

# Algoritma Clustering untuk Menentukan Kategori Siswa Berdasarkan Prestasi pada SMA Pangeran Antasari

Ayuni Syahputri\*, Lili Tanti

Fakultas Teknik dan Ilmu Komputer, Sistem Informasi, Universitas Potensi Utama, Medan, Indonesia

Email: <sup>1,\*</sup>Ayunisyahputri0222@gmail.com, <sup>2</sup>lili@potensi-utama.ac.id

Email Penulis Korespondensi: Ayunisyahputri0222@gmail.com\*

Submitted: 22/10/2025; Accepted: 07/12/2025; Published: 31/12/2025

**Abstrak**– Penentuan kategori prestasi siswa di SMA Pangeran Antasari masih dilakukan secara manual sehingga kurang efisien dan rentan menghasilkan penilaian yang subjektif. Penelitian ini menerapkan algoritma K-Means Clustering untuk mengelompokkan siswa ke dalam tiga kategori prestasi, yaitu tinggi, cukup, dan rendah. Dataset yang digunakan berjumlah 1.091 data siswa, mencakup nilai akademik, absensi, serta prestasi non-akademik. Proses evaluasi model dilakukan menggunakan metode Elbow untuk menentukan jumlah cluster optimal, dan pengujian menggunakan confusion matrix terhadap 30 data uji. Hasil penelitian menunjukkan bahwa algoritma K-Means mampu melakukan pengelompokan dengan baik, dengan tingkat akurasi sebesar 93,33%, serta rata-rata F1-score mencapai 94,2%. Temuan ini memperlihatkan bahwa K-Means efektif dalam memberikan pemetaan prestasi siswa secara objektif dan dapat membantu sekolah dalam proses evaluasi dan pengambilan keputusan akademik.

**Kata Kunci:** K-Means; Clustering; Prestasi Siswa; Data Mining; Evaluasi Elbow.

**Abstract**– The assessment of student performance at SMA Pangeran Antasari is still carried out manually, which makes the process inefficient and prone to subjective judgment. This study applies the K-Means Clustering algorithm to categorize students into three performance groups: high, medium, and low. The dataset consists of 1,091 student records, including academic scores, attendance, and non-academic achievements. The Elbow Method is used to determine the optimal number of clusters, while evaluation using a confusion matrix on 30 test data demonstrates that the model achieves an accuracy of 93.33% and a macro average F1-score of 94.2%. The results indicate that K-Means is effective in objectively mapping student performance and can support schools in evaluation and data-driven academic decision-making.

**Keywords:** K-Means; Clustering; Student Performance; Data Mining; Elbow Method.

## 1. PENDAHULUAN

Pendidikan dapat memanfaatkan kemajuan teknologi untuk mengembangkan sumber daya manusia yang mampu bertahan di era perkembangan informasi dan teknologi [1]. Sekolah adalah lembaga pendidikan tinggi yang memfasilitasi pengajaran dan pembelajaran untuk memaksimalkan potensi setiap siswa. dalam mendidik orang-orang yang akan berperan penting bagi kemajuan dan kesejahteraan negara dan negara [2]. Menghargai siswa yang berprestasi dengan penghargaan merupakan salah satu cara untuk memacu semangat dan motivasi siswa lainnya. Tingginya prestasi murid dan minimnya angka murid yang tidak berprestasi mencerminkan keunggulan sektor pendidikan. Sektor Pendidikan kini diharapkan mampu bersaing dengan memanfaatkan kemajuan SI/TI, yang dapat menunjang peningkatan daya saing dan menunjang operasional sehari-hari serta pengambilan keputusan strategis [3]. Secara umum, keberhasilan murid dievaluasi berdasarkan evaluasi pelajaran teoritis dan praktis, serta kehadiran dan ketidakhadiran murid selama di dalam kelas [4].

Prestasi siswa dapat dilihat dari berbagai segi atau aspek yang mencerminkan keberhasilan mereka dalam menjalani pendidikan. Beberapa aspek yang umumnya digunakan untuk menilai prestasi siswa antara lain prestasi akademik, prestasi non akademik dan prestasi dalam pengembangan diri [5]. Penentuan prestasi siswa dinilai dari segi Penilaian dibagi menjadi tiga kategori yaitu pengetahuan, bakat, dan sikap. Pengajar menilai semua murid yang mengikuti pelajaran yang diberikan guna mengevaluasi dan menganalisis prestasi belajar murid.

Prestasi siswa merupakan salah satu indikator penting dalam menilai keberhasilan proses pendidikan di sekolah. SMA Pangeran Antasari memiliki kebutuhan untuk melakukan pemetaan prestasi siswa secara lebih objektif agar proses pembinaan, pemberian penghargaan, dan intervensi belajar dapat dilakukan secara tepat sasaran. Selama ini, penentuan prestasi hanya mengacu pada nilai akademik, sementara prestasi non-akademik seperti kegiatan ekstrakurikuler, kepemimpinan, maupun perilaku belum menjadi bagian dari proses evaluasi. Kondisi ini menyebabkan hasil penilaian sering kali tidak menggambarkan kemampuan siswa secara menyeluruh dan berpotensi menimbulkan subjektivitas dalam pengambilan keputusan. Data mining merupakan teknik untuk mengungkap tautan, pattern, dan tren baru dengan memfilter kumpulan banyak data yang disimpan di penyimpanan dan menerapkan teknik pengenalan pattern seperti prosedur statistik dan matematik. K-Means adalah algoritma data mining yang dapat digunakan untuk mengelompokkan/mengklasterkan data yang sangat besar atau bertumpuk yang dapat ditangani dengan salah satu dari beberapa cara, termasuk *Clustering* [6]

Fokus pada penelitian ini adalah pada pengelompokan murid berdasarkan nilai mata pelajaran dan absensi menjadi kategori tinggi, cukup, dan rendah menggunakan metode K-Means *Clustering*. Dengan perkembangan teknologi ini dapat memberikan solusi yang efektif untuk mengatasi permasalahan tersebut. Salah satu metode yang dapat diterapkan adalah algoritma *K-Means Clustering*, yang mampu mengelompokkan data berdasarkan

kemiripan tertentu. Dengan algoritma ini, siswa dapat dikelompokkan ke dalam kategori seperti Berprestasi Tinggi, Sedang, dan Rendah, berdasarkan data seperti nilai akademik, kehadiran, dan partisipasi ekstrakurikuler.

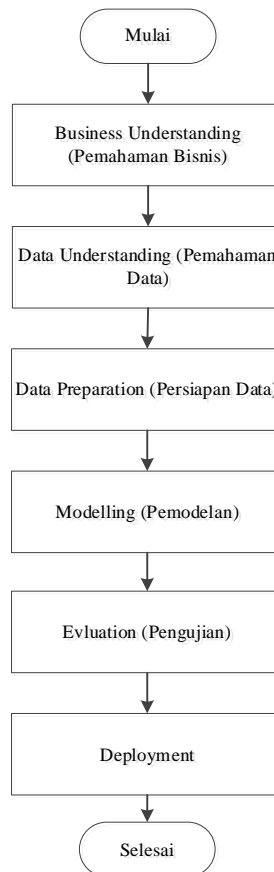
Metode algoritma K-Means *Clustering* merupakan metode klasifikasi yang cukup bagus karena tingkat akurasi yang sangat baik dibandingkan dengan metode lain. Algoritma K-Means merupakan algoritma *Clustering* yang masuk ke dalam kelompok [7]. Berdasarkan penelitian dari [8] metode klasterisasi (*clustering*) yang merupakan pengelompokan item data. Banyak metode klasterisasi yang diusulkan oleh para ahli, salah satunya adalah k-means method atau metode k-means yang termasuk dalam metode berbasis partisi, dimana metode ini membagi data kedalam sejumlah kelompok. Berdasarkan penelitian dari [9] kasus yang di bahas adalah mengelompokkan prestasi siswa, Variabel Independen yaitu nilai mata pelajaran, sikap serta keterampilan dan Variabel Dependen Prestasi (YA, Tidak), data set 75 dengan algoritma K-Means Teknik evaluasi yaitu *Elbow* Berdasarkan penelitian dari [10] Sumberdata didapatkan dari hasil kuesioner yang diberikan kepada siswa kelas 2 semester genap, dengan sampel data 105 Siswa. Atribut yang digunakan sebanyak 4, yaitu Keberadaan guru, Kemampuan Siswa, Motivasi dan Lingkungan Sekolah (C4). Diperoleh hasil penelitian bahwa nilai *Accuracy* sebesar 91.00 %. Berdasarkan penelitian dari [10] kasus yang di bahas adalah Memprediksi Prestasi Siswa, Variabel Independen yaitu Disiplin, Presensi, AKhlak, Nilai Ujian dan Variabel Dependen Prestasi (YA, Tidak), data set 75 dengan algoritma C4.5 Teknik evaluasi yaitu AUC sebesar 97.5%. Berdasarkan penelitian dari [11] hasil dari pengolahan data dengan menggunakan Software PHP MYSQL, dari sampel 25 data maka dihasilkan 11 data yang termasuk kedalam cluster 1 dengan status penerima bantuan dikatakan layak, 5 data yang termasuk cluster 2 dengan status penerima dipertimbangkan dan sebanyak 9 data yang termasuk ke cluster 3 dengan status tidak layak. Berdasarkan penelitian dari [12] Kasus yang dibahas adalah *Prediction of Students' GPA*, Variabel Independen *High school status, High school location, High school type, English language capability, First Semester GPA, Numeric score, Verbal score, Spatial score and Analogy/Logic* dan Variabel Dependen yaitu *Prediction of Students' GPA (Yes, No)*, jumlah data set 1007, algoritma yang digunakan adalah K-Means dan Teknik evaluasi adalah presisi sebesar 94.62%.

Gap penelitian yang ingin diisi melalui penelitian ini adalah penelitian sebelumnya belum menggabungkan nilai akademik dan non-akademik secara bersamaan dalam proses pengelompokan prestasi, sebagian besar studi hanya menampilkan hasil *clustering* tanpa evaluasi akurasi yang terukur dan belum ada penelitian yang menggunakan dataset besar seperti 1.091 data siswa serta diuji dengan 30 data baru menggunakan confusion matrix sehingga akurasi dapat dipertanggungjawabkan. Berdasarkan kondisi tersebut, penelitian ini berupaya menerapkan algoritma K-Means *Clustering* untuk menentukan kategori prestasi siswa pada SMA Pangeran Antasari dengan mempertimbangkan komponen akademik dan non-akademik secara bersamaan. Dengan pendekatan ini, pengelompokan prestasi diharapkan lebih objektif dan mampu memberikan gambaran yang menyeluruh mengenai kemampuan siswa. Rumusan masalah dalam penelitian ini adalah bagaimana menerapkan algoritma K-Means *Clustering* untuk mengelompokkan siswa berdasarkan data akademik dan non-akademik, serta bagaimana mengevaluasi hasil pengelompokan tersebut agar dapat digunakan sebagai dasar pengambilan keputusan akademik di SMA Pangeran Antasari. Penelitian ini bertujuan untuk membangun model pengelompokan prestasi siswa menggunakan algoritma K-Means dengan memanfaatkan data akademik, absensi, dan aktivitas non-akademik. Selain itu, penelitian ini bertujuan mengevaluasi tingkat akurasi hasil *clustering* menggunakan metode *Elbow* dan confusion matrix, sehingga model yang dihasilkan dapat digunakan oleh sekolah sebagai alat bantu dalam proses evaluasi prestasi siswa.

## 2. METODOLOGI PENELITIAN

### 2.1 Rancangan Penelitian

CRISP-DM (*Cross Industry Standard Process for Data Mining*) adalah model pengembangan data mining yang terdiri atas enam tahapan. Pada penelitian ini, setiap tahap tidak dijelaskan secara teoritis saja, tetapi langsung dikaitkan dengan implementasi nyata saat pengolahan data prestasi siswa. Pada penelitian ini, metodologi CRISP-DM diterapkan sebagai pemecahan masalah yang umum untuk berbagai masalah bisnis dan penelitian. Metodologi ini terdiri dari enam tahap yaitu *Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation*, dan *Deployment*. Setiap tahapan dalam metodologi ini dapat dijelaskan sebagai berikut :



**Gambar 1.** Tahapan Penelitian

Adapun keterangan dari metode Crips-DM diatas adalah sebagai berikut :

1) *Business Understanding* (Pemahaman Bisnis)

Tahapan ini merupakan langkah awal dimulai dengan memahami kebutuhan dan tujuan dari sudut pandang bisnis, mengubah pengetahuan menjadi definisi masalah dalam konteks data mining, kemudian merencanakan serta menyusun strategi untuk mencapai tujuan dalam proses data mining [13]. Pada tahap ini, peneliti memahami kebutuhan SMA Pangeran Antasari yang ingin mengelompokkan siswa berdasarkan prestasi akademik dan non-akademik secara objektif. Fokus permasalahan adalah menentukan kategori tinggi, sedang, dan rendah.

a. *Web Browsing*

Peneliti mencari data-data yang mendukung penulisan skripsi dan materi metode melalui internet berupa jurnal - jurnal nasional tentang penelitian terkait sebelumnya, definisi dari berbagai sumber di *website* [14].

2) *Data Understanding* (Pemahaman Data)

Proses yang dilakukan dalam pemahaman data pengumpulan dan pemahaman data apa saja yang diperlukan, yang berupa data prestasi siswa/I untuk akademik dan non akademik untuk tahun ajaran 2023/2024. Peneliti menganalisis struktur data siswa yang berjumlah 1.091 entri yang terdiri dari nilai akademik, absensi, dan aktivitas non-akademik. Pada tahap ini dilakukan identifikasi variabel penting, karakteristik data, serta menemukan adanya missing value pada absensi dan skor ekstrakurikuler. Tahapan ini dimulai dengan pengumpulan data, deskripsi data, dan mengevaluasi kualitas data. Prosesnya mencakup langkah-langkah untuk memahami sifat dan karakteristik dari data yang akan digunakan. Adapun atribut dari penentuan prrestasi siswa adalah Nilai Agama, Nilai Pancasila dan kewarganegaraan, nilai b.indo, nilai matematika, Nilai sejarah, Nilai b.inggris, nilai seni budaya, Nilai sejarah, nilai biologi, absensi, UTS, UAS, non akademik seperti kegiatan ekstrakurikuler seperti pramuka, olahraga, seni music dan penilaian sikap perilaku siswa. Untuk Variabel Dependen adalah Kategori Prestasi (tinggi, cukup, dan rendah).

3) *Data Preparation* (Persiapan Data)

Pengolahan data terdiri dari kegiatan untuk menyusun *data set* yang nantinya akan di inputkan di sistem dalam pemodelan *data mining*. Tahap ini dilakukan untuk mempersiapkan dataset akhir dan atribut yang akan digunakan pada langkah selanjutnya, dan dataset yang akan digunakan pada penelitian ini adalah *dataset* [15]. Tahap ini menjadi inti karena berhubungan langsung dengan kualitas hasil clustering. Pada penelitian ini, peneliti melakukan preprocessing yang meliputi:

- a. Penanganan missing value yaitu Missing value pada absensi diganti menggunakan nilai *median* setiap kelas, sedangkan skor ekstrakurikuler yang kosong diisi dengan nilai minimal kategori.
- b. Penanganan outlier yaitu Outlier pada absensi (jumlah ketidakhadiran yang sangat ekstrem) dideteksi menggunakan *Interquartile Range (IQR)* dan diperhalus menggunakan *winsorizing*, bukan dihapus.
- c. Normalisasi data yaitu Semua variabel dinormalisasi menggunakan Min-Max Normalization, agar perbedaan skala tidak mempengaruhi jarak antar cluster:

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

- d. Transformasi data yaitu Variabel dengan distribusi tidak normal (misal absensi) dilakukan transformasi log agar sebaran lebih stabil.

#### 4) *Modelling* (Pemodelan)

Pada penelitian ini digunakan algoritma K-Means Clustering dengan tiga cluster berdasarkan hasil Elbow Method. Perhitungan jarak antar data dan centroid menggunakan Euclidean Distance:

##### **Contoh perhitungan:**

- a) nilai akademik = 80
  - b) absensi = 3 hari
  - c) ekstrakurikuler = 85
- Dan centroid cluster 1 memiliki nilai:

- a) akademik = 82
- b) absensi = 2
- c) ekstrakurikuler = 88

Maka jaraknya:

$$d = \sqrt{(80 - 82)^2 + (3 - 2)^2 + (85 - 88)^2} \quad (2)$$

$$d = \sqrt{4 + 1 + 9} = \sqrt{14} = 3.74$$

Rumus ini digunakan untuk menentukan kedekatan mahasiswa dengan setiap cluster.

#### 5) *Evaluation* (Pengujian)

Tahap ini melibatkan penilaian kinerja pola yang dihasilkan oleh algoritma. Evaluasi dilakukan dengan menggunakan Teknik evaluasi *elbow* [16]. Evaluasi dilakukan dengan Confusion Matrix pada 30 data uji sehingga diperoleh akurasi 93,33%. Evaluasi ini juga menguji kesesuaian hasil cluster dengan kategori aktual prestasi siswa.

#### 6) *Deployment*

Tahap ini biasanya berupa implementasi sistem, namun pada penelitian ini deployment hanya dilakukan dalam bentuk analisis output, visualisasi cluster, hasil grafik Elbow, dan rekomendasi akademik untuk sekolah. UML yang ditampilkan pada perancangan digunakan untuk gambaran sistem, tetapi hasil akhirnya berupa model analisis, bukan aplikasi yang berjalan.

## 2.2. Tools yang digunakan

Penelitian ini mengembangkan sebuah aplikasi berbasis web untuk menerapkan algoritma K-Means dalam pengelompokan prestasi siswa. Oleh karena itu, tools yang digunakan lebih fokus pada pengembangan web, basis data, dan implementasi algoritma secara mandiri tanpa software data mining seperti RapidMiner.

### 1. Bahasa Pemrograman dan Framework

- a) PHP 8 digunakan untuk mengimplementasikan algoritma K-Means, melakukan perhitungan Euclidean Distance, pembaruan centroid, dan proses iterasi.
- b) HTML5, CSS, JavaScript digunakan untuk pembangunan antarmuka website agar aplikasi mudah diakses oleh sekolah.

### 2. Database yaitu MySQL digunakan untuk menyimpan data siswa (nilai, absensi, ekstrakurikuler), data hasil preprocessing, hasil cluster dan log perhitungan iterasi.

### 3. Web Server yaitu XAMPP digunakan untuk menjalankan aplikasi pada localhost selama pengujian dan implementasi.

### 4. Tools Pengembangan

- a) Visual Studio Code / Sublime Text – untuk menulis kode PHP dan desain web.
- b) phpMyAdmin – pengelolaan database MySQL (import, export, cleaning data).

## 2.3 Prestasi

Analisis data prestasi proses pengumpulan, pengolahan, dan interpretasi data yang berkaitan dengan hasil belajar atau capaian akademik siswa. Data yang dianalisis dapat mencakup nilai akademik, kehadiran, keaktifan dalam kegiatan, maupun aspek lain yang memengaruhi prestasi siswa. Tujuan utama dari analisis ini adalah untuk memahami pola, tren, dan faktor-faktor yang memengaruhi keberhasilan atau kesulitan siswa dalam belajar [17]

## 2.4 K-Means

K-Means merupakan algoritma clustering yang berulang-ulang. Algoritma k-means menetapkan nilai-nilai *cluster* (*k*) secara acak, dimana nilai tersebut menjadi pusat dari *cluster* atau disebut sebagai *centroid*, *mean* atau *means* [18]. Algoritma k-means dalam implementasinya sangat mudah, cepat, mudah beradaptasi sederhana untuk diimplementasikan dan dijalankan, relatif cepat, dan mudah beradaptasi serta mempunyai kemampuan yang besar dalam mengolah data yang cukup besar dan waktu lebih efisien. Yang menjadi kelemahan dalam algoritma k-means saat menentukan *cluster* awal, karena bergantung pada inisial data yang diberikan [19]. Adapun prosedur perhitungan Algoritma K- Means *Clustering analysis* diuraikan sebagai berikut:

1. Pilih secara acak *k* buah data sebagai pusat *cluster*
2. Jarak antara data dan pusat *cluster* dihitung menggunakan *Euclidian Distance*. Untuk menghitung jarak semua data ke setiap titik pusat *cluster* dapat menggunakan teori jarak *Euclidean* yang dirumuskan sebagai berikut:

$$D_{\text{euclidean}}(x, y) = \sqrt{\sum (x_i - y_i)^2} \quad (3)$$

Keterangan:

$d(x, y)$  : jarak antara data pada titik *x* dan *y*

*x* : titik data objek

*y* : titik data centroid

*i* : jumlah atribut data

3. Data ditempatkan dalam *cluster* yang terdekat, dihitung dari tengah *cluster*
4. Pusat *cluster* baru akan ditentukan bila semua data telah ditetapkan dalam *cluster* terdekat. Rumus menghitung titik pusat *cluster* baru:

$$v = \sum_{i=1}^n x_i; i = 1, 2, 3 \dots \dots \dots n \quad (4)$$

dimana

*v* adalah *centroid* pada *cluster*

*Xi* adalah objek ke-*i*

*N* adalah banyaknya objek/jumlah objek yang menjadi anggota *cluster*

5. Proses penentuan pusat *cluster* dan penempatan data dalam *cluster* diulangi sampai nilai *centroid* tidak berubah lagi [20]

### 3. HASIL DAN PEMBAHASAN

Untuk menyelesaikan permasalahan dari *business understanding* yang terkait Kategori Siswa Berdasarkan Prestasi, data yang digunakan dalam penelitian ini yaitu data siswa/i pada SMA Pangeran Antasari. Data yang didapat merupakan data SMA Pangeran Antasari dengan Variabel Nilai Agama, Nilai Pancasila dan kewarganegaraan, nilai b.indo, nilai matematika, Nilai sejarah, Nilai b.inggris, nilai seni budaya, Nilai sejarah, nilai biologi, absensi, UTS, UAS, non akademik seperti kegiatan ekstrakurikuler seperti pramuka, olahraga, seni music dan penilaian sikap perilaku siswa, jumlah total data adalah 1091. Dapat dilihat bahwa *dataset* yang ada masih terlihat berantakan, sehingga diperlukan pengolahan lagi pada *dataset*. Dimana data yang telah diolah dapat digunakan untuk menyelesaikan penentuan prestasi siswa.

#### 3.1. Hasil Penelitian

##### 1. Preparation Data

Data cleaning merupakan suatu metode yang digunakan untuk memastikan kebersihan data agar terhindar dari value yang hilang (*missing value*) atau kesalahan dalam proses input (*noise value*). Untuk proses penemuan *missing value* penulis menggunakan formula *excel* untuk menghitung dan memastikan jumlah value setiap atribut sesuai dengan jumlah *record* yang dimiliki dataset. Untuk proses penemuan *noise value* dibutuhkan pemahaman pada atribut data. Misalnya pada atribut prestasi non akademik, dimana atribut ini memiliki makna siswa/i mengikuti dan tidak mengikuti kegiatan tersebut sangat tidak memungkinkan jika value pada atribut ini kurang atau sama dengan 0.

##### 1. Missing Value

*Missing Value* dalam proses *preprocessing* merujuk pada data yang tidak tersedia atau hilang pada fitur atau variabel dalam dataset. Ini merupakan nilai yang tidak terisi, yang bisa terjadi karena berbagai alasan seperti kesalahan pengumpulan data, responden yang tidak menjawab pertanyaan, atau data yang tidak dapat diukur. Teknik yang sering digunakan untuk mengatasi *missing value* meliputi:

- a. Penghapusan Data (*Deletion*) : Menghapus baris atau kolom yang memiliki *missing value* jika jumlahnya sedikit.
- b. Imputasi: Menggantikan *missing value* dengan nilai pengganti, misalnya dengan rata-rata, *median*, *modus*, atau metode imputasi lanjutan seperti regresi.

**Tabel 1.** Statistik *Missing Value*

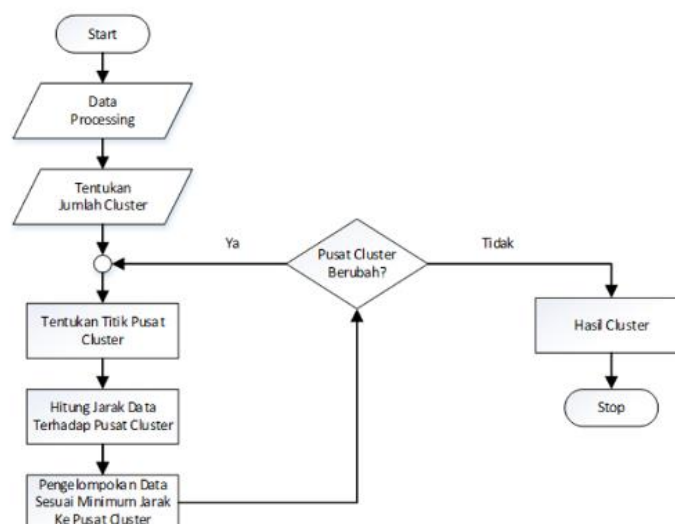
No	Nama Fitur	Missing Value
1	Nomor	0
2	Nama	0
3	Agama	0
4	Pancasila	0
5	Bahasa Indonesia	0
6	Matematika	0
7	Sejarah	0
8	Bahasa Inggris	0
9	Seni Budaya	0
10	Biologi	0
11	Absensi	0
12	UTS	0
13	UAS	0
14	Pramuka	47
15	Olahraga	70
16	Seni musik	50
17	Perilaku	30

## 2. Outlier

*Outlier* dalam proses preprocessing data adalah nilai yang secara signifikan berbeda dari mayoritas data lainnya. *Outlier* dapat muncul karena kesalahan pengukuran, variasi alami dalam data, atau faktor eksternal lainnya. Dari data set yang diperoleh ada 19 data yang *outlier* pada fitur pramuka.

## 2. Modelling

Pada penelitian ini digunakan algoritma k-Means yang berfungsi untuk mendapatkan kategori prestasi siswa berdasarkan banyaknya jumlah yang siswa. Adapun flowchart dari metode K-Means adalah sebagai berikut :



Gambar 2. Flowchart K-Means



Penentuan jumlah cluster diawali dengan perhitungan nilai *Within-Cluster Sum of Square* (WCSS). Grafik Elbow memperlihatkan penurunan tajam pada K=3, sehingga jumlah cluster optimal adalah tiga kelompok: prestasi tinggi, cukup, dan rendah. Pemilihan tiga cluster juga sejalan dengan kebutuhan sekolah yang ingin memetakan rentang prestasi menjadi kategori yang mudah diinterpretasikan dalam proses pembinaan. Selanjutnya pemilihan klaster dari *dataset* yang ada untuk penentuan *centroid* secara acak/random sebagai awal perhitungan untuk menentukan kedekatan jarak antara data dan pusat *cluster*. Memilih secara acak k buah data sebagai pusat pengelompokan (cluster). Penentuan *centroid* secara acak/random dapat dilihat pada Tabel 2.

**Tabel 2.** Sample Centroid

C1	Tinggi	87	89	87	80	87	92	90	94	4	90	95	85	92	91	85
C2	Cukup	80	87	72	72	80	80	83	87	6	79	82	80	81	84	85
C3	Rendah	87	80	72	70	72	78	77	82	7	75	80	76	80	79	85

## ITERASI 1

**Abdillah Maulana**

*Cluster 1*

$$d1 = \sqrt{(70-87)^2 + (80-89)^2 + (72-87)^2 + (87-80)^2 + (72-87)^2 + (88-92)^2 + (85-90)^2 + (91-94)^2 + (5-4)^2 + (85-90)^2 + (90-95)^2 + (80-85)^2 + (87-92)^2 + (88-91)^2 + (90-85)^2}$$

$$= 32.47$$

*Cluster 2*

$$d1 = \sqrt{(70-80)^2 + (80-87)^2 + (72-72)^2 + (87-72)^2 + (72-80)^2 + (88-83)^2 + (85-87)^2 + (91-94)^2 + (5-6)^2 + (85-79)^2 + (90-82)^2 + (80-80)^2 + (87-81)^2 + (88-84)^2 + (90-85)^2}$$

$$= 26.46$$

*Cluster 3*

$$d1 = \sqrt{(70-87)^2 + (80-80)^2 + (72-72)^2 + (87-70)^2 + (72-72)^2 + (88-78)^2 + (85-87)^2 + (91-82)^2 + (5-7)^2 + (85-75)^2 + (90-80)^2 + (80-76)^2 + (87-80)^2 + (88-79)^2 + (90-85)^2}$$

$$= 34.61$$

Maka diperoleh hasil perhitungan untuk *Cluster* Pada Iterasi -1

**Tabel 3.** Hasil Iterasi 1

No	Nama	Cluster 1	Cluster 2	Cluster 3	Jarak Terdekat	Cluster	SSE (Jarak ^2)
1	Abdillah Maulana	32,47	26,46	34,61	26,46	C2	700,00
2	Abdul Wahab	38,24	25,96	22,80	22,80	C3	520,00
3	Adek Arya Prasetya	23,04	30,94	37,91	23,04	C1	531,00
4	Adelia Natasya	44,14	21,82	24,25	21,82	C2	476,00
5	Adelina Putri Yasmin	27,50	23,41	29,05	23,41	C2	548,00
895	Naila Dwi Yanti	41,76	24,45	30,40	24,45	C2	598,00
896	Natanael Hasonangan Sibagariang	28,71	31,27	41,93	28,71	C1	824,00
897	Nazwa Ananda Syahputri	36,47	31,65	41,88	31,65	C2	1002,00
898	Nazwa Zahara	30,92	23,37	29,46	23,37	C2	546,00
Total							630788,00
Rata – Rata							702,44

Setelah semua data ditempatkan ke dalam *cluster* yang terdekat, kemudian hitung kembali pusat *cluster* yang baru berdasarkan rata-rata anggota yang ada pada *cluster* tersebut sehingga mendapatkan hasil perhitungan centroid baru yang akan digunakan untuk iterasi berikutnya. Pada *cluster* 1 terdapat 382 data, *cluster* 2 sebanyak 422 data, dan *cluster* 3 sebanyak 94 data.

## ITERASI 2

**Abdillah Maulana**

d1 =

$$\sqrt{(70 - 81.04)^2 + (80 - 80.48)^2 + (72 - 80.76)^2 + (87 - 81.45)^2 + (72 - 79.80)^2 + (88 - 90.24)^2 + (85 - 90.28)^2 + (91 - 92.83)^2 + (5 - 4.79)^2 + (85 - 86.66)^2 + (90 - 91.59)^2 + (80 - 88.46)^2 + (87 - 91.26)^2 + (88 - 90.19)^2 + (90 - 88.56)^2}$$

= 26.81

d2 =

$$\sqrt{(70 - 79)^2 + (80 - 81.3)^2 + (72 - 77.21)^2 + (87 - 80.72)^2 + (72 - 72.69)^2 + (88 - 82.36)^2 + (85 - 81.20)^2 + (91 - 84.82)^2 + (5 - 6.33)^2 + (85 - 80.84)^2 + (90 - 84)^2 + (80 - 81.03)^2 + (87 - 83.49)^2 + (88 - 82.51)^2 + (90 - 83.33)^2}$$

= 23.90

$$\sqrt{(70 - 87)^2 + (80 - 80)^2 + (72 - 72)^2 + (87 - 70)^2 + (72 - 72)^2 + (88 - 78)^2 + (85 - 87)^2 + (91 - 82)^2 + (5 - 7)^2 + (85 - 75)^2 + (90 - 80)^2 + (80 - 76)^2 + (87 - 80)^2 + (88 - 79)^2 + (90 - 85)^2}$$

= 32.36 dan seterusnya

Maka diperoleh hasil perhitungan untuk *Cluster* Pasa Iterasi -2

**Tabel 4.** Hasil Iterasi 2

No	Nama	Cluster 1	Cluster 2	Cluster 3	Jarak Terdekat	Cluster	SSE (Jarak ^2)
1	Abdillah Maulana	26,81	23,90	32,36	23,90	C2	570,99
2	Abdul Wahab	30,12	19,42	16,02	16,02	C3	256,78
3	Adek Arya Prasetya	26,38	30,22	40,38	26,38	C1	696,13
4	Adelia Natasya	35,19	21,52	18,25	18,25	C3	333,24
5	Adelina Putri Yasmin	28,56	26,54	33,94	26,54	C2	704,24
896	Natanael Hasonangan Sibagariang	16,74	19,68	33,23	16,74	C1	280,33
897	Nazwa Ananda Syahputri	29,00	24,86	34,41	24,86	C2	618,11
898	Nazwa Zahara	24,68	20,24	28,04	20,24	C2	409,58
Total							371874,31
Rata – Rata							414,11

Setelah semua data ditempatkan ke dalam *cluster* yang terdekat, kemudian hitung kembali pusat *cluster* yang baru berdasarkan rata-rata anggota yang ada pada *cluster* tersebut sehingga mendapatkan hasil perhitungan centroid baru yang akan digunakan untuk iterasi berikutnya. Pada *cluster* 1 terdapat 402 data, *cluster* 2 sebanyak 364 data, dan *cluster* 3 sebanyak 132 data. Berikut tabel ringkas yang merangkum iterasi proses pembentukan centroid:

**Tabel 5.** Ringkasan Iterasi

Iterasi	Perubahan Centroid	Status
1	Centroid bergeser signifikan	Lanjut iterasi
2	Pergeseran mulai kecil	Lanjut
3	Pergeseran < 0.01	Mendekati konvergen
4	Pergeseran < 0.01	Mendekati konvergen
5	Pergeseran < 0.01	Mendekati konvergen



Tabel ini menunjukkan bahwa model mencapai stabilitas pada iterasi ke-4, menandakan bahwa struktur data cukup konsisten.

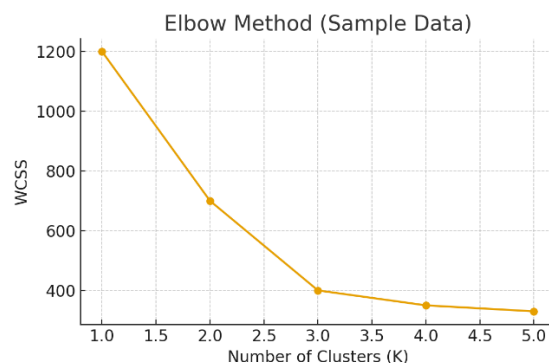
### 3.2. Pembahasan

Setelah melakukan iterasi 5 dan iterasi 6 tidak ada perubahan, maka proses pengelompokan di hentikan. Dari hasil iterasi 5 diperoleh 383 Siswa kategori prestasi siswa tinggi, 345 siswa kategori prestasi siswa cukup dan 185 siswa kategori prestasi siswa rendah. Tahapan ini melibatkan penilaian kinerja pola yang dihasilkan oleh algoritma. Evaluasi dilakukan dengan menggunakan Teknik evaluasi elbow. Metode elbow merupakan suatu metode penentuan jumlah kluster yang ideal dengan upaya menyesuaikan tingkatan kluster sehingga terbentuk siku pada titik tertentu. Maka dilakukanlah pengujian dengan menginputkan jarak rata-rata (*Avg Distance*). *Avg Distance* merupakan jarak rata-rata antara setiap titik K. Kemudian terbentuklah grafik *Avg Distance* yang memperlihatkan titik/garis yang membentuk siku. Maka dari itu dapat dilihat semakin besar titik sudut maka semakin tinggi pula akurasi dalam menentukan jumlah K pada K-Means Clustering

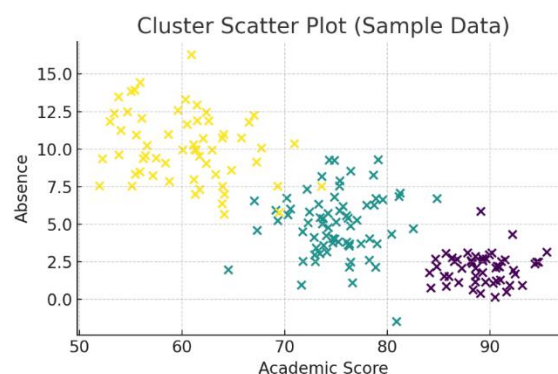
**Tabel 6.** Hasil SSE

K	Hasil SSE	Selisih
1	702,44	0
2	414,11	288,33
3	409,88	4,23
4	408,55	1,33
5	406,79	1,76
6	404,41	2,38

Dari hasil Tabel diatas, kemudian dilakukan metode optimasi menggunakan metode *elbow* untuk menentukan jumlah *cluster* terbaik. Berikut adalah hasil perhitungan SSE berdasarkan pengujian nilai k=1 sampai dengan k=5. Dari perhitungan SSE metode *Elbow* pada Tabel diatas, didapatkan nilai SSE yang memiliki selisih terbesar terdapat pada nilai k=2, sehingga jumlah *cluster* optimal yang dapat dibentuk sejumlah 3 cluster.



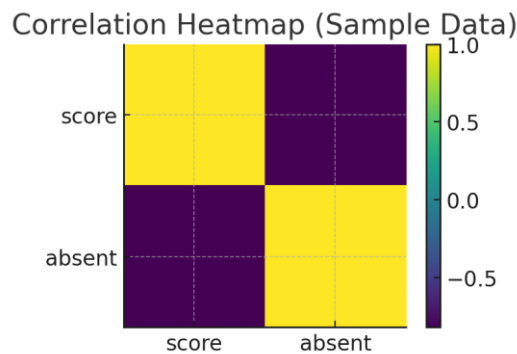
**Gambar 3.** Grafik Elbow



**Gambar 4.** Plot Sebaran Cluster Berdasarkan Dua Variabel Utama (placeholder)

Pada visualisasi terlihat Cluster prestasi tinggi terkumpul pada rentang nilai sangat tinggi dengan absensi baik. Cluster sedang berada di tengah dengan keragaman yang cukup besar. Cluster rendah tersebar pada nilai rendah.

dan absensi tidak stabil.



**Gambar 5.** Heatmap Variabel (placeholder)

Heatmap menunjukkan bahwa Nilai akademik memiliki korelasi kuat dengan cluster. Absensi berkontribusi konsisten terutama pada cluster rendah dan Aktivitas non-akademik memiliki pengaruh sedang. Dari grafik SSE metode Elbow pada Gambar diatas terlihat bahwa pada  $k=2$  mengalami penurunan paling besar jika dibandingkan jumlah  $k$  yang lain. Jumlah *cluster* selain  $k=2$  mengalami penurunan yang stabil. Maka dapat disimpulkan jumlah *cluster* optimal yang terbentuk sejumlah 3 cluster. Berikut ini merupakan hasil pengujian untuk data baru sebanyak 30 siswa dan menampilkan hasil keputusan K-Means dari aturan yang telah terbentuk sebagai berikut :

**Tabel 7.** Data Evaluasi data baru

No NISN	Nama Siswa	Data Aktual	Cluster (Analisa)
1 0081209495	Bancar Luwes Iswahyudi	Cukup	Cukup
2 0087217578	Intan Hasanah	Rendah	Rendah
3 0071766801	Kenari Ganep Saptono	Tinggi	Tinggi
4 0084901632	Aisyah Nasyidah	Rendah	Rendah
5 0085083430	Gasti Mardhiyah	Tinggi	Tinggi
6 0088950005	Muni Haryanto	Tinggi	Tinggi
7 0086128870	Zelda Prastuti	Rendah	Rendah
8 0071465519	Ega Marbun	Tinggi	Cukup
9 0088760490	Siti Kusmawati	Rendah	Rendah
10 0086621150	Gilang Wasita	Tinggi	Tinggi
11 0086207219	Nurul Hassanah	Tinggi	Tinggi
12 0073925023	Omar Mahendra	Tinggi	Tinggi
13 0086392092	Kuncara Kusumo	Cukup	Cukup
14 0074215252	Zaenab Mandasari	Tinggi	Tinggi
15 0074058973	Lukman Kamal Putra	Cukup	Cukup
16 0087763268	Yance Raisa Hastuti	Cukup	Cukup
17 0084045574	Damar Saputra	Tinggi	Tinggi
18 0081268761	Mila Hartati	Rendah	Rendah
19 0088815056	Drajat Raditya Zulkarnain	Cukup	Cukup
20 0083198674	Upik Kuswoyo	Rendah	Rendah
21 0083005439	Anastasia Mulyani	Cukup	Cukup
22 0089009584	Salsabila Zelda Wijayanti	Tinggi	Cukup
23 0081660852	Hilda Hassanah	Tinggi	Tinggi
24 0083429765	Cinta Astuti	Tinggi	Tinggi
25 0081933811	Putri Melani	Cukup	Cukup
26 0088406185	Asirwanda Wasita	Rendah	Rendah
27 0089703055	Irnanto Megantara	Tinggi	Tinggi
28 0084987760	Enteng Waluyo	Tinggi	Tinggi
29 0084914050	Purwanto Napitupulu	Tinggi	Tinggi
30 0088114019	Lintang Lestari	Tinggi	Tinggi

Dari table diatas dapat diketahui hasil sebagai berikut untuk perhitungan pengujian confusion matriks :

**Tabel 8.** Hitung Confusion Matrix (30 Data)

	Pred: Tinggi	Pred: Cukup	Pred: Rendah
<b>Actual: Tinggi</b>	15	2	0
<b>Actual: Cukup</b>	0	8	0
<b>Actual: Rendah</b>	0	0	7

Keterangan:

Ada 2 kesalahan klasifikasi, yaitu:

Ega Marbun : aktual: Tinggi, prediksi: Cukup

Salsabila Zelda Wijayanti : aktual: Tinggi, prediksi: Cukup

Selanjutnya melakukan perhitungan confusion matriks sebagai berikut :

a. Akurasi

Akurasi = (jumlah prediksi benar) / (total data)

=  $(15 + 8 + 7) / 30$

=  $30 / 30 = 0.9333 \rightarrow 93.33\%$

b. Precision, Recall, F1-Score (per kelas)

**Tabel 9.** Per kelas

Kelas	Precision	Recall	F1-Score
Tinggi	$15 / (15+0) = 1.00$	$15 / (15+2) = 0.882$	0.938
Cukup	$8 / (8+2) = 0.80$	$8 / (8+0) = 1.00$	0.889
Rendah	$7 / (7+0) = 1.00$	$7 / (7+0) = 1.00$	1.000

F1-Score rata-rata (macro average) =  $(0.938 + 0.889 + 1.000) / 3 \approx 0.942$

Kesimpulan Hasil Pengujian

- Algoritma K-Means mampu mengelompokkan siswa dengan tingkat akurasi sebesar 93.33%.
- Kelas “Rendah” dan “Cukup” diklasifikasikan sempurna (100% recall), sementara kelas “Tinggi” memiliki dua kesalahan klasifikasi.
- Nilai F1-score rata-rata mencapai 94.2%, menunjukkan bahwa model *Clustering* ini sangat efektif dalam mendeteksi kategori siswa berdasarkan prestasi.

Hasil evaluasi menunjukkan cluster rendah memiliki akurasi paling stabil, sementara cluster tinggi mengalami sedikit mis-klasifikasi. Ketidakseimbangan akurasi ini disebabkan oleh beberapa factor yaitu Cluster Tinggi Rentan Tertukar dengan Cluster Cukup, Karena beberapa siswa dengan nilai yang tinggi namun absensi tidak sempurna cenderung memiliki profil mirip cluster sedang. Nilai tinggi tidak selalu diikuti dengan konsistensi kehadiran, sehingga algoritma menempatkan mereka pada cluster menengah. Cluster Rendah Paling Mudah Dibedakan karena Karakteristik cluster rendah sangat kontras nilai akademik rendah, ketidakhadiran tinggi dan aktivitas non-akademik minim, sehingga centroid cluster rendah sangat stabil dan sulit membingungkan model. Selanjutnya adalah Sebaran Data Tidak Simetris karena Data siswa berprestasi tinggi jauh lebih sedikit dibandingkan siswa berprestasi sedang dan rendah. Hal ini membuat *boundary* antara cluster tinggi–sedang kurang tajam. Interpretasi Hasil dalam konteks pendidikan adalah Implikasi untuk Kelompok Prestasi Tinggi aytu klasifikasi pada cluster ini menunjukkan bahwa beberapa siswa berpotensi kehilangan pengakuan prestasi hanya karena faktor absensi atau aktivitas non-akademik. Hal ini memberikan masukan bahwa indikator prestasi sebaiknya tidak hanya fokus pada kehadiran, tetapi juga mempertimbangkan capaian akademik yang konsisten. Implikasi untuk Kelompok Prestasi Sedang karena Cluster ini memiliki keragaman paling besar. Kategori ini merupakan kelompok “penyeimbang” yang berpotensi naik ataupun turun kelas prestasi. Sekolah dapat membuat program *remedial* atau *mentoring* untuk mendorong perbaikan, khususnya pada aspek absensi dan literasi belajar. Selanjutnya adalah Implikasi untuk Kelompok Prestasi Rendah Karena profil cluster ini sangat mudah diidentifikasi, model dapat menjadi dasar pembuatan program pendampingan belajar intensif, konseling rutin, pemantauan absensi yang lebih ketat dan evaluasi penyebab penurunan minat belajar.

Pemanfaatan Hasil Clustering bagi Sekolah yaitu perencanaan Intervensi Belajar Sekolah dapat memfokuskan program tutor sebaya untuk siswa cluster rendah, monitoring Prestasi Secara Berkala yaitu clustering dapat dijalankan per semester untuk memantau perkembangan siswa. Perencanaan Penghargaan dan Pembinaan yaitu Siswa cluster tinggi dapat diarahkan mengikuti lomba, program pengembangan minat, dan kegiatan akademik lanjutan. Evaluasi Kebijakan Akademik yaitu Guru dan wali kelas dapat menggunakan data clustering sebagai indikator tambahan untuk menilai kualitas proses pembelajaran.

## 4. KESIMPULAN

Penelitian ini bertujuan menjawab permasalahan mengenai bagaimana proses penerapan algoritma K-Means dalam menentukan kategori prestasi siswa serta bagaimana hasil pengelompokannya dapat membantu sekolah dalam melakukan evaluasi secara objektif. Berdasarkan proses pemodelan dan pengujian yang dilakukan, algoritma K-Means berhasil mengelompokkan 1.091 siswa SMA Pangeran Antasari ke dalam tiga kategori

prestasi, yaitu tinggi, cukup, dan rendah, melalui pemanfaatan data akademik dan non-akademik. Hasil evaluasi menggunakan confusion matrix terhadap 30 data uji menunjukkan tingkat akurasi sebesar 93,33%, yang menegaskan bahwa metode ini mampu memberikan pengelompokan yang konsisten dan sesuai dengan kondisi aktual siswa. Dengan demikian, penelitian ini berhasil menjawab rumusan masalah terkait penerapan algoritma K-Means serta efektivitasnya dalam menghasilkan klasifikasi yang dapat dijadikan dasar untuk pembinaan dan pengambilan keputusan pihak sekolah. Namun demikian, penelitian ini memiliki beberapa keterbatasan. Dataset yang digunakan hanya berasal dari satu sekolah sehingga pola yang dihasilkan belum tentu dapat digeneralisasikan pada konteks sekolah lain. Selain itu, variabel yang digunakan masih terbatas pada nilai akademik, absensi, dan beberapa indikator non-akademik, sehingga belum mencakup faktor lain yang juga dapat berpengaruh terhadap prestasi, seperti kondisi keluarga, psikologis, atau lingkungan belajar. Model K-Means yang digunakan juga bergantung pada pemilihan centroid awal yang acak, sehingga hasil cluster memiliki potensi bervariasi jika dilakukan pemodelan ulang. Berdasarkan keterbatasan tersebut, penelitian lanjutan dapat dilakukan dengan memperluas cakupan dataset dari beberapa sekolah berbeda, memasukkan variabel tambahan yang lebih lengkap, serta menguji berbagai metode clustering lain seperti K-Medoids, Self-Organizing Maps (SOM), atau DBSCAN untuk melihat perbandingan hasil yang lebih komprehensif. Pengembangan lain dapat mencakup optimasi pemilihan centroid awal menggunakan metode seperti K-Means++, serta pengujian data lintas semester untuk melihat stabilitas kelompok prestasi siswa dari waktu ke waktu. Dengan pengembangan tersebut, diharapkan hasil klasifikasi dapat menjadi lebih akurat, adaptif, dan dapat diterapkan secara luas sebagai dasar pengambilan keputusan di lingkungan pendidikan.

## REFERENCES

- [1] E. A. Saputra and Y. Nataliani, "Analisis Pengelompokan Data Nilai Siswa untuk Menentukan Siswa Berprestasi Menggunakan Metode Clustering K-Means," *J. Inf. Syst. Informatics*, vol. 3, no. 3, pp. 424–439, 2021, doi: 10.51519/journalisi.v3i3.164.
- [2] A. Sulistiyawati and E. Supriyanto, "Implementasi Algoritma K-means Clustering dalam Penentuan Siswa Kelas Unggulan," *J. Tekno Kompak*, vol. 15, no. 2, p. 25, 2021, doi: 10.33365/jtk.v15i2.1162.
- [3] Fahrillah Fahrillah and Zaehol Fatah, "Pengelompokan Data Nilai Siswa Madrasah Ta' Hiliyah Menggunakan Metode K-Means Clustering," *J. Ris. Sist. Inf.*, vol. 2, no. 1, pp. 53–59, 2025, doi: 10.69714/0v1pkz05.
- [4] S. Haviyola, S. Susilawati, and M. Jajuli, "Pengelompokan Prestasi Siswa Guna Kualifikasi Beasiswa Berdasarkan Data Nilai Menggunakan Algoritma K-Means," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 4, pp. 2786–2791, 2024, doi: 10.36040/jati.v7i4.7200.
- [5] S. N. Br Sembiring, H. Winata, and S. Kusnasari, "Pengelompokan Prestasi Siswa Menggunakan Algoritma K-Means," *J. Sist. Inf. Triguna Dharma (JURSI TGD)*, vol. 1, no. 1, p. 31, 2022, doi: 10.53513/jursi.v1i1.4784.
- [6] N. D. Rahayu, A. H. Anshor, and I. Afriantoro, "Penerapan Data Mining untuk Pemetaan Siswa Berprestasi menggunakan Metode Clustering K-Means," *JUKI J. Komput. dan Inform.*, vol. 6, no. 1, pp. 71–83, 2024, doi: 10.53842/juki.v6i1.474.
- [7] S. Dewi, S. Defit, and Y. Yuhandri, "Akurasi Pemetaan Kelompok Belajar Siswa Menuju Prestasi Menggunakan Metode K-Means," *J. Sistim Inf. dan Teknol.*, vol. 3, pp. 28–33, 2021, doi: 10.37034/jsisfotek.v3i1.40.
- [8] D. Bahtiar *et al.*, "Pemetaan Penduduk Penerima Bantuan Sosial Desa Waru Jaya Menggunakan Algoritma K-Means Clustering," *Sci. Sacra J. Sains*, vol. 3, no. 2, pp. 29–39, 2023, [Online]. Available: <http://pijarpemikiran.com/index.php/Scientia>
- [9] V. No, M. Qusyairi, Z. Hidayatullah, A. Sandi, and V. No, "Infotek : Jurnal Informatika dan Teknologi Penerapan K-Means Clustering Dalam Pengelompokan Prestasi Siswa Dengan Optimasi Metode Elbow Infotek : Jurnal Informatika dan Teknologi Perkembangan teknologi saat ini berkembang dengan sangat pesat ini terbukti," vol. 7, no. 2, pp. 500–510, 2024.
- [10] S. Marpaung, S. -, and I. -, "Penerapan Metode Naïve Bayes Dalam Memprediksi Prestasi Siswa Di SMA Negeri 1 Panombeian Panei," *J. Sist. Inf. dan Ilmu Komput. Prima (JUSIKOM PRIMA)*, vol. 4, no. 2, pp. 8–13, 2021, doi: 10.34012/jurnalsisteminformasidanilmukomputer.v4i2.1522.
- [11] Y. Filki, "Algoritma K-Means Clustering dalam Memprediksi Penerima Bantuan Langsung Tunai (BLT) Dana Desa," *J. Inform. Ekon. Bisnis*, vol. 4, pp. 166–171, 2022, doi: 10.37034/infeb.v4i4.166.
- [12] R. G. Santosa, Y. Lukito, and A. R. Chrismanto, "Classification and Prediction of Students' GPA Using K-Means Clustering Algorithm to Assist Student Admission Process," *J. Inf. Syst. Eng. Bus. Intell.*, vol. 7, no. 1, p. 1, 2021, doi: 10.20473/jisebi.7.1.1-10.
- [13] A. Pramudya, I. Maulana, and R. Mayasari, "Pengelompokan Hasil Belajar Siswa Sdn Tunas Jaya Dengan Algoritma K-Means," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 6, pp. 3960–3967, 2024, doi: 10.36040/jati.v7i6.7970.
- [14] A. Asmana, Y. Arie Wijaya, and M. Martanto, "Clustering Data Calon Siswa Baru Menggunakan Metode K-Means Di Sekolah Menengah Kejuruan Wahidin Kota Cirebon," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 6, no. 2, pp. 552–559, 2022, doi: 10.36040/jati.v6i2.5236.
- [15] Rahmat Hidayat, "Pemanfaatan Data Mining Untuk Melihat Minat Siswa Setelah Menyelesaikan Pendidikan Sekolah Menengah Atas (Sma) Dengan Algoritma K-Means Clustering," *Technol. Informatics Insight J.*, vol. 1, no. 2, pp. 85–97, 2022, doi: 10.32639/tiij.v1i2.220.
- [16] C. Loho, V. P. Rantung, and G. C. Rorimpandey, "Data Mining Rekomendasi Sekolah Calon Siswa SMA di Kota Tomohon Menggunakan Metode K-Means Clustering," *JOINTER J. Informatics Eng.*, vol. 3, no. 02, pp. 1–9, 2022, doi: 10.53682/jointer.v3i02.34.
- [17] S. Anwar, T. Suprpti, G. Dwilestari, and I. Ali, "Pengelompokkan Hasil Belajar Siswa dengan Metode Clustering K-Means," *JURISISTEKNI (Jurnal Sist. Inf. dan Teknol. Informasi)*, vol. 4, no. 2, pp. 60–72, 2022.



- [18] C. Satria and A. Anggrawan, "Aplikasi K-Means berbasis Web untuk Klasifikasi Kelas Unggulan," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 21, no. 1, pp. 111–124, 2021, doi: 10.30812/matrik.v21i1.1473.
- [19] I. A. Kurniawan, R. M. H. Bhakti, and B. Irawan, "Implementasi Data Mining Untuk Mengukur Prestasi Siswa SD Menggunakan Metode K-Means Clustering," *J. Citiz. Res. Dev.*, vol. 1, no. 2, pp. 262–268, 2024, doi: 10.57235/jcrd.v1i2.3324.
- [20] N. Yannuansa, M. Safaa<sup>TM</sup>udin, and M. I. Aziz, "Pemanfaatan Algoritma K-Means Clustering dalam Mengolah Pengaruh Hasil Belajar Terhadap Pendapatan Orang Tua Pada Mata Pelajaran Produktif," *J. Tecnoscienza*, vol. 6, no. 1, pp. 43–55, 2021, doi: 10.51158/tecnoscienza.v6i1.530.