

Penerapan Teknologi *Machine Learning* Untuk Otomatisasi Pengelompokan Siswa Berdasarkan Kemampuan Akademik Di SMK Negeri 8 Kab Tangerang Dengan Metode *Principal Component Analysis* Dan Metode *K-means Clustering*

Muhamad Fahmi Idris^{1*}, Wasis Haryono²

¹Ilmu Komputer, Teknik Informatika, Universitas Pamulang, Tangerang Selatan, Indonesia

E-mail : ¹fahmiidris356@gmail.com, ²wasis@unpam.ac.id

(* : coresponding author)

Abstrak - Di era digital, pemanfaatan teknologi dalam dunia pendidikan menjadi kebutuhan penting, termasuk dalam pengelolaan data akademik siswa. Penelitian ini mengembangkan sistem otomatisasi pengelompokan siswa berdasarkan kemampuan akademik menggunakan metode *Principal Component Analysis (PCA)* dan *K-Means Clustering* di SMK Negeri 8 Kabupaten Tangerang. Pengelompokan konvensional yang dilakukan guru cenderung memakan waktu lama dan rentan terhadap subjektivitas, sementara sistem yang dikembangkan mampu meningkatkan efisiensi waktu hingga lebih dari 80% dengan hasil yang lebih objektif dan konsisten. PCA digunakan untuk mereduksi dimensi data akademik seperti kehadiran, nilai tugas, UTS, dan UAS, sehingga mempercepat proses clustering tanpa kehilangan informasi penting. Selanjutnya, metode *K-Means Clustering* membagi siswa ke dalam kelompok kemampuan tinggi, sedang, dan rendah berdasarkan nilai akademik. Sistem ini dibangun dengan pendekatan *Waterfall*, melalui tahapan analisis kebutuhan, desain, implementasi, pengujian, hingga pemeliharaan. Hasil pengujian menunjukkan bahwa aplikasi yang dikembangkan mudah digunakan dan memberikan output yang relevan untuk mendukung pengambilan keputusan strategis dalam pembelajaran. Penelitian ini memberikan kontribusi dalam meningkatkan kualitas proses belajar mengajar melalui integrasi teknologi *machine learning* dalam sistem akademik sekolah.

Kata Kunci : *Machine Learning*, *Principal Component Analysis (PCA)*, *K-Means Clustering*, Pengelompokan Siswa, Prestasi Akademik, Sistem Otomatisasi, SMK Negeri 8 Kabupaten Tangerang, Aplikasi Berbasis Web, Reduksi Dimensi Data.

Abstract - In the digital era, the integration of technology in education has become essential, particularly in managing student academic data. This study developed an automated student grouping system based on academic performance using *Principal Component Analysis (PCA)* and *K-Means Clustering* at SMK Negeri 8 Kabupaten Tangerang. Traditional grouping methods conducted by teachers are time-consuming and prone to subjectivity, whereas the proposed system improves time efficiency by over 80% with more objective and consistent results. PCA is employed to reduce the dimensionality of academic data such as attendance, assignment scores, midterm exams, and final exams, allowing for faster clustering without significant information loss. The *K-Means Clustering* algorithm then classifies students into high, medium, and low academic performance groups. The system was developed using the *Waterfall* model, involving stages of requirement analysis, system design, implementation, testing, and maintenance. Final testing shows that the application is user-friendly and provides valuable outputs to support academic decision-making. This research contributes to enhancing the teaching and learning process by integrating machine learning technology into the school's academic system.

Keywords: *Machine Learning*, *Principal Component Analysis (PCA)*, *K-Means Clustering*, Student Grouping, Academic Performance, Automation System, SMK Negeri 8 Kabupaten Tangerang, Web-Based Application, Data Dimensionality Reduction.

1. PENDAHULUAN

Di era digital saat ini, analisis data telah menjadi elemen penting dalam berbagai bidang, termasuk pendidikan. Teknologi modern memungkinkan pengelolaan dan analisis data akademik siswa di Smk Negeri 8 Kab tangerang yang lebih efektif untuk meningkatkan kualitas pembelajaran. Pengelompokan siswa secara konvensional sangat tidak efisien karena memerlukan waktu, dan tenaga, namun hasil yang diperoleh belum tentu optimal. Proses ini biasanya dilakukan oleh guru dengan menilai dan mengelompokkan siswa berdasarkan pengamatan terhadap nilai akademik secara individu, selain memakan waktu yang lama, metode ini juga kurang konsisten karena bergantung pada persepsi masing-masing guru. ketidakefisienan ini juga dapat menghambat

pengambilan keputusan yang cepat dan akurat dalam merancang strategi pembelajaran yang sesuai, karena tidak adanya dukungan sistem yang mampu menganalisis data secara otomatis. Salah satu aplikasi dari teknologi ini adalah otomatisasi pengelompokan siswa berbasis web, yang bertujuan untuk membantu pendidik memahami pola belajar siswa dan mengambil tindakan proaktif untuk meningkatkan hasil belajar mereka. Faktor-faktor seperti kehadiran, nilai tugas, nilai ujian tengah semester (UTS), dan nilai ujian akhir semester (UAS) dikenal sebagai indikator utama mempengaruhi prestasi akademik siswa. (Rismaya, Yuniarto, and Setiadi 2025)

Sejumlah penelitian sebelumnya telah mengkaji penerapan metode system otomatisasi pengelompokan siswa berbasis web. Misalnya, penelitian oleh (No et al. 2024) menunjukkan bahwa metode *elbow* dapat dikelompokkan dengan penerapan clustering dalam pengelompokan prestasi siswa dengan sangat baik. Selain itu, penelitian oleh (Fadlisyah 2021) dengan menggunakan Metode *Naive Bayes* juga menunjukkan hasil yang positif dalam pengelompokan siswa. Namun, belum banyak penelitian yang mengintegrasikan metode *principal component analysis* dan *k-means clustering*, dalam konteks pengelompokan siswa di sekolah, sehingga masih ada ruang untuk eksplorasi lebih lanjut.

Berdasarkan permasalahan yang telah diidentifikasi serta tinjauan literatur yang relevan, penelitian ini bertujuan untuk mengembangkan sistem otomatisasi pengelompokan siswa berdasarkan kemampuan akademik Metode *Principal Component Analysis (PCA)* digunakan untuk mereduksi dimensi data, sedangkan *K-Means Clustering* diterapkan untuk mengelompokkan siswa ke dalam kelompok-kelompok yang sesuai dari nilai yang diperoleh dari nilai terbesar hingga terkecil. Sistem ini diharapkan dapat membantu pihak sekolah dalam pengambilan keputusan yang lebih tepat, objektif, dan efisien. terkait strategi pembelajaran yang sesuai dengan karakteristik akademik siswa. Dengan demikian, penelitian ini memberikan kontribusi dalam meningkatkan efektivitas proses belajar mengajar melalui pemanfaatan teknologi *machine learning*. (Rosyada and Utari 2024)

2. METODOLOGI PENELITIAN

2.1 Pengertian Machine Learning

Menurut (Chyan, Arni, and Thayf n.d.), *machine learning* merupakan bagian dari kecerdasan buatan yang menginstruksikan komputer untuk mempelajari informasi dari data dan mengeluarkan prediksi atau tindakan berdasarkan interpretasi terhadap data tersebut. dengan machine learning, komputer dapat mengambil keputusan atau melakukan tindakan dari analisis data yang diperoleh.

2.2 Metode Principal Component Analysis

Menurut (Diba, Lydia, and Sihombing 2023) *Principal Component Analysis* adalah sebuah metode matematis yang mentransformasikan data ke dalam ranah baru dengan menghasilkan sejumlah principal component yang lebih penting. Dalam dunia Machine Learning dan Statistika, PCA sering digunakan untuk mengatasi masalah dimensionality curse, yaitu ketika data memiliki terlalu banyak fitur sehingga sulit dianalisis secara efektif. Dengan PCA, data yang awalnya memiliki banyak variabel dapat direpresentasikan dalam jumlah yang lebih sedikit tetapi tetap mempertahankan informasi utama.

Berikut adalah langkah-langkah dalam Metode PCA :

1. Standarisasi Data

Data harus distandarisasi agar setiap fitur memiliki skala yang sama.

Biasanya digunakan metode Z-score normalization:

$$X' = \frac{x - \mu}{\sigma}$$

Di mana:

X' = data setelah normalisasi

X = nilai asli

μ = rata-rata fitur

σ = standar deviasi fitur

Standarisasi ini penting karena PCA berbasis pada variansi, sehingga fitur dengan skala lebih besar tidak mendominasi analisis.

2. Membentuk Matriks Kovarians

Matriks kovarians digunakan untuk mengukur hubungan antar fitur dalam dataset.

Persamaan matriks kovarians:

$$COV(X, Y) = \frac{1}{n-1} \sum (X_i - uX)(Y_i - uY)$$

Jika nilai kovarians tinggi, berarti kedua fitur memiliki korelasi kuat.

PCA akan mengidentifikasi fitur yang berkorelasi tinggi dan menggabungkannya ke dalam komponen utama.

3. Menghitung Eigenvalues dan Eigenvectors

Dari matriks kovarians, hitung eigenvalues (nilai eigen) dan eigenvectors (vektor eigen).

Eigenvalues menunjukkan seberapa banyak informasi (variansi) yang dapat dijelaskan oleh setiap komponen utama.

Eigenvectors menentukan arah dari komponen utama dalam ruang data.

Persamaan dasar untuk mencari eigenvalues dan eigenvectors:

$$Av = \lambda v$$

Di mana:

A = matriks kovarians

v = eigenvector

λ = eigenvalue

4. Memilih *Principal Components*

Urutkan *eigenvalues* dari yang terbesar ke yang terkecil.

Pilih sejumlah k komponen utama yang menjelaskan persentase variansi terbesar.

Gunakan aturan cumulative explained variance untuk menentukan jumlah komponen:

$$\frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^n \lambda_i} \times 100\%$$

Jika ingin mempertahankan 95% informasi, maka pilih komponen hingga persentase variansi kumulatif mencapai $\geq 95\%$.

5. Membentuk Ruang Dimensi Baru

Transformasikan dataset awal menggunakan *principal components* yang telah dipilih:

$$X' = X \times W$$

X' = dataset baru dengan dimensi lebih kecil

X = dataset asli

W = matriks eigenvectors yang dipilih

Dataset baru ini memiliki lebih sedikit fitur tetapi tetap mewakili pola utama dalam data.

6. Menggunakan Data yang Telah Direduksi

Dataset hasil PCA dapat digunakan untuk:

Visualisasi data dalam 2D atau 3D.

Input dalam model Machine Learning untuk meningkatkan efisiensi dan mengurangi overfitting.

Analisis klustering (misalnya K-Means Clustering) dengan fitur yang lebih sederhana.

2.3 Metode K-means Clustering

Menurut (Informasi, Teknologi, and Mulia 2025) Metode *K-Means* adalah algoritma *clustering* yang mengelompokkan data ke dalam sejumlah *k cluster* tertentu tanpa memerlukan label atau target variabel. digunakan untuk mengelompokkan data berdasarkan kesamaan karakteristik. Algoritma ini bekerja dengan cara membagi dataset ke dalam sejumlah kelompok atau klaster berdasarkan jarak antar data. Pengguna harus menentukan jumlah klaster yang diinginkan, yang disebut sebagai nilai *K*. Algoritma kemudian secara iteratif menyesuaikan posisi pusat klaster (*centroid*) hingga mencapai kondisi optimal di mana data dalam satu kelompok memiliki kemiripan yang tinggi dan berbeda dengan kelompok lainnya.

Gunakan rumus jarak Euclidean untuk menghitung jarak setiap data ke masing-masing centroid:

$$d(X_i, C_j) = \sqrt{\sum (X_i - C_j)^2}$$

Di mana:

- $d(X_i, C_j)$ = jarak antara data x_i dan *centroid* c_j
- X_i = titik data
- C_j = *centroid*
- \sum = penjumlahan untuk semua fitur dalam data

Setelah semua data dikelompokkan ke klaster, hitung kembali centroid baru berdasarkan rata-rata dari semua titik dalam klaster:

$$C_j = \frac{1}{n} \sum_{i=1}^n X_i$$

Di mana:

- C_j = *centroid* baru dari klaster j
- n = jumlah data dalam klaster
- X_i = titik data dalam klaster

2.4 Metode Waterfall

Metode *Waterfall Model* SDLC air terjun (*Waterfall*) sering juga di sebut model sekuensial linier (*sequential linier*) atau alur hidup klasik (*classic life cycle*), model air terjun menyediakan pendekatan alur hidup perangkat lunak secara sekuensial atau terurut dimulai dari analisis, desain, pengkodean, pengujian dan tahapan pendukung (*support*), berikut ini model air terjun. (Wasis Haryono 2024).

3. ANALISA DAN PEMBAHASAN

3.1 Analisa Sistem Berjalan

Karena nilai masih masih ditangani secara manual menggunakan teknik konvensional, nilai masih tetap terkendali dalam kendali. Siswa diajar diajar dengan tangan, atau diawasi dengan tangan, atau dengan menggunakan metode tradisional, karena perlu memiliki guru yang cocok untuk membimbing mereka. (Mesak and Wasis Haryono 2025) Dalam tahap ini proses untuk memahami bagaimana sistem yang sedang digunakan bekerja dalam konteks sebuah organisasi atau proyek. Dalam konteks judul di atas, "Penerapan Teknologi *Machine Learning* untuk Otomatisasi Pengelompokan Siswa Berdasarkan Kemampuan Akademik di SMK Negeri 8 Kab Tangerang dengan Dengan *Metode Principal Component Analysis* Dan *Metode K-means Clustering*. analisis sistem berjalan meliputi langkah-langkah untuk memahami proses manual atau semi-otomatis yang digunakan sekolah untuk mengelompokkan siswa berdasarkan kemampuan akademik mereka sebelum adanya penerapan *machine learning*.

3.2 Analisa Sistem Usulan

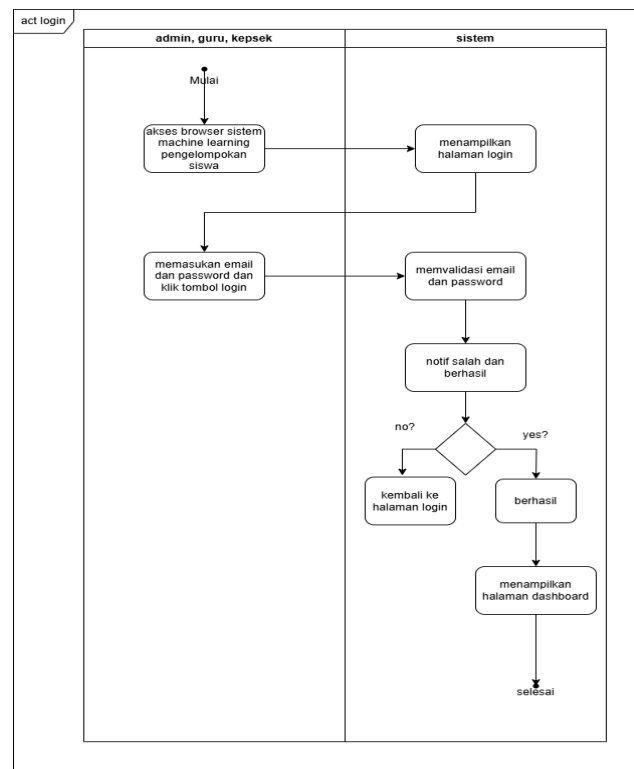
Activity diagram sistem usulan adalah jenis diagram yang digunakan untuk memodelkan alur kerja atau aktivitas dalam sistem yang diusulkan atau sistem baru yang dirancang untuk menggantikan atau memperbaiki sistem yang sedang berjalan. (Putra et al. and Wasis Haryono. 2024) Sistem yang diusulkan menggunakan teknologi *machine learning*, yaitu metode *Principal Component Analysis* dan *metode K-means Clustering* untuk mengotomatisasi pengelompokan siswa berdasarkan kemampuan akademik di SMK Negeri 8 Kab Tangerang. Tujuannya adalah untuk meningkatkan efisiensi dan akurasi pengelompokan, menghilangkan subjektivitas, dan membantu guru dalam merancang strategi pembelajaran yang lebih tepat sasaran. Sistem ini bekerja melalui proses pengumpulan dan *preprocessing* data, reduksi dimensi menggunakan PCA, pengelompokan dengan K-Means, evaluasi hasil *clustering*, dan penyajian laporan pengelompokan yang mendukung keputusan pengajaran yang lebih efektif

3.3 Perancangan Sistem

Perancangan Sistem adalah kumpulan dari elemen-elemen yang saling berhubungan yang tidak dapat dipisahkan untuk mencapai suatu tujuan tertentu. Secara sederhana, sistem dapat diartikan sebagai sekumpulan komponen atau variabel yang tersusun, terintegrasi, dan saling bergantung. Suatu sistem terdiri dari bagian-bagian atau komponen-komponen yang terintegrasi untuk satu tujuan. (Ubaydillah, Fuji, Muhamad Mahmud, Sinta Rahmawati, and Wasis Haryono. 2023)

a. Activity Diagram

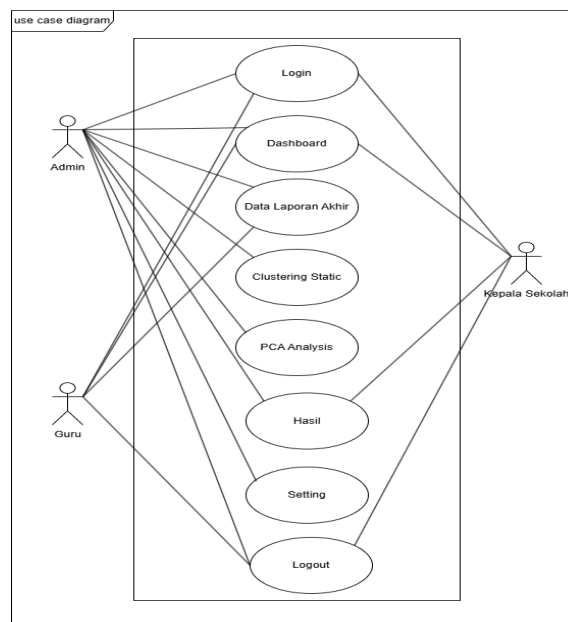
Diagram aktivitas adalah jenis diagram dalam UML yang memungkinkan untuk memodelkan proses yang terjadi dalam diagram sistem. Diagram aktivitas adalah representasi visual dari alur kerja yang berisi aktivitas dan tindakan, dan juga dapat berisi opsi, loop, dan konkurensi. Dalam UML, diagram aktivitas dibuat untuk menggambarkan aktivitas komputer dan aliran aktivitas ke suatu organisasi. (Ubaydillah, ahmzud, Sinta Rahmawati, and Wasis Haryono. 2023.) berikut adalah activity diagram login :



Gambar 1. Activity Diagram

b. Use Case Diagram

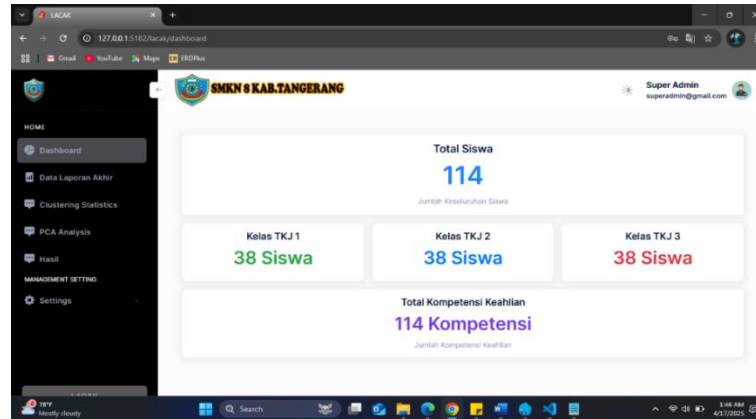
Use case adalah sekumpulan rangkaian atau deskripsi yang saling berhubungan dan tersusun secara sistematis, dilaksanakan atau diawasi oleh seorang aktor. *Use case* berfungsi untuk menggambarkan perilaku objek dalam suatu model dan diimplementasikan melalui sebuah kolaborasi. (Sachrul Sidiq et al. and Wasis Haryono 2024), berikut adalah *use case* diagram penerapan machine learning :



Gambar 2. Use Case Diagram

4. IMPLEMENTASI

4.1 Implementasi Aplikasi



Gambar 3. *Dashboard Admin*

Halaman login ini antarmuka sederhana dengan dua pilihan role yaitu *Admin/Regular User* yang dapat dipilih menggunakan *checkbox*, dilengkapi kolom input untuk Email (superadmin@gmail.com) dan Password (tersembunyi dengan asterisk), serta opsi *Reply Password* untuk pemulihan kata sandi, sementara bahasa antarmuka *default* ditampilkan dalam english, dirancang untuk memudahkan proses autentikasi pengguna dengan tampilan yang jelas dan fungsional.

4.2 Hasil Pengujian Metode PCA

a. Data Siswa (TKJ 1)

Tabel 1. Perhitungan Metode PCA(Data Siswa)

Nama Siswa	Nilai Agama (X_1)	Nilai PPKN (X_2)	Nilai B.indo (X_3)
AMELIAH AGUSTIN	87	85	82
ANTONIUS MAYORIKUS	88	85	95
AYUNI RIZKIANA	87	88	90
BAYU	87	90	98
DIMAS SATRIO	88	88	98

1. Standarisasi Data

Rumus :

$$X^1 = \frac{X - \mu}{\sigma}$$

Hitung Rata-Rata (μ) dan Standar Deviasi (σ) :

a. Agama (X_1)

$$\mu = \frac{87+88+87+87+88}{5} = 87.4$$

$$\sigma = \frac{\sqrt{(87-87.4)^2+(88-87.4)^2+(87-87.4)^2+(87-87.4)^2+(88-87.4)^2}}{5} = 0.55$$

b. PKN (X_2)

$$\mu = \frac{85+85+88+90+88}{5} = 87.2$$

$$\sigma = \frac{\sqrt{(85-87.2)^2+(85-87.2)^2+(88-87.2)^2+(90-87.2)^2+(88-87.2)^2}}{5} = 2.04$$

c. B.Indo (X_3)

$$\mu = \frac{82+95+90+98+98}{5} = 92.6$$

$$\sigma = \frac{\sqrt{(82-92.6)^2+(95-92.6)^2+(90-92.6)^2+(98-92.6)^2+(98-92.6)^2}}{5} = 6.54$$

Data Standar (Z-Score)

Tabel 2. Perhitungan Metode PCA (Z-Score)

Nama Siswa	Z-Agama (X_1)	Z-PPKN (X_2)	Z-B.indo (X_3)
AMELIAH AGUSTIN	$\frac{87 - 87.4}{0.55} = 0.73$	$\frac{85 - 87.2}{2.04} = 1.08$	$\frac{82 - 92.6}{6.54} = 1.62$
ANTONIUS MAYORIKUS	$\frac{88 - 87.4}{0.55} = 1.09$	-1.08	$\frac{95 - 92.6}{6.54} = 0.37$
AYUNI RIZKIANA	-0.73	$\frac{88 - 87.2}{2.04} = 0.39$	$\frac{90 - 92.6}{6.54} = 0.40$
BAYU	-0.73	$\frac{90 - 87.2}{2.04} = 1.37$	$\frac{98 - 92.6}{6.54} = 0.83$
DIMAS SATRIO	1.09	0.39	0.83

2. Matriks Kovarians

Rumus :

$$Cov(X, Y) = \frac{1}{n-1} \sum_{i=1}^n (X_i^1 \cdot Y_i^1)$$

Hitung Kovarians :

$$a) \quad Cov(X_1, X_2) = \frac{(-0.73 \times -1.08) + (1.09 \times -1.08) + (-0.73 \times 0.39) + (-0.73 \times 1.37) + (1.09 \times 0.39)}{4} = 0.20$$

$$b) \quad Cov(X_1, X_3) = \frac{(-0.73 \times -1.62) + (1.09 \times -0.37) + (-0.73 \times -0.40) + (-0.73 \times 0.83) + (1.09 \times 0.83)}{4} = 0.68$$

$$c) \quad Cov(X_2, X_3) = \frac{(1.08 \times -1.62) + (1.08 \times -0.37) + (0.39 \times -0.40) + (1.37 \times 0.83) + (0.39 \times 0.83)}{4} = 0.74$$

Matriks Kovarians :

$$\begin{pmatrix} VAR(X_1) & COV(X_1, X_2) & COV(X_1, X_3) \\ COV(X_2, X_1) & Var(X_2) & COV(X_2, X_3) \\ COV(X_3, X_1) & COV(X_3, X_2) & Var(X_3) \end{pmatrix} = \begin{pmatrix} 1 & 0,20 & 0,68 \\ 0,20 & 1 & 0,74 \\ 0,68 & 0,74 & 1 \end{pmatrix}$$

Var (X) =1 Karena data sudah distandarasi

3. Eigenvalues dan Eigenvectors

Persamaan Karakteristik :

$$\text{Det}(\text{Cov} - \lambda I) = 0$$

Hitung Eigenvalues (λ):

$$\begin{pmatrix} 1-\lambda & 0.20 & 0.68 \\ 0.20 & 1-\lambda & 0.74 \\ 0.68 & 0.74 & 1-\lambda \end{pmatrix} = 0$$

Solusi dengan algoritma

$$\lambda_1 = 2.04 \text{ (PC utama)}$$

$$\lambda_2 = 0.63$$

$$\lambda_3 = 0.33$$

Eigenvectors untuk $\lambda_1 = 2.04$:

$$\begin{pmatrix} 1-2.04 & 0.20 & 0.68 \\ 0.20 & 1-2.04 & 0.74 \\ 0.68 & 0.74 & 1-2.04 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = 0$$

Hasil : $v_1 = 0.52, v_2 = 0.57, v_3 = 0.64$ (Vektor Utama)

4. Transformasi ke Principal Component

Rumus :

$$\text{PC} = X^T \times \text{Eigenvectors}$$

$$\text{PC1} = 0.52, X_1 + 0.57, X_2 + 0.64, X_3$$

Tabel 3. Perhitungan Metode PCA (PC1)

Nama Siswa	PC1 (Komponen Utama)
Amelia Agustin	$0.52(-0.73)+0.57(-1.08)+0.64(-1.62) = -1.92$
Antonius Mayorikus	$0.52(1.09)+0.57(-1.08)+0.64(0.37) = -0.03$
Ayuni Rizkiana	$0.52(-0.73)+0.57(0.39)+0.64(0.40) = -0.33$
Bayu	$0.52(-0.73)+0.57(1.37)+0.64(0.83) = 1.05$
Dimas Satrio	$0.52(1.09)+0.57(0.39)+0.64(0.83) = 1.23$

a. Interpretasi Hasil :

- 1) PC1 (63% Variansi) merepresentasikan prestasi akademik
- 2) Siswa dengan PC1 Tinggi (Bayu, Dimas) Unggul di semua mata pelajaran
- 3) Siswa dengan PC1 Rendah (Amelia) memiliki nilai di bawah rata-rata

b. PC1 Scores:

$$\text{Amelia } (-1.92) < \text{Ayuni } (-0.33) < \text{Antoni } (-0.03) < \text{Bayu } (1.05) < \text{Dimas } (1.23)$$

c. Analisa Korelasi :

Nilai Bahasa Indonesia (X_3) paling berpengaruh pada PC1 (koefisien eigenvector terbesar : 0.64).

4.3 Hasil Pengujian Metode K-Means Clustering

a. Data awal

Tabel 4. Perhitungan metode k-means (data awal)

Nama Siswa	Nilai Agama (X_1)	Nilai PPKN (X_2)	Nilai B.indo (X_3)	Rata-Rata
AMELIAH AGUSTIN	87	85	82	84.67
ANTNIUS MAYORIKUS	88	85	95	89.33
AYUNI RIZKIANA	87	88	90	88.33
BAYU	87	90	98	91.67
DIMAS SATRIO	88	88	98	91.33

1. Tentukan Jumlah Klaster ($K=2$)

- Klaster 1 : Nilai Terendah
- Klaster 2 : Nilai Tertinggi

2. Inisialisasi Centroid

Pilih 2 Siswa Sebagai Centroid Awal :

- Centroid 1 (C_1) : Amelia (Nilai Terendah : 87,85,82)
- Centroid 2 (C_2): Bayu (Nilai Tertinggi : 87,90,98)

3. Hitung Jarak Eucliden ke Centroid

$$\text{Rumus Jarak : } d = \sqrt{(X_1 - C_1)^2 + (X_2 - C_2)^2 + (X_3 - C_3)^2}$$

Tabel 5. Perhitungan Metode K-Means (Hitung Jarak)

Nama Siswa	Jarak Ke C_1 (Amelia)	Klaster
AMELIAH	0	Klaster 1
	Jarak ke C_2 (Bayu)	
	$87-87+ 85-90+ 82-98=16.76$	
ANTNIUS	Jarak Ke C_1 (Amelia)	Klaster 2
	$88-87+ 85-85+ 95-82=13.04$	
	Jarak ke C_2 (Bayu)	
	$\sqrt{(88-87)^2 + (85-90)^2 + (95-98)^2} = 5.92$	
AYUNI RIZKIANA	Jarak Ke C_1 (Amelia)	Klaster 2
	$\sqrt{(87-87)^2 + (88-85)^2 + (90-82)^2} = 8.54$	
	Jarak ke C_2 (Bayu)	

	$\sqrt{(87 - 87)^2 + (88 - 90)^2 + (90 - 98)^2} = 8.25$	
BAYU	Jarak Ke C1 (Amelia)	Klaster 2
	$\sqrt{(87 - 87)^2 + (90 - 85)^2 + (98 - 82)^2} = 16.76$	
	Jarak ke C2 (Bayu)	
	0	
DIMAS SATRIO	Jarak Ke C1 (Amelia)	Klaster 2
	$\sqrt{(88 - 87)^2 + (88 - 85)^2 + (98 - 82)^2} = 16.31$	
	Jarak ke C2 (Bayu)	
	$\sqrt{(88 - 87)^2 + (88 - 90)^2 + (98 - 98)^2} = 2.24$	

4. Hitung Ulang Centroid

a) klaster 1 : (Amelia)

Centroid Baru : Tetap (87,85,82) karena hanya 1 anggota

b) klaster 2 : (antonius, ayuni, bayu, dimas)

$$\text{Agama} = \frac{88+87+87+88}{4} = 87.5$$

$$\text{Pkn} = \frac{85+88+90+88}{4} = 87.75$$

$$\text{B.indo} = \frac{95+90+98+98}{4} = 95.25$$

5. Iterasi ke 2 (Hitung Jarak Baru)

Centroid Baru:

1) C1 : (87, 85, 82)

2) C2 : (87.5, 87.75, 95.25)

Perhitungan Jarak:

Tabel 6. Perhitungan Manual K-Means (Iterasi)

Nama Siswa	Jarak ke C1	Jarak ke C2	Klaster
Amelia Agutin	0	$\sqrt{(87 - 87.5)^2 + (85 - 87.75)^2 + (82 - 95.25)^2} = 13.54$	Klaster 1
Antonius Mayorikus	13.04	$\sqrt{(88 - 87.5)^2 + (85 - 87.75)^2 + (95 - 95.25)^2} = 2.80$	Klaster 2
Ayuni Rizkiana	8.54	$\sqrt{(87 - 87.5)^2 + (88 - 87.75)^2 + (90 - 95.25)^2} = 5.32$	Klaster 2
Bayu	16.76	$\sqrt{(87 - 87.5)^2 + (90 - 87.75)^2 + (98 - 95.25)^2} = 3.91$	Klaster 2
Dimas Satrio	16.31	$\sqrt{(88 - 87.5)^2 + (88 - 87.75)^2 + (98 - 95.25)^2} = 2.87$	Klaster 2

6. Konvergen

Hasil Klaster Tidak Berubah dan Proses berhenti (konvergen), karena tidak ada perubahan anggota klaster antara iterasi pertama dan kedua.

Tabel 7. Perhitungan Metode K-Means (Konvergen)

Iterasi	Klaster 1	Klaster 2
1	Amelia	Antonius, Ayuni, Bayu, Dimas
2	Amelia	Antonius, Ayuni, Bayu, Dimas

Tidak ada perpindahan anggota → Algoritma berhenti.

7. Hasil Klasterisasi :

Tabel 8. Perhitungan Manual K-Means (Hasil Clustering)

Klaster	Anggota	Rata-Rata Nilai	Kategori
Klaster 1	Amelia Agustin	84.67	Nilai Terendah
Klaster 2	Antonius, Ayuni, Bayu, Dimas	89.67-91.67	Nilai Tertinggi

5. KESIMPULAN

Sistem *machine learning* untuk otomatisasi pengelompokan siswa yang menggunakan *metode principal component analysis* dan *metode k-means clustering* telah berhasil dirancang dan diimplementasikan di SMK Negeri 8 Kab Tangerang. Proses pengelompokan siswa secara konvensional yang dilakukan oleh guru biasanya memerlukan waktu yang cukup lama, terutama ketika jumlah siswa dan mata pelajaran yang dianalisis cukup banyak. Untuk mengelompokkan 30 siswa berdasarkan nilai dari 6 mata pelajaran inti, guru membutuhkan waktu sekitar 5–10 menit per siswa untuk membaca, menilai, membandingkan, dan memutuskan kelompok yang sesuai. Dengan demikian, total waktu yang dibutuhkan dapat mencapai 150–300 menit atau sekitar 2,5 hingga 5 jam. Selain memakan waktu, proses ini juga rentan terhadap subjektivitas guru. Dengan menerapkan sistem otomatis berbasis *Principal Component Analysis* dan *K-Means Clustering*, proses pengelompokan dapat diselesaikan jauh lebih cepat.

REFERENCE

- Chyan, Phie, Sitti Arni, and Mohammad Sofyan S Thayf. *Pengantar Machine Learning PT . MIFANDI MANDIRI DIGITAL*.
- Diba, Farah, Maya Silvi Lydia, and Poltak Sihombing. 2023. "Analisis Random Forest Menggunakan Principal Component Analysis Pada Data Berdimensi Tinggi." *The Indonesian Journal of Computer Science* 12(4): 2152–60.
- Fadlisyah, Fadlisyah. 2021. "Pengelompokan Siswa Penyandang Disabilitas Berdasarkan Tingkat Tunagrahita Menggunakan Metode Naïve Bayes." *Jurnal Teknologi Terapan and Sains* 4.0 2(1): 337.
- Haryono, Wasis. 2024. "Sistem Informasi Pemberkasan Order Marketplace Pada Pt. Flexofast Indonesia." *Jurnal Jaringan Sistem Informasi Robotik (JSIR)* 8(1): 21–26. <http://ojsamik.amikmitragama.ac.id>.
- Informasi, Sistem, Fakultas Teknologi, and Universitas Bunda Mulia. 2025. "IMPLEMENTASI MACHINE LEARNING DALAM PENGELOMPOKAN MUSIK MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING." 8: 74–83.
- Mesak, Maria Inasensia, and Wasis Haryono. 2025. "Aplikasi Rekap Nilai Siswa Berbasis Web Di SDN Tajurhalang 03 Bogor Web-Based Student Grade Recap Application at SDN Tajurhalang 03 Bogor." : 2361–71.
- No, Vol et al. 2024. "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 ." 7(2):



500–510.

- Putra, Muhammad Rizki Pratama, Nouval Aulia Ahmad, Ramadhan Makarim Ulya, and Wasis Haryono. 2024. "Rancangan Website Delivery Order Minimarket Vimart Dengan GPS (Global Positioning)." *SANTIKA: Jurnal Ilmiah Sains dan Teknologi* 14: 2088–5407.
- Rismaya, Riska, Dwi Yuniarto, and David Setiadi. 2025. "Penerapan Algoritma Machine Learning Dalam Prediksi Prestasi Akademik Mahasiswa."
- Rosyada, Istina Alya, and Dina Tri Utari. 2024. "Penerapan Principal Component Analysis Untuk Reduksi Variabel Pada Algoritma K-Means Clustering." *Jambura Journal of Probability and Statistics* 5(1): 6–13.
- Sachrul Sidiq, Fa'asya Vladimar Kasidin, Shaffly Fawwaz Fadhlullah, and Wasis Haryono. 2024. "Implementasi Sistem Aplikasi Pembayaran Sekolah Dan Pendaftaran Siswa Berbasis Web." *Switch : Jurnal Sains dan Teknologi Informasi* 3(1): 27–36.
- Ubaydillah, Fuji, Muhamad Mahmud, Sinta Rahmawati, and Wasis Haryono. 2023. "Analisa Dan Perancangan Sistem Informasi Berbasis Website Menggunakan Metode Agile Di Sd Negeri Pamulang 01." *Journal Information & Computer* 1(1): 33–40.