* yang lain, atau apabila perubahan pada nilai *centroid* dirasa masih di atas nilai *threshold* (nilai ambang) yang ditentukan.

2.2. Pra Proses *k-Means*

k-Means menentukan jumlah *cluster* dengan melakukan inisiasi *k* terhadap pusat *cluster* yang diberi dengan nilai secara *random* berdasarkan jarak ke dua objek terdekat, dengan menggunakan rumus sebagai berikut:

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

Dimana $x = x_1, x_2, x_3, \dots, x_n$, $y = y_1, y_2, y_3, \dots, y_n$, dan *n* menyatakan banyaknya nilai atribut dari 2 data. Selanjutnya akan dihitung nilai *centroid* masing-masing *cluster* dengan menggunakan rumus berikut:

$$v_{ij} = \frac{\sum_{k=1}^{N_i} x_{kj}}{N_i}$$

dimana:

N_i : Jumlah data yang menjadi anggota *cluster* ke-*i*

Mekanisme penelitian ini adalah dengan melakukan pemilahan terhadap 99 *dataset* yang berisikan data dari nilai rapor siswa yang diambil satu semester terakhir. Sehingga dibentuklah 2 *cluster* besar yaitu *cluster* kelas umum dan *cluster* kelas khusus. *Cluster* kelas khusus ini merupakan kelompok yang digolongkan ke dalam kelompok siswa yang perlu mendapatkan Pendalaman Materi (PM) terhadap 4 mata pelajaran yang diujikan dalam Ujian Nasional.

Tahapan *k-Means* selanjutnya adalah menentukan nilai *mean* atau nilai rata-rata dalam kelompok mata pelajaran. Nilai *mean* ini akan dijadikan nilai *centroid* selanjutnya sehingga didapatkanlah *cluster* model seperti pada tabel1 berikut:

Tabel 1. Nilai *Mean* dengan 2 *Cluster*

Attribute:	Cluster Umum:	Cluster Khusus:
Bhs. Indonesia	80,109	73,067
Matematika	75,250	70,985
IPA	76,281	71,403
Bhs. Inggris	77,031	72,112
View	32 Items	67 Items
Total Number of Items	99 Item	



DOI: 10.52362/jisamar.v6i1.670

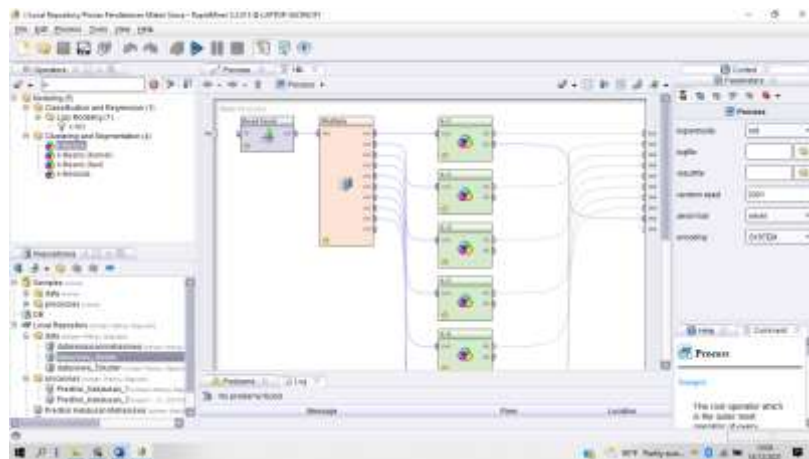
Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).

Dari tabel 1 tersebut dapat disimpulkan bahwa hasil perhitungan *mean* dari *clustering* terhadap 99 *dataset*, menunjukkan rata-rata nilai siswa yang tercatat memiliki nilai >75,00 sebanyak 32 *items* termasuk ke dalam *cluster umum* dan pada *cluster khusus* tercatat siswa yang memiliki nilai <75,00 sebanyak 67 *items*. Artinya pada *cluster khusus* perlu adanya penanganan khusus terhadap mata pelajaran yang akan diujikan.

III. PEMBAHASAN DAN HASIL

Setelah *cluster* di setiap *dataset* telah ditentukan seperti pada tabel 1 diatas, maka langkah selanjutnya adalah dengan memperbaharui nilai *centroid*, yang mana sebelumnya nilai *centroid cluster* didapatkan sebanyak 67 *items*. Selanjutnya akan dilakukannya perhitungan *dataset* untuk menemukan 10 *cluster* di *cluster kelas khusus Rapid Miner 5.3*.

Rapid Miner 5.3 memiliki fitur yang sangat memudahkan dalam pengolahan *data mining*. Proses pada tahapan ini akan dilakukan perhitungan untuk mengalokasikan 67 *items* pada *cluster khusus* menjadi 10 *cluster* terkecil, dengan tujuan agar dapat terlihat lebih jelas *cluster* mana yang teridentifikasi sebagai *cluster* dengan nilai *centroid* terendah. Pada Gambar1 berikut ini menampilkan *main process* dari *clustering data 67 items* :



Gambar 1. Perhitungan *Rapid Miner 5.3. Cluster Khusus*.

Proses selanjutnya dari tabel nilai *centroid* pada tabel 1, ditemukan adanya data nilai *centroid* yang berpindah ke kelompok lain dan mengakibatkan pengelompokan pada iterasi yang belum *konvergen*.

Tabel 2. Nilai *Centroid Model 10 Cluster*

Attribute	Cluster_1	Cluster_2	Cluster_3	Cluster_4	Cluster_5	Cluster_6	Cluster_7	Cluster_8	Cluster_9	Cluster_10
Bhs. Indonesia	75,179	78,5	75,5	72,7	79,85	86	71,065	74,071	83,5	78,429
Matematika	70	76,125	73,75	76,9	76,2	71	70	70,286	80	70,286
IPA	70,714	76,875	70	74,1	73,5	75	70,516	75,143	85	74,571
Bhs. Inggris	72,179	78,694	74,708	72,728	73,839	81,267	71,43	73,222	78,333	74,817
View	3 Items	5 Items	9 Items	7 Items	4 Items	19 Items	3 Items	8 Items	2 Items	7 Items
Total Number of Items	67 Item									

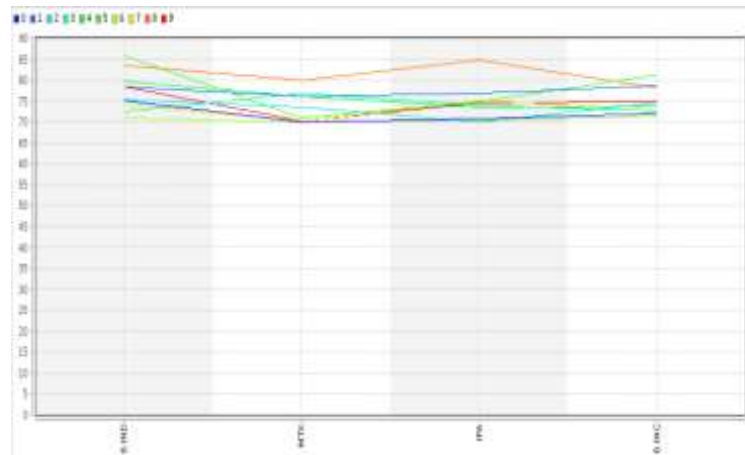
Dari tabel 2 tersebut dapat dilihat bahwa tidak terjadinya adanya perpindahan data kelompok disetiap data dan hal ini dinyatakan bahwa pengelompokan sudah konvergen atau telah dianggap optimal sesuai dengan nilai *threshold* (batas ambang) yang ditentukan.



DOI: 10.52362/jisamar.v6i1.670

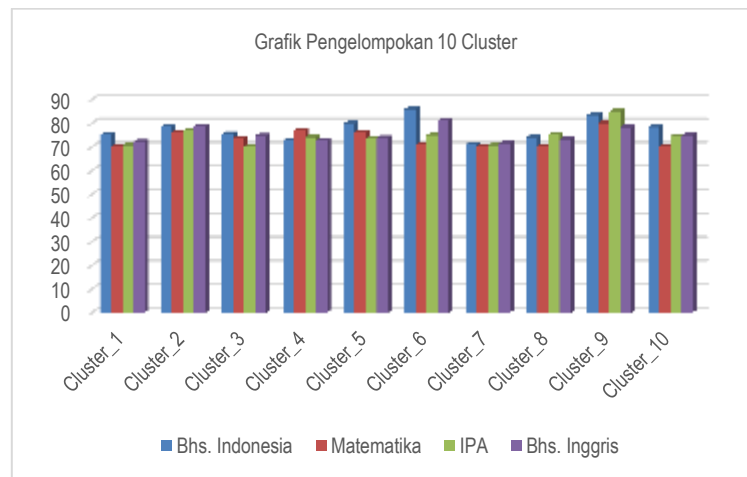
Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).

Dengan adanya banyak kejadian iterasi yang berpindah pada saat pemilihan acak pada tabel 1, maka dapat disimpulkan bahwa dalam tabel *centroid* tersebut memiliki selisih jarak yang cukup jauh sehingga dapat mewakili *cluster* data yang akan dikelompokkan kembali. Terlihat bahwa pembentukan kelompok berdasarkan jarak dari masing-masing dataset dengan pembentukan kelompok *cluster_1*, *cluster_2*, *cluster_3*, *cluster_4*, *cluster_5*, *cluster_6*, *cluster_7*, *cluster_8*, *cluster_9*, *cluster_10* dengan nilai *centroid* seperti pada gambar 2 berikut:



Gambar 2. Tampilan *Centroid Plot*.

Terbentuknya kelompok berdasarkan jarak dari masing-masing dataset seperti pada gambar 2, dimana adanya proses alokasi masing-masing item data ke rata-rata nilai terdekat. Data yang berpindah *cluster* ataupun adanya perubahan nilai *centroid* menggunakan nilai *threshold* (batas ambang) seperti yang telah digambarkan pada grafik berikut:



Gambar 3. Grafik Pengelompokan 10 *Cluster*

Dari grafik pada gambar 3 dapat disimpulkan bahwa nilai *centroid* dari *clustering* terhadap 67 *items* pada *cluster* khusus, menunjukkan nilai terbaik ada di *cluster 2* dan *cluster 9*. Nilai ***threshold* “Tertinggi”** berada di ***cluster 9*** dimana mata pelajaran Bhs. Indonesia **83,5**, Matematika **80**, IPA **85**, dan Bhs. Inggris **78.3**. Untuk nilai ***threshold* yang menunjukkan angka “Baik”** ada di ***cluster 2*** dimana mata pelajaran Bhs. Indonesia **78,5**, Matematika **76,1**, IPA **76,8**, dan Bhs. Inggris **78.6**. Artinya siswa yang berada di *cluster 2* dan *cluster 9* **Tidak Perlu** adanya kelompok pendalaman materi terhadap mata pelajaran yang akan diujikan. Dan siswa

yang beara pada *cluster 1, 3, 4, 5, 6, 7, 8*, dan *cluster 10 Disarankan* untuk membentuk kelompok Pendalaman Materi guna menghadapi Ujian Nasional.

IV. KESIMPULAN

Dari penjabaran tersebut maka dapat disimpulkan bahwa dengan menerapkan metode *k-Means* dapat membantu untuk menentukan kelompok Pendalaman Materi siswa terhadap 4 mata pelajaran yang akan diujikan dalam Ujian Nasional yaitu Bhs.Indonesia, Matematika, IPA, dan Bhs.Ingggris. Hasil penelitian inimenemukan bahwa siswa yang berada pada *cluster 2* dan *cluster 9*. Dengan nilai **threshold “Tertinggi”** berada pada **cluster 9** dan menunjukkan nilai **“Baik”** ada di **cluster 2**. Artinya siswa yang berada di *cluster 2* dan *cluster 9 Tidak Perlu* adanya kelompok pendalaman materi terhadap mata pelajaran yang akan diujikan, dan siswa yang beara pada *cluster 1, 3, 4, 5, 6, 7, 8*, dan *cluster 10 Disarankan* untuk membentuk kelompok Pendalaman Materi guna menghadapi Ujian Nasional.

REFERENSI

- [1] Kementerian Pendidikan Nasional, “Peraturan menteri pendidikan dan kebudayaan republik indonesia nomor 4 tahun 2018 tentang penilaian hasil belajar oleh satuan pendidikan dan penilaian hasil belajar oleh pemerintah,” *Permendikbud*, pp. 1–16, 2018.
- [2] K. I. Kasanah, B. Suyadi, and S. Sukidin, “Implementasi Program Intensif Belajar Untuk Menghadapi Ujian Nasional Kelas Xii Ips Di Man 1 Jember Tahun Ajaran 2016/2017,” *Ilm. Ilmu Pendidikan, Ilmu Ekon. Dan Ilmu Sos.*, vol. 12, no. 1, p. 128, 2018, doi: <https://doi.org/10.19184/jpe.v12i1.7623>.
- [3] Asfihan, “Pengertian Data Mining,” *ruangpengetahuan*, 2021. <https://ruangpengetahuan.co.id/pengertian-data-mining/>.
- [4] Sutrisno, Afriyudi, and Widiyanto, “Penerapan Data Mining Pada Penjualan Menggunakan Metode Clustering Study Kasus Pt . Indomarco,” *Penerapan Data Min. Pada Penjualan Menggunakan Metod. Clust.*, vol. Vol.x No.x, no. Data Mining, pp. 1–11, 2013, [Online]. Available: <http://eprints.binadarma.ac.id/78/1/PENERAPAN DATA MINING PADA PENJUALAN MENGGUNAKAN METODE CLUSTERING STUDY KASUS PT. INDOMARCO PALEMBANG.pdf>.
- [5] M. A. Muslim *et al.*, *Data Mining Algoritma C4.5*, Pertama. Semarang: unnes, 2019.
- [6] C. Study and D. David, “Penerapan Educational Data Mining Untuk Memprediksi Hasil Belajar Siswa SMAK Ora et Labora,” *Ilmu Komput.*, vol. XII, no. 2, pp. 73–82, 2019, doi: <https://doi.org/2622-321X>.
- [7] N. Cahyana and A. Aribowo, “Metode Data Mining K-Means Untuk Klasterisasi Data Penanganan Dan Pelayanan Kesehatan Masyarakat,” *Semin. Nas. Inform. Medis*, no. 5, pp. 24–31, 2018.
- [8] A. R. Jannah, D. Arifianto, and M. Kom, “Penerapan Metode Clustering dengan Algoritma K-Means untuk Prediksi Kelulusan Mahasiswa Jurusan Teknik Informatika di Universitas Muhammadiyah Jember,” *J. Manaj. Sist. Inf. dan Teknol.*, vol. 1, no. 1210651237, pp. 1–10, 2015.
- [9] R. A. Indraputra and R. Fitriana, “K-Means Clustering Data COVID-19,” *J. Tek. Ind.*, vol. 10, no. 3, p. 3, 2020.
- [10] Y. P. Santoso, M. Marlina, and H. Agung, “Implementasi Metode K-Means Clustering pada Sistem Rekomendasi Dosen Tetap Berdasarkan Penilaian Dosen,” *J. Inform. Univ. Pamulang*, vol. 3, no. 4, p. 228, 2018, doi: [10.32493/informatika.v3i4.2133](https://doi.org/10.32493/informatika.v3i4.2133).



DOI: 10.52362/jisamar.v6i1.670

Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).