

OPTIMALISASI MODEL CLUSTERING DATA PENDUDUK DESA FOHOEKA MENGUNAKAN ALGORITMA K-MEANS

Ignasius Anjelino Emanuel¹, Ade Irma Purnamasari², Denni Pratama³, Irfan Ali⁴.

Program Studi Teknik Informatika¹²
Program Studi Komputerisasi Akuntansi³
Program Studi Rekayasa Perangkat Lunak⁴

STMIK IKMI Cirebon
<https://ikmi.ac.id/page/18/?lang=de>
jinoemanuel@gmail.com

(*) Corresponding Author : jinoemanuel@gmail.com.
Published : 30 Maret 2026

Abstract—This study develops and evaluates an optimized K-Means algorithm for clustering population data in Fohoeka Village, with the goal of improving accuracy, stability, and interpretability of demographic segmentation as a basis for evidence-based village planning. The dataset consists of 1,690 individual records with 16 initial attributes (expanded to 59 numerical features after preprocessing and one-hot encoding). The methodological steps include data cleaning (imputation of missing values and correction of entries), transformation of ordinal variables into numerical scales, one-hot encoding for categorical variables, feature standardization using Standard Scaler, and outlier detection and handling prior to model training. To reduce sensitivity to centroid initialization and increase the likelihood of converging to better solutions, advanced initialization strategies—such as K-Means++ and multiple restarts—were applied, along with validation of the optimal number of clusters using the Elbow Method, Silhouette Score, and WSS visualization. Internal evaluation indicates that the highest Silhouette Score occurs at $k = 10$, leading to a final model consisting of ten interpretable demographic clusters. Centroid analysis reveals heterogeneous patterns: several large clusters are dominated by farmers, housewives, and students, indicating majority groups with essential service needs, while smaller clusters represent individuals with relatively higher income or professional occupations. The findings highlight that improved preprocessing and initialization significantly enhance cluster coherence and reduce intra-cluster dispersion, making the clustering results more reliable for prioritizing social programs, resource allocation, and identifying vulnerable groups at the village level. Future research is recommended to explore robust K-Means variants to mitigate outlier effects and integrate spatial data for location-based aggregation analysis.

Keywords: K-Means, demographic clustering, preprocessing, cluster validity, Fohoeka Village.

Abstrak— Penelitian ini mengembangkan dan menguji model optimalisasi algoritma K-Means untuk klusterisasi data kependudukan di Desa Fohoeka dengan tujuan meningkatkan akurasi, stabilitas, dan keterbacaan hasil segmentasi demografis sebagai dasar perencanaan desa berbasis bukti. Dataset yang digunakan berisi 1.690 rekaman individu dengan 16 atribut awal (setelah pra-pemrosesan dan one-hot encoding menjadi 59 fitur numerik). Tahapan metodologis meliputi pembersihan data (imputasi nilai hilang dan koreksi entri), transformasi variabel ordinal menjadi skala numerik, one-hot encoding untuk variabel kategorikal, standarisasi fitur (Standard Scaler), serta deteksi dan penanganan outlier sebelum pelatihan model. Untuk mengurangi sensitivitas inisialisasi centroid dan meningkatkan peluang konvergensi ke solusi yang lebih baik, diterapkan strategi inisialisasi cerdas (mis. K-Means++ dan multiple restarts) serta validasi pemilihan jumlah kluster melalui kombinasi Elbow Method, Silhouette Score, dan visualisasi WSS. Hasil evaluasi internal menunjukkan nilai Silhouette tertinggi pada $k = 10$; model akhir menghasilkan sepuluh kluster yang dapat diinterpretasikan secara demografis. Analisis centroid mengungkap pola heterogen: beberapa kluster besar didominasi oleh petani, ibu rumah tangga, dan pelajar (menandakan kelompok mayoritas dengan kebutuhan layanan dasar), sedangkan kluster kecil menunjukkan segmen berpendapatan relatif lebih tinggi atau pekerjaan profesional. Temuan menegaskan bahwa optimasi pra-pemrosesan dan inisialisasi signifikan meningkatkan koherensi kluster dan menurunkan dispersi intra-kluster, sehingga hasil klusterisasi lebih layak dipakai untuk penentuan prioritas program sosial, alokasi sumber daya, dan identifikasi kelompok rentan di tingkat desa. Rekomendasi penelitian mencakup pengujian varian robust

K-Means untuk mengurangi pengaruh outlier dan integrasi data spasial untuk analisis agregasi berbasis lokasi.

Kata Kunci: K-Means, klasterisasi demografis, pra-pemrosesan, validitas kluster, Desa Fohoeka.

INTRODUCTION

Pengelolaan data penduduk di tingkat desa masih menghadapi tantangan signifikan karena data sering tersebar pada berbagai sumber dan disimpan dalam format yang tidak terstruktur sehingga menghambat upaya analisis untuk mendukung perencanaan berbasis bukti [1], [2]. Ketidakkonsistenan penulisan, anomali entri, dan pencatatan yang tidak standar memperburuk kualitas dataset demografi sehingga memerlukan teknik pendeteksian kesalahan yang lebih andal, terutama ketika volume data terus meningkat [3], [4]. Kompleksitas tersebut menunjukkan bahwa data kependudukan tidak dapat dikelola secara efektif hanya dengan pendekatan manual, mengingat tingginya variasi atribut dan pertumbuhan jumlah data dari waktu ke waktu [5], [6]

Selain tantangan kualitas data, variasi administratif dan perbedaan batas wilayah menimbulkan masalah agregasi spasial seperti modifiable areal unit problem (MAUP) yang menyebabkan bias estimasi populasi dan menutupi heterogenitas lokal yang penting dalam penentuan kebijakan desa [7], [8]. Bias tersebut dapat mengakibatkan kesalahan dalam memahami kebutuhan masyarakat karena karakteristik wilayah menjadi tereduksi oleh agregasi statistik yang tidak tepat sasaran [9], [10]. Kondisi ini memperkuat urgensi penerapan metode analitis yang mampu memberikan segmentasi penduduk secara lebih akurat dan mempertimbangkan variasi data yang kompleks pada tingkat lokal [11], [12].

Metode klustering menjadi pendekatan penting dalam mengatasi tantangan tersebut karena mampu mengungkap pola laten melalui pengelompokan berdasarkan kesamaan atribut tanpa memerlukan label sebelumnya [13], [14].

K-Means merupakan algoritma yang paling sering digunakan karena kesederhanaan mekanisme partisi dan efisiensinya dalam mengolah dataset besar, sehingga sering diterapkan dalam berbagai konteks demografi [15]. Berbagai penelitian menunjukkan bahwa K-Means mampu mengidentifikasi pola pembangunan infrastruktur, tingkat ketertinggalan wilayah, serta variasi kepadatan penduduk, sehingga mendukung pengambilan keputusan yang lebih tepat sasaran [16], [17].

Pemanfaatan K-Means dalam menganalisis harapan hidup, kebutuhan keluarga berencana, dan keterlambatan administrasi kelahiran menunjukkan fleksibilitas metode ini dalam menangani data demografi yang beragam [18], [19]

Walaupun banyak digunakan, algoritma K-Means memiliki beberapa keterbatasan yang berpotensi menurunkan kualitas hasil kluster apabila tidak dioptimalkan, seperti sensitivitas terhadap inisialisasi centroid yang sering menyebabkan konvergensi pada solusi suboptimal. Asumsi bentuk kluster yang bulat dan ukuran yang seragam menjadikan algoritma ini kurang sesuai untuk data demografi yang tidak seimbang atau tidak berbentuk sferis. Kehadiran outlier atau kesalahan pencatatan dapat mengengaruhi posisi centroid secara signifikan dan menghasilkan kluster yang menyimpang dari pola populasi sebenarnya. Selain itu, kebutuhan untuk menentukan jumlah kluster di awal membuat proses analisis menjadi subjektif dan memerlukan teknik validasi tambahan seperti Elbow dan Silhouette untuk memperoleh hasil yang [20], [21].

MATERIALS AND METHODS



Gambar 1. Alur Penelitian

1. Pengumpulan Data (Data Collection)
Tahap awal penelitian dimulai dengan proses pengumpulan dataset yang akan digunakan dalam proses clustering. Data dapat berasal dari berbagai sumber seperti database, dataset publik, maupun data hasil pengukuran atau survei. Data yang diperoleh kemudian diperiksa kelengkapannya untuk memastikan bahwa data tersebut dapat digunakan pada tahap analisis selanjutnya.

2. Pra-pemrosesan Data (Data Preprocessing)
Pada tahap ini dilakukan proses pembersihan dan persiapan data sebelum diterapkan algoritma K-Means. Proses preprocessing meliputi beberapa langkah seperti pembersihan data (data cleaning), penanganan nilai hilang (missing value), deteksi dan penghapusan outlier, serta normalisasi atau scaling data agar setiap atribut memiliki skala yang sama. Tahapan ini penting karena kualitas data sangat mempengaruhi hasil clustering yang dihasilkan.

3. Penentuan Jumlah Cluster Optimal (k)
Setelah data siap digunakan, langkah berikutnya adalah menentukan jumlah cluster yang optimal. Penentuan nilai k dapat dilakukan menggunakan beberapa metode seperti Elbow Method, Silhouette Coefficient, atau Gap Statistic. Metode ini bertujuan untuk menemukan jumlah cluster yang menghasilkan pembagian data paling optimal berdasarkan tingkat kedekatan antar data dalam cluster dan jarak antar cluster.

4. Inisialisasi Centroid Awal
Pada tahap ini ditentukan titik centroid awal yang akan digunakan dalam proses K-Means. Pemilihan centroid awal sangat penting karena dapat mempengaruhi hasil clustering. Beberapa pendekatan yang digunakan dalam penelitian antara lain K-Means++, metode berbasis Minimum Spanning Tree, atau metode inisialisasi centroid berbasis diferensiasi atribut.

5. Proses Clustering Menggunakan Algoritma K-Means

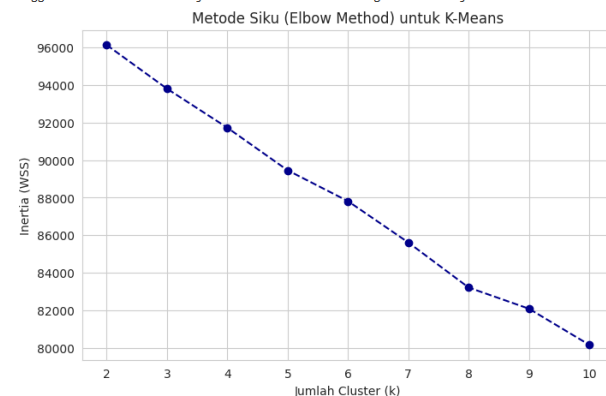
Tahap ini merupakan inti dari penelitian, yaitu melakukan proses pengelompokan data menggunakan algoritma K-Means. Prosesnya dimulai dengan menghitung jarak antara data dan centroid, kemudian setiap data akan dimasukkan ke cluster dengan jarak terdekat. Setelah itu centroid akan diperbarui berdasarkan rata-rata data dalam cluster. Proses ini dilakukan secara iteratif hingga centroid tidak lagi berubah atau telah mencapai kondisi konvergen.

6. Evaluasi dan Optimasi Hasil Clustering
Tahap terakhir adalah mengevaluasi kualitas cluster yang dihasilkan. Evaluasi dilakukan menggunakan beberapa metrik seperti Silhouette Score, Davies-Bouldin Index (DBI), dan Sum of Squared Error (SSE). Hasil evaluasi ini digunakan untuk mengetahui seberapa baik model clustering bekerja serta untuk membandingkan performa metode yang digunakan dengan metode lainnya.

RESULTS AND DISCUSSION

Dalam proses klusterisasi menggunakan algoritma K-Means, pemilihan jumlah cluster (k) yang optimal merupakan langkah penting agar hasil pengelompokan lebih akurat dan representatif. Salah satu teknik yang umum digunakan untuk menentukan nilai k terbaik adalah *Elbow Method*, yang memvisualisasikan hubungan antara jumlah cluster dan nilai *Within-Cluster Sum of Squares (WSS)*

Menggunakan inisialisasi lanjutan K-Means++ untuk meningkatkan kinerja.



Dari plot di atas, nilai k optimal adalah titik 'siku' (elbow).

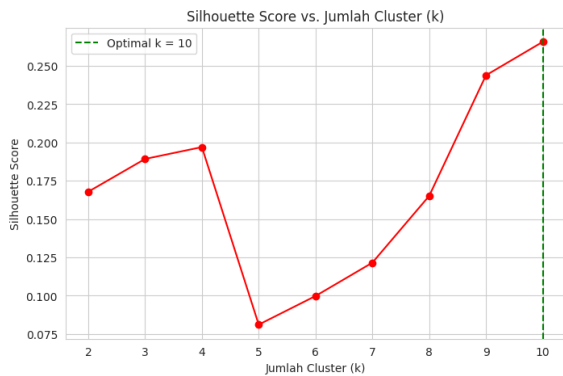
Gambar 2 Hasil Jumlah Kluster Optimal

Gambar tersebut menampilkan grafik *Elbow Method* untuk menentukan jumlah cluster optimal pada algoritma K-Means. Pada sumbu horizontal ditunjukkan jumlah cluster (k) dari 2 hingga 10, sedangkan sumbu vertikal menggambarkan nilai inertia atau WSS, yaitu total variasi dalam cluster. Terlihat bahwa seiring bertambahnya jumlah cluster, nilai WSS terus menurun secara bertahap, menunjukkan bahwa penambahan cluster mampu mengurangi kesalahan pengelompokan. Namun, penurunan tersebut mulai melambat setelah titik tertentu. Pola garis putus-putus pada grafik memperlihatkan adanya perubahan sudut atau "siku" yang menjadi indikasi nilai k terbaik. Dari visual tersebut, titik siku menjadi acuan untuk menentukan jumlah cluster yang paling optimal, karena setelah titik itu penurunan WSS tidak lagi

signifikan. Dengan demikian, grafik ini membantu mengidentifikasi bahwa nilai k optimal adalah pada titik di mana terjadi perubahan kelengkungan paling mencolok.

Evaluasi Kinerja K-Means

Setelah menentukan kandidat jumlah cluster menggunakan *Elbow Method*, langkah berikutnya adalah memvalidasi pilihan tersebut dengan menggunakan *Silhouette Score*. Metode ini sangat penting karena memberikan ukuran seberapa baik objek berada dalam klusternya dibandingkan dengan cluster lain. Semakin tinggi nilai silhouette, semakin jelas pemisahan antar cluster dan semakin baik kualitas klasterisasi.



Nilai K Optimal berdasarkan Silhouette Score tertinggi: k = 10 (Score: 0.2657)

Gambar 3 Hasil Evaluasi Kinerja K-Means

Gambar tersebut menunjukkan grafik hubungan antara jumlah cluster (k) dan nilai Silhouette Score pada proses klasterisasi K-Means. Pada sumbu horizontal ditampilkan jumlah cluster dari k = 2 hingga k = 10, sedangkan sumbu vertikal menunjukkan nilai silhouette yang mengukur kualitas pengelompokan. Dari grafik terlihat bahwa nilai silhouette mengalami fluktuasi: meningkat pada k = 3 dan k = 4, kemudian turun tajam pada k = 5, dan perlahan meningkat kembali pada k berikutnya. Peningkatan terbesar terjadi pada k = 9 dan mencapai nilai tertinggi pada k = 10 dengan skor sekitar 0.2657. Garis putus-putus berwarna hijau menegaskan bahwa k = 10 dipilih sebagai jumlah cluster optimal berdasarkan nilai *Silhouette Score* tertinggi. Hal ini menunjukkan bahwa pembagian data ke dalam 10 cluster memberikan struktur pengelompokan yang paling baik dan paling terpisah secara konsisten dibandingkan nilai k lainnya.

Ringkasan Cluster (Rata-rata Fitur Terskala - Centroid)

Berikut ini merupakan interpretasi naratif dari tampilan ringkasan centroid hasil pengelompokan (*clustering*). Setiap cluster menunjukkan nilai rata-rata fitur yang telah ditransformasi dalam bentuk skala, sehingga dapat memberikan gambaran karakteristik umum dari masing-masing kelompok responden.

6.2.1 Ringkasan Cluster (Rata-rata Fitur Terskala - Centroid):

Cluster	PENDAPATAN_SKALA	PENDIDIKAN_SKALA	STATUS_PERKAWINAN_Belum Kawin	STATUS_PERKAWINAN_Ela Kawin	STATUS_PERKAWINAN_Coral Mati	STATUS_PERKAWINAN_Duda	STATUS_PERKAWINAN_Janda
0	0.98	3.26	1.0	0.00	0.00	0.00	0.00
1	0.97	1.39	0.0	0.00	0.00	0.00	1.0
2	0.33	1.95	0.0	0.00	0.02	0.26	0.0
3	1.00	6.00	0.0	1.00	0.00	0.00	0.0
4	0.91	3.49	0.0	0.97	0.00	0.03	0.0
5	0.94	2.65	0.0	0.00	0.00	0.00	0.0
6	0.00	1.45	0.0	0.85	0.00	0.05	0.0
7	0.00	1.36	0.0	0.00	0.00	0.00	0.0
8	0.20	3.00	0.0	1.00	0.00	0.00	0.0
9	1.00	1.00	0.0	1.00	0.00	0.00	0.0

10 rows x 59 columns

STATUS_PERKAWINAN_Kawin	PEKERJAAN_ART	PEKERJAAN_Aparat Desa	...	PEKERJAAN_Tutor Paed	PEKERJAAN_Veteran	PEKERJAAN_huruh	PEKERJAAN_karyawan	PEKERJAAN_karyawan Susta	PEKERJAAN_karyawan smasta
0.00	0.0	0.0	...	0.0	0.0	0.00	0.0	0.00	0.0
0.00	0.0	0.0	...	0.0	0.0	0.00	0.0	0.00	0.0
0.71	0.0	0.0	...	0.0	1.0	0.00	0.0	0.00	0.0
0.00	0.0	0.0	...	0.0	0.0	0.00	0.0	0.00	0.0
0.00	0.0	0.0	...	0.0	0.0	0.01	0.0	0.01	0.0
1.00	0.0	0.0	...	0.0	0.0	0.00	0.0	0.00	0.0
0.10	0.0	0.0	...	0.0	0.0	0.00	0.0	0.00	0.0
1.00	0.0	0.0	...	0.0	0.0	0.00	0.0	0.00	0.0
0.00	0.0	0.0	...	0.0	0.0	0.00	0.0	0.00	0.0
0.00	0.0	0.0	...	0.0	0.0	0.00	0.0	0.00	0.0

Gambar 4 Rata-rata Fitur terskala-centroid

Gambar tersebut menampilkan ringkasan centroid dari sepuluh cluster yang terbentuk berdasarkan variabel-variabel demografis seperti pendapatan, pendidikan, status perkawinan, dan pekerjaan. Setiap baris mewakili satu cluster, sementara setiap kolom berisi nilai rata-rata dari fitur yang digunakan dalam proses clustering. Nilai-nilai ini menggambarkan kecenderungan dominan dalam tiap cluster.

Secara umum, beberapa cluster tampak berbeda secara mencolok berdasarkan status perkawinan. Misalnya, Cluster 0 memiliki nilai tinggi pada kategori Belum Kawin, sedangkan Cluster 1 menunjukkan dominasi pada kategori Janda. Cluster lainnya seperti Cluster 3 dan 8 menunjukkan nilai maksimum pada kategori Kawin, sehingga dapat ditafsirkan sebagai kelompok yang sebagian besar berstatus menikah. Variasi ini mengindikasikan bahwa status perkawinan menjadi salah satu variabel yang cukup membedakan karakteristik kelompok.

Selain itu, perbedaan terlihat pula pada variabel pendapatan dan pendidikan. Beberapa cluster seperti Cluster 0 dan 4 memiliki skor pendapatan skala yang relatif tinggi, sedangkan cluster lain seperti Cluster 8 menunjukkan nilai yang lebih rendah. Variabel pendidikan pun mengalami variasi

serupa, di mana cluster tertentu cenderung memiliki tingkat pendidikan lebih tinggi dibandingkan cluster lainnya, hal ini terlihat dari nilai rata-rata pendidikan yang bervariasi cukup lebar.

Pada bagian pekerjaan, tampak bahwa sebagian besar cluster memiliki nilai rendah atau mendekati nol pada sebagian besar kategori pekerjaan, menandakan bahwa pekerjaan-pekerjaan tertentu hanya muncul secara dominan pada cluster spesifik. Sebagai contoh, Cluster 2 menampilkan kecenderungan lebih tinggi pada kategori Veteran, sementara cluster lain lebih menonjol pada kategori seperti Karyawan Swasta atau Buruh meskipun dengan proporsi yang kecil.

Secara keseluruhan, tabel centroid ini memberikan gambaran bahwa setiap cluster memiliki karakteristik khas yang ditentukan oleh kombinasi status perkawinan, tingkat pendidikan, pendapatan, dan jenis pekerjaan. Hal ini penting untuk memahami segmentasi responden serta untuk menyusun strategi atau analisis lanjutan berdasarkan tiap kelompok yang terbentuk.

CONCLUSION

Berikut ini adalah kesimpulan dalam penelitian ini, antara lain :

1. Penelitian berhasil merancang alur optimalisasi K-Means melalui pembersihan data, standarisasi, *encoding*, dan penentuan jumlah klaster menggunakan metode *Elbow* dan *Silhouette*. Tahapan ini terbukti meningkatkan akurasi klaster, mengurangi distorsi jarak, serta menghasilkan segmentasi demografis yang stabil untuk data penduduk Fohoea.
2. Hasil penelitian menunjukkan bahwa K-Means rentan terhadap pemilihan *centroid* awal dan keberadaan *outlier* sehingga stabilitas hasil belum sepenuhnya optimal. Meskipun K-Means++ dan normalisasi telah membantu meningkatkan kualitas, diperlukan teknik lanjutan untuk mencapai stabilitas yang lebih tinggi.
3. Klaster demografis yang dihasilkan termasuk klaster mayoritas (petani, pelajar) serta klaster minoritas berpendapatan tinggi menyediakan gambaran sosial ekonomi yang dapat digunakan pemerintah desa dalam perumusan program bantuan, alokasi sumber daya, dan identifikasi kelompok rentan. Namun pemanfaatannya masih terbatas pada data tabular tanpa validasi eksternal.

REFERENCE

- [1] E. Luthfi, A. Wahyu Wijayanto, And P. Statistika, "Analisis Perbandingan Metode Hierarchy, K-Means, Dan K-Medoids Clustering Dalam Pengelompokan Indeks Pembangunan Manusia Indonesia," No. 4, Pp. 761-773, 2021, [Online]. Available: [Http://journal.feb.unmul.ac.id/index.php/Inovasi](http://journal.feb.unmul.ac.id/index.php/Inovasi)
- [2] A. E. Pramitasari And Y. Nataliani, "Perbandingan Clustering Karyawan Berdasarkan Nilai Kinerja Dengan Algoritma K-Means Dan Fuzzy C-Means," *Jatasi (Jurnal Teknik Informatika Dan Sistem Informasi)*, Vol. 8, No. 3, Pp. 1119-1132, 2021, Doi: 10.35957/Jatasi.V8i3.957.
- [3] S. Kurniawan, A. M. Siregar, And H. Y. Novita, "Penerapan Algoritma K-Means Dan Fuzzy C-Means Dalam Mengelompokan Prestasi Siswa Berdasarkan Nilai Akademik," Vol. Iv, No. 1, 2023.
- [4] R. A. Sotar, Putri Nancy, Suardinata, "Analisa Penjualan Menggunakan Algoritma K-Means," *Stmik Indonesia Padang*, Vol. 6, No. 1, P. 404, 2020, [Online]. Available: [Http://ljcs.stmikindonesia.ac.id/index.php/ljcs/article/view/390/131](http://ljcs.stmikindonesia.ac.id/index.php/ljcs/article/view/390/131)
- [5] D. Ramdhan, G. Dwilestari, R. D. Dana, A. Ajiz, And K. Kaslani, "Clustering Data Persediaan Barang Dengan Menggunakan Metode K-Means," *Means (Media Informasi Analisa Dan Sistem)*, Vol. 7, No. 1, Pp. 1-9, 2022, Doi: 10.54367/Means.V7i1.1826.
- [6] D. Anggarwati, O. Nurdiawan, I. Ali, And D. A. Kurnia, "Penerapan Algoritma K-Means Dalam Prediksi Penjualan," *Jurnal Data Science & Informatika (Jdsi)*, Vol. 1, No. 2, Pp. 58-62, 2021.
- [7] W. Prihartono, "Analisis Segmentasi Pelanggan Menggunakan Algoritma K-Means," *Jurnal Informatika Dan Teknik Elektro Terapan*, Vol. 13, No. 1, Pp. 665-673, 2025.
- [8] A. K. Jain, "Data Clustering: 50 Years Beyond K-Means," *Pattern Recognit. Lett.*, Vol. 145, Pp. 4-16, 2021.
- [9] D. Anggarwati, O. Nurdiawan, I. Ali, And D. A. Kurnia, "Penerapan Algoritma K-Means Dalam Prediksi Penjualan Karoseri," Vol. 1, No. 2, Pp. 58-62, 2021.
- [10] Y. Dharma Putra, M. Sudarma, And I. B. A. Swamardika, "Clustering History Data Penjualan Menggunakan Algoritma K-Means," *Majalah Ilmiah Teknologi Elektro*, Vol. 20, No. 2, P. 195, 2021, Doi: 10.24843/Mite.2021.V20i02.P03.

- [11] A. Yudhistira And R. Andika, "Pengelompokan Data Nilai Siswa Menggunakan Metode K-Means Clustering," *Journal Of Artificial Intelligence And Technology Information (Jaiti)*, Vol. 1, No. 1, Pp. 20–28, Feb. 2023, Doi: 10.58602/Jaiti.V1i1.22.
- [12] Ragil Prasojo And Yustina Retno, "Implementasi K-Means Clustering Pada Pengelompokan Potensi Kerjasama Pelanggan," *Jurnal Tikomsin*, 2020.
- [13] E. Suprpto, "Pengelompokan Potensi Padi Di Indonesia Menggunakan K-Means Cluster," *Jurnal Ilmiah Populer*, Vol. 5, 2021.
- [14] M. G. Pradana, A. C. Nurcahyo, P. H. Saputro, U. A. Ata, And U. A. Ata, "Penerapan Metode K-Means Klustering Untuk Menentukan Kepuasan Pelanggan," *Citec Journal*, Vol. 7, No. 1, Pp. 42–50, 2020.
- [15] I. A. Thaher, A. Septiariani, And N. Puspitasari, "Pengelompokan Kualitas Kinerja Pegawai Menggunakan Metode K-Means Grouping The Quality Of Employee Performance For Position Recommendations Using The K-Means Clustering Method," Vol. 11, No. 28, Pp. 131–141, 2022, Doi: 10.34010/Komputika.V11i2.5518.
- [16] M. A. Ghofar And Y. I. Kurniawan, "Aplikasi Pengelompokan Pelanggan Pada Ums Store Menggunakan Algoritma K-Means," 2018.
- [17] V. Purwayoga And B. Susanto, "Pengelompokan Daerah Berdasarkan Ketersediaan Masjid Muhammadiyah Dengan Algoritma K-Means," Vol. 13, No. 1, Pp. 75–80, 2021.
- [18] S. Sonang, A. T. Purba, And F. O. I. Pardede, "Pengelompokan Jumlah Penduduk Berdasarkan Kategori Usia Dengan Metode K-Means," *Jurnal Teknik Informasi Dan Komputer (Tekinkom)*, Vol. 2, No. 2, P. 166, 2019, Doi: 10.37600/Tekinkom.V2i2.115.
- [19] Y. Hartati, S. Defit, And G. W. Nurcahyo, "Klasterisasi Bibit Terbaik Menggunakan Algoritma K-Means Dalam Meningkatkan Penjualan," *Jurnal Informatika Ekonomi Bisnis*, Vol. 3, Pp. 4–10, 2021, Doi: 10.37034/Infek.V3i1.56.
- [20] I. Reisandi, Daryana, F. Sri Mulyati, And M. Fauzi, "Implementasi Clustering K-Means Terhadap Penilaian Kinerja Karyawan Pt Xyz," *Jurnal Sosial Dan Teknologi (Sostech)*, Vol. 1, No. 8, Pp. 757–767, 2021.
- [21] D. Gustian And M. S. Al-Farits, "Data Mining Untuk Melihat Minat Belajar Siswa Menerapkan Metode K-Means," *Journal Of Information System Research (Josh)*, Vol. 4, No. 3, Pp. 775–784, Apr. 2023, Doi: 10.47065/Josh.V4i3.3218.