Progresif: Jurnal Ilmiah Komputer
https://ojs.stmik-banjarbaru.ac.id/index.php/progresif/index
Jl. Ahmad Yani, K.M. 33,5 - Kampus STMIK Banjarbaru
Loktabat - Banjarbaru (Tlp. 0511 4782881), e-mail: puslit.stmikbjb@gmail.com
e-ISSN: 2685-0877

Penerapan *K-Means Clustering* dalam Segmentasi Siswa Berdasarkan Status Sosial Ekonomi

DOI: http://dx.doi.org/10.35889/progresif.v21i2.2809

Dewi Eka Putri[1*], Eka Praja Wiyata Mandala[2]
Teknik Informatika, Universitas Putra Indonesia YPTK, Padang, Indonesia
*e-mail *Corresponding Author:* dewieka@upiyptk.ac.id

Abstract

The accuracy of educational aid distribution remains a challenge, especially when it is not based on structured socioeconomic data. This study aims to group students at SMP Negeri 1 Lunang based on socioeconomic status using the K-Means Clustering algorithm as a segmentation approach. The data used includes parents' income and occupation, number of dependents, social assistance, certificates of poverty, and distance from home to school. After data normalization, clustering and visualization were performed using Principal Component Analysis (PCA). The clustering results yielded three main groups representing different socioeconomic levels: low, medium, and high. Validation using the Silhouette Score yielded a value of 0.2592, indicating that the cluster separation was adequate. These findings suggest that K-Means can serve as a decision-making tool for data-driven aid distribution. This study offers a new approach to student segmentation that simultaneously considers geographical and socioeconomic indicators.
Keywords: *K-Means; Socioeconomic status; Student segmentation; PCA; Silhouette score*

Abstrak

Ketepatan penyaluran bantuan pendidikan masih menjadi tantangan, terutama ketika tidak berbasis pada data sosial ekonomi yang terstruktur. Penelitian ini bertujuan untuk mengelompokkan siswa SMP Negeri 1 Lunang berdasarkan status sosial ekonomi menggunakan algoritma *K-Means Clustering* sebagai pendekatan segmentasi. Data yang digunakan mencakup penghasilan dan pekerjaan orang tua, jumlah tanggungan, bantuan sosial, surat keterangan tidak mampu, dan jarak rumah ke sekolah. Setelah data dinormalisasi, dilakukan klasterisasi dan visualisasi menggunakan *Principal Component Analysis (PCA)*. Hasil *clustering* menghasilkan tiga kelompok utama yang merepresentasikan tingkatan sosial ekonomi berbeda yaitu rendah, menengah dan tinggi. Validasi menggunakan *Silhouette Score* menunjukkan nilai sebesar 0,2592, menandakan bahwa pemisahan klaster cukup baik. Temuan ini menunjukkan bahwa *K-Means* dapat menjadi alat bantu pengambilan keputusan untuk penyaluran bantuan berbasis data. Penelitian ini menawarkan pendekatan baru dalam segmentasi siswa yang mempertimbangkan indikator geografis dan sosial secara bersamaan.
Kata kunci: *K-Means; Status sosial ekonomi; Segmentasi siswa; PCA; Silhouette score*

1. Pendahuluan

Pendidikan menjadi kebutuhan penting yang tidak dapat dikesampingkan bagi masyarakat Indonesia [1] dan menjadi faktor yang sangat penting dalam mengembangkan kualitas sumber daya manusia. Pendidikan berkembang dengan signifikan sehingga pemerintah harus menaikkan kualitas dan kuantitas pendidikan di Indonesia [2]. Namun, terjadi ketimpangan terhadap pendidikan sehingga menjadi masalah yang serius di seluruh penjuru Indonesia. Pemerintah Indonesia telah meluncurkan berbagai program untuk meningkatkan pemerataan pendidikan, seperti Bantuan Operasional Sekolah (BOS) [3], Kartu Indonesia Pintar (KIP), Program Indonesia Pintar (PIP) dan sebagainya [4]. Program-program ini merupakan upaya pemerintah untuk menambah dan mengupayakan pendidikan bagi siswa yang kurang mampu [5]. Bantuan pendidikan memberikan jaminan dan peluang bagi anak yang kurang mampu.

Program Indonesia Pintar (PIP) dan Bantuan Operasi Sekolah (BOS) merupakan dua program unggulan dan prioritas dari pemerintah [6]. Program-program yang telah diluncurkan pemerintah merupakan cara agar pendidikan merata bagi seluruh masyarakat Indonesia, khususnya pada keterbatasan di sektor ekonomi dan geografis [7]. Bantuan pendidikan ini dapat meringankan beban finansial keluarga siswa, dapat menenangkan pikiran agar bisa fokus pada pelajaran dengan tidak memikirkan biaya tambahan [8].

Penelitian ini dilakukan di SMP Negeri 1 Lunang yang terletak di Kecamatan Lunang, Kabupaten Pesisir Selatan, Provinsi Sumatera Barat. Salah satu permasalahan utama dalam penyaluran bantuan pendidikan di SMP Negeri 1 Lunang adalah kurangnya data yang akurat dan terstruktur mengenai kondisi sosial ekonomi siswa. Hal ini menyebabkan ketidaktepatan dalam pemberian bantuan, di mana siswa yang seharusnya menerima bantuan tidak mendapatkannya, sementara siswa yang kurang membutuhkan justru mendapatkannya. Untuk mengatasi permasalahan ini, diperlukan pendekatan yang dapat mengelompokkan siswa berdasarkan kondisi sosial ekonomi mereka secara objektif dan sistematis.

Untuk mengatasi permasalahan diatas, perlu dilakukan penganalisaan terhadap data pokok siswa untuk mengelompokkan siswa yang berhak mendapatkan bantuan pendidikan dari pemerintah. Penganalisaan data pokok siswa dilakukan dengan menggunakan pendekatan *data mining*. *Data mining* diterapkan untuk menguji data dalam jumlah besar untuk mendapatkan pola baru yang lebih bermanfaat [9]. *Data mining* mengidentifikasi korelasi, pola dan tren yang tidak terlihat dalam data yang besar [10]. Salah satu teknik dalam *data mining* adalah teknik *clustering.* Penelitian ini menggunakan teknik *clustering* untuk melakukan segmentasi siswa. *Clustering* dipakai untuk membagi kumpulan data ke dalam beberapa kelompok sesuai dengan kemiripan masing-masing atribut yang dimiliki [11]. Algoritma yang akan dipakai pada penelitian ini adalah *Algoritma K-Means. Algoritma K-Means* merupakan metode dalam teknik clustering yang dipakai untuk mengelompokkan data yang sesuai dengan kesamaan karakteristik, salah satunya untuk analisis data pendidikan karena mampu mengelompokkan data yang kompleks [12].

Penelitian ini membantu pihak SMP Negeri 1 Lunang dalam klasterisasi siswa yang berhak menerima bantuan pendidikan dari pemerintah. Tujuan dari penelitian ini adalah untuk menerapkan metode *K-Means Clustering* dalam segmentasi siswa berdasarkan status sosial ekonomi, sehingga dapat membantu pihak sekolah dalam menyalurkan bantuan pendidikan secara lebih tepat sasaran.

2. Tinjauan Pustaka

Pada penelitian sebelumnya yang dilakukan oleh Eka Indriati pada tahun 2023 membahas pengelompokan status calon penerima beasiswa KIP Kuliah di Universitas Papua, yang selama ini dilakukan secara manual dan kurang optimal akibat banyaknya pendaftar. Objek penelitian adalah data calon penerima KIP Kuliah di Universitas Papua. Metode yang digunakan adalah algoritma K-Means Clustering dengan bantuan bahasa pemrograman Python pada platform Jupyter Notebook, untuk mengelompokkan data berdasarkan kemiripan karakteristik. Hasil penelitian menunjukkan terbentuknya dua cluster, yaitu 687 data yang dikelompokkan sebagai penerima KIP Kuliah dan 547 data yang tidak diterima sebagai penerima [13].

Pada penelitian sebelumnya yang dilakukan oleh Eviana Nahak dkk pada tahun 2024 tentang penentuan siswa yang akan menerima beasiswa Program Indonesia Pintar. Penelitian ini menjelaskan bahwa permasalahan utama pada SMPN Satu Atap Nununamat sebagai objek penelitian yaitu penentuan calon penerima beasiswa sering memakan waktu yang lama dan banyak siswa yang memiliki kesamaan kriteria sehingga pemberian beasiswa sering tidak tepat sasaran. Penelitian ini menggunakan pendekatan *MOORA.* Hasil dari penelitian ini dijelaskan dapat memberikan kemudahan bagi staf tata usaha dalam tahap seleksi penerima beasiswa PIP sehingga tidak salah sasaran dan menghasilkan keputusan dengan tepat dan cepat [14].

Penelitian lainnya yang dilakukan oleh Ida Bagus Adisimakrisna Peling dkk pada tahun 2024 membahas permasalahan dalam penentuan penerima beasiswa yang sering kali tidak objektif dan memerlukan pertimbangan lebih lanjut. Objek penelitian adalah mahasiswa calon penerima beasiswa dengan kriteria seperti IPK, presensi kehadiran, dan UKT. Penelitian ini menggunakan metode *K-Means* untuk melakukan pengelompokan mahasiswa ke dalam tiga cluster berdasarkan karakteristik data, serta metode *SAW (Simple Additive Weighting)* untuk memberikan peringkat prioritas terhadap masing-masing cluster. Hasilnya, *Cluster* 1 menempati peringkat tertinggi dengan anggota yang memiliki IPK dan kehadiran tinggi serta UKT sedang; *Cluster* 2 di peringkat kedua dengan seluruh kriteria bernilai tinggi; dan *Cluster* 3 di peringkat

terakhir dengan karakteristik data yang tidak konsisten. Pendekatan ini membantu institusi pendidikan dalam pengambilan keputusan penerima beasiswa secara lebih tepat dan efisien [15].

Penelitian selanjutnya yang dilakukan Nana Suarna dkk pada tahun 2025 membahas upaya optimalisasi prestasi akademik siswa melalui pengelompokan indeks prestasi menggunakan metode *K-Means Clustering*. Objek penelitian adalah data indeks prestasi siswa dari beberapa semester, yang dianalisis untuk mengidentifikasi kelompok siswa berdasarkan kesamaan nilai akademik. Metode K-Means dipilih karena mampu mengelompokkan data secara efisien ke dalam klaster homogen. Proses dilakukan dengan menetapkan jumlah klaster awal dan mengoptimalkan posisi *centroid* hingga konvergen. Hasil penelitian menunjukkan bahwa siswa dapat dikelompokkan menjadi tiga kategori utama, yaitu berprestasi tinggi, sedang, dan rendah. Pengelompokan ini membantu pihak sekolah dalam menyusun strategi pembinaan dan intervensi akademik yang lebih tepat sasaran [16].

Penelitian lainnya yang dilakukan oleh Eneng Okta Srirahmawati pada tahun 2025 membahas pengelompokan prestasi akademik siswa di SDN Lebakwangi dengan menggunakan algoritma *K-Means*. Objek penelitian adalah 171 data nilai rapor siswa kelas 1 hingga 5 semester genap tahun 2024, yang mencakup berbagai mata pelajaran. Penelitian ini menerapkan metode *data mining* untuk mengidentifikasi pola pengelompokan berdasarkan kesamaan karakteristik nilai akademik. Hasil analisis menunjukkan bahwa pada percobaan dengan nilai k = 2, diperoleh hasil terbaik dengan nilai *Davies-Bouldin Index (DBI)* sebesar 0,738, yang menghasilkan dua cluster: cluster_0 dengan 141 siswa kategori “baik” (rata-rata nilai 76,242) dan cluster_1 dengan 27 siswa kategori “sangat baik” (rata-rata nilai 85,181). Temuan ini dapat membantu guru dalam merancang strategi pembelajaran yang lebih sesuai dengan kebutuhan tiap kelompok siswa [17].

Penelitian lainnya yang dilakukan oleh Amanda Salsabila pada tahun 2025 membahas pengelompokan kemampuan calistung (membaca, menulis, dan berhitung) siswa sekolah dasar sebagai dasar dalam menentukan strategi pembelajaran yang tepat sasaran. Objek penelitian adalah data nilai siswa dari dua Sekolah Dasar di Kota Lubuklinggau. Penelitian menggunakan metode *hybrid* berbasis *machine learning* dengan menggabungkan algoritma *K-Means Clustering* untuk pengelompokan awal dan *K-Nearest Neighbors (KNN)* untuk klasifikasi. Proses analisis meliputi *preprocessing* data, validasi klaster dengan *Silhouette Score*, serta evaluasi model klasifikasi menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Hasilnya, *K-Means* berhasil mengelompokkan siswa ke dalam tiga klaster: Rendah, Menengah, dan Tinggi, sementara model *KNN* dengan nilai k=3 memberikan akurasi klasifikasi hingga 97%, menunjukkan bahwa metode hybrid ini efektif dalam mengklasifikasikan kemampuan calistung siswa [18].

Penelitian yang berkaitan dengan segmentasi siswa berdasarkan status sosial ekonomi sudah banyak dilakukan menggunakan teknik klasterisasi. Beberapa studi sebelumnya menggunakan algoritma *K-Means Clustering* karena mampu mengelompokkan data dengan efisien. Namun, sebagian besar penelitian terdahulu belum mengintegrasikan teknik reduksi dimensi seperti *Principal Component Analysis (PCA)* untuk mengatasi masalah kompleksitas data berdimensi tinggi. Selain itu, penilaian efektivitas performa model pada penelitian sebelumnya masih terbatas pada visualisasi. Dalam penelitian ini, dilakukan integrasi antara *PCA* dan *K-Means Clustering* untuk meningkatkan kualitas hasil segmentasi, serta digunakan *Silhouette Score* untuk mengukur konsistensi dan kekompakan antar klaster yang terbentuk. Pendekatan ini memberikan keunggulan dibandingkan penelitian sebelumnya, karena mampu memberikan hasil segmentasi yang lebih optimal dan terukur secara objektif. Kombinasi metode ini menjadi *novelty* utama dalam penelitian, yang berkontribusi pada peningkatan akurasi pengambilan keputusan segmentasi siswa berdasarkan variabel sosial ekonomi.

3. Metodologi

Agar penelitian lebih terstrukur, maka penelitian ini memerlukan metodologi yang akan menjadi landasan dalam penelitian ini. Penelitian ini melalui beberapa tahapan yang harus dilakukan. Tahapan penelitian dijelaskan dalam beberapa tahap sebagai berikut:

1) Pengumpulan Data

   Dataset pada penelitian ini didapatkan dari SMP Negeri 1 Lunang. Penelitian ini menggunakan data 35 orang siswa yang mengajukan beasiswa. Masing-masing siswa memiliki data yang terdiri dari data penghasilan orang tua, pekerjaan orang tua, status bantuan lainnya, jumlah tanggungan yang masih sekolah, jarak rumah ke sekolah (km) dan surat keterangan tidak mampu.

2) Pra-pemrosesan Data
Tahapan ini terdiri dari pembersihan data yang dilakukan untuk menghapus data duplikat atau kosong, transformasi data yang dilakukan untuk merubah data numerik dan kategorikal ke dalam bentuk yang sesuai untuk pemrosesan algoritma dan normalisasi data agar setiap variabel memiliki skala yang sebanding. Untuk memastikan bahwa seluruh variabel memiliki skala yang sebanding dan tidak mendominasi dalam proses klasterisasi, dilakukan normalisasi menggunakan metode *Min-Max Scaling.*

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

3) Penerapan Algoritma *K-Means*
Algoritma *K-Means* diawali dengan menentukan jumlah klaster dan tentukan titik pusat klaster secara acak. Menghitung jarak antar data dengan titik pusat klaster dengan *Ecludean Distances* untuk memperoleh jarak terdekat.

$$D_{11} = \sqrt{(M_{1a} - C_{1a})^2 + (M_{1b} - C_{1b})^2 + \cdots + \left(M_{1f} - C_{1f}\right)^2} \quad (2)$$

Jarak terdekat dimasukkan ke dalam masing-masing klaster. Hitung titik pusat klaster baru, lakukan perhitungan jarak dan kelompokkan kembali. Jika ada perubahan anggota klaster maka proses diulang kembali dari menentukan titik pusat klaster baru, jika tidak ada perubahan anggota klaster, maka proses selesai.

4) Mereduksi Dimensi Data
Penelitian ini menggunakan banyak dimensi data yaitu 6 dimensi, maka perlu dilakukan pengurangan dimensi dengan melakukan reduksi dimensi data menggunakan *Principal Component Analysis (PCA)* agar hasil pengelompokkan lebih mudah divisualisasikan.

5) Mengukur Validasi Klaster
Mengukur kualitas klaster dengan mempertimbangkan seberapa baik data yang sama berkumpul dalam satu klaster dan seberapa jauh perpisahan antara klaster menggunakan *Silhoutte Score.*

6) Evaluasi dan Interpretasi Hasil
Pada tahap ini dilakukan pengamatan dan evaluasi terhadap distribusi anggota pada tiap klaster. Kemudian dilanjutkan dengan memberikan penerjemahan untuk tiap klaster. Hasilnya, klaster 0 menunjukkan kelompok siswa dengan status sosial ekonomi menengah. Klaster 1 menggambarkan siswa dengan status sosial ekonomi yang lebih tinggi. Klaster 2 merupakan siswa dengan status sosial ekonomi yang lebih rendah.

4. Hasil dan Pembahasan

4.1 Implementasi Algoritma

Penelitian ini dimulai dengan mengumpulkan dataset dari SMP Negeri 1 Lunang berupa data siswa sebanyak 35 orang siswa yang mengajukan untuk mendapat bantuan pendidikan yang dapat dilihat pada Tabel 1.

Table 1. Dataset siswa yang mengajukan bantuan pendidikan

No	Nama	Penghasilan orang tua (Rp.)	Pekerjaan orang tua	Bantuan lainnya	Jumlah tang-gungan	Jarak rumah ke sekolah (km)	Surat keterangan tidak mampu
1	APF	1.500.000	Petani	Ya	2	7	Ada
2	CP	2.000.000	Petani	Ya	2	1	Ada
3	KKI	1.500.000	Pedagang kecil	Ya	2	1	Ada
4	MDF	2.500.000	Buruh	Ya	2	4	Ada
5	NA	2.000.000	Petani	Ya	1	1	Ada
6	ZNF	2.000.000	Wiraswasta	Tidak	2	1	Tidak
7	ZJP	2.000.000	Petani	Tidak	3	2	Tidak
8	WRN	1.500.000	Petani	Tidak	2	6	Tidak
9	RNR	5.000.000	Karyawan swasta	Tidak	2	10	Tidak
10	RF	1.500.000	Karyawan swasta	Tidak	2	2	Tidak
….	….	….	….	….	….	….	….
35	YS	3.000.000	Petani	Tidak	2	1	Tidak

Dataset masing-masing siswa terdiri dari penghasilan orang tua, pekerjaan orang tua, status bantuan lainnya, jumlah tanggungan yang masih sekolah, jarak rumah ke sekolah (km) dan status surat keterangan tidak mampu. Dataset yang digunakan pada penelitian ini tidak ditampilkan nama siswa (nama diinisialkan) karena permintaan pihak sekolah, namun kriteria yang digunakan merupakan data yang benar dari siswa tersebut. Dari Tabel 1 dapat dilihat bahwa data yang digunakan dalam penelitian ini terdiri dari 35 siswa yang berasal dari SMP Negeri 1 Lunang, dengan variabel-variabel yang mencerminkan status sosial ekonomi masing-masing siswa. Variabel yang diamati meliputi penghasilan orang tua, pekerjaan orang tua, kepemilikan bantuan pendidikan lainnya, jumlah tanggungan dalam keluarga, jarak tempat tinggal ke sekolah, serta kepemilikan surat keterangan tidak mampu.

Sebagian besar siswa berasal dari keluarga dengan penghasilan di bawah Rp. 2.500.000 per bulan, dengan pekerjaan orang tua yang didominasi oleh sektor informal seperti petani, buruh, dan pedagang kecil. Pekerjaan sebagai petani mendominasi dengan lebih dari separuh siswa berasal dari keluarga petani. Ada siswa yang tercatat sebagai penerima bantuan lain dari pemerintah, sementara sisanya tidak menerima bantuan tambahan. Mayoritas orang tua siswa memiliki 2–3 tanggungan, yang menunjukkan beban ekonomi keluarga yang cukup tinggi. Sebagian besar siswa tinggal dalam radius 1–3 km, namun terdapat pula beberapa siswa yang tinggal cukup jauh, hingga 10 km dari sekolah. Ada siswa yang memiliki SKTM, ada ada juga yang tidak memilikinya. Ini mengindikasikan bahwa tidak semua siswa dengan kondisi ekonomi rendah secara administratif terdata sebagai penerima bantuan formal. Secara keseluruhan, data ini memberikan gambaran mengenai kondisi sosial ekonomi siswa SMP Negeri 1 Lunang dan menjadi dasar yang relevan untuk penerapan metode *K-Means Clustering* dalam segmentasi penerima bantuan pendidikan secara lebih objektif dan tepat sasaran.

Langkah selanjutnya adalah melakukan pra-pemrosesan data dari dataset diatas. Sebelum data digunakan dalam proses analisis menggunakan algoritma *K-Means Clustering*, dilakukan tahapan pra-pemrosesan untuk memastikan mutu dan kesiapan data. Tahapan pra-pemrosesan yang dilakukan terdiri dari pembersihan data, transformasi data dan normalisasi data.

1) Pembersihan Data
   Pada tahap ini, data diperiksa untuk memastikan tidak terdapat data kosong *(missing values)*, duplikat, atau inkonsistensi penulisan. Dari hasil pengecekan terhadap 35 data siswa, seluruh atribut terisi dengan lengkap dan tidak ditemukan data ganda.
2) Transformasi Data
   Beberapa atribut bersifat kategorikal dan perlu dikonversi ke dalam bentuk numerik agar dapat diproses oleh algoritma *K-Means Clustering*. Transformasi dilakukan sebagai berikut:
   a) Kolom Nama tidak digunakan dalam pemodelan dan dihapus karena tidak berkontribusi terhadap analisis klasterisasi.
   b) Kolom Penghasilan orang tua diubah dari format teks menjadi format angka numerik. Misalnya merubah “Rp. 1.500.000” menjadi 1500000.
   c) Pekerjaan orang tua dikodekan ke dalam angka sesuai kategori. Semua pekerjaan diberikan kode seperti “Petani” menjadi 0.1, “Buruh” menjadi 0.2, “Pedagang kecil” menjadi 0.3, “Pedagang besar” menjadi 0.4, “Karyawan swasta” menjadi 0.5 dan “Wiraswasta” menjadi 0.6.
   d) Kolom Bantuan lainnya dikodekan ke dalam angka. Bantuan dengan atribut “Ya” dirubah menjadi 1 dan bantuan dengan atribut “Tidak” dirubah menjadi 0.
   e) Kolom Surat keterangan tidak mampu juga dikodekan ke dalam angka. Surat dengan atribut “Ada” dirubah menjadi 1 dan surat dengan atribut “Tidak” dirubah menjadi 0.
3) Normalisasi Data
   Metode *Min-Max Scaling* digunakan untuk menormalisasikan variabel numerik berikut:
   a) Penghasilan orang tua
      Penghasilan orang tua memiliki nilai minimum Rp. 1.000.000 dan maksimum Rp. 5.000.000, maka penghasilan Rp. 2.000.000 akan dinormalisasi menjadi:

$$X_{norm} = \frac{2.000.000 - 1.000.000}{5.000.000 - 1.000.000} = 0.250$$

   b) Jumlah tanggungan
      Jumlah tanggungan orang tua memiliki nilai minimum 1 dan maksimum 3, maka jumlah tanggungan 2 akan dinormalisasi menjadi:

$$X_{norm} = \frac{2-1}{3-1} = 0.500$$

c) Jarak rumah ke sekolah
Jarak rumah ke sekolah memiliki nilai minimum 1 km dan maksimum 10 km, maka jarak rumah ke sekolah 7 km akan dinormalisasi menjadi:

$$X_{norm} = \frac{7-1}{10-1} = 0.667$$

Setelah dilakukan pembersihan, transformasi, dan normalisasi, data siap untuk diproses menggunakan algoritma *K-Means Clustering* dalam tahap analisis.

Setelah melalui tahapan pembersihan, transformasi, dan normalisasi data, maka diperoleh data yang telah siap untuk dianalisis menggunakan metode *K-Means Clustering*. Tabel 2 menyajikan hasil akhir dari proses pra-pemrosesan, di mana seluruh atribut kategorikal telah dikonversi menjadi bentuk numerik, dan atribut numerik telah dinormalisasi ke dalam rentang nilai 0 hingga 1.

Table 2. Dataset setelah melalui pra-pemrosesan

No	Nama Siswa (Inisial)	Penghasilan orang tua	Pekerjaan orang tua	Bantuan lainnya	Jumlah tang-gungan	Jarak rumah ke sekolah (km)	Surat keterangan tidak mampu
1	APF	0.125	0.1	1	0.500	0.667	1
2	CP	0.250	0.1	1	0.500	0.000	1
3	KKI	0.125	0.3	1	0.500	0.000	1
4	MDF	0.375	0.2	1	0.500	1.500	1
5	NA	0.250	0.1	1	0.000	0.000	1
6	ZNF	0.250	0.6	0	0.500	0.000	0
7	ZJP	0.250	0.1	0	1.000	0.500	0
8	WRN	0.125	0.1	0	0.500	2.500	0
9	RNR	1.000	0.5	0	0.500	4.500	0
10	RF	0.125	0.5	0	0.500	0.500	0
….	….	….	….	….	….	….	….
35	YS	0.500	0.1	0	0.500	0.000	0

Tabel 2 juga menunjukkan bahwa semua atribut telah disatukan dalam satu format numerik yang seragam dan tidak memiliki nilai kosong maupun duplikat. Dengan data yang telah terstandarisasi dan terstruktur dengan baik ini, proses klasterisasi dapat berjalan lebih optimal dan menghasilkan segmentasi yang akurat sesuai dengan karakteristik sosial ekonomi masing-masing siswa.

Pada tahap selanjutnya dilakukan proses klasterisasi data siswa berdasarkan status sosial ekonomi menggunakan algoritma *K-Means Clustering*. Algoritma ini merupakan metode *unsupervised learning* yang efektif dalam pengelompokan data berdasarkan kedekatan karakteristik. Dalam penelitian ini, nilai jumlah klaster (k) ditentukan sebanyak 3 klaster, dengan asumsi pembagian segmentasi siswa ke dalam tiga kelompok status sosial ekonomi, yaitu rendah, menengah, dan tinggi.

Langkah pertama adalah menentukan tiga titik *centroid* awal secara acak dari data yang telah dinormalisasi. Titik *centroid* awal ini menjadi pusat sementara untuk masing-masing klaster, dan selanjutnya akan diperbarui berdasarkan rata-rata anggota klaster yang terbentuk. Berikut tiga centroid yang menjadi centroid awal:

C_0 = {0.125; 0.1; 1; 0.500; 0.667; 1}
C_1 = {0.000; 0.2; 0; 0.500; 0.000; 1}
C_2 = {0.875; 0.5; 0; 0.500; 0.500; 0}

Seluruh data siswa dihitung jaraknya ke masing-masing centroid menggunakan metode *Euclidean Distance*. Hasil perhitungan jarak antara data dengan *centroid* dapat dilihat pada Tabel 3.

Table 3. Jarak data siswa ke masiing-masing centroid

No	Nama	Jarak ke C_{0lama}	Jarak ke C_{1lama}	Jarak ke C_{2lama}
1	APF	**0.000**	1.457	2.340
2	CP	**0.569**	1.036	2.495
3	KKI	**0.644**	1.013	2.516
4	MDF	**0.964**	3.318	3.158
5	NA	**0.819**	1.286	2.745
6	ZNF	2.569	1.472	**0.883**
7	ZJP	2.286	1.769	**0.992**
8	WRN	5.361	7.410	**4.850**
9	RNR	17.082	22.294	**16.125**
10	RF	2.105	1.575	**0.750**
….	….	….	….	….
35	YS	2.512	1.510	**0.798**

Dari Tabel 3 dilakukan pengelompokkan dengan membandingkan jarak masing-masing data dengan *centroid.* Jarak terpendek akan masuk ke dalam masing-masing klaster. Misalnya Siswa 1 memiliki jarak terpendek ke klaster 0 (C_0) maka Siswa 1 masuk ke dalam klaster 0, dan begitu seterusnya. Jika semua sudah dikelompokkan, maka akan dihitung *centroid* baru dengan menghitung rata-rata dari setiap anggota masing-masing klaster. Dari hasil perhitungan maka diperoleh *centroid* baru sebagai berikut:

C_{0baru} = {0.236; 0.133; 1.000; 0.444; 0.630; 0.889}
C_{1baru} = {0.125; 0.200; 0.000; 0.214; 0.071; 0.857}
C_{2baru} = {0.382; 0.274; 0.000; 0.579; 1.447; 0.000}

Langkah selanjutnya adalah menghitung kembali jarak masing-masing data siswa dengan *centroid* baru yang sudah diperoleh. Perhitungan jarak ini dilakukan untuk mengelompokkan anggota ke masing-masing klaster. Hasil perhitungan jarak dengan *centrod* baru dapat dilihat pada Tabel 4.

Table 4. Jarak data siswa ke masing-masing centroid baru

No	Nama	Jarak ke C_{0baru}	Jarak ke C_{1baru}	Jarak ke C_{2baru}
1	APF	**0.133**	1.461	2.663
2	CP	**0.448**	1.120	4.125
3	KKI	**0.612**	1.112	4.134
4	MDF	**0.927**	3.174	2.012
5	NA	**0.642**	1.084	4.454
6	ZNF	2.293	**1.241**	2.453
7	ZJP	2.116	1.696	**1.293**
8	WRN	5.298	6.814	**1.424**
9	RNR	17.084	21.354	**9.983**
10	RF	1.881	1.300	**1.246**
….	….	….	….	….
35	YS	2.224	**1.210**	2.311

Dari Tabel 4 dapat dilihat bahwa ada perubahan anggota klaster yang terjadi pada Siswa 6. Siswa 6 sebelumnya berada pada klaster 2 yang tampak pada Tabel 3 berpindah ke klaster 1 yang dapat dilihat pada Tabel 4. Karena terjadi perubahan anggota, maka perhitungan berlanjut ke iterasi berikutnya. Perhitungan terus dilakukan, sampai perhitungan berhenti pada iterasi ke-4 karena tidak terjadi lagi perubahan anggota klaster. Hasil perhitungan jarak data siswa dengan *centroid* baru pada iterasi ke-4 dapat dilihat pada Tabel 5.

Table 5. Jarak data siswa ke masing-masing centroid baru pada iterasi ke-4

No	Nama	Jarak ke C_{0baru}	Jarak ke C_{1baru}	Jarak ke C_{2baru}
1	APF	**0.291**	1.577	7.734
2	CP	**0.115**	1.621	11.328
3	KKI	**0.282**	1.615	11.360
4	MDF	**1.655**	2.769	4.348
5	NA	**0.302**	1.792	11.703
6	ZNF	2.179	**0.602**	9.851
7	ZJP	2.370	**0.600**	7.055
8	WRN	6.982	4.815	**0.787**
9	RNR	20.191	18.006	**2.622**
10	RF	2.127	**0.395**	7.069
….	….	….	….	….
35	YS	2.109	**0.564**	9.602

Setelah proses iterasi selesai yang hasilnya ditampilkan pada Tabel 5, maka diperoleh tiga klaster akhir sebagai berikut:

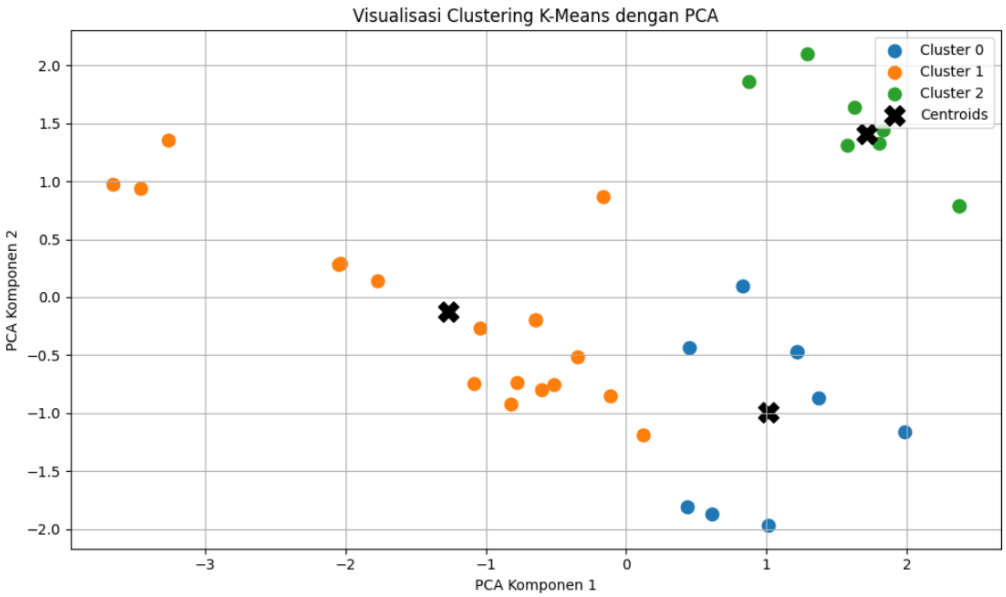
a) Klaster 0 diberi nama “Status Sosial Ekonomi Menengah” yaitu siswa dengan penghasilan orang tua menengah, pekerjaan lebih variatif (petani, wiraswasta, karyawan), dan jarak rumah ke sekolah sedang.
b) Klaster 1 diberi nama “Status Sosial Ekonomi Rendah” yaitu mayoritas siswa dengan penghasilan orang tua rendah, pekerjaan informal (petani/buruh), tidak memiliki bantuan lain, dan memiliki surat keterangan tidak mampu.
c) Klaster 2 diberi nama “Status Sosial Ekonomi Tinggi” yaitu siswa dengan penghasilan orang tua tinggi, orang tua bekerja sebagai karyawan swasta atau pedagang besar, tanpa bantuan pemerintah dan tidak memiliki surat keterangan tidak mampu.

Distribusi klaster ini dapat digunakan pihak sekolah dalam menentukan siswa yang paling layak menerima bantuan pendidikan berdasarkan karakteristik sosial ekonomi dari masing-masing kelompok. Model ini membantu dalam pengambilan keputusan yang lebih objektif dan tepat sasaran. Hasil pengelompokkan siswa ke dalam masing-masing klaster dapat dilihat pada Tabel 6.

Table 6. Anggota dari masing-masing klaster status sosial ekonomi

Klaster 0 Menengah	Klaster 1 Rendah	Klaster 2 Tinggi
APF, CP, KKI, MDF, NA, MS, RMA, ALN	ZNF, ZJP, RF, BDA, HDP, MAA, UIK, AN, AS, AAT, FSD, HA, MDY, MSO, SWR, ZA, ZR, VNP, YS	WRN, RNR, RN, NPS, RNP, SER, VES, ZAK

Untuk memahami pola pengelompokan siswa berdasarkan status sosial ekonominya, dilakukan proses clustering menggunakan algoritma *K-Means* dengan jumlah klaster sebanyak tiga (k = 3). Seluruh data yang digunakan sebelumnya telah dinormalisasi agar berada dalam skala yang seragam. Agar hasil pengelompokan lebih mudah divisualisasikan secara dua dimensi, diterapkan teknik *Principal Component Analysis (PCA)* untuk mereduksi dimensi data dari enam variabel menjadi dua komponen utama, yaitu PCA Komponen 1 (PCA1) dan PCA Komponen 2 (PCA2). Kedua komponen ini mewakili variansi terbesar dari seluruh data dan mampu mempertahankan sebagian besar informasi penting. Hasil scatter plot dua dimensi dapat dilihat pada Gambar 1.



Gambar 1. Visualisasi *clustering* K-Means dengan PCA

Gambar 1 menunjukkan hasil visualisasi dari proses pengelompokan menggunakan algoritma *K-Means Clustering* yang direduksi ke dalam dua dimensi melalui teknik *Principal Component Analysis (PCA)*. Proses ini dilakukan untuk menyederhanakan visualisasi dari enam variabel utama menjadi dua komponen utama, yakni PCA Komponen 1 dan PCA Komponen 2, yang secara kolektif mewakili variansi terbesar dalam dataset.

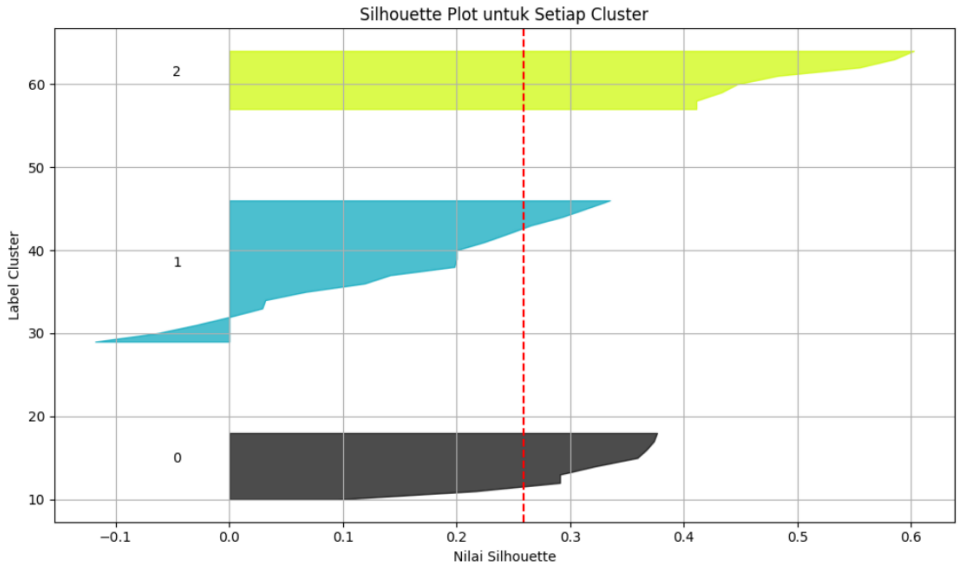
Setiap titik pada scatter plot merepresentasikan satu siswa, yang diwarnai berdasarkan hasil klasterisasi:

a) *Cluster* 0 berwarna biru yang merepresentasikan kelompok siswa dengan status sosial ekonomi menengah. Distribusi titik-titik dalam klaster ini tersebar di wilayah kanan bawah hingga tengah atas plot.
b) *Cluster* 1 berwarna hijau yang menggambarkan siswa dengan indikasi status sosial ekonomi yang lebih tinggi, ditandai oleh posisi yang cenderung mengarah ke kuadran kanan atas dari plot.
c) *Cluster* 2 berwarna oranye yang merupakan kelompok siswa dengan status sosial ekonomi yang lebih rendah, dengan konsentrasi titik yang terkonsentrasi di sisi kiri grafik.
d) Tanda X menunjukkan posisi centroid dari masing-masing klaster, yang menjadi pusat dari kelompok berdasarkan karakteristik data. Jarak antar centroid yang cukup signifikan memperkuat validitas pemisahan antar klaster.

4.2 Pengukuran Validasi Efektifitas Kinerja Algoritma

Validasi efektivitas algoritma *K-Means Clustering* dalam proses pengelompokan dilakukan menggunakan metrik *Silhouette Score*, yaitu salah satu ukuran evaluasi yang umum dipakai untuk menilai kualitas klaster. *Silhouette score* menggambarkan seberapa dekat data dalam suatu klaster terhadap data lain dalam klaster yang sama dibandingkan dengan klaster lain. Rentang *silhouette score* berada antara -1 hingga 1, di mana nilai mendekati 1 menandakan bahwa objek berada dalam klaster yang sangat sesuai, sedangkan nilai mendekati 0 mengindikasikan bahwa objek berada di batas antara dua klaster. Nilai negatif menandakan kemungkinan bahwa objek telah salah diklasifikasikan ke dalam klaster yang tidak tepat.

Berdasarkan hasil pengujian yang dilakukan terhadap data yang telah dinormalisasi dan dikelompokkan menggunakan algoritma K-Means dengan jumlah klaster sebanyak tiga (k=3), diperoleh *silhouette score* rata-rata sebesar 0.2592. Nilai ini menunjukkan bahwa struktur klaster yang terbentuk berada dalam kategori cukup baik, di mana mayoritas data berada pada posisi yang tepat dalam klasternya masing-masing, meskipun terdapat beberapa data yang berada di dekat batas antar klaster. Hasil pengukuran dapat divisualisasikan melalui *silhouette plot* yang dapat dilihat pada Gambar 2.



Gambar 2. Visualisasi *Silhouette Score* dengan *Silhouette Plot*

Gambar 2 memberikan gambaran visual sebaran *silhouette score* untuk masing-masing klaster. *Plot* ini menunjukkan bahwa ukuran dan kepadatan masing-masing klaster relatif bervariasi, namun secara umum masih mempertahankan kohesi internal yang memadai. Garis vertikal merah pada grafik menunjukkan *silhouette score* rata-rata keseluruhan. Dengan demikian, berdasarkan nilai *Silhouette Score* dan hasil visualisasinya, dapat disimpulkan bahwa algoritma *K-Means* berhasil melakukan pengelompokan data dengan struktur yang cukup jelas dan terpisah antar klaster, sehingga dapat diterima sebagai metode segmentasi yang valid untuk membantu SMP Negeri 1 Lunang dalam mengelompokkan siswa yang berhak menerima bantuan pendidikan secara lebih tepat sasaran.

4.3 Pembahasan

Penelitian ini berhasil menerapkan algoritma *K-Means clustering* pada data siswa SMP Negeri 1 Lunang berdasarkan variabel sosial ekonomi seperti penghasilan orang tua, pekerjaan, jumlah tanggungan, jarak ke sekolah, bantuan lain, dan surat keterangan tidak mampu. Hasil segmentasi mengelompokkan siswa ke dalam tiga klaster yang representatif yaitu rendah, menengah, dan tinggi untuk status sosial ekonomi. Visualisasi dalam bentuk *scatter plot* berbasis *PCA* dan pembuktian kinerja menggunakan *Silhouette Score* memperlihatkan bahwa klaster terbentuk dengan cukup baik dan dapat menjadi dasar valid untuk penyaluran bantuan pendidikan secara tepat sasaran.

Penelitian sebelumnya meneliti segmentasi siswa SMK berdasarkan prestasi akademik dan kehadiran menggunakan *K-Means* [19], sementara pada penelitian ini meneliti segmentasi siswa SMP yang fokus pada bantuan pendidikan menggunakan *K-Means* karena belum banyak penelitian khusus pada tingkat SMP untuk distribusi bantuan berdasarkan analisis sosial ekonomi. Penelitian sebelumnya memanfaatkan kombinasi variabel sosial ekonomi dan psikologi (motivasi, stres, akses internet) untuk membentuk 3 klaster [20], sementara pada penelitian ini menggunakan kombinasi variabel dari sisi sosial ekonomi dan akses pendidikan dengan membentuk 3 klaster.

Penelitian sebelumnya menggunakan *K-Means* untuk mengukur tingkat keterlibatan siswa dalam pembelajaran online yang divalidasi dengan *Silhouette Score* [21], sementara pada penelitian ini menggunakan kombinasi visualisasi *PCA* dan metrik *Silhouette Score* memberikan gambaran yang lebih solid terkait efektivitas klasterisasi sehingga langsung dapat dimanfaatkan oleh pihak sekolah untuk merancang strategi distribusi bantuan pendidikan yang lebih objektif.

5. Simpulan

Penelitian ini berhasil menerapkan algoritma *K-Means clustering* untuk mengelompokkan siswa berdasarkan indikator sosial ekonomi, seperti penghasilan dan pekerjaan orang tua, jumlah tanggungan, bantuan sosial, surat keterangan tidak mampu, dan jarak rumah ke sekolah. Proses normalisasi data dan pemrosesan lanjutan dengan *PCA* menghasilkan visualisasi klaster yang

jelas, yang kemudian divalidasi menggunakan nilai *Silhouette Score* sebesar 0,2592, menandakan bahwa pemisahan antar klaster berada dalam kategori cukup baik. Hasil ini menunjukkan bahwa algoritma *K-Means* efektif dalam mengidentifikasi tiga kelompok utama siswa: status sosial ekonomi rendah, menengah, dan tinggi. Klasterisasi ini dapat digunakan sebagai dasar dalam pengambilan keputusan yang lebih objektif untuk penyaluran bantuan pendidikan secara tepat sasaran di lingkungan sekolah.

Secara spesifik, kontribusi utama penelitian ini terletak pada penerapan *unsupervised learning* untuk membantu kebijakan distribusi bantuan sosial di jenjang pendidikan menengah, yaitu SMP. Penelitian ini juga memperluas pendekatan segmentasi siswa dengan mempertimbangkan faktor geografis (jarak ke sekolah), yang masih jarang digunakan dalam studi sejenis. Dengan demikian, hasil penelitian tidak hanya memperkuat efektivitas metode *K-Means* dalam konteks pendidikan, tetapi juga memberikan pendekatan praktis berbasis data yang relevan bagi pengelola sekolah khususnya SMP Negeri 1 Lunang.

Namun demikian, penelitian ini masih memiliki keterbatasan, terutama pada jumlah sampel yang relatif kecil dan belum adanya validasi lapangan atas keakuratan klaster yang terbentuk. Selain itu, penelitian ini belum mengkaji dimensi temporal, seperti perubahan kondisi sosial ekonomi siswa dari waktu ke waktu. Oleh karena itu, penelitian selanjutnya direkomendasikan untuk melibatkan dataset yang lebih besar, memperluas cakupan sekolah atau wilayah, serta mengintegrasikan pendekatan lokasi untuk mengamati dinamika status sosial ekonomi siswa secara berkala. Penggunaan metode clustering lain seperti *DBSCAN* atau *Agglomerative Clustering* juga dapat menjadi bahan pembanding untuk menguji hasil segmentasi.

Daftar Referensi

[1] M. F. Firdani, G. A. Wibowo, dan B. O. Lubis, “Sistem Pendukung Keputusan Penerimaan Bantuan dari Pemerintah untuk Siswa Tidak Mampu dengan Metode Simple Additive Weighting (SAW) pada SMP Permata Depok,” *METHODIKA*, vol. 11, no. 1, pp. 20–29, 2025, doi: doi.org/10.46880/mtk.v11i1.3549.

[2] A. Abrianto, “Application of K-Means Clustering Algorithm for Determining PIP Scholarship Recipients at SMPN 9 Blitar,” *JOSAR*, vol. 9, no. 1, pp. 204–214, 2024, doi: doi.org/10.35457/josar.v9i1.3105.

[3] G. G. Dawous, S. S. Oktaviany, dan M. R. Ashari, “Dana Bos Dan Pemerataan Layanan Pendidikan Dasar Di Daerah Timur Indonesia,” *J. Al Burhan Staidaf*, vol. 2, no. 2, pp. 32–41, 2022, [Daring]. Tersedia pada: https://jurnal.staidaf.ac.id/index.php/jab/article/view/79.

[4] S. H. H. Lubis, W. Pangaribuan, S. T. Ahmad, dan S. Arif, “Kebijakan Pemerataan dan Perluasan Akses Pendidikan dan Dampaknya Terhadap Sekolah Swasta,” *Syntax Lit. J. Ilm. Indones.*, vol. 6, no. 7, pp. 6172–6182, 2022, [Daring]. Tersedia pada: https://jurnal.syntaxliterate.co.id/index.php/syntax-literate/article/view/7135.

[5] M. A. M. P dan M. Qibtiyah, “Pemanfaatan Algoritma K-Means Clustering dalam Penentuan Prioritas Penerima Program Bantuan Sosial Pendidikan,” *J. Inform. dan Komput. J. Inform. dan Komput.*, vol. 15, no. 2, pp. 68–76, 2024, [Daring]. Tersedia pada: https://journal.unmaha.ac.id/index.php/jik/article/view/386.

[6] M. Ivan, “Evaluasi Kebijakan Bantuan Pendidikan (Program Indonesia Pintar/Bantuan Operasional Sekolah) dalam Mengatasi Anak Tidak Sekolah (ATS) dan Peningkatan Angka Partisipasi Kasar/Angka Partisipasi Murni (APK/APM) di Indonesia,” *J. Transform. Adm.*, vol. 14, no. 01, pp. 80–92, 2024, doi: 10.59098/talim.v2i02.1255.

[7] H. Sahila, S. Maesaroh, dan Baharuddin, “Efektivitas program bantuan operasional sekolah dalam meningkatkan akses pendidikan,” *Idarah Tarb. J. Manag. Islam. Educ.*, vol. 5, no. 3, pp. 307–315, 2024, doi: 10.32832/itjmie.v5i3.16633.

[8] I. Batuk, A. M. Baitanu, A. R. L. Atu, dan Y. Benu, “Evaluasi Pemberian Bantuan Pendidikan bagi Siswa Siswi Kurang Mampu,” *J. Stud. Multidisipliner*, vol. 8, no. 6, pp. 159–162, 2024, [Daring]. Tersedia pada: https://oaj.jurnalhst.com/index.php/jsm/article/view/3279.

[9] A. Fauzan, W. Witanti, dan F. R. Umbara, “Prediksi Bantuan Operasional Raudhatul Athfal di Tingkat Kabupaten Menggunakan Metode Support Vector Machine – Regression,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 9, no. 1, pp. 773–779, 2025, doi: doi.org/10.36040/jati.v9i1.12505.

[10] T. V. Sandiva, S. Defit, dan G. W. Nurcahyo, “Implementasi Algoritma C4.5 untuk Prediksi Penerima Beasiswa Program Indonesia Pintar,” *J. KomtekInfo*, vol. 11, no. 4, pp. 354–362,

2024, doi: 10.35134/komtekinfo.v11i4.582.
[11] W. Agustin dan A. Bahtiar, “Implementasi Algoritma Naïve Bayes Terhadap Penerima Kartu Indonesia Pintar,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 2, pp. 1521–1528, 2024, doi: 10.36040/jati.v8i2.8981.
[12] B. A. Prayudha, R. Kurniawan, Y. Wijaya, dan U. Hayati, “Algoritma K-Means untuk Meningkatkan Model Klasterisasi Data Siswa SMK Samudra Nusantara Kabupaten Cirebon Berdasarkan Nilai Akademik,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 9, no. 1, pp. 1314–1321, 2025, doi: doi.org/10.36040/jati.v9i1.12689.
[13] E. Indriati, N. Suharyani Azisa, E. Ivo Sihombing, dan Z. Sukma Dewi Mokodompit, “Implementasi Algoritma K-Means Clustering Untuk Pengelompokan Status Penerima Kip Kuliah Mahasiswa Universitas Papua,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 6, pp. 3458–3463, 2023, doi: 10.36040/jati.v7i6.8222.
[14] E. Nahak, F. Tedy, Y. C. H. Siki, E. Ngaga, E. Jando, dan S. D. B. Mau, “Implementasi Metode MOORA dalam Sistem Pendukung Keputusan bagi Calon Penerima Beasiswa Program Indonesia Pintar di SMPN Satu Atap Nununamat,” *KONSTELASI Konvergensi Teknol. dan Sist. Inf.*, vol. 4, no. 1, pp. 83–98, 2024, doi: 10.24002/konstelasi.v4i1.8972.
[15] I. B. A. Peling, M. P. A. Ariawan, dan G. B. Subiksa, “Analisis Cluster Mahasiswa Penerima Beasiswa dengan Metode K-means dan SAW,” *Sist. J. Sist. Inf.*, vol. 13, no. 4, pp. 1334–1343, 2024, doi: doi.org/10.32520/stmsi.v13i4.2914.
[16] N. Suarna, N. Rahaningsih, dan A. A. Suarna, “Optimalisasi Prestasi Akademik Siswa Melalui Pengelompokan Indeks Prestasi dengan K-Means Clustering,” *J. Kecerdasan Buatan dan Teknol. Inf.*, vol. 4, no. 2, pp. 198–207, 2025, doi: doi.org/10.69916/jkbti.v4i2.321.
[17] E. O. Srirahmawati, A. I. Purnamasari, A. Bahtiar, dan E. Tohidi, “Pengelompokan Prestasi Akademik Siswa SD Menggunakan Algoritma K-Means,” *JIRE (Jurnal Inform. Rekayasa Elektron.*, vol. 8, no. 1, pp. 80–86, 2025, doi: doi.org/10.36595/jire.v8i1.1358.
[18] A. Salsabila, A. A. T. Susilo, dan N. K. Daulay, “Metode Hybrid Dalam Pengelompokkan Kemampuan Calistung Siswa Berbasis Machine Learning,” *J. Informatics Manag. Inf. Technol.*, vol. 5, no. 2, pp. 112–119, 2025, doi: 10.47065/jimat.v5i2.500.
[19] S. A. Mukhsyi, A. I. Purnamaari, A. Bahtiar, dan Kaslani, “Improving Student Achievement Clustering Model Using K-Means Algorithm in Pasundan Majalaya Vocational School,” *J. Artif. Intell. Eng. Appl.*, vol. 4, no. 2, hal. 977–985, 2025, doi: doi.org/10.59934/jaiea.v4i2.793.
[20] M. Iqbal, S. P. Sipayung, A. R. Sinaga, dan P. M. Hasugian, “Analysis of Student Achievement with K-Means on Socioeconomic , Behavioral , and Psychological Factors,” *J. Info Sains Inform. dan Sains*, vol. 14, no. 04, pp. 715–728, 2024, doi: 10.54209/infosains.v14i04.
[21] S. Kim, S. Cho, J. Y. Kim, dan D. J. Kim, “Statistical Assessment on Student Engagement in Asynchronous Online Learning Using the k-Means Clustering Algorithm,” *Sustain.*, vol. 15, no. 3, pp. 1–14, 2023, doi: 10.3390/su15032049.