

Reduksi Dimensi untuk Meningkatkan Kinerja Pengklasteran Perilaku Siswa pada Sistem *e-Learning*

(*Dimension Reduction to Improve the Clustering of Students' Behavior on e-Learning*)

Yuni Yamasari¹, Naim Rochmawati², Anita Qoiriah³, Asmunin⁴, Atik Wintarti⁵

Abstract—The corona pandemic has changed the learning process from face-to-face (offline) to online learning. However, this online learning has caused difficulties in monitoring student behavior by teachers due to reduced direct interaction. Additionally, students often feel isolated. Therefore, this situation causes failure in their learning achievement. This problem encourages a lot of research on modeling related to student behavior. However, previous research did not focus much on improving the model's performance or system being built. In fact, the performance of this model significantly affects the result's quality of this student behavior mapping. Therefore, this study focuses on improving the performance of student behavior clustering when they interact with the e-Learning system. Performance improvement was made by reducing dimensions of student data with Principal Component Analysis (PCA). Furthermore, two techniques for the centroid initialization were explored to obtain optimal results: random and K-means++. For measuring cluster quality, this study employed the silhouette index. The experimental results show that the clusters with the highest quality are achieved by applying PCA with seven components. In addition, the cluster number for all centroid initialization techniques is three to four. This quality cluster can assist teachers in monitoring student behavior in the e-Learning system.

Intisari—Pandemi *corona* telah mengubah proses pembelajaran dari yang semula tatap muka secara langsung (*offline*) menjadi pembelajaran secara daring (*online*). Pembelajaran daring ini menyebabkan kesulitan dalam pemantauan perilaku siswa oleh guru karena berkurangnya interaksi secara langsung. Lebih dari itu, siswa seringkali merasa terisolasi sehingga jika dibiarkan, situasi ini akan menyebabkan kegagalan dalam prestasi belajarnya. Permasalahan ini mendorong banyak dilakukannya penelitian tentang pemodelan yang berkaitan dengan perilaku siswa. Namun, para peneliti sebelumnya tidak banyak yang fokus pada peningkatan kinerja model atau sistem yang dibangun, padahal kinerja model ini sangat berpengaruh terhadap kualitas hasil pemetaan perilaku siswa. Untuk itu, makalah ini berfokus pada peningkatan kinerja pengklasteran perilaku siswa ketika berinteraksi dengan sistem *e-Learning*. Peningkatan kinerja dilakukan dengan reduksi dimensi

pada data siswa dengan *Principal Component Analysis* (PCA). Selanjutnya, dua teknik inisialisasi titik pusat kluster dieksplorasi untuk mendapatkan hasil yang optimal, yaitu: *random* dan *K-means++*. Untuk pengukuran kualitas kluster, makalah ini menggunakan *silhouette index*. Hasil pengujian menunjukkan bahwa kluster dengan kualitas tertinggi dicapai oleh penerapan PCA dengan tujuh komponen dan banyaknya kluster tiga sampai empat untuk semua teknik inisialisasi titik pusat. Kluster yang berkualitas ini dapat membantu guru dalam memonitor perilaku siswa pada pembelajaran secara daring.

Kata Kunci—Kluster, PCA, Reduksi Dimensi, *e-Learning*, Perilaku.

I. PENDAHULUAN

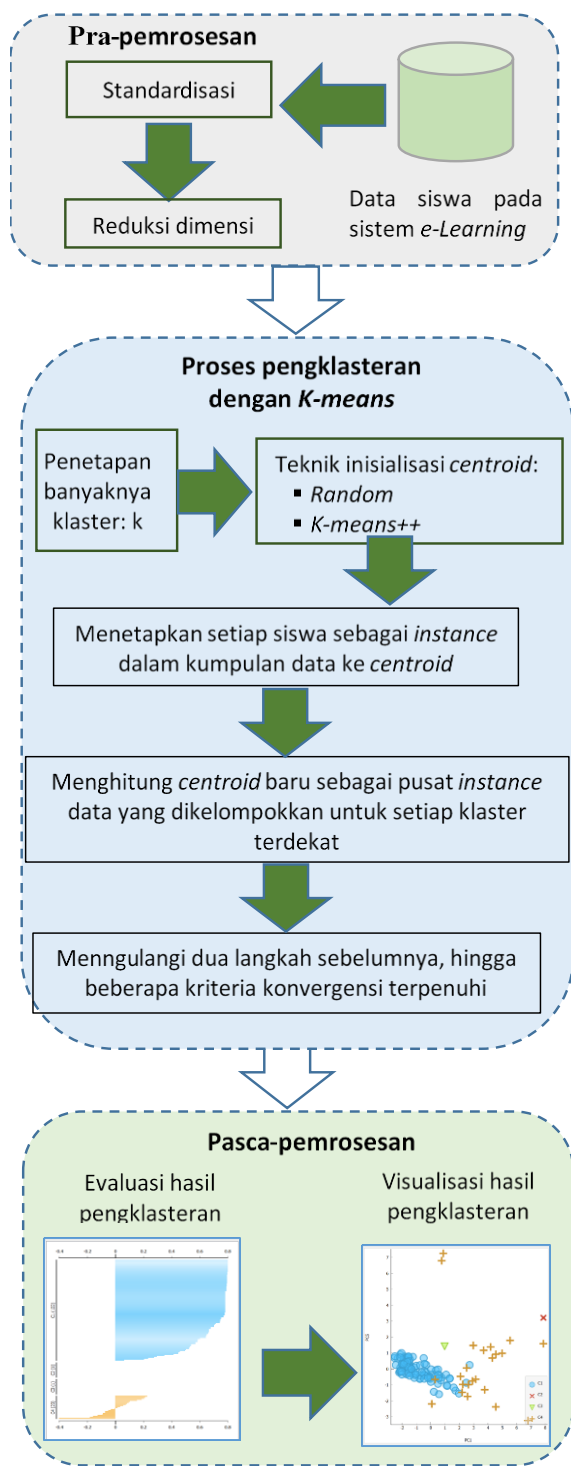
Saat ini pemanfaatan kemajuan Teknologi Informasi dan Komunikasi (TIK) telah diimplementasikan oleh hampir semua instansi pendidikan. Terlebih saat pandemi *corona* seperti saat ini, proses pembelajaran harus dilakukan secara daring untuk memutus rantai penularan virus *corona* [1]. Proses pembelajaran daring ini menurunkan tingkat keterlibatan siswa dalam proses pembelajaran karena berkurangnya interaksi secara langsung antara guru dan siswa [2]. Lebih lanjut, pembelajaran seperti ini memunculkan kesulitan bagi guru dalam memantau perilaku siswa pada saat proses pembelajaran berlangsung.

Implementasi proses pembelajaran daring ini tentu saja berdampak pada terkumpulnya data dari waktu ke waktu yang salah satunya data siswa. Data siswa ini dapat dieksplorasi dengan metode-metode *Educational Data Mining* (EDM) untuk menghasilkan informasi [3].

Beberapa penelitian sebelumnya melakukan penambahan terhadap data siswa untuk menemukan informasi yang berkaitan dengan prestasi belajar [4]-[8], kinerja akademik [9]-[15], perilaku [16]-[20], dan sikap [21], [22]. Berkaitan dengan domain perilaku, telah dilakukan pengklasteran terhadap perilaku daring pada *blended learning* yang mengombinasikan pembelajaran secara daring dan *face-to-face* [17]. Pendekatan pengklasteran yang digunakan adalah metode *Latent Class Analysis*. Penelitian ini mampu mengidentifikasi empat tipe perilaku siswa, yaitu tipe *inactive* (50%), tipe *communication/collaboration* (24,3%), tipe *delivery/discussion* (18,0%), dan tipe *sharing/submission* (7,2%). Penelitian berikutnya melakukan analisis terhadap perilaku dari *platform* pembelajaran *autonomous* dengan *decision tree* [18]. Hasil analisis menemukan faktor-faktor yang memengaruhi perilaku daring siswa. Pada penelitian yang lain, perilaku siswa pada sistem *Online Chinese Reading Assessment* dianalisis

^{1,2,3,4} Jurusan Teknik Informatika Fakultas Teknik Universitas Negeri Surabaya, Gedung A10, Jl. Ketintang Wiyata, Ketintang, Gayungan, Kota Surabaya, Jawa Timur 60231 (e-mail: ¹yuniyamasari@unesa.ac.id; ²naimrochmawati@unesa.ac.id, ³anitaqoiriah@unesa.ac.id, ⁴asmunin@unesa.ac.id)

⁵ Jurusan Matematika Fakultas MIPA Universitas Negeri Surabaya, Jl. Ketintang Wiyata No.36, Ketintang, Gayungan, Kota Surabaya, Jawa Timur 60231 (e-mail: atikwintarti@unesa.ac.id)



Gbr. 1 Metodologi yang diusulkan.

menggunakan kluster [19]. Metode yang digunakan adalah *K-means* dan hasil yang diperoleh digunakan untuk meningkatkan kinerja dari kemampuan membaca siswa. Penelitian selanjutnya menganalisis pola perilaku siswa pada pembelajaran bahasa pemrograman dengan menggunakan *sequential pattern mining tool* [23]. Hasil penelitian menemukan pola perilaku siswa berprestasi tinggi dan siswa berprestasi rendah.

Penelitian-penelitian sebelumnya ini belum berfokus pada peningkatan kinerja sistem ataupun model yang dibangun, padahal kinerja dari sistem ini, dalam konteks pengklasteran, sangat berpengaruh terhadap validitas atau kualitas kluster yang dihasilkan. Kualitas dari kluster biasanya dievaluasi dengan menggunakan ukuran *silhouette index* yang mempertimbangkan evaluasi interkluster dan intrakluster. Dalam dunia pendidikan, jika kluster yang dihasilkan mempunyai validitas yang tinggi, informasi yang dihasilkan tentu saja juga mempunyai tingkat validitas yang tinggi. Hal ini sangat membantu guru dalam mengambil tindakan yang tepat untuk memperbaiki proses belajar mengajar dan lain sebagainya.

Beberapa penelitian sebelumnya berupaya untuk meningkatkan validitas kluster dengan seleksi fitur [24]-[26] Kemudian, sebuah lain melakukan peningkatan akurasi pada kluster perilaku siswa dengan pemilihan inisialisasi titik pusat dan banyaknya kluster [27]. Sedangkan, penelitian lain melakukan ekstraksi fitur untuk meningkatkan akurasi pada pengklasteran prestasi belajar siswa [28].

Oleh karena itu, makalah ini juga berfokus pada peningkatan proses pengklasteran dari perilaku siswa ketika berinteraksi dengan sistem *e-Learning*. Peningkatan kinerja dari proses ini dilakukan dengan reduksi dimensi pada data siswa. Peningkatan ini dimaksudkan untuk menghasilkan informasi terkait kluster perilaku siswa dengan level validitas yang tinggi. Informasi ini dapat membantu guru dalam melakukan pemantauan perilaku siswa pada pembelajaran daring di masa pandemi *corona* ini.

Akhirnya, artikel ini disusun menjadi beberapa bagian sebagai berikut: bagian I menguraikan pendahuluan, bagian II menjelaskan tentang metodologi yang diusulkan, bagian III membahas tentang hasil pengujian dan analisis, dan bagian terakhir merupakan kesimpulan dari penelitian yang telah dilakukan.

II. METODOLOGI

Bagian ini menjelaskan metodologi yang diusulkan. Metodologi terdiri atas tiga tahapan utama yang masing-masing terdiri atas beberapa langkah, sebagaimana digambarkan pada Gbr. 1.

A. Tahap Prapemrosesan

Tahap ini terdiri atas dua prosedur utama, yaitu standarisasi dan reduksi dimensi dengan menggunakan *Principle Component Analysis (PCA)*. Tentu saja prosedur-prosedur ini dapat dilakukan setelah data siswa yang berasal dari ekstraksi aktivitas siswa pada sistem *e-Learning* terbangun.

Data siswa dikumpulkan dari dua SMK Negeri di Surabaya. Data ini diperoleh pada saat siswa berinteraksi dengan sistem *e-Learning*. Partisipan dalam penelitian ini berjumlah 127 siswa. Aktivitas siswa dicatat kemudian disimpan di basis data (*database*). Selanjutnya, data diekstraksi untuk menghasilkan data siswa yang terdiri atas fitur-fitur. Dalam makalah ini, proses pengklasteran melibatkan semua fitur, yaitu tiga puluh fitur. Deskripsi dan tipe data dari data siswa disajikan pada Tabel I.

TABEL I
FITUR DAN TIPE DATA DARI DATA SISWA

No	Fitur	Tipe Data	Deskripsi
1	<i>DownloadIndFar</i>	Numerik	Banyaknya aktivitas siswa yang berkaitan dengan pengunduhan materi hukum induksi Faraday.
2	<i>DownloadMedMag</i>	Numerik	Banyaknya aktivitas siswa yang berkaitan dengan pengunduhan materi medan magnet.
3	<i>ForumIndFar</i>	Numerik	Banyaknya aktivitas siswa saat mengikuti forum yang berkaitan dengan materi hukum induksi Faraday.
4	<i>ForumMedMag</i>	Numerik	Banyaknya aktivitas siswa saat mengikuti forum yang berkaitan dengan materi medan magnet.
5	<i>FailLogin</i>	Numerik	Banyaknya aktivitas siswa yang berhubungan dengan berhasil atau gagal dalam proses <i>login</i> .
6	<i>LearnVideo1</i>	Numerik	Jumlah aktivitas siswa yang berkaitan dengan materi pembelajaran menggunakan video1.
7	<i>LearnVideo2</i>	Numerik	Jumlah aktivitas siswa yang berkaitan dengan materi pembelajaran menggunakan video2.
8	<i>LearnVideo3</i>	Numerik	Jumlah aktivitas siswa yang berkaitan dengan materi pembelajaran menggunakan video3.
9	<i>LearnVideo4</i>	Numerik	Jumlah aktivitas siswa yang berkaitan dengan materi pembelajaran menggunakan video4.
10	<i>LearnVideo5</i>	Numerik	Jumlah aktivitas siswa yang berkaitan dengan materi pembelajaran menggunakan video5.
11	<i>LearnVideo6</i>	Numerik	Jumlah aktivitas siswa yang berkaitan dengan materi pembelajaran menggunakan video6.
12	<i>LearnVideo7</i>	Numerik	Jumlah aktivitas siswa yang berkaitan dengan materi pembelajaran menggunakan video7.
13	<i>LearnVideo8</i>	Numerik	Jumlah aktivitas siswa yang berkaitan dengan materi pembelajaran menggunakan video8.
14	<i>LearnVideo9</i>	Numerik	Jumlah aktivitas siswa yang berkaitan dengan materi pembelajaran menggunakan video9.
15	<i>LearnVideo10</i>	Numerik	Jumlah aktivitas siswa yang berkaitan dengan materi pembelajaran menggunakan video10.
16	<i>LearnVideo11</i>	Numerik	Jumlah aktivitas siswa yang berkaitan dengan materi pembelajaran menggunakan video11.
17	<i>LearnVideo12</i>	Numerik	Jumlah aktivitas siswa yang berkaitan dengan materi pembelajaran menggunakan video12.
18	<i>LearnVideo13</i>	Numerik	Jumlah aktivitas siswa yang berkaitan dengan materi pembelajaran menggunakan video13.
19	<i>LearnVideo14</i>	Numerik	Jumlah aktivitas siswa yang berkaitan dengan materi pembelajaran menggunakan video14.
20	<i>LearnVideo15</i>	Numerik	Jumlah aktivitas siswa yang berkaitan dengan materi pembelajaran menggunakan video15.
21	<i>LearnVideo16</i>	Numerik	Jumlah aktivitas siswa yang berkaitan dengan materi pembelajaran menggunakan video16.
22	<i>LearnVideo17</i>	Numerik	Jumlah aktivitas siswa yang berkaitan dengan materi pembelajaran menggunakan video17.
23	<i>LearnVideo18</i>	Numerik	Jumlah aktivitas siswa yang berkaitan dengan materi pembelajaran menggunakan video18.
24	<i>SuccessLogin</i>	Numerik	Jumlah aktivitas siswa yang berhubungan dengan <i>login</i> dengan status berhasil.
25	<i>Logout</i>	Numerik	Banyaknya aktivitas siswa yang berkaitan dengan proses keluarnya siswa dari sistem <i>e-Learning</i> .
26	<i>Examination</i>	Numerik	Banyaknya aktivitas siswa selama mengerjakan ujian.
27	<i>ExerIndFar</i>	Numerik	Banyaknya aktivitas siswa yang berhubungan dengan latihan materi hukum induksi Faraday.
28	<i>ExcerMedMag</i>	Numerik	Banyaknya aktivitas siswa yang berhubungan dengan latihan materi medan magnet.
29	<i>Sum</i>	Numerik	Jumlah total aktivitas siswa setelah masuk hingga mereka keluar dari <i>e-Learning</i> .
30	<i>Average</i>	Numerik	Jumlah aktivitas siswa setelah <i>log in</i> sebelum mereka <i>logout</i> dari <i>e-Learning</i> dibagi dengan jumlah aktivitas.

Berkaitan dengan prosedur standardisasi, prosedur ini dilakukan untuk penyekalaan ulang data agar memiliki rata-rata 0 dan standar deviasi 1. Prosedur selanjutnya adalah PCA, yang diterapkan untuk melakukan reduksi dimensi. PCA dapat didefinisikan sebagai proyeksi ortogonal dari data ke ruang linier berdimensi lebih rendah, yang dikenal sebagai subruang utama, sehingga varian dari data yang diproyeksikan dimaksimalkan [29].

Secara detail, PCA merupakan operasi yang diterapkan ke data siswa yang dapat dijelaskan sebagai berikut: Y didefinisikan sebagai matriks $n \times m$ yang hasilnya diproyeksikan dari Y ke B . Lalu, n terdiri atas 127 baris

(sebanyak siswa yang terlibat dalam penelitian ini) dan m merupakan kolom yang terdiri atas tiga puluh fitur. Operasi-operasinya dapat dijelaskan dengan langkah-langkah berikut.

1) *Langkah 1*: Menghitung nilai *mean* μ dari masing-masing kolom (fitur).

$$M = \mu(Y). \quad (1)$$

2) *Langkah 2*: Melakukan *centering* nilai pada masing-masing kolom (fitur) dengan melakukan pengurangan dengan nilai *mean*.

$$C = Y - M. \quad (2)$$

3) *Langkah 3*: Menghitung *covariance* dari matriks yang telah dikenai proses *centering*.

$$V = \text{cov}(C). \quad (3)$$

4) *Langkah 4*: Menghitung *eigencomposition* dari matriks V untuk menghasilkan *list* dari *eigenvalues* dan dari *eigenvectors*.

$$\text{values, vectors} = \text{eig}(V). \quad (4)$$

5) *Langkah 5*: Memilih komponen untuk membentuk subruang. *Eigenvectors* merepresentasikan arah atau komponen-komponen untuk subruang dari B yang tereduksi dan *eigenvalues* merepresentasikan besaran arah tersebut. *Eigenvectors* diurutkan dengan *eigenvalues* dalam urutan dari nilai terbesar ke nilai terkecil. Hal ini dilakukan untuk menyediakan *components ranking* atau *axis* dari subruang baru dari Y .

Jika semua nilai Eigen memiliki nilai yang sama, representasi yang ada sudah cukup padat atau padat. Jika ada nilai Eigen mendekati nol, nilai tersebut mewakili komponen atau sumbu B yang dapat dibuang. Sebanyak m matriks kolom atau komponen yang lebih kecil harus dipilih untuk membentuk subruang yang dipilih. Idealnya, akan dipilih k vektor Eigen, yang disebut sebagai komponen utama dengan nilai Eigen terbesar.

$$B = \text{select}(\text{values, vectors}). \quad (5)$$

6) *Langkah 6*: Memproyeksikan komponen yang dipilih ke subruang via multiplikasi matriks.

$$P = B^T \cdot Y \quad (6)$$

dengan Y adalah data original yang akan diproyeksikan, B^T merupakan *transpose* dari *principle components* yang dipilih, dan P adalah proyeksi dari Y .

B. Tahap Proses Pengklasteran dengan *K-means*

Tahap pemrosesan melakukan pengklasteran dengan *K-means*. Langkah awal pada tahap ini adalah menentukan banyaknya klaster yang disimpan pada variabel k . Langkah selanjutnya, inisialisasi sekumpulan *centroid*, dilakukan dengan menggunakan dua teknik, yaitu: *random* dan *K-means++*.

1) *Teknik Random*: Klaster ditugaskan secara acak atau *random* pada awalnya dan kemudian diperbarui dengan iterasi lebih lanjut.

2) *Teknik K-means++*: Pusat pertama dipilih secara acak dan selanjutnya dipilih dari titik yang tersisa dengan probabilitas proporsional dengan jarak kuadrat dari pusat terdekat.

Langkah selanjutnya, seperti dijelaskan pada Gbr. 1, yaitu menetapkan setiap siswa dalam kumpulan data ke *centroid*. Kemudian, *centroid* baru dihitung sebagai pusat *instance* data yang nantinya dikelompokkan ke klaster terdekat. Langkah terakhir pada tahap ini adalah mengulangi kedua langkah sebelumnya hingga kriteria konvergensi terpenuhi.

TABEL II
PELABELAN SKENARIO

Label	Skenario
P	Standardisasi; PCA 5 komponen; $k = 4$; inisialisasi titik awal <i>random</i>
O	Standardisasi; PCA 5 komponen; $k = 4$; inisialisasi titik awal <i>K-means++</i>
N	Standardisasi; PCA 5 komponen; $k = 3$; inisialisasi titik awal <i>random</i>
M	Standardisasi; PCA 5 komponen; $k = 3$; inisialisasi titik awal <i>K-means++</i>
L	Standardisasi; PCA 7 komponen; $k = 4$; inisialisasi titik awal <i>random</i>
K	Standardisasi; PCA 7 komponen; $k = 4$; inisialisasi titik awal <i>K-means++</i>
J	Standardisasi; PCA 7 komponen; $k = 3$; inisialisasi titik awal <i>random</i>
I	Standardisasi; PCA 7 komponen; $k = 3$; inisialisasi titik awal <i>K-means++</i>
H	Standardisasi; $k = 4$; inisialisasi titik awal <i>random</i>
G	Standardisasi; $k = 4$; inisialisasi titik awal <i>K-means++</i>
F	Standardisasi; $k = 3$; inisialisasi titik awal <i>random</i>
E	Standardisasi; $k = 3$; inisialisasi titik awal <i>K-means++</i>
D	$k = 4$; inisialisasi titik awal <i>random</i>
C	$k = 4$; inisialisasi titik awal <i>K-means++</i>
B	$k = 3$; inisialisasi titik awal <i>random</i>
A	$k = 3$; inisialisasi titik awal <i>K-means++</i>

C. Tahap Pascapemrosesan

Tahap terakhir adalah tahap pascapemrosesan. Tahap ini terdiri atas dua langkah, yaitu evaluasi hasil pengklasteran dan kemudian dilanjutkan dengan langkah visualisasi dari hasil pengklasteran. Kemudian evaluasi dilakukan untuk mengukur kualitas atau validitas klaster yang dihasilkan. Dalam makalah ini, klaster-klaster yang dihasilkan diukur dengan *silhouette index*.

III. HASIL PENGUJIAN DAN ANALISIS

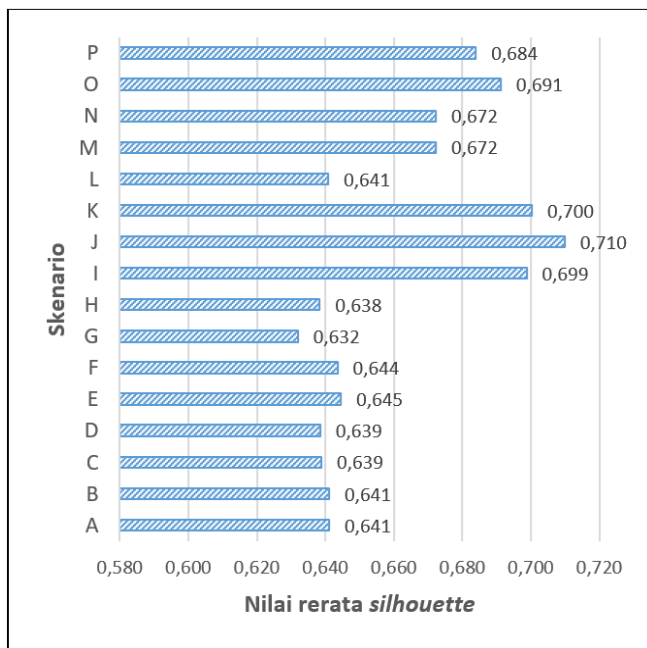
Hasil uji coba dan analisis dijabarkan pada bagian ini. Penjelasan terdiri atas dua bagian, yaitu hasil pengklasteran evaluasi serta visualisasi hasil pemetaan.

A. Hasil Pengklasteran dan Evaluasi

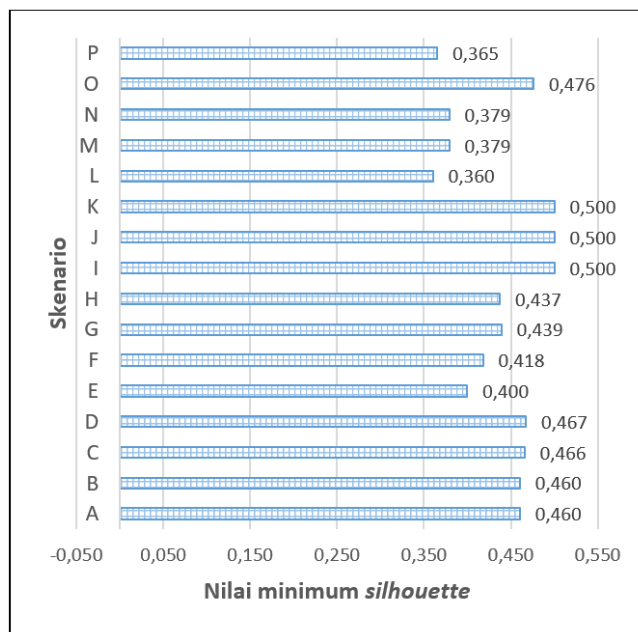
Pada subbagian ini, data siswa yang telah terbentuk ditambah dengan *K-means*. Namun, sebelum *K-means* diimplementasikan, beberapa langkah diterapkan pada tahap prapemrosesan sebagaimana yang telah dijelaskan pada bagian sebelumnya, yaitu standardisasi dan reduksi dimensi. Pada standardisasi, teknik ini akan melakukan perubahan skala yang nilainya terpusat pada *mean* 0 dan standar deviasi 1.

Selanjutnya, PCA dilakukan untuk mereduksi dimensi sehingga dimensi menjadi lima dan tujuh komponen saja, yang semula sebanyak tiga puluh fitur. Untuk tahap pemrosesan dengan *K-means*, banyaknya klaster ditentukan dengan nilai 3 dan 4.

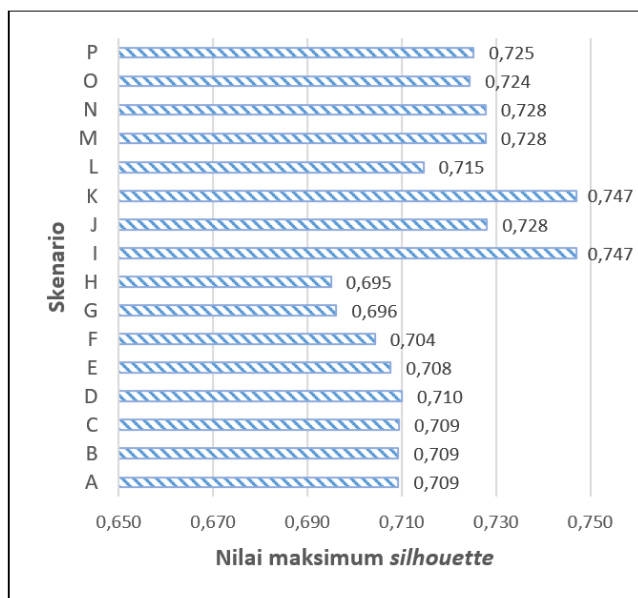
Kemudian, pada langkah inisialisasi titik awal, makalah ini mengeksplorasi dua teknik, yaitu *random* dan *K-means++*.



Gbr. 2 Perbandingan nilai rerata silhouette pada semua skenario.



Gbr. 4 Perbandingan nilai minimum silhouette pada semua skenario.



Gbr. 3 Perbandingan nilai maksimum silhouette pada semua skenario.

TABEL III
KOMPOSISI KLASTER DARI SKENARIO K

Klaster	Indeks Siswa	Banyaknya Siswa
1	2,3,4,7,8,9,10,11,12,13,14,15,16,17,18,19,21,23,24,25,26,27,28,29,31,35,36,37,38,39,40,41,42,43,45,46,47,48,49,50,51,52,53,54,55,56,57,58,59,60,61,62,64,65,66,67,69,70,71,72,75,76,77,78,79,81,82,83,86,87,88,89,91,92,93,96,97,98,99,101,102,103,104,105,106,107,108,109,110,112,113,114,115,116,117,118,119,120,121,122,124,125	102
2	126	1
3	127	1
4	1,5,6,20,22,30,32,33,34,44,63,68,73,74,80,84,85,90,94,95,100,111,123	23

Dengan mengombinasikan teknik-teknik tersebut, secara keseluruhan terdapat enam belas skenario yang dilabelkan, seperti dideskripsikan pada Tabel II. Keseluruhan skenario diuji coba untuk memperoleh klaster dengan kualitas yang terbaik.

Setelah proses pengklasteran dilakukan dengan menggunakan keenam belas skenario tersebut, nilai silhouette dihitung terhadap keseluruhan anggota klaster dalam rangka mengevaluasi kualitas klaster. Untuk nilai silhouette ini, nilai ditinjau berdasarkan tiga ranah perbandingan, yaitu rerata, maksimum, dan minimum dari nilai silhouette. Berdasarkan keenam belas skenario tersebut, ketiga ranah perbandingan nilai silhouette diilustrasikan pada Gbr. 2 sampai Gbr. 4.

Berdasarkan ketiga gambar tersebut, dapat dilihat bahwa skenario pengujian K, J, dan I mendominasi nilai silhouette yang lebih tinggi dari skenario pengujian lainnya. Pencapaian ini terjadi pada semua perbandingan nilai silhouette, baik pada rerata, minimum, maupun maksimum.

Secara detail, berdasarkan nilai rerata silhouette yang ditunjukkan pada Gbr. 2, nilai tertinggi dicapai oleh skenario J dengan nilai 0,710. Nilai ini cukup tinggi mengingat nilai tersebut mendekati nilai tertinggi silhouette, yaitu 1. Hal ini menunjukkan bahwa masing-masing perilaku siswa berada tepat pada klaster tersebut. Sebaliknya, nilai rerata silhouette terendah terjadi pada skenario G, yaitu sebesar 0,623.

Nilai tertinggi dicapai oleh skenario K dan I, yaitu sekitar 0,747, berdasarkan perbandingan nilai maksimum silhouette

TABEL IV
KOMPOSISI KLASTER DARI SKENARIO J

Klaster	Indeks Siswa	Banyaknya Siswa
1	84	1
2	1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,41,42,43,44,45,46,47,48,49,50,51,52,53,54,55,56,57,58,59,60,61,62,63,64,65,66,67,68,69,70,71,72,73,74,75,76,77,78,79,80,81,82,83,85,86,87,88,89,90,91,92,93,94,95,96,97,98,99,100,101,102,103,104,105,106,107,108,109,110,111,112,113,114,115,116,117,118,119,120,121,122,123,124,125,126,127	125
3	40	1

TABEL V
KOMPOSISI KLASTER DARI SKENARIO I

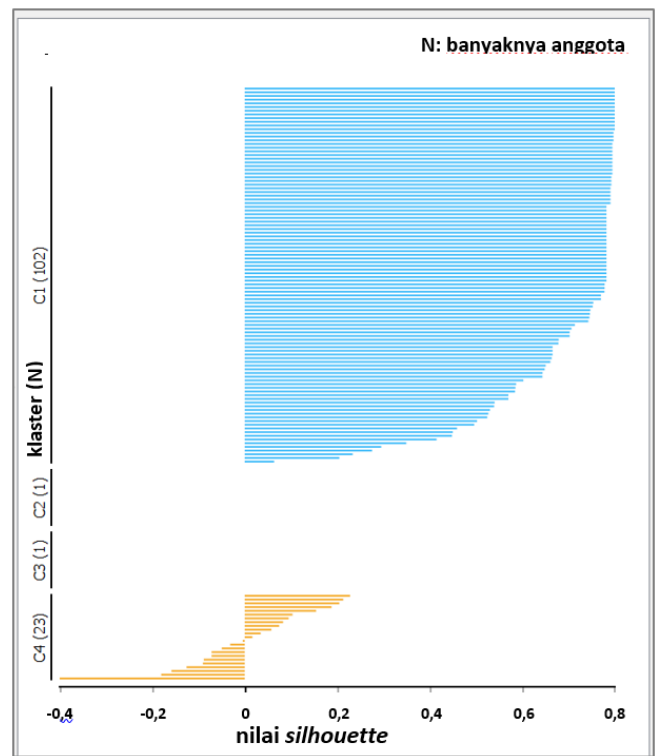
Klaster	Indeks Siswa	Banyaknya Siswa
1	1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50,51,52,53,54,55,56,57,58,59,60,61,62,63,64,65,66,67,68,69,70,71,72,73,74,75,76,77,78,79,80,81,82,83,84,85,86,87,88,89,90,91,92,93,94,95,96,97,98,99,100,101,102,103,104,105,106,107,108,109,110,111,112,113,114,115,116,117,118,119,120,121,122,123,124,125	125
2	126	1
3	127	1

yang disajikan pada Gbr. 3. Sebaliknya, nilai terendah terjadi pada H, yaitu sekitar 0,695.

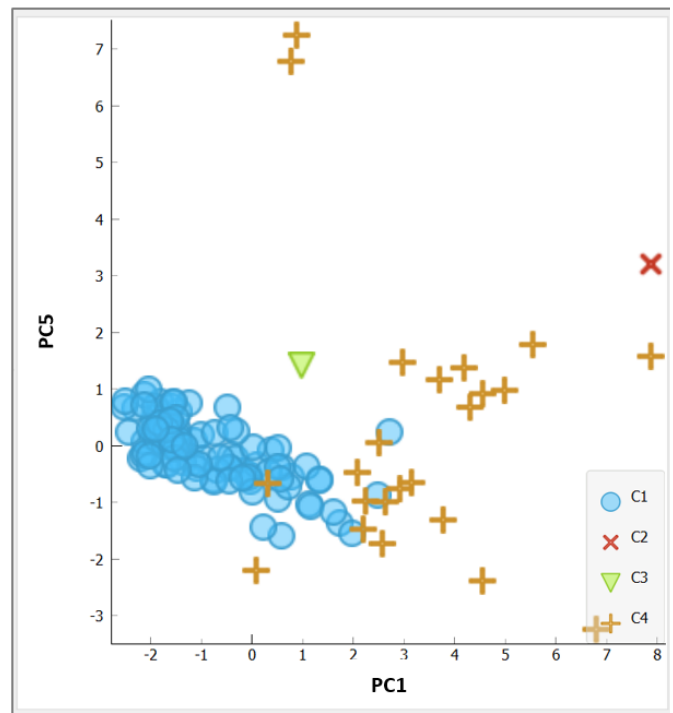
Sementara itu, dengan mengacu pada Gbr. 4, nilai minimum tertinggi dicapai oleh skenario I, J, dan K, dengan nilai sekitar 0,5. Kondisi kontras, yaitu nilai minimum *silhouette*, terjadi pada skenario L, dengan nilai 0,36.

Namun, secara keseluruhan, ketiga skenario, yaitu K, I, dan J masih mendominasi dengan capaian nilai tertinggi pada nilai rerata, nilai maksimum, dan nilai minimum dari *silhouette*.

Berdasarkan konversi label untuk semua skenario pada Tabel II, dapat diterjemahkan bahwa ketiga skenario dapat mencapai nilai yang lebih tinggi dengan melakukan standardisasi dan reduksi dimensi dengan menggunakan PCA tujuh komponen. Secara detail, ketiga skenario dapat dijelaskan sebagai berikut.



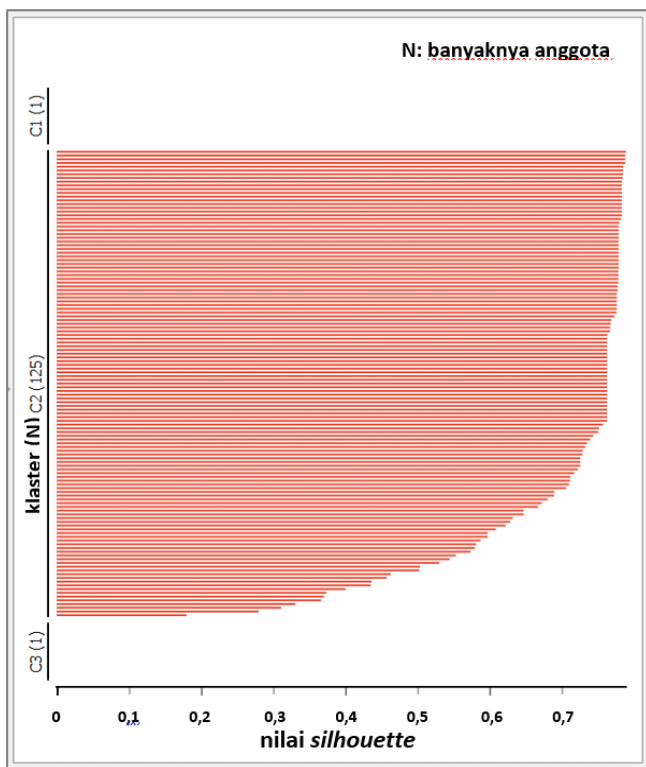
(a)



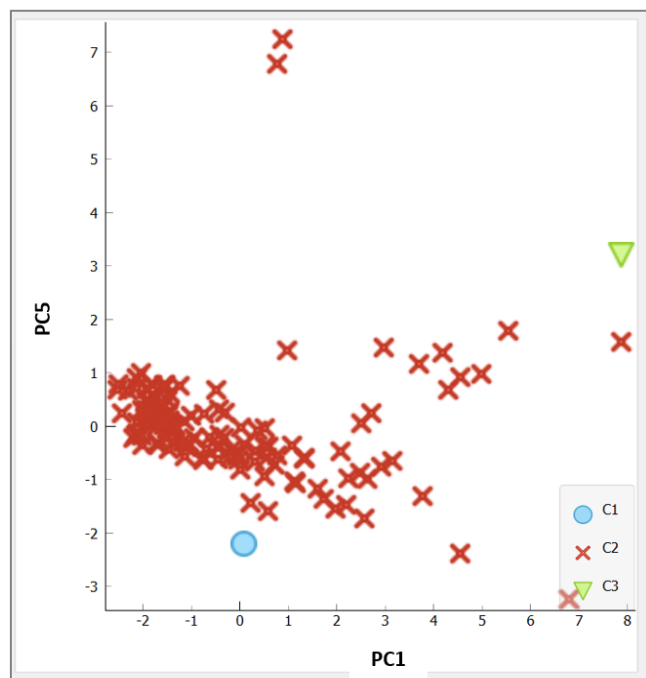
(b)

Gbr. 5 Visualisasi hasil dari skenario K, (a) nilai *silhouette*, (b) hasil pemetaan perilaku siswa.

1) *Skenario K*: Pada tahap prapemrosesan dilakukan standardisasi dan PCA tujuh komponen. Kemudian, pada tahap pemrosesan, banyaknya klaster pada *K-means* ditetapkan $k = 4$



(a)

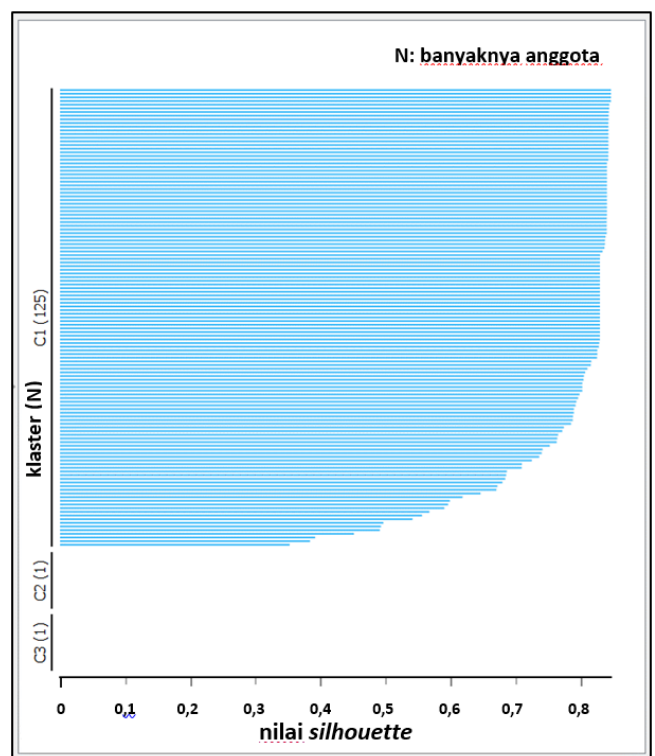


(b)

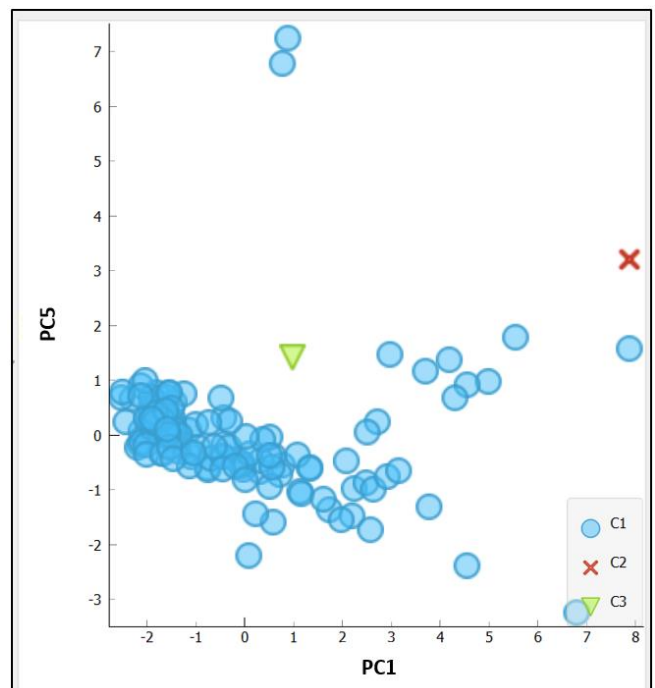
Gbr. 6 Visualisasi hasil skenario J, (a) nilai *silhouette*, (b) hasil pemetaan perilaku siswa.

dan langkah inialisasi titik awal menggunakan teknik *K-means++*.

2) *Skenario J*: Pada tahap prapemrosesan, dilakukan standarisasi dan PCA tujuh komponen. Kemudian, pada tahap



(a)



(b)

Gbr. 7 Visualisasi hasil skenario I, (a) nilai *silhouette*, (b) hasil pemetaan perilaku siswa.

pemrosesan, banyaknya klaster pada *K-means* ditetapkan $k = 3$ dan langkah inialisasi titik awal menggunakan teknik *random*.

3) *Skenario I*: Pada tahap prapemrosesan dilakukan standarisasi dan PCA tujuh komponen. Kemudian, pada tahap

pemrosesan, banyaknya kluster pada *K-means* ditetapkan $k = 3$ dan langkah inialisasi titik awal menggunakan teknik *K-means++*.

Dengan mengacu pada ketiga skenario tersebut, komposisi kluster dari perilaku siswa pada sistem *e-Learning* dideskripsikan pada Tabel III sampai Tabel V. Berdasarkan ketiga tabel tersebut, terlihat dalam hasil pemetaan bahwa pada skenario K dan I terjadi irisan, yaitu pada kluster 2 dan kluster 3. Anggota klasternya berturut-turut adalah siswa dengan indeks 126 dan 127. Ini menunjukkan bahwa indeks-indeks siswa tersebut memang telah berada pada kluster yang tepat, yaitu kedua skenario tersebut mempunyai kinerja yang tinggi. Ini berarti bahwa kluster yang telah dihasilkan dalam penelitian ini merupakan kluster dengan kualitas yang cukup tinggi dengan level validitas sekitar 0,747, ketika levelnya dibandingkan dengan penelitian sebelumnya yang hanya mencapai level validitas sebesar 0,637 [25].

B. Visualisasi dari Hasil Evaluasi dan Pengklasteran

Pada bagian ini, visualisasi dilakukan pada hasil evaluasi dari nilai *silhouette* dan visualisasi hasil pengklasteran. Visualisasi pada kedua hasil hanya dilakukan pada ketiga skenario yang menghasilkan kluster yang paling berkualitas. Visualisasi dalam bentuk grafik dari skenario K, J, dan I diperlihatkan secara berturut-turut pada Gbr. 5 sampai Gbr. 7.

Berkaitan dengan visualisasi nilai *silhouette*, visualisasi dari skenario K, J, dan I disajikan berturut-turut pada Gbr. 5(a), Gbr. 6(a), dan Gbr. 7(a). Berdasarkan visualisasi tersebut, terlihat bahwa nilai *silhouette* tertinggi dicapai oleh skenario K dan I.

Visualisasi hasil pengklasteran dari skenario K, J, dan I disajikan pada Gbr. 5(b), Gbr. 6(b), dan Gbr. 7(b). Visualisasi dari ketiga skenario menggunakan sumbu x dengan PC1 dan sumbu y dengan PC5. Hasil visualisasi memperlihatkan keterpisahan antar kluster cukup bagus. Hal ini mengindikasikan bahwa kluster cukup berkualitas berdasarkan intrakluster dan interkluster.

IV. KESIMPULAN

Proses pengklasteran perilaku siswa pada sistem *e-Learning* dapat ditingkatkan kinerjanya dengan mereduksi dimensi dari data siswanya. Reduksi dimensi dapat dilakukan dengan PCA. Peningkatan ini diindikasikan dengan level validitas kluster yang cukup tinggi yang didasarkan pada *silhouette index* sebagai alat ukurnya. Kluster dengan level validitas atau kualitas tertinggi dicapai dengan penerapan PCA tujuh komponen dan banyaknya kluster $k = 3$ dan $k = 4$ untuk semua teknik inialisasi *centroid*, baik *random* maupun *K-means++*.

Hasil penelitian ini hanya melakukan pemetaan perilaku siswa dengan kinerja model yang cukup tinggi tanpa melakukan identifikasi dari tipe perilaku siswanya. Oleh karena itu, penelitian selanjutnya dapat berfokus pada proses pelabelan dari perilaku siswa yang telah dikelompokkan ini.

UCAPAN TERIMA KASIH

Terima kasih disampaikan kepada Universitas Negeri Surabaya, khususnya Jurusan Teknik Informatika, yang telah memberikan dukungan atas terlaksananya penelitian ini.

REFERENSI

- [1] D. Evayanti (2020) "Efektivitas Pembelajaran Melalui Metode Daring (Online) dalam Masa Darurat Covid-19 – STIT Al-Kifayah Riau," [Online], <https://www.stit-alkifayahriau.ac.id/efektivitas-pembelajaran-melalui-metode-daring-online-dalam-masa-darurat-covid-19/>, tanggal akses: 3-Feb-2021.
- [2] S.S. Kusumawardani dan S.A.I. Alfarozi, "Kajian Penggunaan Data Log Mahasiswa untuk Berbagai Permasalahan Analisis Pembelajaran," *J. Nas. Tek. Elektro dan Teknol. Inf.*, Vol. 9, No. 4, hal. 365–374, 2020.
- [3] A. Peña-Ayala, *Educational Data Mining: Applications and Trends*. Cham, Switzerland: Springer, 2014.
- [4] R. Cerezo, M. Sánchez-Santillán, M.P. Paule-Ruiz, dan J.C. Núñez, "Students' LMS Interaction Patterns and Their Relationship with Achievement: A Case Study in Higher Education," *Comput. Educ.*, Vol. 96, hal. 42–54, Mei 2016.
- [5] B. Sen dan E. Ucar, "Evaluating the Achievements of Computer Engineering Department of Distance Education Students with Data Mining Methods," *Procedia Technol.*, Vol. 1, hal. 262–267, 2012.
- [6] L.S.R. Pedrozo dan M. Rodriguez-Artacho, "A Cluster-based Analysis to Diagnose Students' Learning Achievements," *2013 IEEE Global Engineering Education Conference (EDUCON)*, 2013, hal. 1118–1123.
- [7] J.N. Purwaningsih dan Y. Suwarno, "Predicting Students Achievement Based on Motivation in Vocational School Using Data Mining Approach," *2016 4th International Conference on Information and Communication Technology (ICoICT)*, 2016, hal. 1–5.
- [8] N. Buniyamin, U. bin Mat, dan P.M. Arshad, "Educational Data Mining for Prediction and Classification of Engineering Students Achievement," *2015 IEEE 7th International Conference on Engineering Education (ICEED)*, 2015, hal. 49–53.
- [9] L. Rahman, N.A. Setiawan, dan A.E. Permanasari, "Feature Selection Methods in Improving Accuracy of Classifying Students' Academic Performance," *2017 2nd International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*, 2017, hal. 267–271.
- [10] A. Marwaha dan S. Ahuja, "A Review on Identifying Influencing Factors and Data Mining Techniques Best Suited for Analyzing Students' Performance," *2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC)*, 2017, hal. 373–378.
- [11] A.M. de Moraes, J.M.F.R. Araujo, dan E.B. Costa, "Monitoring Student Performance Using Data Clustering and Predictive Modelling," *2014 IEEE Front. Educ. Conf. Proc.*, 2014, hal. 1–8.
- [12] Z. Li, C. Shang, dan Q. Shen, "Fuzzy-Clustering Embedded Regression for Predicting Student Academic Performance," *2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 2016, hal. 344–351.
- [13] A.I. Adekitan dan E. Noma-Osaghae, "Data Mining Approach to Predicting the Performance of First Year Student in a University Using the Admission Requirements," *Educ. Inf. Technol.*, Vol. 24, No. 2, hal. 1527–1543, Mar. 2019.
- [14] A.I. Adekitan dan O. Salau, "The Impact of Engineering Students' Performance in the First Three Years on Their Graduation Result Using Educational Data Mining," *Heliyon*, Vol. 5, No. 2, hal. 1-21, Feb. 2019.
- [15] A.B. El Din Ahmed dan I. Elaraby, "Data Mining: A Prediction for Student's Performance Using Classification Method," *World J. Comput. Appl. Technol.*, Vol. 2, No. 2, hal. 43–47, 2014.
- [16] L. Juhaňák, J. Zounek, dan L. Rohlíková, "Using Process Mining to Analyze Students' Quiz-Taking Behavior Patterns in a Learning Management System," *Comput. Human Behav.*, Vol. 92, hal. 496-506, Des. 2017.
- [17] Y. Park, J.H. Yu, dan I.-H. Jo, "Clustering Blended Learning Courses by Online Behavior Data: A Case Study in a Korean Higher Education Institute," *Internet High. Educ.*, Vol. 29, hal. 1–11, Apr. 2016.
- [18] W. Jie, L. Hai-yan, C. Biao, dan Z. Yuan, "Application of Educational Data Mining on Analysis of Students' Online Learning Behavior," *2017 2nd International Conference on Image, Vision and Computing (ICIVC)*, 2017, hal. 1011–1015.
- [19] L. Jia, H.N.H. Cheng, S. Liu, W.-C. Chang, Y. Chen, dan J. Sun, "Integrating Clustering and Sequential Analysis to Explore Students'

- Behaviors in an Online Chinese Reading Assessment System,” *2017 6th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI)*, 2017, hal. 719–724.
- [20] M. Jovanovic, M. Vukicevic, M. Milovanovic, dan M. Minovic, “Using Data Mining on Student Behavior and Cognitive Style Data for Improving e-Learning Systems: A Case Study,” *Int. J. Comput. Intell. Syst.*, Vol. 5, No. 3, hal. 597–610, Jun. 2012.
- [21] Y. Promdee, S. Kasemvilas, N. Phangsuk, dan R. Yodthasarn, “Predicting Persuasive Message for Changing Student’s Attitude Using Data Mining,” *2017 International Conference on Platform Technology and Service (PlatCon)*, 2017, hal. 1–5.
- [22] P. Provelengios dan G. Fesakis, “Educational Applications of Serious Games: The Case of the Game Food Force in Primary Education Students,” *Proc. 5th Eur. Conf. Games Based Learn.*, 2011, hal. 476–485.
- [23] W.-C. Shih, “Mining Sequential Patterns to Explore Users’ Learning Behavior in a Visual Programming App,” *2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC)*, 2018, hal. 126–129.
- [24] Y. Yamasari, S.M.S. Nugroho, R. Harimurti, dan M.H. Purnomo, “Improving the Cluster Validity on Student’s Psychomotor Domain Using Feature Selection,” *2018 International Conference on Information and Communications Technology, ICOIACT 2018*, 2018, hal. 460–465.
- [25] Y. Yamasari, A. Qoiriah, H.P.A. Tjahyaningtjas, R.E. Putra, A. Prihanto, dan Asmunin, “Improving the Quality of the Clustering Process on Students’ Performance Using Feature Selection,” *2020 International Seminar on Application for Technology of Information and Communication (iSemantic)*, 2020, hal. 454–458.
- [26] M. Anusha dan J.G.R. Sathiasaelan, “Feature Selection Using K-Means Genetic Algorithm for Multi-objective Optimization,” *Procedia Computer Science*, Vol. 57, hal. 1074–1080, 2015.
- [27] D. Ding, J. Li, H. Wang, dan Z. Liang, “Student Behavior Clustering Method Based on Campus Big Data,” *2017 13th International Conference on Computational Intelligence and Security (CIS)*, 2017, hal. 500–503.
- [28] N. Rochmawati, H.B. Hidayati, Y. Yamasari, W. Yustanti, L. Rakhmawati, H.P.A. Tjahyaningtjas, dan Y. Anistyasari, “Covid Symptom Severity Using Decision Tree,” in *2020 Third International Conference on Vocational Education and Electrical Engineering (ICVEE)*, 2020, hal. 1–5.
- [29] C.M. Bishop, *Pattern Recognition and Machine Learning*, New York, USA: Springer, 2006.