JURNAL ILMIAH INFORMATIKA DAN ILMU KOMPUTER (JIMA-ILKOM)
E-ISSN 2988-747X P-ISSN 2988-7461
Volume 3, Nomor 1, Maret 2024, Page 46-56
DOI: https://doi.org/10.58602/jima-ilkom.v3i1.26

Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa

Nirwana Hendrastuty
Sistem Informasi, Universitas Teknokrat Indonesia, Indonesia
nirwanahendrastuty@teknokrat.ac.id

Kata Kunci:
Clustering;
Data *Mining;*
Evaluasi;
K-Means;
Pembelajaran Siswa;

Abstrak

Evaluasi hasil pembelajaran siswa merupakan proses kritis dalam pendidikan yang bertujuan untuk mengukur pencapaian tujuan pembelajaran. Melalui berbagai metode seperti tes, proyek, dan observasi, guru dapat menilai pemahaman, keterampilan, dan kemajuan siswa dalam materi pelajaran. Tujuan penerapan data *mining* menggunakan algoritma *K-Means Clustering* dalam evaluasi hasil pembelajaran siswa adalah untuk mengidentifikasi pola-pola yang mungkin tersembunyi dalam data hasil pembelajaran, membagi siswa ke dalam kelompok-kelompok berdasarkan tingkat pencapaian atau karakteristik pembelajaran mereka, serta memberikan wawasan yang berharga kepada guru dan *stakeholder* pendidikan. Hasil *clustering* data penilaian belajar siswa dapat mengungkap pola yang bermanfaat bagi pendidik dan administrator sekolah. Analisis terhadap *cluster-cluster* ini dapat mengungkapkan informasi tentang tren prestasi, kecenderungan keberhasilan atau kesulitan dalam mata pelajaran tertentu, serta memungkinkan identifikasi siswa yang memerlukan bantuan tambahan. Pengelompokan hasil *cluster* berdasarkan data penilaian siswa dengan *K-Means* memperoleh 2 kelompok siswa yaitu siswa Rajin dengan kelompok C0 dan kelompok siswa Sangat Rajin dengan kelompok C1. Kelompok C0 siswa Rajin terdiri dari 63 siswa dan kelompok C1 terdiri dari 91 siswa Sangat Rajin. Hasil pengujian *silhouette score* untuk klaster 2 yang paling tinggi sebasar 0,9168 dan menunjukkan bahwa pengelompokan data menjadi kelompok-kelompok tersebut lebih baik, penggunaan *silhouette score* sebagai metrik evaluasi memberikan panduan yang berguna dalam menentukan jumlah kluster yang optimal dalam analisis *clustering* dan interpretasi data.

Keywords:
Clustering;
Data Mining;
Evaluation;
K-Means;
Student Learning;

Abstract

Evaluation of student learning outcomes is a critical process in education that aims to measure the achievement of learning objectives. Through various methods such as tests, projects, and observations, teachers can assess students' understanding, skills, and progress in the subject matter. The purpose of applying data mining using the K-Means Clustering algorithm in evaluating student learning outcomes is to identify patterns that may be hidden in learning outcome data, divide students into groups based on their level of achievement or learning characteristics, and provide valuable insights to teachers and education stakeholders. The results of clustering student learning assessment data can uncover patterns that are beneficial to educators and school administrators. Analysis of these clusters can reveal information about achievement trends, trends in success or difficulty in specific subjects, as well as allow identification of students who need additional help. Grouping of cluster results based on student assessment data with k-means obtained 2 groups of students, namely Diligent students with group C0 and group students Very Diligent with group C1.

Nirwana Hendrastuty: *Penulis Korespondensi

JURNAL ILMIAH INFORMATIKA DAN ILMU KOMPUTER (JIMA-ILKOM)
E-ISSN 2988-747X P-ISSN 2988-7461
Volume 3, Nomor 1, Maret 2024, Page 46-56
DOI: https://doi.org/10.58602/jima-ilkom.v3i1.26

The C0 group of Diligent students consists of 63 students and the C1 group consists of 91 Very Diligent students. The silhouette score test results for cluster 2 are as high as 0.9168 and show that grouping data into these groups is better, the use of silhouette score as an evaluation metric provides useful guidance in determining the optimal number of clusters in clustering analysis and data interpretation.

1.PENDAHULUAN

Evaluasi hasil pembelajaran siswa merupakan proses kritis dalam pendidikan yang bertujuan untuk mengukur pencapaian tujuan pembelajaran. Melalui berbagai metode seperti tes, proyek, dan observasi, guru dapat menilai pemahaman, keterampilan, dan kemajuan siswa dalam materi pelajaran. Evaluasi ini tidak hanya mengukur pengetahuan akademis, tetapi juga mengidentifikasi kebutuhan individual siswa serta memperbaiki proses pembelajaran. Hasil evaluasi juga memberikan umpan balik yang berharga kepada siswa untuk meningkatkan kinerja mereka. Dengan menggunakan data evaluasi secara efektif, guru dapat merancang strategi pengajaran yang lebih efisien dan berfokus pada kebutuhan individual siswa, sehingga meningkatkan kesempatan mereka untuk mencapai potensi penuh dalam pembelajaran. Selain itu, evaluasi hasil pembelajaran siswa juga memungkinkan sekolah dan sistem pendidikan untuk mengevaluasi efektivitas kurikulum, metode pengajaran, dan strategi pembelajaran yang digunakan. Dengan menganalisis data evaluasi secara menyeluruh, stakeholder pendidikan dapat mengidentifikasi area yang memerlukan perbaikan dan membuat keputusan yang berdasarkan bukti untuk meningkatkan mutu pendidikan secara keseluruhan. Evaluasi hasil pembelajaran siswa bukan hanya tentang memberikan nilai, tetapi juga tentang menciptakan lingkungan pembelajaran yang mendukung pertumbuhan dan perkembangan siswa secara holistik. Tujuan evaluasi pembelajaran siswa adalah untuk mengukur pemahaman dan penguasaan siswa terhadap materi pelajaran, memberikan umpan balik yang konstruktif kepada siswa untuk meningkatkan kinerja belajar mereka, mengidentifikasi kebutuhan individual siswa, serta menilai efektivitas kurikulum dan metode pengajaran. Melalui evaluasi, guru dapat mengembangkan strategi pembelajaran yang lebih efektif dan menyesuaikan pendekatan pengajaran sesuai dengan kebutuhan siswa, sehingga membantu memastikan bahwa setiap siswa dapat mencapai potensi belajar mereka yang optimal. Selain itu, evaluasi juga merupakan alat penting bagi sekolah dan sistem pendidikan untuk mengukur dan meningkatkan kualitas pendidikan secara keseluruhan.

Data *mining* merupakan proses eksplorasi dan analisis data yang bertujuan untuk menemukan pola-pola yang bermanfaat, hubungan tersembunyi, dan informasi yang dapat diambil dari sekumpulan data besar[1]–[3]. Dengan menggunakan berbagai teknik dan algoritma, data mining membantu dalam mengidentifikasi tren, mengklasifikasikan data, dan membuat prediksi yang berguna untuk pengambilan keputusan. Ini melibatkan tahap-tahap seperti *preprocessing* data, pemodelan, evaluasi, dan interpretasi hasil. Data *mining* memiliki beragam aplikasi di berbagai bidang, termasuk bisnis, ilmu pengetahuan, kesehatan, keuangan, dan lainnya, membantu organisasi untuk memanfaatkan informasi yang tersimpan dalam data mereka untuk tujuan analisis dan pengambilan keputusan yang lebih baik[3], [4]. Dengan terus berkembangnya teknologi dan kemampuan analitik, peran data mining menjadi semakin penting dalam memanfaatkan potensi besar dari data yang tersedia untuk kemajuan dan inovasi di berbagai bidang. Salah satu teknik dalam data mining yaitu *clustering*.

Clustering merupakan salah satu teknik dalam data *mining* yang bertujuan untuk mengelompokkan data ke dalam kelompok-kelompok yang memiliki kesamaan berdasarkan pada karakteristik tertentu. Teknik ini mencari struktur dalam data tanpa memerlukan label kelas yang telah ditentukan sebelumnya[5], [6]. Dengan menggunakan berbagai metode seperti *K-Means*, *Hierarchical Clustering*, atau *DBSCAN*, *clustering* memungkinkan pengelompokan data yang serupa berdasarkan pada atribut atau fitur tertentu, sehingga memungkinkan analisis lebih lanjut terhadap kelompok-kelompok ini untuk mengungkap pola-pola yang tersembunyi atau hubungan yang signifikan di antara data. *Clustering* digunakan dalam berbagai bidang, termasuk segmentasi pelanggan, analisis data genetik, pengelompokan dokumen, serta pemrosesan gambar dan video. Melalui proses *clustering*, data yang kompleks dapat disederhanakan menjadi kelompok-kelompok yang lebih terstruktur, memungkinkan

Nirwana Hendrastuty: *Penulis Korespondensi

JURNAL ILMIAH INFORMATIKA DAN ILMU KOMPUTER (JIMA-ILKOM)
E-ISSN 2988-747X P-ISSN 2988-7461
Volume 3, Nomor 1, Maret 2024, Page 46-56
DOI: https://doi.org/10.58602/jima-ilkom.v3i1.26

pemahaman yang lebih baik terhadap hubungan antara data dan pengambilan keputusan yang lebih efektif. Setiap kelompok yang dihasilkan oleh algoritma *clustering* mewakili pola atau karakteristik tertentu dalam data, yang dapat membantu dalam mengidentifikasi tren, memprediksi perilaku, atau mengelompokkan entitas yang serupa untuk analisis lebih lanjut. Namun, pemilihan metode *clustering* yang tepat dan interpretasi yang akurat terhadap hasil *clustering* merupakan langkah kritis dalam memastikan keberhasilan penerapan teknik ini dalam berbagai aplikasi. Salah satu metode dalam *clustering* yaitu *K-Means Clustering*.

K-Means Clustering adalah salah satu teknik clustering yang paling umum digunakan dalam analisis data. Algoritma ini bekerja dengan cara membagi data ke dalam K kelompok yang telah ditentukan sebelumnya, di mana K merupakan jumlah kelompok yang diinginkan. Proses dimulai dengan memilih secara acak K titik pusat kelompok (*centroid*) di dalam ruang data, lalu mengelompokkan setiap titik data ke dalam kelompok yang memiliki *centroid* terdekat[7], [8]. Kemudian, titik pusat setiap kelompok dihitung kembali berdasarkan rata-rata titik data dalam kelompok tersebut, dan proses ini diulangi hingga tidak ada lagi perubahan dalam penempatan titik data ke dalam kelompok atau hingga batasan iterasi yang ditentukan tercapai. *K-Means Clustering* sering digunakan dalam segmentasi pelanggan, pengelompokan dokumen, analisis citra, dan bidang lainnya karena kecepatan dan skalabilitasnya, meskipun hasilnya dapat dipengaruhi oleh inisialisasi awal *centroid*[9], [10]. Selain kecepatan dan skalabilitasnya, *K-Means Clustering* juga memiliki kemampuan untuk menangani data yang besar dengan baik, membuatnya cocok untuk analisis pada dataset yang besar seperti yang sering ditemui dalam aplikasi bisnis dan ilmiah. Namun, perlu dicatat bahwa keberhasilan *K-Means* sangat tergantung pada pemilihan jumlah kelompok yang tepat (nilai K) dan inisialisasi awal *centroid* yang baik.

Penelitian terkait dengan *K-Means Clustering* dilakukan oleh Hutagalung (2022) penerapan data mining dengan algoritma k-means sebagai bahan pertimbangan pendidik dalam proses pengambilan keputusan[11]. Penelitian dari Mawarni (2022) implementasi algoritma k-means clustering pada disiplin ilmu mahasiswa dibagi menjadi tiga cluster, hasil pengelompokan tingkat kedisiplinan siswa dengan metode k-means dapat dijadikan acuan atau penilaian terhadap kedisiplinan setiap siswa[12]. Penelitian dari Rohma (2021) menganalisis tingkat hambatan pembelajaran daring di SMK YASPIM dengan menggunakan algoritma *k-means clustering analysis*[13]. Penelitian oleh Dewi (2022) hasil penelitian ini dapat dikelompokkan menjadi 3 cluster diantaranya cluster 0 menunjukkan siswa aktif sebanyak 30 siswa, cluster 1 menunjukkan siswa tidak aktif sebanyak 8 siswa, dan cluster 2 menunjukkan siswa kurang aktif sebanyak 21 siswa[14].

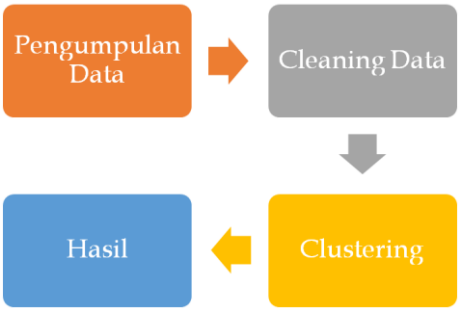
Tujuan penerapan data *mining* menggunakan algoritma *K-Means Clustering* dalam evaluasi hasil pembelajaran siswa adalah untuk mengidentifikasi pola-pola yang mungkin tersembunyi dalam data hasil pembelajaran, membagi siswa ke dalam kelompok-kelompok berdasarkan tingkat pencapaian atau karakteristik pembelajaran mereka, serta memberikan wawasan yang berharga kepada guru dan *stakeholder* pendidikan. Penerapan data mining menggunakan algoritma *K-Means Clustering* dapat memungkinkan pendekatan pembelajaran yang lebih diferensial dan responsif terhadap kebutuhan individual siswa, sehingga meningkatkan kualitas pembelajaran dan pencapaian akademis secara keseluruhan.

2.METODE PENELITIAN

Penelitian dalam bidang data *mining* merupakan suatu proses yang melibatkan serangkaian tahapan yang terstruktur untuk mengeksplorasi dan menganalisis data secara mendalam. Tahapan ini dimulai dengan pengumpulan data yang relevan, kemudian dilanjutkan dengan tahap *pre-processing* di mana data tersebut dibersihkan, diintegrasikan, dan dipersiapkan untuk analisis lebih lanjut. Selanjutnya, tahapan pemodelan data dilakukan, di mana berbagai teknik dan algoritma data *mining* diterapkan untuk mengidentifikasi pola, tren, dan hubungan yang tersembunyi dalam data[1], [15]. Hasil dari pemodelan dievaluasi dan divalidasi untuk memastikan keakuratan serta kehandalannya. Tahapan penelitian yang dilakukan dapat lihat Gambar. 1.

Nirwana Hendrastuty: *Penulis Korespondensi

JURNAL ILMIAH INFORMATIKA DAN ILMU KOMPUTER (JIMA-ILKOM)
E-ISSN 2988-747X P-ISSN 2988-7461
Volume 3, Nomor 1, Maret 2024, Page 46-56
DOI: https://doi.org/10.58602/jima-ilkom.v3i1.26



Gambar 1. Tahapan Penelitian

Pengumpulan Data

Pengumpulan data dalam evaluasi penilaian siswa merupakan langkah penting untuk mendapatkan pemahaman yang komprehensif tentang kemajuan belajar dan pencapaian siswa. Proses ini melibatkan berbagai metode dan instrumen yang dirancang untuk mengukur berbagai aspek kinerja siswa. Pentingnya pengumpulan data yang berkualitas dalam evaluasi penilaian siswa tidak hanya untuk mengevaluasi kemajuan individu siswa, tetapi juga untuk mengevaluasi efektivitas kurikulum, strategi pengajaran, dan program pembelajaran secara keseluruhan. Dengan pengumpulan data yang komprehensif dan beragam, pengambilan keputusan yang berkualitas dapat dibuat untuk meningkatkan kualitas pendidikan dan membantu setiap siswa mencapai potensi maksimal mereka.

***Cleaning* Data**

Pembersihan data, atau yang sering disebut sebagai tahap *pre-processing*, merupakan langkah penting dalam penelitian data mining. Pada tahap ini, data yang telah dikumpulkan dievaluasi untuk mengidentifikasi dan mengatasi masalah seperti data yang hilang, duplikat, tidak konsisten, atau tidak relevan. Proses pembersihan data melibatkan beberapa langkah, termasuk identifikasi dan penghapusan entri data yang kosong atau tidak lengkap, penanganan outlier yang mungkin mengganggu analisis, dan penggabungan atau pemisahan entri data yang redundan atau tidak konsisten. Selain itu, normalisasi atau standarisasi data juga dapat dilakukan untuk memastikan konsistensi dalam format dan skala data. Langkah-langkah ini bertujuan untuk mempersiapkan data yang bersih dan terstruktur agar analisis data mining dapat dilakukan dengan tepat dan hasil yang dihasilkan menjadi lebih akurat dan bermakna. Dengan melakukan pembersihan data yang cermat, peneliti dapat meminimalkan bias dan mengoptimalkan kualitas data yang digunakan dalam proses penelitian.

Clustering

Clustering merupakan salah satu teknik penting dalam data *mining* yang digunakan untuk mengelompokkan data menjadi subset yang serupa berdasarkan karakteristik atau pola tertentu. Tujuan utama dari clustering adalah untuk mengidentifikasi struktur yang tersembunyi dalam data, yang dapat membantu dalam pemahaman lebih lanjut tentang kelompok atau kategori yang ada di dalamnya. Algoritma *K-Means* adalah salah satu algoritma *clustering* yang paling umum dan sederhana yang digunakan dalam analisis data. Algoritma ini bekerja dengan cara mengelompokkan titik-titik data ke dalam k kelompok (*clusters*), di mana setiap titik data tergabung dalam kelompok yang memiliki pusat (*centroid*) yang terdekat dengannya. Langkah-langkah utama dalam algoritma *K-Means* adalah sebagai berikut:

a. Inisialisasi: Pilih secara acak k titik awal sebagai pusat kelompok (*centroid*). Titik-titik ini bisa dipilih dari data atau secara acak.
b. Pembentukan Kelompok: Setiap titik data dikelompokkan ke dalam kelompok yang memiliki pusat terdekat (*centroid*) dengannya. Ini dapat dihitung menggunakan jarak *Euclidean* atau metrik jarak lainnya dengan menggunakan persamaan berikut ini.

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad (1)$$

Nirwana Hendrastuty: *Penulis Korespondensi

JURNAL ILMIAH INFORMATIKA DAN ILMU KOMPUTER (JIMA-ILKOM)
E-ISSN 2988-747X P-ISSN 2988-7461
Volume 3, Nomor 1, Maret 2024, Page 46-56
DOI: https://doi.org/10.58602/jima-ilkom.v3i1.26

c. Pembaruan Pusat: Hitung ulang pusat setiap kelompok dengan mengambil rata-rata dari semua titik data yang termasuk dalam kelompok tersebut.
d. Iterasi: Langkah 2 dan 3 diulangi hingga tidak ada perubahan yang signifikan dalam posisi pusat kelompok atau titik-titik data tidak berpindah kelompok.
e. Penyelesaian: Proses dihentikan ketika kriteria penghentian telah terpenuhi, misalnya, ketika tidak ada lagi perubahan dalam posisi pusat kelompok atau ketika jumlah iterasi maksimum telah tercapai.

$$R_k = \frac{1}{N_k}(X_{1k} + \cdots X_{nk}) \quad (2)$$

Hasil akhir dari algoritma K-Means adalah k kelompok di mana setiap kelompok memiliki pusatnya sendiri. Algoritma ini bergantung pada inisialisasi pusat kelompok yang baik dan dapat menghasilkan solusi yang berbeda tergantung pada titik awal yang dipilih. Oleh karena itu, sering kali disarankan untuk menjalankan algoritma beberapa kali dengan inisialisasi yang berbeda dan memilih solusi terbaik berdasarkan kriteria tertentu, seperti nilai inersia atau jarak antara titik data dan pusat kelompok.

Hasil

Hasil dari algoritma *K-Means* adalah sebuah model yang dapat digunakan untuk memprediksi kelompok mana suatu data baru akan termasuk, serta posisi pusat dari masing-masing kelompok. Evaluasi model dapat dilakukan menggunakan metrik seperti *within-cluster sum of squares* (WCSS) dan *silhouette score* untuk mengevaluasi seberapa baik pembagian kelompok yang dihasilkan oleh algoritma ini.

3.HASIL DAN PEMBAHASAN

Penerapan data *mining* menggunakan algoritma *K-Means clustering* dalam evaluasi hasil pembelajaran siswa dapat memberikan wawasan yang berharga bagi pendidik untuk memahami pola dan karakteristik dari pencapaian siswa. Langkah-langkah penerapan algoritma *K-Means clustering* dalam evaluasi hasil pembelajaran siswa dapat diuraikan sebagai berikut:

a. Pengumpulan Data
Data hasil pembelajaran siswa meruapakan data nilai akhir dari mata pelajaran yang didapat oleh siswa dikumpulkan dari berbagai sumber, data hasil pembelajaran siswa seperti pada tabel 1.

Tabel 1. Dataset Hasil Pembelajaran Siswa

Nama	PAI	PPKn	BINA	MTK	IPA	IPS	BING	BALAM	SENI	PJOK	PRAK	TIK
Abdullah Ibnu Aziz Hsb	87	82	81	78	81	86	82	79	82	83	81	83
Afreza Peneduh	76	74	73	73	73	76	76	77	77	78	75	75
Zalika Nurul Fadilla	80	73	74	78	77	77	80	76	76	81	78	80
Zaskia Indira Kurniawan	81	82	80	80	85	86	90	89	79	82	78	85
Ziyan Azzahra	82	81	76	81	79	82	89	87	79	83	78	82

b. *Pre-processing* Data
Pre-processing data adalah tahap penting dalam proses data *mining* yang melibatkan persiapan dan pembersihan data mentah sebelum dilakukan analisis lebih lanjut. Langkah-langkah dalam pre-processing data mencakup identifikasi dan penanganan nilai yang hilang atau tidak lengkap, deteksi dan penanganan outlier, normalisasi atau standarisasi data untuk menghilangkan perbedaan skala, serta penghapusan atau transformasi fitur yang tidak relevan atau redundan. Tujuan dari *pre-processing* data adalah untuk memastikan kualitas data yang baik, meminimalkan noise atau gangguan, dan mengoptimalkan data untuk analisis selanjutnya, sehingga memungkinkan hasil yang lebih akurat dan bermakna dari proses data *mining*. Dalam pre-processing tetap merupakan langkah penting dalam mempersiapkan data untuk analisis lebih

Nirwana Hendrastuty: *Penulis Korespondensi

JURNAL ILMIAH INFORMATIKA DAN ILMU KOMPUTER (JIMA-ILKOM)
E-ISSN 2988-747X P-ISSN 2988-7461
Volume 3, Nomor 1, Maret 2024, Page 46-56
DOI: https://doi.org/10.58602/jima-ilkom.v3i1.26

lanjut, membantu memastikan bahwa data siap untuk digunakan dalam proses data *mining* dengan hasil yang lebih akurat dan dapat diandalkan.

Algoritma *K-Means Clustering*

Algoritma *K-Means Clustering* merupakan salah satu metode *clustering* yang paling umum digunakan dalam analisis data. Tujuan utamanya adalah untuk mengelompokkan data ke dalam sejumlah kategori atau klaster yang berbeda, di mana setiap titik data termasuk ke dalam klaster dengan pusat terdekat. Berikut adalah langkah-langkah utama dalam algoritma *K-Means*:

a. Inisialisasi

Tahap ini kita menentukan jumlah klaster yang diinginkan (K) dan acak pilih K titik sebagai pusat awal klaster. Jumlah klister yang akan kita gunakan yaitu 2 klaster dengan nilai *centroid* awal seperti pada tabel 2.

Tabel 2. Inisialisasi *Centroid*

Cluster	PAI	PPKn	BINA	MTK	IPA	IPS	BING	BALAM	SENI	PJOK	PRAK	TIK
C1	76	77	74	74	76	81	81	77	78	78	77	75
C2	81	82	80	80	85	86	90	89	79	82	78	85

b. *Assignment*

Setiap titik data didistribusikan ke klaster yang memiliki pusat terdekat dengan menggunakan suatu metrik jarak, biasanya jarak *Euclidean* dengan menggunakan persamaan (1), hasil perhitungan sebagai berikut.

Perhitungan data atas nama siswa Abdullah Ibnu Aziz Hsb ke *cluster* 0 yaitu.

$$d(1,1) = \sqrt{\begin{matrix}(87-76)^2-(82-77)^2-(81-74)^2-(78-74)^2-(81-76)^2-(86-81)^2 \\ -(82-81)^2-(79-77)^2-(82-78)^2-(83-78)^2-(81-77)^2-(83-75)^2\end{matrix}}$$

Hasil perhitungan *cluster* dapat dilihat pada tabel 3.

Tabel 3. Iterasi Awal

Nama	C1	C2	Cluster
Abdullah Ibnu Aziz Hsb	20,14	15,62	C2
Afreza Peneduh	12,60	30	C1
Zalika Nurul Fadilla	12,88	23,89	C1
Zaskia Indira Kurniawan	23,60	0	C2
Ziyan Azzahra	19,82	9,27	C2

Selanjutnya menghitung ulang pusat setiap kelompok dengan mengambil rata-rata dari semua titik data yang termasuk dalam kelompok tersebut. Hasil perhitungan *centroid* yaitu sebagai berikut.

Tabel 2. *Centroid* Kedua

Cluster	PAI	PPKn	BINA	MTK	IPA	IPS	BING	BALAM	SENI	PJOK	PRAK	TIK
C1	78	73,5	73,5	75,5	75	76,5	78	76,5	76,5	79,5	76,5	77,5
C2	83,33	81,67	79	79,67	81,67	84,67	87	85	80	82,67	79	83,33

Hasil perhitungan *cluster* untuk *centroid* kedua dapat dilihat pada tabel 4.

Tabel 4. Iterasi Kedua

Nama	C1	C2	Cluster
Abdullah Ibnu Aziz Hsb	22,58318	9,581352	C2
Afreza Peneduh	10,3923	25,38863	C1

Nirwana Hendrastuty: *Penulis Korespondensi

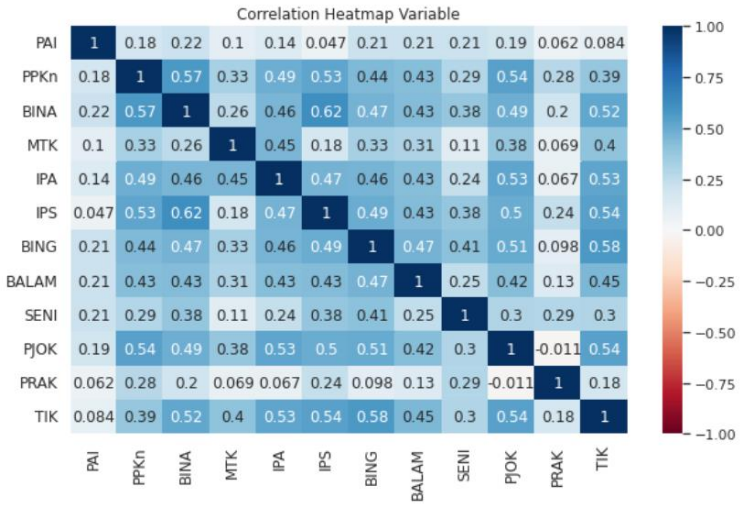
JURNAL ILMIAH INFORMATIKA DAN ILMU KOMPUTER (JIMA-ILKOM)
E-ISSN 2988-747X P-ISSN 2988-7461
Volume 3, Nomor 1, Maret 2024, Page 46-56
DOI: https://doi.org/10.58602/jima-ilkom.v3i1.26

Zalika Nurul Fadilla	10,44031	18,85636	C1
Zaskia Indira Kurniawan	27,11088	7,052822	C2
Ziyan Azzahra	22,22611	6,254782	C2

Pada iterasi kedua tidak ada perubahan dalam *cluster* pada saat iterasi pertama dan kedua, maka proses dihentikan dengan jumlah cluster sebanyak 2 dengan melakukan 2 iterasi.

Implementasi Algoritma *K-Means Clustering* Dalam Hasil Pembelajaran Siswa

Implementasi algoritma *K-Means clustering* dalam hasil pembelajaran siswa dapat memberikan wawasan yang berharga bagi pendidik untuk memahami pola-pola dalam pencapaian akademis siswa dan mendukung pengambilan keputusan yang lebih baik dalam perencanaan dan pengembangan kurikulum serta strategi pembelajaran. Dalam algoritma k-means, korelasi antar variabel memiliki dampak yang signifikan terhadap pembentukan kelompok data. K-means mengasumsikan bahwa variabel yang digunakan untuk mengukur kesamaan antar data memiliki hubungan yang tidak terlalu kuat, karena algoritma ini bergantung pada perhitungan jarak Euclidean. Jika terdapat korelasi tinggi antar variabel, hal ini dapat menyebabkan distorsi dalam pengelompokan data, di mana pengaruh dominan dari satu variabel dapat mengaburkan peran variabel lainnya dalam pembentukan kelompok. Oleh karena itu, sebelum menerapkan k-means, penting untuk melakukan analisis korelasi antar variabel dan mungkin mempertimbangkan teknik reduksi dimensi atau pemilihan fitur untuk mengurangi dampak dari korelasi yang tinggi tersebut. Hasil korelati antar variabel dapat dilihat pada Gambar 2.



Gambar 2. *Correlation Heatmap Variable*

Dalam k-means, korelasi antar variabel mengacu pada tingkat hubungan atau interdependensi antara variabel yang digunakan dalam proses pengelompokan data. Algoritma k-means mengoperasikan data dalam ruang berdimensi n, di mana setiap variabel mewakili satu dimensi dalam ruang tersebut. Ketika terdapat korelasi yang tinggi antara variabel, artinya ada hubungan yang kuat antara nilai variabel satu dengan yang lainnya.

Hasil dari proses clustering data nilai siswa dapat memberikan wawasan yang berharga tentang pola-pola dalam prestasi akademik mereka. Dengan menggunakan algoritma k-means, siswa dapat dikelompokkan ke dalam cluster berdasarkan kesamaan dalam nilai-nilai yang mereka peroleh dalam berbagai mata pelajaran. Interpretasi hasil clustering ini dapat memberikan pemahaman yang lebih mendalam tentang karakteristik setiap kelompok siswa, seperti tingkat kemampuan akademik yang serupa atau kecenderungan dalam prestasi di bidang tertentu. Informasi ini dapat digunakan untuk mengidentifikasi siswa yang membutuhkan perhatian tambahan atau untuk merancang strategi pembelajaran yang lebih sesuai dengan kebutuhan individu dari masing-masing kelompok siswa.

Nirwana Hendrastuty: *Penulis Korespondensi

JURNAL ILMIAH INFORMATIKA DAN ILMU KOMPUTER (JIMA-ILKOM)
E-ISSN 2988-747X P-ISSN 2988-7461
Volume 3, Nomor 1, Maret 2024, Page 46-56
DOI: https://doi.org/10.58602/jima-ilkom.v3i1.26

Dalam penelitian ini cluster yang dibentuk berjumlah 2 cluster, hasil nilai *cluster* seperti pada gambar 3.

Segment K-means	PAI	PPKn	BINA	MTK	IPA	IPS	BING	BALAM	SENI	PJOK	PRAK	TIK
0	76.934066	76.10989	75.087912	74.373626	76.824176	79.021978	79.021978	77.978022	77.571429	78.868132	78.659341	76.945055
1	80.507937	81.52381	80.936508	76.000000	81.301587	83.523810	84.619048	82.650794	80.095238	81.984127	79.333333	80.714286

Gambar 3. *Correlation Heatmap Variable*

Segmentasi K-Means pada Gambar 3 merupakan salah satu teknik analisis klaster yang digunakan dalam analisis data untuk mengelompokkan data ke dalam segmen atau kelompok yang berbeda berdasarkan kesamaan fitur tertentu. Algoritma K-Means bekerja dengan cara membagi data ke dalam K klaster, di mana setiap data tergabung dalam klaster yang memiliki rata-rata fitur yang paling dekat dengan data tersebut.

Hasil dari clustering data penilaian belajar siswa dapat memberikan wawasan yang signifikan dalam pemahaman terhadap pola-pola dalam prestasi akademik mereka. Dengan menerapkan metode seperti K-Means, siswa dapat dikelompokkan ke dalam segmen berdasarkan kesamaan dalam kinerja mereka dalam berbagai mata pelajaran. Interpretasi dari hasil clustering ini memungkinkan untuk mengidentifikasi kelompok siswa dengan karakteristik serupa, seperti tingkat kemampuan yang sebanding atau kecenderungan dalam pencapaian akademik di bidang tertentu. Informasi ini dapat sangat bermanfaat dalam merancang program pembelajaran yang disesuaikan secara individual, memungkinkan pendidik untuk memberikan perhatian khusus pada setiap kelompok siswa sesuai dengan kebutuhan mereka, serta meningkatkan efektivitas pengajaran secara keseluruhan. Hasil dari *clustering* data penilaian siswa dapat dilihat pada Gambar 4.

NAMA	PAI	PPKn	BINA	MTK	IPA	IPS	BING	BALAM	SENI	PJOK	PRAK	TIK	Cluster
ABDULLAH IBNU AZIZ HSB	87	82	81	78	81	86	82	79	82	83	81	83	1
AFREZA PENEDUH	76	74	73	73	73	76	76	77	77	78	75	75	0
ALDILA WATI	77	82	83	76	78	86	80	85	79	82	78	76	1
Aldo Hartadiansyah	76	77	74	74	76	81	81	77	78	78	77	75	0
Aulia Puspita Sari	78	82	90	75	81	87	82	82	81	82	80	86	1
...	...	...	...	...	...	...	...	...	...	...	...	...	...
Windi	75	73	73	77	76	75	79	77	75	78	77	75	0
Zahra Alliya Dama Elvaretta	80	77	76	79	83	81	86	79	81	82	78	81	1
Zalika Nurul Fadilla	80	73	74	78	77	77	80	76	76	81	78	80	0
Zaskia Indira Kurniawan	81	82	80	80	85	86	90	89	79	82	78	85	1
ZIYAN AZZAHRA	82	81	76	81	79	82	89	87	79	83	78	82	1

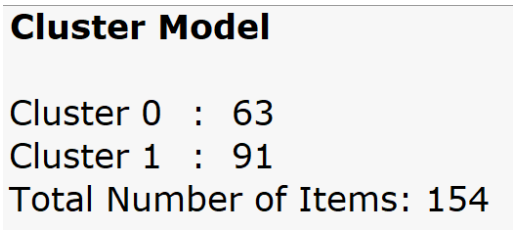
Gambar 4. Hasil *Clustering*

Hasil *clustering* data penilaian belajar siswa dapat mengungkap pola yang bermanfaat bagi pendidik dan administrator sekolah. Analisis terhadap *cluster-cluster* ini dapat mengungkapkan informasi tentang tren prestasi, kecenderungan keberhasilan atau kesulitan dalam mata pelajaran tertentu, serta memungkinkan identifikasi siswa yang memerlukan bantuan tambahan.

Model *clustering* dari data penilaian belajar siswa merupakan alat analisis yang kuat dalam memahami pola-pola kompleks dalam prestasi akademik siswa. Dengan menerapkan teknik clustering seperti K-Means, data penilaian belajar siswa dapat dikelompokkan ke dalam segmen-segmen yang berbeda berdasarkan kesamaan nilai mereka dalam berbagai mata pelajaran. Hasil *clustering* model seperti pada Gambar 5.

Nirwana Hendrastuty: *Penulis Korespondensi

JURNAL ILMIAH INFORMATIKA DAN ILMU KOMPUTER (JIMA-ILKOM)
E-ISSN 2988-747X P-ISSN 2988-7461
Volume 3, Nomor 1, Maret 2024, Page 46-56
DOI: https://doi.org/10.58602/jima-ilkom.v3i1.26

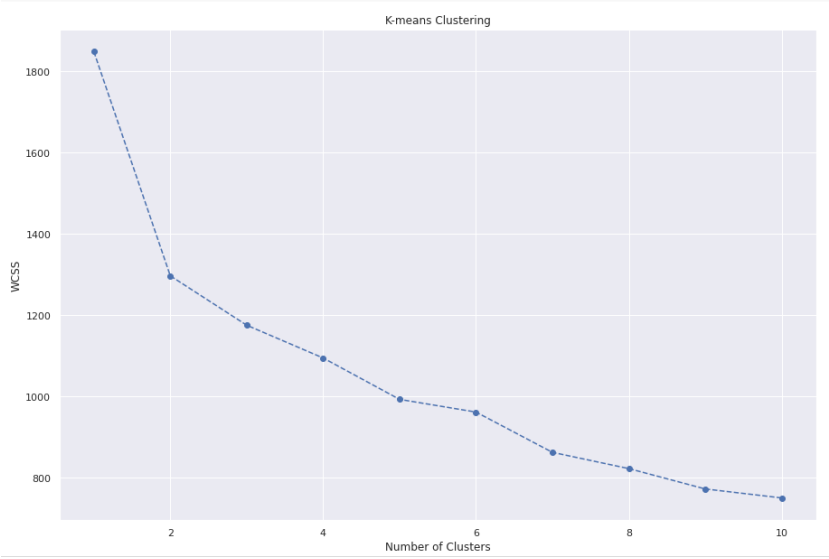


Gambar 5. Hasil Jumlah *Clustering*

Pengelompokan hasil *cluster* berdasarkan data penilaian siswa dengan k-means memperoleh 2 kelompok siswa yaitu siswa Rajin dengan kelompok C0 dan kelompok siswa Sangat Rajin dengan kelompok C1. Kelompok C0 siswa Rajin terdiri dari 63 siswa dan kelompok C1 terdiri dari 91 siswa Sangat Rajin.

Pengujian *Within-Cluster Sum of Squares* (WCSS)

Pengujian *Within-Cluster Sum of Squares* (WCSS) adalah metrik evaluasi yang umum digunakan dalam analisis *clustering*, khususnya pada algoritma *K-Means*. WCSS mengukur seberapa jauh setiap titik data dalam sebuah kluster berjarak dari pusat kluster (*centroid*) yang sesuai. Metrik ini menghitung jumlah kuadrat jarak antara setiap titik data dan *centroid* kluster tempat titik tersebut berada, kemudian menjumlahkan hasilnya untuk semua titik dalam kluster. Tujuan dari *K-Means clustering* adalah untuk meminimalkan nilai WCSS, dengan cara memindahkan *centroid* kluster sedemikian rupa sehingga total jarak kuadrat dari setiap titik data ke *centroid* kluster yang sesuai diminimalkan. Dengan demikian, WCSS digunakan sebagai kriteria untuk mengevaluasi seberapa baik model *k-Means* telah memisahkan data ke dalam kelompok yang sesuai dengan karakteristiknya. Semakin rendah nilai WCSS, semakin baik pula clustering yang dihasilkan. Hasil pengujian WCSS seperti pada gambar berikut.



Gambar 5. Pengujian WSCC

Hasil pengujian menggunakan WCSS diatas menunjukkan bahwa jumlah kluster yang baik yaitu berjumlah 2 klaster, hal ini sesuai dengan yang telah ditetapkan dalam penelitian ini yaitu sebanyak 2 klaster.

Pengujian *Silhouette Score*

Pengujian *silhouette score* adalah proses evaluasi yang digunakan untuk mengukur seberapa baik suatu algoritma clustering mampu memisahkan data menjadi kelompok-kelompok yang berbeda. Metode ini melibatkan perhitungan *silhouette score* untuk setiap titik data, yang mencerminkan seberapa baik titik tersebut cocok dengan kelompoknya sendiri dibandingkan dengan kelompok lainnya.

Nirwana Hendrastuty: *Penulis Korespondensi

JURNAL ILMIAH INFORMATIKA DAN ILMU KOMPUTER (JIMA-ILKOM)
E-ISSN 2988-747X P-ISSN 2988-7461
Volume 3, Nomor 1, Maret 2024, Page 46-56
DOI: https://doi.org/10.58602/jima-ilkom.v3i1.26

Pengujian *silhouette score* dilakukan dengan membagi data menjadi kelompok-kelompok dengan berbagai jumlah cluster yang berbeda dan membandingkan nilai *silhouette score* untuk masing-masing pengelompokan. Nilai *silhouette score* yang tinggi menandakan bahwa *clustering* tersebut baik, sementara nilai yang rendah menunjukkan bahwa ada penyebaran yang tidak homogen dalam satu atau lebih kelompok. Oleh karena itu, pengujian *silhouette score* membantu dalam pemilihan jumlah kluster yang optimal untuk clustering suatu dataset. Setelah menghitung *silhouette score* untuk berbagai skenario clustering dengan jumlah kluster yang berbeda, langkah selanjutnya dalam pengujian adalah memilih jumlah kluster yang memberikan nilai *silhouette score* tertinggi. Jumlah kluster ini merupakan pilihan yang optimal untuk membagi data menjadi kelompok-kelompok yang saling terpisah dengan baik. Pengujian *silhouette score* juga dapat membantu dalam mengevaluasi performa algoritma *clustering* yang digunakan, memungkinkan penyesuaian parameter atau pemilihan metode *clustering* yang lebih baik jika hasilnya tidak memuaskan. Hasil pengujian *silhouette score* memberikan panduan yang berharga dalam proses pengelompokan data yang efektif dan informatif seperti pada gambar 6.

```
from sklearn.metrics import silhouette_score
from sklearn.cluster import KMeans

for n_cluster in range(2, 6):
    kmeans = KMeans(n_clusters=n_cluster).fit(X)
    label = kmeans.labels_
    sil_coeff = silhouette_score(X, label, metric='euclidean')
    print("For n_clusters={}, The Silhouette Coefficient is {}".format(n_cluster, sil_coeff))
```

```
For n_clusters=2, The Silhouette Coefficient is 0.9168280530665532
For n_clusters=3, The Silhouette Coefficient is 0.4893575334428582
For n_clusters=4, The Silhouette Coefficient is 0.4832492381375905
For n_clusters=5, The Silhouette Coefficient is 0.4039407450993525 7
```

Gambar 6. Pengujian *Silhouette Score*

Berdasarkan hasil pengujian *silhouette score* gambar 6, dapat disimpulkan bahwa pemilihan jumlah kluster sebanyak 2 sangat optimal sangat penting dalam pembentukan kelompok-kelompok yang homogen dalam sebuah dataset. Nilai *silhouette score* untuk klaster 2 yang paling tinggi sebasar 0,9168 dan menunjukkan bahwa pengelompokan data menjadi kelompok-kelompok tersebut lebih baik, penggunaan *silhouette score* sebagai metrik evaluasi memberikan panduan yang berguna dalam menentukan jumlah kluster yang optimal dalam analisis *clustering*, serta membantu dalam peningkatan kualitas hasil clustering dan interpretasi data.

4.KESIMPULAN

Tujuan penerapan data *mining* menggunakan algoritma *K-Means Clustering* dalam evaluasi hasil pembelajaran siswa adalah untuk mengidentifikasi pola-pola yang mungkin tersembunyi dalam data hasil pembelajaran, membagi siswa ke dalam kelompok-kelompok berdasarkan tingkat pencapaian atau karakteristik pembelajaran mereka, serta memberikan wawasan yang berharga kepada guru dan *stakeholder* pendidikan. Hasil *clustering* data penilaian belajar siswa dapat mengungkap pola yang bermanfaat bagi pendidik dan administrator sekolah. Analisis terhadap *cluster-cluster* ini dapat mengungkapkan informasi tentang tren prestasi, kecenderungan keberhasilan atau kesulitan dalam mata pelajaran tertentu, serta memungkinkan identifikasi siswa yang memerlukan bantuan tambahan.Pengelompokan hasil *cluster* berdasarkan data penilaian siswa dengan k-means memperoleh 2 kelompok siswa yaitu siswa Rajin dengan kelompok C0 dan kelompok siswa Sangat Rajin dengan kelompok C1. Kelompok C0 siswa Rajin terdiri dari 63 siswa dan kelompok C1 terdiri dari 91 siswa Sangat Rajin. Hasil pengujian *silhouette score* dapat disimpulkan bahwa pemilihan jumlah kluster sebanyak 2 sangat optimal sangat penting dalam pembentukan kelompok-kelompok yang homogen dalam sebuah dataset. Nilai *silhouette score* untuk klaster 2 yang paling tinggi sebasar 0,9168 dan menunjukkan bahwa pengelompokan data menjadi kelompok-kelompok tersebut lebih baik, penggunaan *silhouette score* sebagai metrik evaluasi memberikan panduan yang berguna dalam menentukan jumlah kluster yang optimal dalam analisis *clustering*, serta membantu dalam peningkatan kualitas hasil clustering dan interpretasi data

Nirwana Hendrastuty: *Penulis Korespondensi

E-ISSN 2988-747X P-ISSN 2988-7461
Volume 3, Nomor 1, Maret 2024, Page 46-56
DOI: https://doi.org/10.58602/jima-ilkom.v3i1.26

5.REFERENSI

[1] M. Syahril, K. Erwansyah, and M. Yetri, "Penerapan Data Mining untuk menentukan pola penjualan peralatan sekolah pada brand wigglo dengan menggunakan algoritma apriori," *J-SISKO TECH (Jurnal Teknol. Sist. Inf. dan Sist. Komput. TGD)*, vol. 3, no. 1, pp. 118–136, 2020.

[2] M. J. Zaki and W. J. Meira, *Data Mining and Machine Learning Fundamental Concepts and Algorithms*, vol. 53, no. 9. 2020.

[3] G. Gustientiedina, M. H. Adiya, and Y. Desnelita, "Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan," *J. Nas. Teknol. dan Sist. Inf.*, vol. 5, no. 1, pp. 17–24, 2019, doi: 10.25077/teknosi.v5i1.2019.17-24.

[4] M. S. Sungkar and M. T. Qurohman, "Penerapan Algoritma C5.0 Untuk Prediksi Kelulusan Pembelajaran Mahasiswa Pada Matakuliah Arsitektur Sistem Komputer," *J. Media Inform. Budidarma*, vol. 5, no. 3, p. 1166, 2021, doi: 10.30865/mib.v5i3.3116.

[5] A. Zheng, J. Cai, H. Yang, and X. Zhao, "CPGAN: Curve Clustering Architecture Based on Projected Latent Vector of Generative Adversarial Network," *IEEE Access*, vol. 8. Institute of Electrical and Electronics Engineers (IEEE), pp. 86765–86776, 2020. doi: 10.1109/access.2020.2992887.

[6] M. A. Syakur, B. K. Khotimah, E. M. S. Rochman, and B. D. Satoto, "Integration k-means clustering method and elbow method for identification of the best customer profile cluster," in *IOP conference series: materials science and engineering*, 2018, vol. 336, no. 1, p. 12017.

[7] M. Cui, "Introduction to the k-means clustering algorithm based on the elbow method," *Accounting, Audit. Financ.*, vol. 1, no. 1, pp. 5–8, 2020.

[8] A. Yudhistira and R. Andika, "Pengelompokan Data Nilai Siswa Menggunakan Metode K-Means Clustering," *J. Artif. Intell. Technol. Inf.*, vol. 1, no. 1, pp. 20–28, 2023.

[9] N. Noviyanto, "Penerapan Data Mining dalam Mengelompokkan Jumlah Kematian Penderita COVID-19 Berdasarkan Negara di Benua Asia," *Paradig. - J. Komput. dan Inform.*, vol. 22, no. 2, pp. 183–188, 2020, doi: 10.31294/p.v22i2.8808.

[10] A. A. Aldino, D. Darwis, A. T. Prastowo, and C. Sujana, "Implementation of K-Means Algorithm for Clustering Corn Planting Feasibility Area in South Lampung Regency," in *Journal of Physics: Conference Series*, 2021, vol. 1751, no. 1, p. 12038.

[11] J. Hutagalung, "Pemetaan Siswa Kelas Unggulan Menggunakan Algoritma K-Means Clustering," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 1, pp. 606–620, 2022.

[12] Q. I. Mawarni and E. S. Budi, "Implementasi Algoritma K-Means Clustering Dalam Penilaian Kedisiplinan Siswa," *J. Sist. Komput. dan Inform.*, vol. 3, no. 4, pp. 522–528, 2022.

[13] A. Rohmah, F. Sembiring, and A. Erfina, "Implementasi Algoritma K-Means Clustering Analysis Untuk Menentukan Hambatan Pembelajaran Daring (Studi Kasus: Smk Yaspim Gegerbitung)," in *Prosiding Seminar Nasional Sistem Informasi dan Manajemen Informatika Universitas Nusa Putra*, 2021, vol. 1, no. 01, pp. 290–298.

[14] F. P. Dewi, P. S. Aryni, and Y. Umaidah, "Implementasi Algoritma K-Means Clustering Seleksi Siswa Berprestasi Berdasarkan Keaktifan dalam Proses Pembelajaran," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 7, no. 2, pp. 111–121, 2022.

[15] Z. Nabila, A. R. Isnain, P. Permata, and Z. Abidin, "ANALISIS DATA MINING UNTUK CLUSTERING KASUS COVID-19 DI PROVINSI LAMPUNG DENGAN ALGORITMA K-MEANS," *J. Teknol. dan Sist. Inf.*, vol. 2, no. 2, pp. 100–108, 2021.

Nirwana Hendrastuty: *Penulis Korespondensi