# IDENTIFIKASI POLA BELAJAR DAN ABSENSI SISWA DENGAN TEKNIK CLUSTERING DAN ANALISIS KORELASI

**Cinta Renita, Tesa Nur Padilah, Muhammad Adrian Eka Wahyu**
Informatika, Universitas Singaperbangsa Karawang
Jalan HS Ronggo Waluyo, Indonesia
*2110631170007@student.unsika.ac.id*

**ABSTRAK**

Penelitian ini dilakukan untuk mengidentifikasi pola belajar dan absensi siswa menggunakan teknik clustering dan analisis korelasi. Latar belakang penelitian ini berfokus pada pentingnya memahami hubungan antara waktu belajar, kehadiran, dan kinerja akademik untuk membantu pendidik merancang strategi pembelajaran yang lebih efektif. Dengan menggunakan algoritma K-Means, data mengenai waktu belajar siswa, kehadiran, dan nilai ujian diolah untuk menemukan pola tertentu. Metode siku digunakan untuk menentukan jumlah cluster yang optimal, dan ditemukan bahwa tiga cluster adalah yang paling sesuai. Hasil analisis korelasi menunjukkan bahwa terdapat korelasi lemah antara waktu belajar dan hasil ujian, sementara terdapat korelasi yang sangat kuat antara hasil ujian pertama, kedua, dan nilai akhir. Meskipun skor Silhouette menunjukkan kualitas pengelompokan yang moderat, hasil penelitian ini memberikan wawasan berharga bagi pendidik untuk merancang strategi pembelajaran yang lebih disesuaikan dengan kebutuhan siswa. Saran untuk penelitian selanjutnya termasuk penggunaan algoritma clustering lainnya dan mempertimbangkan variabel tambahan yang mungkin mempengaruhi prestasi akademik siswa.

***Kata Kunci:*** *Clustering, K-Means, Analisis Korelasi, Silhouette Score, Pola Belajar*

## 1. PENDAHULUAN

Penilaian hasil belajar siswa merupakan suatu proses penting dalam pendidikan yang bertujuan untuk mengukur ketercapaian tujuan pembelajaran [7]. Penilaian ini memberikan gambaran yang jelas mengenai pemahaman, keterampilan, dan kemajuan siswa terhadap materi pelajaran yang diberikan. Penggunaan data penilaian secara efektif memungkinkan guru untuk mengembangkan strategi pembelajaran yang lebih tepat sasaran dan disesuaikan dengan kebutuhan individu siswa, sehingga meningkatkan kemungkinan siswa mencapai potensi belajar mereka sepenuhnya [8].

Pada era digital, pengolahan data dan analisis berbasis teknologi menjadi semakin penting dalam dunia pendidikan. Teknik data mining, seperti clustering, memungkinkan pendidik untuk mengelompokkan siswa berdasarkan pola tertentu, seperti pola belajar dan absensi, yang dapat memberikan wawasan berharga dalam pengambilan keputusan pendidikan [4].

Menggunakan data penilaian secara efektif memungkinkan guru untuk mengembangkan strategi pembelajaran yang lebih efektif dan disesuaikan dengan kebutuhan individu siswa, sehingga meningkatkan kemungkinan bahwa mereka akan mencapai potensi belajar mereka sepenuhnya. Konteks ini menggunakan kumpulan data Student-mat.xlsx, yang berisi informasi tentang waktu belajar, kehadiran, dan nilai siswa. Kumpulan data ini sangat relevan karena berisi variabel-variabel kunci yang mempengaruhi prestasi akademik siswa. Data yang lengkap tentang berbagai aspek pembelajaran memungkinkan analisis komprehensif untuk mengidentifikasi pola belajar siswa. Selain itu, data ini akurat dan dapat diandalkan, yang merupakan aspek penting dalam pengambilan keputusan berdasarkan data. Penggunaan algoritma ini telah terbukti efektif dalam berbagai penelitian sebelumnya dalam mengelompokkan data pendidikan dan membantu pendidik memahami dan meningkatkan hasil belajar siswa [8].

Penelitian ini menggunakan teknik clustering dan analisis korelasi untuk mengidentifikasi pola belajar dan absensi siswa. Penerapan data mining dengan menggunakan algoritma K-means clustering pada penilaian hasil belajar siswa sebagaimana dijelaskan dalam artikel Nirwana Hendrastuty (2024) menunjukkan bahwa algoritma K-means clustering efektif terhadap data hasil belajar siswa pola tersembunyi. Artikel ini juga menyoroti pentingnya penilaian hasil pembelajaran untuk memberikan wawasan berharga kepada guru dan tenaga kependidikan [7].

.

## 2. TINJAUAN PUSTAKA

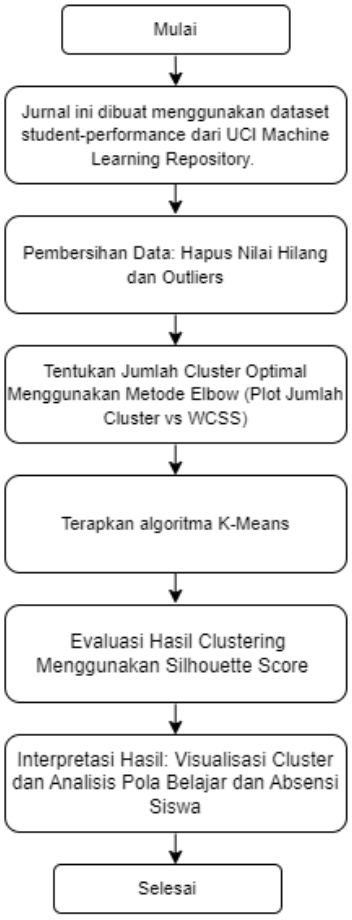
### 2.1. Data Mining dalam Pendidikan

Penambangan data melibatkan eksplorasi dan analisis data untuk menemukan pola, hubungan tersembunyi, dan informasi berguna dalam data dalam jumlah besar. Dalam dunia pendidikan, data mining digunakan untuk mengidentifikasi pola belajar siswa, menilai hasil belajar, dan memprediksi kinerja akademik siswa di masa depan. Teknik ini membantu pendidik dan pengambil kebijakan mengolah data siswa untuk menghasilkan wawasan yang bermanfaat [7].

### 2.2. Teknik Clustering

Clustering adalah teknik data mining yang bertujuan untuk mengelompokkan data ke dalam kelompok-kelompok berdasarkan kesamaan karakteristik [4]. Algoritma K-Means merupakan algoritma clustering yang populer karena kecepatan dan skalabilitasnya ketika mengelompokkan data dalam jumlah besar . Penelitian menunjukkan bahwa K-means clustering efektif dalam mengelompokkan siswa berdasarkan hasil belajar dan dapat digunakan untuk meningkatkan strategi pengajaran [9].

Berikut adalah flowchart yang menggambarkan langkah-langkah yang dilakukan dalam penerapan algoritma ini:




Gambar 1. *Flowchart* Algoritma Clustering

a. Kumpulkan Data Siswa dari Dataset student-performance: Data yang dikumpulkan meliputi waktu belajar (studytime), ketidakhadiran (absences), dan nilai ujian (G1, G2, G3). Data ini diambil dari dataset "student-performance" di UCI Machine Learning Repository.
b. Pembersihan Data: Data yang mengandung nilai hilang atau outliers dibersihkan untuk memastikan kualitas analisis.
c. Normalisasi Data: Variabel-variabel dinormalisasi menggunakan StandardScaler untuk memastikan semua variabel berada dalam skala yang sama.
d. Tentukan Jumlah Cluster Optimal: Menggunakan Metode Elbow untuk menentukan jumlah cluster yang optimal dengan memplot jumlah cluster terhadap WCSS (Within-Cluster Sum of Squares).
e. Terapkan Algoritma K-Means: Algoritma K-Means diterapkan dengan langkah-langkah inisialisasi centroid, pengelompokan data berdasarkan jarak ke centroid, update centroid, dan iterasi hingga mencapai konvergensi.
f. Evaluasi Hasil Clustering: Hasil clustering dievaluasi menggunakan Silhouette Score untuk menilai kualitas pengelompokan.
g. Interpretasi Hasil: Hasil clustering divisualisasikan, dan pola belajar serta absensi siswa dianalisis untuk memberikan wawasan yang bermanfaat bagi pendidik.

### 2.3. Analisis Korelasi

Analisis korelasi digunakan untuk memahami hubungan antar variabel dalam suatu dataset. Dalam lingkungan pendidikan, analisis korelasi membantu mengidentifikasi faktor-faktor yang paling mempengaruhi kinerja akademik siswa. Memahami hubungan antara waktu belajar, ketidakhadiran, dan kinerja akademik memungkinkan pendidik menentukan intervensi yang paling efektif untuk meningkatkan hasil belajar siswa [11].

### 2.4. Penerapan K-Means Clustering

Menerapkan pengelompokan K-means untuk menilai hasil belajar siswa dapat membantu Anda mengidentifikasi pola tersembunyi dalam data hasil belajar Anda. Metode ini mengelompokkan siswa ke dalam kelompok berdasarkan tingkat kinerja dan karakteristik pembelajarannya, sehingga memberikan wawasan berharga bagi guru dan pendidik. Hasil clustering memberikan informasi tentang tren kinerja dan membantu mengidentifikasi siswa yang membutuhkan dukungan tambahan [8].

### 2.5. Metode ELBOW

Metode Elbow merupakan metode yang digunakan untuk mengoptimalkan pemilihan centroid atau titik cluster dalam proses data mining. Tujuan utama dari metode ini adalah untuk memilih nilai cluster (k) yang kecil tetapi masih memiliki nilai within-cluster sum of squares (WCSS) yang rendah.

Dengan metode Elbow, titik optimal dapat diketahui ketika penurunan WCSS mulai melambat, membentuk siku pada grafik. Titik ini menunjukkan jumlah cluster yang optimal, di mana penambahan cluster lebih lanjut tidak lagi memberikan pengurangan signifikan dalam WCSS [1].

### 2.6. *Silhouette Coefficient*

Silhouette Coefficient adalah sebuah metode evaluasi yang digunakan untuk memvalidasi hasil clustering dengan menggabungkan dua metode, yaitu metode kohesi dan metode separasi [2]. Silhouette Coefficient mengukur seberapa mirip suatu objek

dengan cluster-nya sendiri (kohesi) dibandingkan dengan cluster terdekat lainnya (separasi). Adapun kriteria dalam pengukuran *Silhouette Coefficient* yang dapat dilihat pada Tabel 1.

Tabel 1. Kriteria Struktur

| Nilai *Silhouette Coefficient* | Kriteria |
|---|---|
| 0.71 – 1.00 | Struktur Kuat |
| 0.51 – 0.70 | Struktur Baik |
| 0.26 – 0.50 | Struktur Sedang |
| ≥ 0.25 | Struktur Buruk |

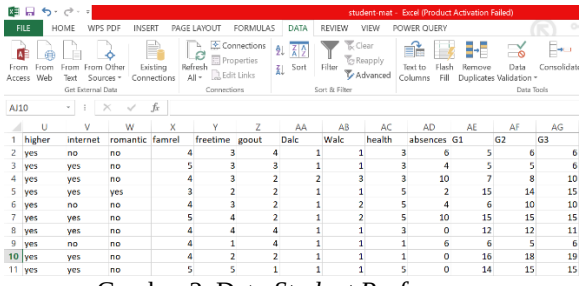
Metode ini sering digunakan karena kemampuannya untuk memberikan gambaran yang jelas tentang seberapa baik objek-objek dalam dataset dikelompokkan. Hasil evaluasi menggunakan Silhouette Coefficient dapat membantu dalam menilai dan membandingkan kualitas berbagai model clustering. Pada kasus penelitian ini, penggunaan Silhouette Coefficient menunjukkan hasil evaluasi yang membantu menentukan cluster yang optimal dengan struktur yang baik [1].

## 3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan teknik data mining untuk mengidentifikasi pola belajar dan absensi siswa. Teknik clustering menggunakan algoritma K-means dan analisis korelasi diterapkan pada kelompok siswa berdasarkan karakteristik belajarnya dan memahami hubungan antar variabel terkait. Desain penelitian ini bertujuan untuk memberikan wawasan yang berguna bagi para pendidik untuk meningkatkan strategi pembelajarannya.

### 3.1. Sumber Data

Data yang digunakan adalah dataset Student-mat.xlsx yang berisi informasi waktu belajar siswa, kehadiran, dan nilai, mencakup variabel seperti waktu belajar, absensi, G1 (hasil ujian pertama), G2 (hasil ujian kedua), dan G3 (hasil akhir).



| | U | V | W | X | Y | Z | AA | AB | AC | AD | AE | AF | AG |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | higher | internet | romantic | famrel | freetime | goout | Dalc | Walc | health | absences | G1 | G2 | G3 |
| 2 | yes | no | no | 4 | 3 | 4 | 1 | 1 | 3 | 6 | 5 | 6 | 6 |
| 3 | yes | yes | no | 5 | 3 | 3 | 1 | 1 | 3 | 4 | 5 | 5 | 6 |
| 4 | yes | yes | no | 4 | 3 | 2 | 2 | 3 | 3 | 10 | 7 | 8 | 10 |
| 5 | yes | yes | yes | 3 | 2 | 2 | 1 | 1 | 5 | 2 | 15 | 14 | 15 |
| 6 | yes | no | no | 4 | 3 | 2 | 1 | 2 | 5 | 4 | 6 | 10 | 10 |
| 7 | yes | yes | no | 5 | 4 | 2 | 1 | 2 | 5 | 10 | 15 | 15 | 15 |
| 8 | yes | yes | no | 4 | 4 | 4 | 1 | 1 | 3 | 0 | 12 | 12 | 11 |
| 9 | yes | no | no | 4 | 1 | 4 | 1 | 1 | 1 | 6 | 6 | 5 | 6 |
| 10 | yes | yes | no | 4 | 2 | 2 | 1 | 1 | 1 | 0 | 16 | 18 | 19 |
| 11 | yes | yes | no | 5 | 5 | 1 | 1 | 1 | 5 | 0 | 14 | 15 | 15 |

Gambar 2. Data *Student Perfomance*

### 3.2. Clustering dengan K-Means

Clustering dengan K-Means adalah metode untuk mengelompokkan data berdasarkan kesamaan karakteristiknya [1]. Algoritma ini membagi data ke dalam sejumlah cluster yang ditentukan sebelumnya, dengan tujuan meminimalkan variasi dalam cluster dan memaksimalkan variasi antar cluster. Dalam penelitian ini, variabel studytime, absences, G1, G2, dan G3 dipilih untuk clustering. Metode Elbow digunakan untuk menentukan jumlah cluster optimal, dan algoritma K-Means diterapkan dengan jumlah cluster optimal tersebut.

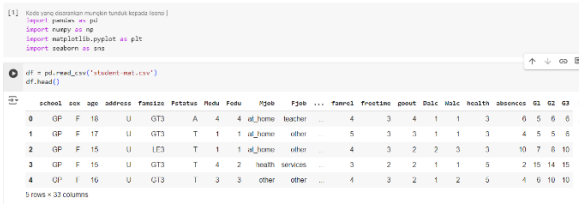
### 3.3. Analisis Korelasi

Analisis Korelasi adalah teknik statistik yang digunakan untuk mengukur dan menganalisis kekuatan serta arah hubungan antara dua variabel [4]. Dalam penelitian ini, koefisien korelasi Pearson digunakan untuk menghitung korelasi antara variabel, seperti waktu belajar, absensi, dan nilai ujian. Visualisasi hasil dilakukan dengan heatmap untuk memvisualisasikan matriks korelasi.

## 4. HASIL DAN PEMBAHASAN

### 4.1. Memuat dan Menampilkan Data

Pada tahap ini, dilakukan langkah-langkah awal dalam proses Exploratory Data Analysis (EDA) dengan memuat dataset dan menampilkan beberapa baris pertama dari data tersebut. Ini bertujuan untuk mendapatkan gambaran awal tentang struktur data, tipe data pada setiap kolom, dan beberapa nilai sampel dari dataset.



```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

df = pd.read_csv('student-mat.csv')
df.head()
```

| | school | sex | age | address | famsize | Pstatus | Medu | Fedu | Mjob | Fjob | ... | famrel | freetime | goout | Dalc | Walc | health | absences | G1 | G2 | G3 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | GP | F | 18 | U | GT3 | A | 4 | 4 | at_home | teacher | ... | 4 | 3 | 4 | 1 | 1 | 3 | 6 | 5 | 6 | 6 |
| 1 | GP | F | 17 | U | GT3 | T | 1 | 1 | at_home | other | ... | 5 | 3 | 3 | 1 | 1 | 3 | 4 | 5 | 5 | 6 |
| 2 | GP | F | 15 | U | LE3 | T | 1 | 1 | at_home | other | ... | 4 | 3 | 2 | 2 | 3 | 3 | 10 | 7 | 8 | 10 |
| 3 | GP | F | 15 | U | GT3 | T | 4 | 2 | health | services | ... | 3 | 2 | 2 | 1 | 1 | 5 | 2 | 15 | 14 | 15 |
| 4 | GP | F | 16 | U | GT3 | T | 3 | 3 | other | other | ... | 4 | 3 | 2 | 1 | 2 | 5 | 4 | 6 | 10 | 10 |

5 rows × 33 columns

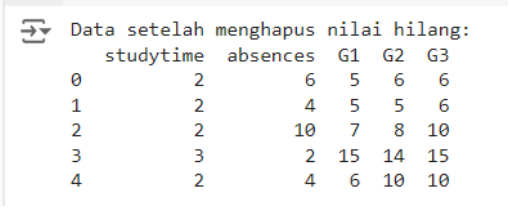
Gambar 3. Memuat Data

Dataset *student-mat.csv* dimuat ke dalam DataFrame df menggunakan *pd.read_csv()*. Data ini berisi informasi tentang siswa, termasuk waktu belajar, kehadiran, dan nilai mereka.

Menampilkan 5 baris pertama dari DataFrame df. Ini berguna untuk mendapatkan gambaran awal tentang struktur data, tipe data pada setiap kolom, dan beberapa nilai sampel dari dataset.

### 4.2. Data *Cleaning*

Proses pembersihan data dimulai dengan menghapus nilai-nilai yang hilang dari dataset student-mat.xlsx. Hal ini penting untuk memastikan integritas data dan keakuratan analisis selanjutnya. Setelah data dibersihkan, tidak ada lagi baris dengan informasi yang tidak lengkap yang dapat mempengaruhi hasil analisis.



```
Data setelah menghapus nilai hilang:
   studytime  absences  G1  G2  G3
0          2         6   5   6   6
1          2         4   5   5   6
2          2        10   7   8  10
3          3         2  15  14  15
4          2         4   6  10  10
```

Gambar 4. Hasil *Handling Missing Value*

Kolom yang ditampilkan dalam data ini meliputi studytime, absences, G1, G2, dan G3. Setiap baris dalam data merepresentasikan seorang siswa dengan informasi mengenai waktu belajar, jumlah ketidakhadiran, dan nilai ujian. Dengan menghapus baris-baris yang memiliki missing values, kita memastikan bahwa semua data yang kita analisis lengkap dan valid.

### 4.3. Normalisasi Data

Pada tahap ini, dilakukan normalisasi data untuk memastikan bahwa semua variabel memiliki skala yang sama. Normalisasi data penting untuk menghindari dominasi variabel tertentu dalam analisis, terutama ketika menggunakan algoritma yang sensitif terhadap skala data, seperti K-Means Clustering.

Normalisasi dilakukan dengan menggunakan metode StandardScaler dari library sklearn, yang mengubah setiap fitur sehingga memiliki rata-rata 0 dan standar deviasi 1.

Data Setelah Normalisasi

| | studytime | absences | G1 | G2 | G3 |
|---|---|---|---|---|---|
| 0 | 0.333333 | 0.080000 | 0.1250 | 0.315789 | 0.30 |
| 1 | 0.333333 | 0.053333 | 0.1250 | 0.263158 | 0.30 |
| 2 | 0.333333 | 0.133333 | 0.2500 | 0.421053 | 0.50 |
| 3 | 0.666667 | 0.026667 | 0.7500 | 0.736842 | 0.75 |
| 4 | 0.333333 | 0.053333 | 0.1875 | 0.526316 | 0.50 |

Gambar 5. Hasil Normalisasi

Setelah normalisasi, setiap fitur memiliki rata-rata 0 dan standar deviasi 1. Ini ditunjukkan oleh nilai-nilai yang tidak lagi berada dalam rentang awal tetapi telah diubah untuk mengikuti distribusi normal standar. Normalisasi membantu dalam menghilangkan bias yang mungkin timbul karena skala yang berbeda-beda pada setiap variabel, memastikan bahwa setiap variabel berkontribusi secara proporsional dalam analisis selanjutnya.

### 4.4. Pemilihan Variabel

Dalam analisis clustering, pemilihan variabel yang tepat sangat penting untuk memastikan bahwa hasil clustering memberikan wawasan yang bermakna dan relevan. Berikut adalah variabel-variabel yang dipilih dan penjelasan lebih mendetail mengenai alasan pemilihannya:

a. Studytime: Waktu yang dihabiskan siswa untuk belajar dalam seminggu. Variabel ini penting karena mencerminkan kebiasaan belajar siswa dan dedikasi waktu yang mereka alokasikan untuk pendidikan mereka. Waktu belajar yang lebih lama biasanya berkorelasi dengan pemahaman yang lebih baik dan kinerja akademik yang lebih tinggi.
b. Absences: Jumlah ketidakhadiran siswa. Kehadiran siswa di sekolah adalah indikator penting dari keterlibatan mereka dalam proses belajar mengajar. Ketidakhadiran yang tinggi dapat menunjukkan masalah dalam keterlibatan atau kesulitan lain yang dapat mempengaruhi kinerja akademik
c. G1: Nilai ujian pertama. Nilai ini memberikan gambaran awal tentang kemampuan akademik siswa pada tahap awal semester. Ini adalah indikator awal yang baik untuk menilai pemahaman siswa terhadap materi yang diajarkan.
d. G2: Nilai ujian kedua. Nilai ini menunjukkan perkembangan siswa setelah beberapa waktu belajar dan memberikan indikasi tentang efektivitas strategi belajar mereka setelah ujian pertama.
e. G3: Nilai akhir siswa. Ini adalah penilaian akhir dari kinerja akademik siswa selama satu semester. Nilai ini sangat penting karena mencerminkan pemahaman kumulatif siswa terhadap seluruh materi yang diajarkan.

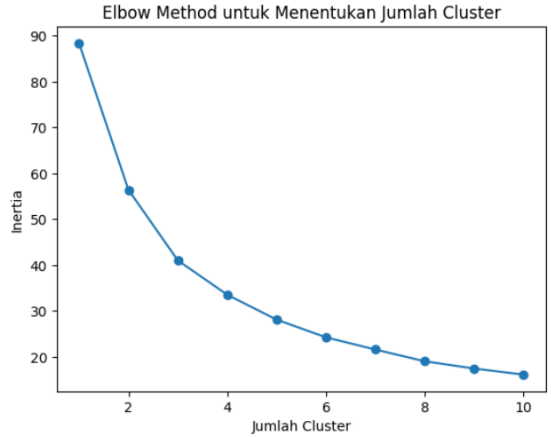
Variabel-variabel tersebut dipilih karena relevansi akademiknya yang langsung berkaitan dengan aspek penting dari pengalaman belajar siswa, yaitu waktu belajar, kehadiran, dan kinerja akademik. Studytime dan absences membantu mengidentifikasi kebiasaan belajar dan kehadiran siswa, yang penting untuk mengetahui siapa yang mungkin membutuhkan intervensi tambahan. Variabel G1, G2, dan G3 memberikan gambaran komprehensif tentang kinerja akademik siswa dari awal hingga akhir semester, memungkinkan kita melacak perkembangan dan memahami pengaruh kebiasaan belajar terhadap hasil akhir.

Dengan menggunakan variabel-variabel ini dalam analisis clustering, kita dapat mengelompokkan siswa berdasarkan kinerja mereka, membantu merancang strategi pembelajaran yang lebih efektif dan disesuaikan. Pemahaman terhadap pola belajar dan ketidakhadiran memungkinkan pendidik mengidentifikasi siswa yang membutuhkan dukungan tambahan. Selain itu, melacak perubahan dalam variabel-variabel ini memungkinkan evaluasi efektivitas intervensi pendidikan dan penyesuaian yang diperlukan, sehingga meningkatkan proses pembelajaran dan kinerja akademik siswa secara keseluruhan.

### 4.5. Penentuan Jumlah Cluster Optimal (Metode Elbow)

Dalam analisis clustering, menentukan jumlah cluster yang optimal sangat penting untuk mendapatkan hasil yang bermakna dan dapat diinterpretasikan. Metode Elbow adalah salah satu teknik yang umum digunakan untuk tujuan ini.

Metode Elbow melibatkan plotting jumlah cluster (k) melawan Within-Cluster Sum of Squares (WCSS). WCSS mengukur total varian dalam cluster, yaitu seberapa dekat data dalam satu cluster dengan centroid-nya. Tujuannya adalah untuk menemukan titik di mana penambahan cluster tambahan tidak lagi memberikan penurunan signifikan dalam WCSS, yang diindikasikan oleh bentuk "siku" atau "elbow" pada grafik.




Gambar 6. Hasil Metode *Elbow*

Dari grafik ini, kita dapat melihat bahwa inertia menurun tajam ketika jumlah cluster meningkat dari 1 hingga 3. Ini menunjukkan bahwa menambah cluster dalam rentang ini secara signifikan mengurangi varian dalam cluster.

Dengan menggunakan tiga cluster, kita dapat mengelompokkan data siswa secara efektif dan efisien berdasarkan pola belajar dan kehadiran mereka. Menentukan jumlah cluster yang optimal penting untuk menghindari overfitting (terlalu banyak cluster) dan underfitting (terlalu sedikit cluster), serta memastikan bahwa hasil clustering memberikan wawasan yang bermakna dan dapat digunakan untuk pengambilan keputusan yang informatif.

### 4.6. Penerapan Algoritma K-Means Clustering

Setelah menentukan jumlah cluster yang optimal, pelatihan model dilakukan dengan menggunakan algoritma K-Means. Dataset sekarang memiliki kolom tambahan bernama Cluster yang menunjukkan label cluster yang dihasilkan oleh algoritma K-Means.

| | school | sex | age | address | famsize | Pstatus | Medu | Fedu | Mjob | Fjob | ... |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | GP | F | 18 | U | GT3 | A | 4 | 4 | at_home | teacher | ... |
| 1 | GP | F | 17 | U | GT3 | T | 1 | 1 | at_home | other | ... |
| 2 | GP | F | 15 | U | LE3 | T | 1 | 1 | at_home | other | ... |
| 3 | GP | F | 15 | U | GT3 | T | 4 | 2 | health | services | ... |
| 4 | GP | F | 16 | U | GT3 | T | 3 | 3 | other | other | ... |

| | freetime | goout | Dalc | Walc | health | absences | G1 | G2 | G3 | Cluster |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 3 | 4 | 1 | 1 | 3 | 6 | 5 | 6 | 6 | 2 |
| 1 | 3 | 3 | 1 | 1 | 3 | 4 | 5 | 5 | 6 | 2 |
| 2 | 3 | 2 | 2 | 3 | 3 | 10 | 7 | 8 | 10 | 1 |
| 3 | 2 | 2 | 1 | 1 | 5 | 2 | 15 | 14 | 15 | 0 |
| 4 | 3 | 2 | 1 | 2 | 5 | 4 | 6 | 10 | 10 | 1 |

[5 rows x 34 columns]

Gambar 7. Hasil Pelatihan Model

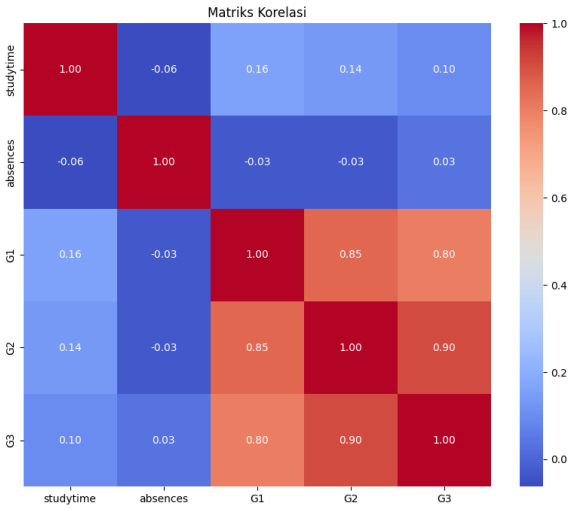
Gambar di atas menunjukkan dataset setelah pelatihan model K-Means, di mana kolom Cluster telah ditambahkan. Kolom ini menunjukkan label cluster untuk setiap siswa berdasarkan pola belajar dan absensi mereka.

a. Cluster 0: Siswa yang memiliki karakteristik tertentu berdasarkan pola belajar dan absensi mereka. Misalnya, siswa dalam cluster ini mungkin memiliki jumlah ketidakhadiran rendah dan waktu belajar yang konsisten.
b. Cluster 1: Siswa yang memiliki karakteristik berbeda dari Cluster 0 dan 2. Siswa dalam cluster ini mungkin menunjukkan pola belajar yang bervariasi dan ketidakhadiran yang moderat.
c. Cluster 2: Siswa yang memiliki karakteristik berbeda dari Cluster 0 dan 1. Siswa dalam cluster ini mungkin memiliki jumlah ketidakhadiran yang tinggi dan waktu belajar yang kurang.

Dengan adanya label cluster ini, kita dapat dengan mudah mengelompokkan siswa berdasarkan karakteristik mereka dan merancang strategi pembelajaran yang disesuaikan dengan kebutuhan setiap kelompok. Hasil clustering ini memberikan informasi yang berharga bagi pendidik untuk mengidentifikasi siswa yang membutuhkan perhatian lebih dan mendukung siswa dalam mencapai potensi akademik mereka secara maksimal.

### 4.7. Analisis Korelasi

Gambar di bawah menunjukkan matriks korelasi antara variabel-variabel yang dipilih untuk analisis clustering, yaitu studytime, absences, G1, G2, dan G3. Matriks korelasi ini membantu kita memahami hubungan antara variabel-variabel tersebut.




Gambar 8. Analisis Korelasi

a. Korelasi antara *Studytime* dan Variabel lain: Korelasi antara studytime dan G1 adalah sebesar 0.16, menunjukkan bahwa semakin banyak waktu yang dihabiskan siswa untuk belajar, semakin tinggi nilai ujian pertama mereka, meskipun korelasi ini lemah. Korelasi antara studytime dan G2 adalah sebesar 0.14, juga menunjukkan korelasi lemah namun positif antara waktu belajar dan nilai ujian kedua. Korelasi antara studytime dan G3 adalah sebesar 0.10, menunjukkan bahwa waktu belajar juga sedikit berkorelasi positif dengan nilai akhir siswa.
b. Korelasi antara *Absences* dan Variabel lain: Korelasi antara *absences* dan G1, G2, G3 adalah negatif yang sangat lemah, menunjukkan bahwa ketidakhadiran tidak memiliki pengaruh signifikan terhadap nilai ujian pertama, kedua, atau nilai

akhir. Hal ini menunjukkan bahwa faktor ketidakhadiran tidak berdampak besar pada kinerja akademik siswa dalam dataset ini.

c. Korelasi antara Nilai Ujian (G1, G2, G3): Korelasi antara G1 dan G2 adalah sebesar 0.85, menunjukkan bahwa nilai ujian pertama berkorelasi sangat tinggi dengan nilai ujian kedua. Siswa yang mendapatkan nilai tinggi pada ujian pertama cenderung mendapatkan nilai tinggi juga pada ujian kedua. Korelasi antara G1 dan G3 adalah sebesar 0.80, menunjukkan bahwa nilai ujian pertama berkorelasi tinggi dengan nilai akhir. Korelasi antara G2 dan G3 adalah sangat kuat sebesar 0.90, menunjukkan bahwa nilai ujian kedua sangat berkorelasi dengan nilai akhir. Ini menunjukkan konsistensi dalam kinerja akademik siswa di berbagai ujian.

### 4.8. Validasi Hasil Clustering

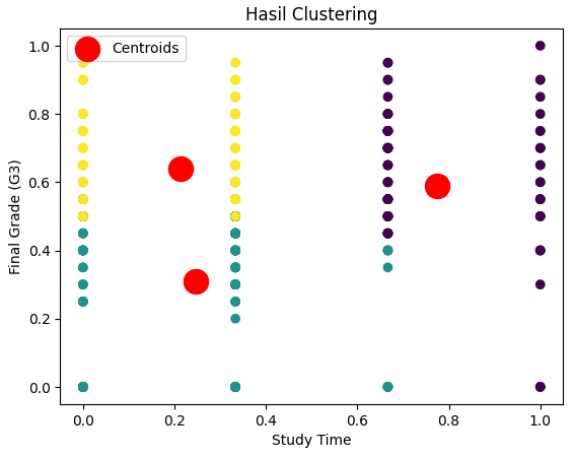
Validasi hasil clustering adalah langkah penting untuk mengevaluasi kualitas dan efektivitas dari model clustering yang telah diterapkan. Pada kasus ini, kita menggunakan dua metode untuk validasi: Silhouette Score dan Visualisasi Cluster.

```
Silhouette Score: 0.33543637221195255
```

Gambar 7. Hasil *Silhoutte Score*

Dari hasil di atas, Silhouette Score yang diperoleh adalah 0.3354. Ini menunjukkan bahwa kualitas clustering yang dilakukan adalah sedang. Nilai ini lebih besar dari 0, artinya titik data dalam suatu cluster biasanya lebih dekat dengan titik data lain di cluster yang sama dibandingkan dengan titik data di cluster lain. Namun, nilai ini mendekati 0.3, yang menunjukkan bahwa masih ada tumpang tindih yang signifikan antar cluster dan pemisahan antar cluster masih dapat ditingkatkan.

Visualisasi cluster dilakukan dengan menggunakan plot yang menunjukkan distribusi data dan posisi centroid dari setiap cluster.




Gambar 9. Hasil Validasi Clustering

Dari visualisasi ini, kita dapat melihat bagaimana data siswa dikelompokkan berdasarkan waktu belajar dan nilai akhir mereka. Setiap warna mewakili cluster yang berbeda. Visualisasi ini membantu untuk memahami struktur dan distribusi data dalam cluster yang terbentuk.

Distribusi titik-titik data menunjukkan bahwa beberapa cluster memiliki tumpang tindih, yang sesuai dengan indikasi dari Silhouette Score.

### 4.9. Interpretasi dan Implikasi Hasil Validasi

Hasil dari validasi ini menunjukkan bahwa model K-Means telah memberikan hasil yang moderat dalam mengelompokkan siswa berdasarkan waktu belajar dan nilai ujian mereka. Namun, adanya tumpang tindih antar cluster mengindikasikan perlunya pemilihan variabel tambahan atau mungkin mempertimbangkan algoritma clustering lain yang bisa memberikan pemisahan cluster yang lebih baik.

## 5. KESIMPULAN DAN SARAN

Penelitian ini berhasil mengidentifikasi pola belajar dan absensi siswa menggunakan teknik clustering dengan algoritma K-Means, yang menghasilkan tiga cluster optimal berdasarkan waktu belajar, kehadiran, dan nilai ujian siswa. Meskipun kualitas clustering dinilai sedang dengan Silhouette Score sebesar 0.3354, hasil ini tetap memberikan wawasan yang berharga mengenai distribusi karakteristik siswa, meskipun terdapat tumpang tindih antar cluster. Analisis korelasi menunjukkan hubungan lemah antara waktu belajar dan nilai ujian, namun korelasi yang sangat kuat antara nilai ujian pertama, kedua, dan nilai akhir siswa.

Hasil ini menunjukkan bahwa performa akademik yang baik cenderung konsisten sepanjang semester. Validasi model melalui visualisasi dan Silhouette Score mengindikasikan bahwa meskipun model berhasil mengelompokkan siswa berdasarkan pola belajar dan absensi mereka, masih terdapat ruang untuk perbaikan, seperti dengan menggunakan algoritma lain atau menambahkan variabel baru. Temuan ini memberikan implikasi penting bagi pendidik dalam merancang strategi pembelajaran yang lebih tepat sasaran, yang dapat disesuaikan dengan kebutuhan setiap kelompok siswa, untuk meningkatkan kinerja akademik mereka secara keseluruhan.

## DAFTAR PUSTAKA

[1] T. Nugroho, U. Enri and I. Maulana, "CLUSTERING MENU MAKANAN BERDASARKAN KEBUTUHAN KALORI HARIAN MENGGUNAKAN ALGORITME K-MEANS," *JATI,* Vols. 8 , no. 4, pp. 4412 - 4413, 2024

[2] A. Maulana, R. D. Dana and N. D. Nuris, "IMPLEMENTASI ALGORITMA K-MEANS CLUSTERING DALAM PENGELOMPOKAN DATA KERUSAKAN RUMAH AKIBAT

BENCANA ALAM DI KABUPATEN CIREBON," *JATI*, Vols. 8, no. 2, p. 1419, 2024

[3] R. A. Sasmoyo, "20245635SISTEM CLUSTERING DAERAH RAWAN KEMATIAN ANAK DI BAWAH UMUR DI WILAYAH JAWA BARAT MENGGUNAKAN ALGORITMA K-MEANS," *JATI*, Vols. 8, no. 4, p. 5653, 2024

[4] D. Apriandi, R. M. Sari and M. I. Sarif, "Analisis Clustering Untuk Menentukan Siswa Berprestasi di SMK Swasta TI Panca Dharma Stungkit Menggunakan Metode K-Means," *Jurnal MINFO POLGAN*, Vols. 13, no. 1, p. 1119, 2024.

[5] Mira, A. C. Nurcahyo, C. Gudiato and Kusnanto, "Implementasi Data Mining Metode K-Means Menggunakan Framework Python Dalam Mengelompokkan Pegawai Berdasarkan Data Presensi," *Jurnal Media Informatika Budidarma*, Vols. 8, no. 3, p. 1335, 2024

[6] S. N. Safitri, H. Setiadi and E. Suryani, "Educational Data Mining Using Cluster Analysis Methods and Decision Trees based on Log Mining," *Jurnal RESTI*, Vols. 6, no. 3, pp. 448 - 449, 2022

[7] N. Hendrastuty, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa," *JIMA-ILKOM*, Vols. 3, no. 1, 2024

[8] S. Anwar, T. Suprapti, G. Dwilestari and I. Ali, "PENGELOMPOKKAN HASIL BELAJAR SISWA DENGAN METODE CLUSTERING K-MEANS," *JURSITEKNI*, Vols. 4, no. 2, pp. 61 - 62, 2022

[9] H. Alexander, Y. Umaidah and M. Jajuli, "MPLEMENTASI CLUSTERING UNTUK MENENTUKAN EFEKTIVITAS NILAI SISWA SESUDAH PANDEMI COVID-19 MENGGUNAKAN ALGORITMA K-MEANS," *JATI*, Vols. 7, no. 3, p. 1495, 2023

[10] N. Yannuansa, M. S. Udin and I. M. Aziz, "Pemanfaatan Algoritma K-Means Clustering dalam Mengolah Pengaruh Pendapatan Orang Tua Terhadap Hasil Belajar Pada Mata Pelajaran Produktif," *TECNOSCIENZA*, Vols. 6, no. 1, pp. 48 - 49, 2021

[11] M. S. Sungkar and M. T. Qurohman, "Penerapan Algoritma C5.0 Untuk Prediksi Kelulusan Pembelajaran Mahasiswa Pada Matakuliah Arsitektur Sistem Komputer," *Media Informatika Budidarma*, Vols. 5, no 3, p. 1166, 2021