PENGELOMPOKAN HASIL BELAJAR SISWA SDN TUNAS JAYA DENGAN ALGORITMA *K-MEANS*

Aditya Pramudya, Iqbal Maulana, Rini Mayasari.
Informatika, Universitas Singaperbangsa Karawang
Jl. HS.Ronggo Waluyo, Puseurjaya, Telukjambe Timur, Karawang, Jawa Barat, Indonesia
aditya.pramudya19153@student.unsika.ac.id

ABSTRAK

Pendidikan memegang peran kunci dalam membentuk karakter individu. Seiring penyebaran COVID-19, Menteri Pendidikan dan Kebudayaan Republik Indonesia merekomendasikan pembelajaran *online* di seluruh sekolah, termasuk SDN Tunas Jaya. Dampak pandemi ini menurunkan capaian hasil belajar siswa secara signifikan. Penelitian ini bertujuan mengelompokkan perkembangan hasil belajar siswa SDN Tunas Jaya menggunakan algoritma *K-Means* dalam *Data Mining*. Metode *Elbow* menentukan optimalitas *cluster*, menghasilkan 4 *cluster* dengan SSE 0,400. Evaluasi Silhouette *Method* (0,399) dan Davies-Bouldin Index (0,749) mengonfirmasi 4 *cluster*. Dari 141 siswa, *cluster* 0 "rendah" (35 siswa), *cluster* 1 "cukup" (46 siswa), *cluster* 2 "baik" (25 siswa), dan *cluster* 3 "sangat baik" (35 siswa). Analisis rata-rata hasil belajar menunjukkan penurunan pada beberapa mata pelajaran. Hasil ini diharapkan memberikan informasi untuk mengidentifikasi strategi pembelajaran yang perlu diperbaiki.

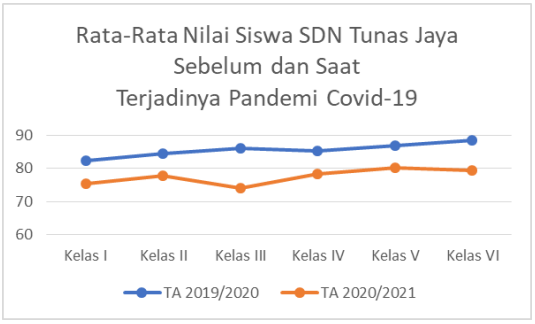
Kata kunci : *:* Algoritma *K-Means, Knowledge Discovery in Database, Data Mining*

1. PENDAHULUAN

Pendidikan merupakan faktor utama yang berperan dalam membentuk pribadi manusia. Dalam Undang-Undang No. 20 Tahun 2003 tentang sistem pendidikan nasional pasal 3 menyebutkan bahwa pendidikan nasional berfungsi mengembangkan kemampuan dan membentuk watak serta peradaban bangsa yang bermartabat dalam rangka mencerdaskan kehidupan bangsa, bertujuan untuk mengembangkan potensi peserta didik agar menjadi manusia yang beriman bertaqwa kepada Tuhan Yang Maha Esa, berakhlak mulia, sehat, berilmu, cakap, kreatif dan menjadi warga Negara yang demokratis serta bertanggung jawab [1].

Berdasarkan pengumuman resmi yang dikeluarkan oleh Menteri Pendidikan dan Kebudayaan Republik Indonesia pada tahun 2020 mengenai pelaksanaan kebijakan pendidikan selama masa darurat penyebaran penyakit *coronavirus disease* (Covid-19), seluruh sekolah yang ada di Indonesia dianjurkan untuk melakukan kegiatan belajar mengajar secara *online*. Pembelajaran secara *online* dinilai sering tidak berjalan dengan maksimal, ditandai dengan munculnya beberapa kendala dalam penyampaian materi pembelajaran, diantaranya seperti masalah kestabilan jaringan ataupun kendala terkait beberapa siswa yang tidak memiliki perangkat memadai untuk mengikuti proses pembelajaran.

Dikarenakan penyebaran covid-19 sudah mereda, saat ini siswa sudah mulai kembali melaksanakan kegiatan pembelajaran secara tatap muka (luring) mulai dari SD, SMP, hingga SMA. Namun dampak dari penerapan sistem pembelajaran *online* tersebut menyebabkan penurunan prestasi siswa. Salah satu sekolah yang terdampak dari pandemi covid-19 ialah SDN Tunas Jaya yang merupakan sekolah berlokasi di Dusun Bungurgede, Desa Sukahaji, Kecamatan Ciasem, Kabupaten Subang. Dampak pandemi covid-19 dapat dilihat dari capaian hasil belajar siswa yang menurun seperti pada Gambar 1.



Gambar 1. Data rata-rata hasil belajar siswa SDN Tunas Jaya

Berdasarkan Gambar 1, dapat kita lihat bahwa sebelum terjadinya pandemi covid-19 rata-rata capaian hasil belajar siswa pada gambar tersebut terbilang cukup baik. Namun pada saat terjadinya pandemi covid-19 rata-rata capaian hasil belajar ini mengalami penurunan secara signifikan dibandingkan dengan sebelum pandemi covid-19.

Oleh karena itu, perlu dilakukan upaya untuk meningkatkan prestasi siswa, salah satu cara yang dapat dilakukan adalah dengan melakukan analisis terhadap hasil belajar siswa untuk mencari tahu pada capaian mata pelajaran mana sajakah yang mengalami penurunan prestasi siswa akibat pandemi covid-19. Teknik yang dapat digunakan dalam analisis tersebut adalah *data mining*, salah satunya adalah dengan menggunakan algoritma *K-Means* untuk melakukan *clustering* hasil belajar siswa.

Data mining adalah proses yang memanfaatkan satu atau lebih teknik pembelajaran mesin (*machine*

learning) guna secara otomatis menganalisis dan mengekstraksi pengetahuan. Tujuan dari data mining adalah untuk mencari tren atau pola yang diinginkan dalam basis data yang luas, sehingga dapat mendukung pengambilan keputusan di masa mendatang [2]. Algoritma *K-Means* yang merupakan salah satu metode *clustering non-hierarchical* dalam *Data Mining* yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih *cluster*. Algoritma ini direkomendasikan untuk digunakan karena memiliki kelebihan dimana mudah diterapkan, kesederhanaannya dalam mengelompokan dataset yang besar, cepat dan mudah untuk diadaptasi, dan paling banyak dipraktekan dalam proses *clustering* pada *data mining* [3].

Beberapa penelitian terdahulu yang menerapkan Algoritma *K-Means* diantaranya penelitian yang dilakukan oleh [2] yang berjudul "Penerapan *Data Mining* Dalam Pengelompokkan Data Nilai Siswa Untuk Penentuan Jurusan Siswa Pada SMA Tamora Menggunakan Algoritma *K-Means Clustering*". Penelitian tersebut menghasilkan bahwa Algoritma *K-Means* dapat diterapkan pada SMA Tamora untuk menganalisis permasalahan yang ada yang berkenaan dengan pengelompokkan data nilai siswa untuk penentuan jurusan. Akan tetapi dalam penelitian ini belum dilakukan uji validitas, baik menggunakan DBI maupun uji validitas yang lain.

Selanjutnya [4] dengan penelitian yang berjudul "Analisa Penentuan *Cluster* Terbaik pada Metode *K-Means* Menggunakan *Elbow* Terhadap Sentra Industri Produksi di Pamekasan". Dari hasil penelitian tersebut dalam penentuan jumlah K menggunakan metode *Elbow* menghasilkan 3 klaster, dengan nilai SEE = 11,45286 dengan selisih 6,20229. Kemudian dalam proses perhitungan *K-Means* yang dilakukan secara manual, memperoleh hasil akhir sampai iterasi ke 9 dengan jumlah klaster 1=710, klaster 2=24 dan klaster 3=28. Dengan nilai *Davies-Bouldin Index* nya adalah 0.515

Penelitian selanjutnya [5] Dalam penelitian berjudul "Perbandingan Algoritma K Means dan K-Medoids untuk Pengelompokan Kelas Siswa Tunagrahita", hasil dari metode K-Means Clustering menunjukkan bahwa terdapat 8 siswa tunagrahita ringan, 14 siswa tunagrahita sedang, dan 14 siswa tunagrahita berat. Sementara itu, metode K-Medoids Clustering menghasilkan informasi bahwa ada 7 siswa tunagrahita ringan, 19 siswa tunagrahita sedang, dan 10 siswa tunagrahita berat. Penelitian ini juga menguji evaluasi dengan menggunakan nilai DBI, dimana nilai DBI untuk validasi K-Means adalah 0,161, sedangkan nilai DBI untuk validasi K-Medoids adalah 0,281. Dengan demikian, dapat disimpulkan bahwa pengelompokan menggunakan metode K-Means memberikan hasil yang lebih baik daripada metode K-Medoids karena menghasilkan nilai DBI yang lebih rendah, yaitu 0,16.

Berdasarkan latar belakang tersebut maka penulis melakukan penelitian dengan judul "Pengelompokkan Hasil Belajar Siswa SDN Tunas Jaya Dengan Algoritma *K-Means*". Penelitian ini menggunakan metode *Elbow* untuk menentukan jumlah *cluster* yang optimal serta pengujian validitas dengan *Silhouette Method* dan *Davies-Bouldin Index* sebagai acuan dalam melakukan pengelompokan *cluster* pada capaian hasil belajar siswa

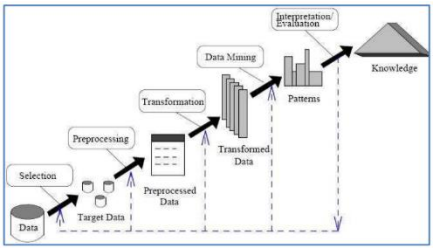
2. TINJAUAN PUSTAKA

2.1. Pendidikan Sekolah Dasar

Pendidikan merupakan upaya yang disengaja dan terencana untuk menciptakan lingkungan belajar dan proses pembelajaran, sehingga peserta didik dapat aktif mengembangkan potensi mereka. Hal ini bertujuan agar mereka memiliki kekuatan spiritual, kemampuan pengendalian diri, kepribadian, kecerdasan, akhlak mulia, serta keterampilan yang diperlukan untuk diri sendiri, masyarakat, bangsa, dan negara. Menurut Undang-undang Nomor 20 Tahun 2003 Tentang Sistem Pendidikan Nasional, pendidikan formal merupakan suatu jalur pendidikan yang terstruktur dan berjenjang, melibatkan pendidikan dasar, menengah, dan tinggi. Pendidikan pada tingkat sekolah dasar ditujukan untuk anak-anak berusia 7 hingga 13 tahun. Pendidikan ini dikembangkan sesuai dengan karakteristik setiap satuan pendidikan, potensi daerah, dan konteks sosial-budaya masyarakat setempat. Di tingkat ini, siswa sekolah dasar mengikuti pembelajaran dalam berbagai bidang studi yang harus mereka kuasai secara menyeluruh.

2.2. KDD

Knowledge Discovery in Database (KDD) merupakan suatu pola kegiatan guna untuk mencari serta mencari pola yang terjadi pada sekumpulan data, dimana pola yang ditemukan bersifat fakta baru, bermanfaat dan dapat dimengerti.



Gambar 2. Tahapan Metodologi KDD (Sumber: [6])

2.3. Algoritma *K-Means*

Algoritma K-Means adalah jenis algoritma pengelompokan non-hierarki yang melakukan pembagian data ke dalam satu atau beberapa cluster. Dengan pendekatan ini, data yang memiliki karakteristik serupa akan ditempatkan dalam satu cluster atau kelompok, sementara data yang memiliki perbedaan karakteristik akan dikelompokkan dalam cluster yang berbeda. [7]. Algoritma K-Means adalah suatu metode dalam teknik pengelompokan (clustering) yang termasuk dalam kategori pemodelan

unsupervised learning pada Data Mining. Algoritma ini digunakan untuk mengelompokkan data dengan sistem partisi [8]. Pada tahap pelaksanaannya, K-Means melakukan pembagian data ke beberapa kelompok, di mana setiap kelompok memiliki karakteristik yang serupa dan berbeda dengan data yang ada di kelompok lainnya.

2.4. Metode *Elbow*

Metode *Elbow* adalah metode yang digunakan untuk menentukan jumlah *cluster* terbaik, khsusnya dalam proses *clustering*. Apabila nilai *cluster* pertama memiliki nilai yang mengalami penurunan drastis terhadap *cluster* kedua, maka *cluster* tersebut menjadi *cluster* terbaik dan membentuk sebuah sudut pada grafik [9].

Menurut [10] Metode *Elbow* adalah suatu pendekatan yang digunakan untuk menentukan jumlah *cluster* optimal dengan memeriksa hasil persentase perbandingan antara jumlah *cluster* (k), yang menghasilkan sudut siku pada suatu titik. Untuk mendapatkan perbandingan tersebut, dilakukan perhitungan *Sum Of Square* (SSE) dari setiap nilai *cluster*, karena semakin tinggi jumlah nilai *cluster* (k), nilai SSE akan semakin menurun.[11].

2.5. *Davies-Bouldin Index*

DBI merupakan salah satu metode evaluasi internal yang mengukur efektivitas suatu metode pengelompokan berdasarkan nilai-nilai kohesi dan separasi. Dalam konteks pengelompokan, kohesi didefinisikan sebagai total kedekatan data terhadap *centroid cluster*, sementara separasi berfokus pada jarak antar *centroid cluster*. Jika jarak di dalam suatu *cluste*r minimal secara signifikan, setiap objek dalam *cluste*r akan memiliki tingkat kemiripan fitur yang tinggi. Semakin rendah nilai DBI (*non-negatif* dan >= 0), semakin baik hasil pengelompokan yang diperoleh melalui algoritma clustering. ***Silhouette Method***

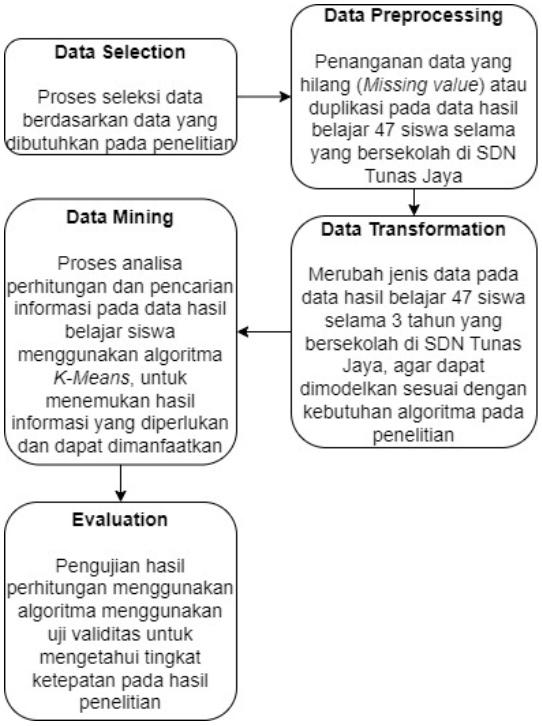
Merupakan salah satu metode yang digunakan untuk mengetahui kualitas dan kekuatan dari suatu *cluster* sera untuk mengukur kedekatan relasi antar objek pada sebuah *cluster*.

Hasil dari *Silhouette* t memiliki nilai yang bervariasi yaitu bernilai antara -1 sampai dengan 1. Suatu model *clustering* dapat disebut baik jika nilai *Silhouette coefficient* memiliki nilai positif (a(i) < b(i)) dan nilai a(i) mendekati 0.

Sedangkan nilai maksimal *Silhouette* yaitu 1 ketika nilai a(i) = 0. Jika nilai *Silhouette* t = 1, maka sebuah *cluster* telah masuk pada *cluster* yang tepat. Namun apabila nilai *Silhouette* = 0, maka objek i berada pada dua *cluster* dan dapat dikatakan bahwa objek tersebut tidak memiliki struktur yang jelas. Jika nilai *Silhouette* = -1, artinya struktur dari *cluster* tersebut memiliki nilai yang *overlapping*, sehingga objek I akan lebih baik jika ditempatkan pada *cluster* lain.

3. METODE PENELITIAN

Metode atau tahapan untuk mencapai tujuan dari penelitian ini melalui metodologi *Knowledge Discovery in Database* (KDD) yang terdiri dari lima tahapan, yaitu *Data Selection*, *Data Preprocessing*, *Data Transformation*, *Data Mining*, dan *Evaluation*. Seperti pada gambar 3.



Gambar 3. Rancangan Penelitian Metodologi KDD

3.1. *Data Selection*

Pada tahapan ini dilakukan pemilihan data yang dilakukan sebelum tahap *data mining*, dengan tujuan untuk memilih atribut atau variabel yang relevan dengan proses penelitian. Data yang digunakan dalam penelitian ini yaitu data perkembangan hasil belajar 47 siswa selama tiga tahun, dimulai dari kelas 4 hingga kelas 6, dengan rentang waktu pada data penelitian ini mencakup tahun ajaran 2020/2021 hingga 2022/2023 di SDN Tunas Jaya.

3.2. *Data Preprocessing*

Tahap kedua ialah *data preprocessing* atau pembersihan data dimana dilakukan untuk menghasilkan data yang berkualitas tinggi. Proses yang terjadi pada tahap ini adalah pembersihan, perbaikan data dari data yang tidak konsisten, nilai yang hilang, dan variabel yang tidak relevan. Pada tahap ini menggunakan bahasa pemrograman python dalam mengatasi *missing value*.

3.3. *Data Transformation*

Pada tahap ini akan dilakukan proses modifikasi pada data yang dipilih, sampai data tersebut dapat di proses untuk tahap *data mining*, proses ini meliputi perubahan yang bertujuan untuk memudahkan dalam proses pengolahan data.

3.4. *Data Mining*

Tahap selanjutnya adalah data mining yaitu proses menemukan pola atau menarik informasi dalam data terpilih dengan menggunakan Algoritma *K-Means* dan untuk menentukan jumlah *cluster* optimal di uji dengan menggunakan metode *elbow*. Proses *data mining* ini nantinya menggunakan bahasa pemrograman python dibantu dengan platform google colaboratory.

3.5. *Evaluation*

Pada tahap ini dilakukan perubahan pola menjadi informasi yang mudah dipahami. Data hasil belajar yang sudah diolah dan dihitung pada tahap sebelumnya akan diuji guna mengukur efektifvitas kinerja hasil perhitungannya menggunakan uji validitas Silhouette *Method* dan Davies-Bouldin Index. Proses evaluasi ini menggunakan bantuan google colab serta bahasa pemrograman pyhton.

4. HASIL DAN PEMBAHASAN

Berdasarkan hasil penelitian *data mining* yang telah dilakukan yaitu dengan menerapkan Algoritma *K-Means* untuk meng-*cluster* hasil belajar siswa guna meningkatkan prestasi SDN Tunas Jaya dimana agar dapat dimanfaatkan oleh pihak sekolah sebagai sumber informasi dalam melakukan *clustering* terhadap prestasi siswa. Penelitian ini menggunakan metode *Knowledge Discovery in Database* (KDD), sebagai metode standar dalam mengolah data yang didapatkan dari hasil observasi, dengan beberapa tahapan seperti *Data Selection*, *Data Transformation*, *Data Mining* dan *Evaluation*.

4.1. *Data Selection*

Beberapa kumpulan data dari hasil observasi perlu melalui proses seleksi terlebih dahulu dalam tahap pencarian informasi dalam KDD, dimana data hasil seleksi tersebut digunakan dalam proses *data mining*. Pada penelitian ini menggunakan data perkembangan hasil belajar 47 siswa selama tiga tahun, dimulai dari kelas 4 hingga kelas 6. Rentang waktu pada data penelitian ini mencakup tahun ajaran 2020/2021 hingga 2022/2023.

Selanjutnya data tersebut dideskripsikan setiap atribut dijelaskan jenis dan deskripsinya untuk masuk kedalam tahap berikutnya. Jenis data yang digunakan dibagi menjadi dua, yakni bertipe numerik dan kategorikal, untuk lebih lengkapnya dapat dilihat pada Tabel 1.

Tabel 1. Atribut data

No	Atribut	Jenis	Keterangan
1	Nama Siswa	Kategorikal	Nama lengkap siswa
2	Kelas	Numerik	Kelas siswa di sekolah
3	PA	Numerik	Nilai Mata Pelajaran Pendidikan Agama dan Budi Pekerti (PABP) siswa
4	PKN	Numerik	Nilai Mata Pelajaran PKN siswa SDN Tunas Jaya
5	IND	Numerik	Nilai Mata Pelajaran Bahasa Indonesia siswa SDN Tunas Jaya
6	MAT	Numerik	Nilai Mata Pelajaran Matematika siswa SDN Tunas Jaya
7	IPAS	Numerik	Nilai Mata Pelajaran Ilmu Pengetahuan Alam dan Sosial siswa SDN Tunas Jaya
8	MUSIK	Numerik	Nilai Mata Pelajaran Seni Musik siswa SDN Tunas Jaya
9	TARI	Numerik	Nilai Mata Pelajaran Seni Tari siswa SDN Tunas Jaya
10	RUPA	Numerik	Nilai Mata Pelajaran Seni Rupa siswa SDN Tunas Jaya
11	TEATER	Numerik	Nilai Mata Pelajaran Seni Teater siswa SDN Tunas Jaya
12	PJOK	Numerik	Nilai Mata Pelajaran Pendidikan Jasmani Olahraga dan Kesehatan siswa SDN Tunas Jaya
13	ING	Numerik	Nilai Mata Pelajaran Bahasa Inggris siswa SDN Tunas Jaya
14	SUND	Numerik	Nilai Mata Pelajaran Bahasa Sunda siswa SDN Tunas Jaya

4.2. *Data Preprocessing*

Tahapan selanjutnya adalah data preprocessing, dimana melibatkan beberapa langkah penting untuk menangani proses *data mining*. Beberapa tahapan yang diperlukan antara lain sebagai berikut:

a. Tipe Dataset
   Memastikan tipe dataset yang benar merupakan hal yang sangat krusial untuk dilakukan karena metode analisis dan pemrosesan data yang akan digunakan dapat berbeda tergantung pada tipe data yang ada. Pada beberapa atribut merupakan tipe data dengan bilangan bulan, sedangkan tipe data float merupakan tipe bilangan desimal. Lalu dikarenakan tidak ditemukannya ketidaksesuaian tipe dataset sehingga data dapat terhindar dari kesalahan dalam menganalisis data.
b. *Missing Value*
   Pada data yang digunakan ialah data nilai maka tidak terdapat missing value, oleh karena itu tidak ada penanganan missing value dan dilanjutkan ke tahap selanjutnya.
c. *Duplicate Data*
   Selanjutnya ialah duplicate data, dimana untuk memastikan kualitas data diperlukan penyatuan data diperlukan integrasi data yang tidak berisi

duplikasi data. Dari hasil pengecekkan ditemukan bahwa tidak ada duplikasi data. Oleh karena itu, tidak diperlukan penanganan khusus untuk kondisi tersebut.

4.3. *Data Transformation*

Tahap *Data Transformation* merupakan proses perubahan pada data kompleks untuk disederhanakan agar lebih mudah dalam proses pengelolaan. Oleh sebab data yang digunakan tidak terdapat missing value dan duplikasi data, pada penelitian ini data transformation tidak dilakukan dan dilanjut ke tahap selanjutnya.

4.4. *Data Mining*

Data Mining berguna untuk melihat pola yang terdapat pada data, serta teknik yang diterapkan dalam pada proses ini adalah *clustering*, dimana pengelompokan data diurutkan berdasarkan tingkat kemiripan dalam *cluster*. Algoritma yang digunakan ialah algoritma *K-Means* dengan optimalisasi penentuan jumlah *cluster* didasarkan pada hasil perhitungan Sum of Square menggunakan Metode *Elbow* serta *Davies-Bouldin Index* dan *Silhouette* sebagai evaluasi dalam penentu tujuan jumlah *cluster* yang optimal. Penentuan titik *cluster* pertama dilanjutkan dengan mencari nilai *Sum of Square Within* (SSW) dan *Sum of Square Between* (SSB) pada dataset yang digunakan untuk mengoptimalisasi berapa jumlah *cluster* yang akan digunakan dalam tahap evaluasi seperti pada Tabel 2.

Tabel 2. Hasil uji SSW dan SSB

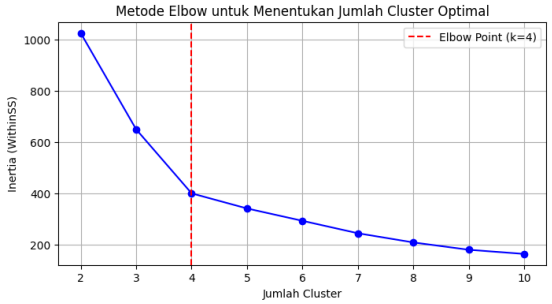
Jumlah *cluster*	SSW	SSB
2	990.4	34.5
3	613.2	36.2
4	360.0	39.9
5	304.2	36.8

Pada Tabel 2, perubahan jumlah *cluster* memengaruhi nilai SSW dan SSB dalam analisis klasterisasi. Secara umum, dengan meningkatnya jumlah klaster, nilai SSW cenderung menurun, menunjukkan bahwa titik data dalam klaster lebih mirip satu sama lain, sementara nilai SSB cenderung meningkat, menunjukkan bahwa *cluster* lebih terpisah satu sama lain dimana titik *cluster* terdapat pada C=4. Untuk memastikan bahwa jumlah *cluster* tersebut tepat, jumlah titik *cluster* tersebut perlu diuji kembali yakni menggunakan metode *Elbow*. Pada metode ini, titik *cluster* yang menunjukkan jumlah klaster yang optimal akan membentuk sudut tajam pada grafik, pemilihan titik optimal didasarkan pada perbandingan perbedaan penurunan yang stabil pada nilai *Sum of Squared Error* (SSE) tanpa mengalami fluktuasi besar pada klaster atau titik berikutnya.

Tabel 3. SSE metode *Elbow*

Jumlah *Cluster*	SSE
2	1025.0
3	649.5
4	400.0
5	341.0

Tabel 3 menunjukkan bahwa penurunan SSE sangat signifikan ketika jumlah klaster meningkat dari 2 ke 4, namun setelah jumlah klaster mencapai 4, penurunan SSE menjadi lebih lambat, mengindikasikan adanya bentuk "siku". Dengan demikian, metode *Elbow* menyarankan bahwa jumlah klaster yang optimal dalam analisis ini adalah C=4, karena setelah titik ini, penurunan SSE tidak lagi signifikan. Visualisasi metode *Elbow* dapat dilihat pada gambar 4.



Gambar 4. Visualisasi *cluster* optimal dengan metode *Elbow*

4.5. *Evaluation*

Setelah melalui tahapan *Data Mining*. Langkah selanjutnya yaitu melakukan tahap evaluasi untuk mengukur hasil atau kualitas dari *cluster* dan model yang dibentuk, dimana uji validitas tersebut menggunakan metode *Silhouette*. Hasil evaluasi *Silhouette Score* dapat dilihat pada tabel 4.

Tabel 4. Hasil evaluasi *Silhouette Method*

Jumlah *cluster*	*Silhouette Score*
2	0.34589073352467004
3	0.3628294740245073
4	0.39964238012881204
5	0.36897616689386936

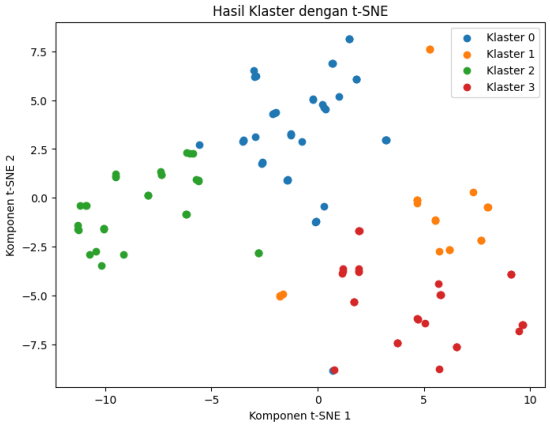
Pada pengamatan *Silhouette* untuk mengetahui *cluster* terbaik adalah dengan mengetahui *Silhouette Score*. Berdasarkan nilai *Silhouette Score* pada tabel 4.6, ketika C=2, nilai *Silhouette Score* adalah sekitar 0.346, yang menunjukkan bahwa data terbagi dengan cukup baik. Ketika C=3, nilai *Silhouette Score* sedikit meningkat menjadi 0.361, menunjukkan pemisahan *cluster* yang lebih baik daripada C=2. Ketika C=4, nilai *Silhouette Score* meningkat lebih signifikan menjadi sekitar 0.399, menunjukkan peningkatan yang signifikan dalam kualitas *clustering*. Namun, ketika jumlah *cluster* dinaikkan menjadi 5, indeks *Silhouette* turun menjadi sekitar 0.368, menandakan adanya overlap antara *cluster*. Oleh karena itu berdasarkan uji validasi dengan metode *Silhouette*, C=4 termasuk kedalam struktur lemah (*weak structure*).

Tabel 5. Uji validitas dengan DBI

Jumlah *Cluster*	DBI *Score*
2	1.1786205411170694
3	0.8624322253305077
4	0.7497457050956073
5	0.8431645967827894

Kemudian dilanjutkan dengan uji validitas menggunakan *Davies-Bouldin Index* (DBI). Hasil Tabel 5, menunjukkan bahwa ketika jumlah *cluster* (C) sama dengan 2, terdapat overlap antara *cluster*, ditandai dengan DBI *Score* sekitar 1.1786. Namun, ketika C=3, DBI *Score* menurun, menunjukkan pemisahan *cluster* yang lebih baik. Saat C=4, DBI *Score* turun lebih rendah lagi, menunjukkan peningkatan pemisahan *cluster*. Namun, ketika C=5, DBI *Score* naik kembali, menandakan bahwa penambahan *cluster* tidak memberikan peningkatan yang signifikan. Pada jumlah *cluster* lebih tinggi, DBI *Score* cenderung stabil di sekitar 0.74 hingga 0.84, menunjukkan bahwa *cluster* terbaik berada pada C=4.

Berdasarkan proses *clustering* yang sudah dilakukan, diantaranya *cluster* 0 dengan jumlah sebanyak 35 data, *cluster* 1 terdiri dari 46 data, *cluster* 2 memiliki 25 data dan *cluster* 3 menghasilkan 35 data. Penyajian visualisasi data pada penelitian ini akan dilakukan menggunakan Scatter Plot. Hasil scatter tersebut tercantum pada Gambar 5.



Gambar 1. Visualisasi *scatter plot* hasil *cluster*

Dari pemaparan hasil sebelumnya, perlu diadakan proses analisis lebih lanjut untuk melihat apakah terdapat pola atau hubungan dari hasil *clustering* yang telah dilakukan, dimana terdapat informasi berupa karakteristik pada setiap *cluster-cluster*nya. Adapun untuk hasil karakteristik yang didapat tersebut berdasarkan rentang minimum dan maksimum dari nilai hasil belajar siswa pada setiap anggota masing-masing *cluster* yang ada. Informasi karakteristik hasil *clustering* dapat dilihat pada Tabel 6. dan Tabel 7.

Tabel 6. Karakteristik *cluster*

Cluster	Karaktersistik					
	PA	PKN	IND	MAT	IPAS	MUSIK
Cluster 0	70 - 82.5	75 - 78	75 - 78	70 - 85	70 - 76	72.5 - 79
Cluster 1	75 - 83.8	76 - 84	75 - 84	75 - 85	72 - 80	73 - 79
Cluster 2	80 – 85	77 - 88	76 - 86	75 - 86	74 - 89	73.5 - 79
Cluster 3	80 - 85.5	81 - 89	78 - 86	78 - 90	70 - 79	77 - 84

Tabel 7. Karakteristik *cluster* (Lanjutan)

Cluster	Karaktersistik					
	TARI	RUPA	TEATER	PJOK	ING	SUND
Cluster 0	72 - 80.5	70 – 77	72.5 - 79	80 - 86	72 - 82	70 - 74
Cluster 1	71 - 96	69.5 – 78	70 - 80.5	80 - 87	80 - 88	72 - 76
Cluster 2	71.5 - 77	71 - 74.5	73 - 78.5	80 - 91	70 - 80	76 - 79
Cluster 3	74 - 98	73 – 82	71.5 - 82.5	80 - 94	75 - 88	76 - 89

Dari tabel 6 dan 7, dapat disimpulkan bahwa dari 4 *cluster* tersebut terbagi ke dalam beberapa kategori kelompok sebagai berikut:

1. *Cluster* 0 (Kurang): *Cluster* ini memiliki rentang nilai yang cenderung rendah dalam hampir semua mata pelajaran, termasuk PA, PKN, IND, MAT, IPAS, MUSIK, TARI, RUPA, TEATER, PJOK, ING, dan SUND. Siswa dalam *cluster* ini memiliki prestasi belajar kurang dalam berbagai aspek.
2. *Cluster* 1 (Cukup): *Cluster* ini memiliki rentang nilai yang sedang dan berada di bawah rata-rata dalam beberapa mata pelajaran, seperti PA, PKN, IND, dan MUSIK. Namun, mereka memiliki prestasi belajar yang cukup baik dalam beberapa mata pelajaran lainnya, seperti TARI, TEATER, dan ING.
3. *Cluster* 2 (Baik): *Cluster* ini memiliki nilai yang tinggi atau di atas rata-rata dalam hampir semua mata pelajaran, menunjukkan variasi prestasi belajar di antara siswa dalam *cluster* ini. Mereka memiliki prestasi belajar baik hampir di semua mata pelajaran.
4. *Cluster* 3 (Sangat Baik): *Cluster* ini memiliki nilai tertinggi dalam hampir semua mata pelajaran, termasuk PA, PKN, IND, MAT, IPAS, dan MUSIK. Mereka adalah kelompok siswa dengan prestasi belajar paling tinggi dalam berbagai aspek.

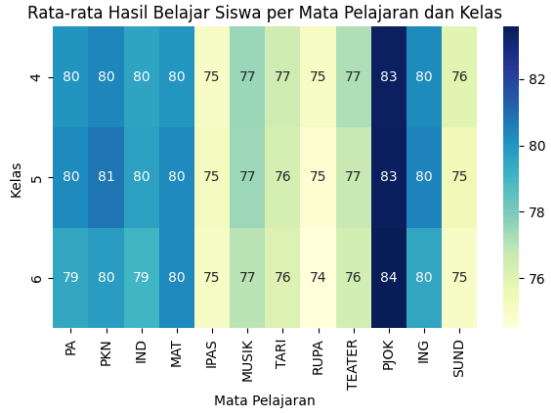
Dengan demikian, analisis ini memberikan gambaran tentang bagaimana siswa di SDN Tunas Jaya dapat dikelompokkan berdasarkan prestasi belajar mereka dalam berbagai mata pelajaran. Setelah

analisis tersebut, maka label pada setiap *cluster* dapat disimpulkan pada tabel 8.

Tabel 8. Label *cluster*

No	*Cluster*	Label
1.	0	Kurang
2.	1	Cukup
3.	2	Baik
4.	3	Sangat Baik

Penelitian dilanjutkan dengan menganalisis rata-rata hasil belajar siswa pada setiap mata pelajaran dengan tujuan untuk mengetahui mata pelajaran yang dapat mempengaruhi hasil *clustering*, lalu didapatlah rata-rata hasil belajar siswa yang divisualisasikan dengan bantuan heatmaps, dapat dilihat pada gambar 6.



Gambar 6. Hasil analisis mata pelajaran kelas 4-6

Pada gambar 6, Tingkat kecerahan spektrum warna dari biru hingga cream mencerminkan penurunan nilai hasil belajar siswa dari karakteristik yang sudah dianalisis sebelumnya. Sebaliknya, tingkat kecerahan spektrum warna dari *cream* hingga biru menandakan peningkatan nilai hasil belajar siswa dari karakteristik yang sudah di analisis sebelumnya.

Dengan bantuan heatmaps, visualisasi dapat menjelaskan bagaimana terjadinya peningkatan hasil belajar siswa seperti pada mata pelajaran PJOK. Hasil belajar siswa yang konsisten dari kelas 4 s/d 6 adalah mata pelajaran MAT, IPAS, MUSIK, ING. Penurunan nilai hasil belajar siswa sejak kelas 4 s/d 6 ada pada mata pelajaran PA, IND, TARI, RUPA, TEATER, dan SUND. Berdasarkan hasil temuan tersebut, pihak sekolah diharapkan dapat lebih memfokuskan perhatian pada mata pelajaran yang memerlukan perbaikan atau pengembangan lebih lanjut, dan diharapkan dapat menjadi rekomendasi untuk sekolah dalam menyusun strategi pembelajaran selanjutnya.

5. KESIMPULAN DAN SARAN

Penelitian ini berhasil mengelompokkan prestasi belajar siswa SDN Tunas Jaya menggunakan algoritma *K-Means* dengan empat *cluster*: "Kurang", "Cukup", "Baik", dan "Sangat Baik". Evaluasi kinerja pengelompokan dilakukan dengan *Silhouette Score* dan *Davies-Bouldien Index*, menunjukkan jumlah *cluster* optimal 4 dengan nilai *Silhouette Score* 0.399 dan DBI *Score* 0.749. Analisis karakteristik prestasi belajar mengidentifikasi penurunan hasil belajar pada mata pelajaran PA, IND, TARI, RUPA, TEATER, dan SUND. Saran meliputi pertimbangan penambahan variabel, seperti aspek sosial, ekonomi, atau lingkungan siswa, serta penerapan praktis hasil analisis klasterisasi dalam merancang program pendukung atau remedial yang lebih efektif.

DAFTAR PUSTAKA

[1] D. Khotimah Nurlaida, "Implementasi Program Penguatan Pendidikan Karakter (PPK) Melalui Kegiatan 5s Di Sekolah," *Inopendas J. Ilm. Kependidikan*, vol. 2, no. 1, pp. 28–31, 2019.

[2] Y. Syahra, "Penerapan *Data Mining* Dalam Pengelompokkan Data Nilai Siswa Untuk Penentuan Jurusan Siswa Pada SMA Tamora Menggunakan Algoritma *K-Means Clustering*," *J. SAINTIKOM (Jurnal Sains Manaj. Inform. dan Komputer)*, vol. 17, no. 2, p. 228, 2018, doi: 10.53513/jis.v17i2.70.

[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] I. Wahyudi, M. B. Sulthan, and L. Suhartini, "Analisa Penentuan *Cluster* Terbaik Pada Metode *K-Means* Menggunakan *Elbow* Terhadap Sentra Industri Produksi Di Pamekasan," *J. Apl. Teknol. Inf. dan Manaj.*, vol. 2, no. 2, pp. 72–81, 2021, doi: 10.31102/jatim.v2i2.1274.

[5] F. Harahap, "Perbandingan Algoritma K Means dan K Medoids Untuk *Clustering* Kelas Siswa Tunagrahita," *TIN Terap. Inform. Nusant.*, vol. 2, no. 4, pp. 191–197, 2021.

[6] Karsito and W. Monika Sari, "Prediksi Potensi Penjualan Produk Delifrance Dengan Metode Naive Bayes Di Pt. Pangan Lestari," *J. Teknol. Pelita Bangsa*, vol. 9, no. 1, pp. 67–78, 2018, [*Online*]. Available: https://jurnal.pelitabangsa.ac.id/index.php/sigma/article/view/465

[7] S. Aulia, "Klasterisasi Pola Penjualan Pestisida Menggunakan Metode *K-Means Clustering* (Studi Kasus Di Toko Juanda Tani Kecamatan Hutabayu Raja)," *Djtechno J. Teknol. Inf.*, vol. 1, no. 1, pp. 1–5, 2021, doi: 10.46576/djtechno.v1i1.964.

[8] W. Kempten, "Overview of Current State of the Art Lane Detection Models and *Method*s and Testing and Training a SCNN on Google Colaboratory Felix Pius Umscheid," *Res. Gate*, no. December, 2020, doi: 10.13140/RG.2.2.25239.44962.

[9] E. Sulistiyawan, A. Hapsery, and L. J. A. Arifahanum, "PERBANDINGAN METODE OPTIMASI UNTUK PENGELOMPOKAN

PROVINSI BERDASARKAN SEKTOR PERIKANAN DI INDONESIA (Studi Kasus Dinas Kelautan dan Perikanan Indonesia)," *J. Gaussian*, vol. 10, no. 1, pp. 76–84, 2021, doi: 10.14710/j.gauss.v10i1.30936.

[10] A. Winarta and W. J. Kurniawan, "Optimasi *cluster K-Means* menggunakan metode *Elbow* pada data pengguna narkoba dengan pemrograman python," *J. Tek. Inform. Kaputama*, vol. 5, no. 1, pp. 113–119, 2021.

[11] D. A. I. C. Dewi and D. A. K. Pramita, "Analisis Perbandingan Metode *Elbow* dan *Silhouette* pada Algoritma *Clustering K-Medoids* dalam Pengelompokan Produksi Kerajinan Bali," *Matrix J. Manaj. Teknol. dan Inform.*, vol. 9, no. 3, pp. 102–109, 2019, doi: 10.31940/matrix.v9i3.1662.