JURTI, Vol.9 No.2, Juni 2025, ISSN: 2579-8790

Pengelompokan Minat Akademik Siswa SMA Negeri 1 Loa Janan Menggunakan Metode *Clustering K-means*

Ammar Nabil Fauzan[1], Faizul Anwar Wandi[2] , Ahmad Zuhair Nur Aiman[3], Masna Wati[4], Haviluddin[5]

[1,2,3,4,5] Program Studi Informatika, Fakultas Teknik, Universitas Mulawarman, Samarinda, Indonesia

e-mail: [1]ammarnabil31@gmail.com, [2]faizulanwarwandi@gmail.com, [3]aimanzuhair05@gmail.com, [4]masnawati@fkti.unmul.ac.id, [5]haviluddin@unmul.ac.id

Abstrak

Penentuan minat akademik masih menjadi tantangan dalam proses mempersiapkan diri sebelum memilih jurusan di perguruan tinggi, terutama jika siswa sendiri masih belum sepenuhnya mengetahui kemampuan dan minat belajarnya. Penelitian ini dilakukan dengan tujuan untuk membentuk kelompok-kelompok siswa kelas XI 3 di SMA Negeri 1 Loa Janan berdasarkan minat akademik mereka dengan menggunakan pendekatan data mining. K-means merupakan algoritma yang dipilih dari metode clustering, dengan menggunakan Knowledge Discovery in Database (KDD) yang dimulai dari seleksi data, kemudian tahap preprocessing data melalui normalisasi. Evaluasi menggunakan Silhouette Score dan Davies Bouldin Index. Hasil menunjukkan bahwa 2 merupakan jumlah cluster yang tepat dengan nilai Silhouette Score 0.74, nilai Davies Bouldin Index sebesar 0.34 dan visualisasi Scatter Plot yang menunjukkan pemisahan cluster yang cukup jelas. Hasil clustering ini bisa menjadi referensi bagi tenaga pengajar seperti guru untuk memudahkan proses penentuan jurusan sebelum masuk perguruan tinggi.

Kata kunci— *Clustering, Data Mining, KDD, K-means, Nilai Akademik.*

1. PENDAHULUAN

Masa remaja adalah fase penting dalam perkembangan manusia, karena pada periode ini individu mulai mengeksplorasi dan membentuk identitas serta jati dirinya. Umumnya, masa remaja dimulai ketika seseorang mengalami kematangan seksual dan berakhir ketika mencapai usia dewasa secara hukum [1]. Remaja mencoba banyak hal baru dan mulai menemukan apa yang mereka sukai atau hal yang ingin mereka capai di masa depan. Setiap siswa dituntut oleh orang tuanya untuk menentukan dan merencanakan hal yang akan dilakukan setelah lulus sekolah. Kebanyakan lulusan SMA yang masih berusia remaja memutuskan untuk melanjutkan pendidikan ke jenjang perguruan tinggi, namun banyak di antara mereka masih merasa ragu atau bingung dalam menentukan program studi yang akan dipilih [2]. Kurangnya informasi yang diperoleh siswa mengenai dunia perguruan tinggi baik dari luar sekolah ataupun dalam sekolah juga dapat menjadi penyebab bingungnya siswa untuk memilih program studi yang sesuai.

Pola pergaulan menjadi pedoman atau contoh yang diikuti dalam interaksi antar siswa, yang mencakup aspek perilaku, emosi, serta pembentukan identitas diri [3]. Pola pergaulan terbagi menjadi dua jenis, yaitu pergaulan yang terarah serta pergaulan yang tidak terarah, yang sering disebut sebagai pergaulan bebas [4]. Di masa remaja, ikatan pertemanan umumnya terbentuk karena adanya kesamaan minat atau hobi antar individu. [5] . Minat merupakan dorongan untuk berinteraksi dengan lingkungan, baik melalui proses mengamati, menyelidiki, mempelajari, maupun melakukan suatu tindakan [6]. Minat untuk melanjutkan pendidikan ke perguruan tinggi merupakan kecenderungan yang mencakup rasa suka, keinginan, perhatian, ketertarikan,

JURTI, Vol.9 No.2, Juni 2025, ISSN: 2579-8790

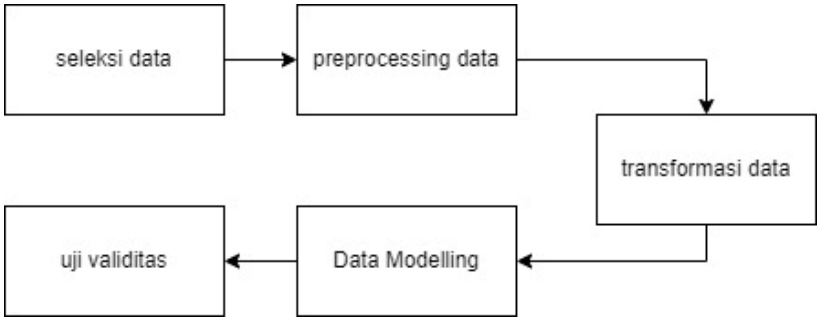
kebutuhan, harapan, motivasi, serta kemauan untuk menempuh pendidikan pada jenjang yang lebih tinggi setelah menyelesaikan sekolah menengah [7].
Rasa kebersamaan, empati, dan sikap saling membantu dapat berperan dalam membentuk hubungan pertemanan yang turut memengaruhi cara siswa dalam menentukan pola hidupnya [8]. Hal ini akhirnya akan mempengaruhi pola pikirnya dalam menentukan pendidikan dan karir yang diinginkan.

Namun, ada kalanya siswa tidak bisa menentukan jurusan kuliah sesuai keinginannya sendiri karena adanya tekanan atau keinginan dari orang tua yang mengharapkan anaknya memilih jurusan tertentu saja [9]. Oleh karena itu, jika siswa tidak menjadikan minat dan bakat sebagai acuan untuk pemilihan jurusan akan muncul kemungkinan bahwa mereka merasa tertekan atau tidak bahagia saat menjalani perkuliahan. Kesalahan memilih jurusan kuliah tidak hanya mengakibatkan ketidakcocokan antara pemikiran mahasiswa serta mata kuliah yang mereka pelajari tidak hanya memengaruhi aspek akademis, tetapi juga berdampak pada motivasi belajar, tingkat kepuasan hidup, kemampuan menghadapi stres, dan keseluruhan pengalaman mereka selama menjalani perkuliahan [10]. Realitanya banyak mahasiswa perguruan tinggi negeri maupun swasta masuk pada jurusan yang tidak sesuai dengan bakat dan minatnya. Hal ini bisa terjadi karena rendahnya informasi yang diperoleh siswa terhadap perguruan tinggi, kurangnya informasi terkait dengan jurusan yang ingin diambil atau bisa jadi mengambil jurusan yang disarankan oleh orang terdekat (orang tua, teman, dan sebagainya) yang mana tidak melihat potensi pada dirinya. Hal tersebut juga dialami oleh siswa Sekolah Menengah Atas (SMA), terutama siswa/i kelas 11 di SMA Negeri 1 Loa Janan. Dengan demikian, penerapan data *mining* dapat menjadi solusi efektif untuk membantu siswa menemukan program studi yang paling sesuai dengan minat dan bakat mereka.

Dalam penelitian ini, algoritma *K-means* dipilih karena memiliki keunggulan dalam mengelompokkan data numerik seperti nilai akademik secara efisien dan sederhana. *K-means* mampu membagi data ke dalam kelompok homogen berdasarkan kemiripan, sehingga hasil *cluster* dapat dengan mudah diinterpretasikan. Selain itu, algoritma ini cukup populer dan banyak digunakan dalam dunia pendidikan karena kecepatan komputasinya serta kemampuannya menangani *dataset* berskala sedang hingga besar [14]. Jadi, *K-means* dianggap tepat untuk digunakan dalam mengelompokkan minat akademik siswa berdasarkan nilai yang dimiliki.

2. METODE PENELITIAN

Pada penelitian ini metode yang digunakan adalah metode *Knowledge Discovery in Databases* (KDD) dengan menggunakan algoritma *K-means Clustering* untuk mengidentifikasi minat akademik siswa kelas XI 3 berdasarkan data nilai dan preferensi mereka terhadap mata pelajaran. Proses KDD terdiri dari seleksi data, *preprocessing* data, transformasi, proses data *mining*, dan interpretasi hasil [11]. Algoritma *K-means* digunakan untuk mengelompokkan siswa ke dalam beberapa *cluster* berdasarkan kesamaan nilai akademik. Urutan alur proses algoritma *K-means* pada konsep bagan seperti yang ditunjukkan Gambar 1.



Gambar 1. Alur metode penelitian

JURTI, Vol.9 No.2, Juni 2025, ISSN: 2579-8790

Penjelasan alur dari metode penelitian dengan KDD pada Gambar 1 sebagai berikut:

1) Seleksi Data
   Mengambil data dari siswa kelas XI 3 SMA Negeri 1 Loa Janan. Data yang dipilih adalah data diri dan nilai siswa.
2) *Preprocessing* Data
   Dilakukan data cleaning untuk membersihkan data-data yg tidak diperlukan dan menangani missing value pada data.
3) Transformasi Data
   Menyesuaikan format pada *dataset* dan melakukan normalisasi data numerik. Rumus standarisasi (z-score) dan normalisasi min-max ditunjukkan pada persamaan(1)

$$x_{\text{stand}} = \frac{x - \text{mean}(x)}{\text{standard deviation}(x)} \} \; x_{\text{norm}} = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (1)$$

   Keterangan:
   ^{X}stand: nilai hasil standarisasi
   x: nilai asli fitur
   standard deviation(x): rata-rata besar penyebaran data
   mean(x): rata-rata seluruh nilai data dari suatu fitur
   min(x): nilai minimum dari suatu fitur
   max(x): nilai maksimum dari suatu fitur
   ^{X}norm: hasil dari proses normalisasi data

4) Data *Modelling*
   Pada tahap ini, algoritma *K-means Clustering* diterapkan untuk mengelompokkan data siswa berdasarkan pola tertentu. Pengelompokan dibagi menjadi 2 yaitu saintek (sains dan teknologi) dan soshum(sosial dan humaniora). Seorang siswa termasuk dalam bidang Saintek jika memiliki nilai Matematika, Fisika, Kimia, dan Biologi yang tinggi, sedangkan termasuk dalam bidang Soshum jika nilai Ekonomi, Sejarah, Geografi, dan Sosiologi yang lebih tinggi. Untuk kelompok 1 berisi data siswa dengan minat akademik Saintek, sedangkan kelompok 2 berisi data siswa dengan minat akademik Soshum.
5) Uji Validitas
   Menganalisis hasil *clustering* untuk mengidentifikasi pola minat akademik siswa berdasarkan kelompok yang terbentuk.

2.1 K-means Clustering

Clustering adalah teknik dalam data *mining* yang digunakan untuk mengelompokkan data ke dalam *cluster* atau segmen, di mana setiap grup dapat dihuni oleh beberapa anggota secara bersamaan [12]. Algoritma *K-means Clustering* adalah salah satu metode yang umum digunakan karena menjadi metode yang sederhana dan memiliki kapabilitas untuk mengelompokkan data dengan waktu komputasi yang cepat dan efisien. Algoritma ini memiliki kelemahan yaitu kebutuhan untuk menentukan jumlah *cluster* (K) sebelum proses *clustering* dimulai, yang dapat menjadi tantangan ketika jumlah *cluster* yang optimal tidak mudah ditentukan atau diprediksi [13]. Algoritma *K-means* adalah metode yang menggunakan jarak untuk mengelompokkan data ke dalam sejumlah *cluster*, di mana pada setiap iterasi objek akan ditempatkan dalam sebuah kelompok, dan pada iterasi berikutnya, objek tersebut bisa dipindahkan ke kelompok lain[14]. Algoritma ini mengelompokkan data ke dalam *cluster*, sehingga data yang memiliki karakteristik serupa dikelompokkan bersama, sementara data dengan karakteristik berbeda akan ditempatkan dalam *cluster* yang terpisah [15]. Langkah-langkah penerapan algoritma *K-means Clustering* sebagai berikut:

JURTI, Vol.9 No.2, Juni 2025, ISSN: 2579-8790

1. Penentuan Awal Pusat *Cluster* Menentukan k buah pusat *cluster* awal dengan memilih secara acak dari data. Pusat awal *cluster* dipilih dari data yang ada, bukan ditentukan secara manual.
2. Perhitungan Jarak Data dengan Pusat *Cluster*
   Menggunakan *Euclidean Distance* untuk mengukur jarak antara tiap data dengan semua pusat *cluster* seperti yang ditunjukkan pada persamaan (2).
   $$d(x_i, \mu_j) = \sqrt{(x_{i1} - \mu_{j1})^2 + (x_{i2} - \mu_{j2})^2 + \cdots + (x_{in} - \mu_{jn})^2} \quad (2)$$
   Keterangan:
   x_i : data ke-i
   μ_j : *centroid* (pusat *cluster* ke-j)
3. Pengelompokan Data ke Cluster
   Tiap data akan dikelompokkan ke *cluster* dengan jarak terkecil ke *centroid* seperti yang ditunjukkan pada persamaan(3).
   $$Cluster(xi) = arg\,min\, j \parallel xi - \mu j \parallel \quad (3)$$
4. Pembaruan Pusat *Cluster*
   Hitung ulang pusat *cluster* sebagai rata-rata dari semua anggota *cluster* seperti yang ditunjukkan pada persamaan(4).
   $$\mu_j(t+1) = \frac{1}{N_{S_j}} \sum_{j \in S_j} {}_j^x \quad (4)$$
   Keterangan:
   μ_j (t+1) : Centroid baru pada iterasi ke 1
   Nsj : Banyak data pada cluster S_j
5. Iterasi Hingga Konvergen
   Ulangi langkah 2–4 hingga pusat *cluster* tidak berubah lagi (konvergen) atau mencapai jumlah iterasi maksimum.

2.2 Davies Bouldin Index (DBI)

Davies Bouldin Index (DBI) adalah metrik yang digunakan untuk mengevaluasi kualitas hasil *clustering*. Metrik ini pertama kali diperkenalkan oleh David L. Davies dan Donald W. Bouldin pada tahun 1979. Skema *clustering* yang dianggap optimal adalah yang menghasilkan nilai DBI serendah mungkin, karena menunjukkan bahwa *cluster* memiliki pemisahan yang baik dan kompak di dalamnya [12]. Langkah-langkah penerapan algoritma DBI sebagai berikut:

1. *Sum Of Square Within-Cluster* (SSW)
   SSW mengukur kohesi (kepadatan) data dalam satu *cluster*, yaitu seberapa dekat data terhadap *centroid*-nya seperti yang ditunjukkan pada persamaan (5).
   $$SSW_i = \frac{1}{m_i} \sum_{j=1}^{m_i} d(x_j, c_i) \quad (5)$$
   Keterangan:
   m_i = jumlah data pada *cluster* ke-i
   x_j = data ke-j pada *cluster* ke-i
   c_i = *centroid* dari *cluster* ke-i
   $d(x_j, c_i)$ = jarak Euclidean dari data ke-j ke *centroid*
2. *Sum Of Square Between Cluster* (SSB)
   SSB mengukur separasi antara dua *cluster* yang berbeda, yaitu jarak antara dua *centroid* seperti yang ditunjukkan pada persamaan(6).

JURTI, Vol.9 No.2, Juni 2025, ISSN: 2579-8790

$$SSB_{i,j} = d(c_i, c_j) \quad (6)$$

Keterangan:

c_i dan c_j = *centroid* dari *cluster* ke-i dan *cluster* ke-j

d (c_i ,c_j) = jarak *Euclidean* antara dua *centroid*

3. Ratio (Rasio Ri,j)

Menghitung rasio antara kohesi dalam *cluster* dengan separasi antar *cluster* seperti yang ditunjukkan pada persamaan(7).

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_{i,j}} \quad (7)$$

Keterangan:

SSWi dan SSWj = *sum of square within-cluster* dari *cluster* ke-i dan j

$SSB_{i,j}$ = *sum of square between-cluster* antara *cluster* ke-i dan j

4. *Davies Bouldin Index* (DBI)

Nilai DBI adalah rata-rata dari rasio tertinggi dari setiap *cluster* terhadap semua *cluster* lain seperti yang ditunjukkan pada persamaan(8).

$$DBI = \frac{1}{k}\sum_{i=1}^{k} \max_{j \neq i}(R_{ij}) \quad (8)$$

Keterangan:

k = jumlah *cluster*

$R_{i,j}$ = rasio antara *cluster* i dan j

Di setiap *cluster* i, diambil rasio maksimum terhadap semua *cluster* j ≠ i

3. HASIL DAN PEMBAHASAN

Pada tahap ini dipaparkan hasil penelitian sesuai dengan metode penelitian yang telah dijabarkan sebelumnya. Bab ini menjelaskan hasil terkait identifikasi minat akademik siswa kelas XI 3 menggunakan metode *K-means Clustering* dalam data *mining*.

3.1 Seleksi Data

Data yang digunakan adalah data siswa dari SMA N 1 Loa Janan, kelas XI 3. Data ini berisi nilai nilai siswa siswi yang disajikan dalam bentuk tabel excel seperti pada Gambar 2. Dari semua kolom yang ada hanya dipilih nilai nilai yang diperlukan saja, seperti No, Nama Siswa, Bahasa Indonesia (B.indo), Bahasa Inggris (B.ing), Matematika Umum (MTK(U)), Sejarah (Sej), Biologi (Bio), Fisika (Fis), Kimia (Kim), Geografi (Geog), Sosiologi (Sos), Ekonomi (Eko) dan Teknologi Informasi dan Komunikasi (Tik).

LEGER NILAI RAPOR SISWA TAHUN PELAJARAN 2023/2024 GENAP

SEKOLAH : SMA NEGERI 1 LOA JANAN

Kelas : XI 3

NO	NAMA SISWA	NISN	NIS	MATA PELAJARAN																						Ketidakhadiran			Ekstra Kurikuler				
				PAI		B.Indo	B.Ing	MTK (U)	Sej	Bio	Fis	Fis	Kim	Kim	Geog	Geog	Sos	Sos	Sos	Eko	Penjas	PKWu	PKWu	PKWu	Tik	Tik	Tik	P5	Sakit	Izin	Alpa	HABSYI	KEPRAMUKAAN	volly	
1	AHMAD SYARIF HIDAYATULLAH	0073116638	1992	85	92	87	90	86	87	90	94		89		82					92	87	85						88	0	1	0		B		88,15
2	AHMAD TRI JULIANSYAH	3071972878	1993	83	92	82	80	81	83	90					89		84			86	84	80	90				88	5	0	0		B		84,92
3	Ahmat Ikbal	0062085050	1994	79	92	82	78	81	83	90					86		84			86	84	83	90				88	0	0	2		B		84,46
4	Andi Arifiyanto .S	0078239721	2003	91	91	80	76	79	81	90					85		80			90	81	80	90				88	4	0	0		B		84,15
5	Aura Nurhaliza Balqis	0077605172	2015	93	92	80	78	82	85	90	80		75							91	83	85			82		88	0	4	5		SB		84,31
6	BAGAS KURNIAWAN	0079946903	2018	91	92	86	82	85	84	90							87			90	88	88	88		85		88	0	0	0		B		87,38
7	CINTA CLAURA HELMALIA	0072746318	2022	88	92	86	80	81	86	90					80		87			91	81	85			83		88	1	1	2	B	SB		85,38
8	Ferdy Apriadi Yudistira	0065665886	2038	84	91	82	75	84	82	90					82		87			91	84	78	85				88	0	3	5		B		84,23
9	Fina Febriana	0073665307	2039	83	92	85	80	91	85	90	85				84		86			88	83	88					88	0	0	0		B		86,15
10	Ibnul Haidar Utomo	0071338552	2048	83	92	83	80	89	83	90							85			88	85	87	88		85		88	0	0	0		B		86,00

JURTI, Vol.9 No.2, Juni 2025, ISSN: 2579-8790

Gambar 2. Data sebelum diseleksi

Gambar 2 menunjukkan bahwa masih banyak kolom-kolom yang tidak dibutuhkan. Maka dari itu perlu dilakukan seleksi secara manual menggunakan excel untuk menghapus beberapa kolom tersebut.

No	Nama Siswa	NIS	B.Indo	B.Ing	MTK (U)	Sej	Bio	Fis	Kim	Geog	Sos	Eko	Tik
1	AHMAD SYARIF HIDAYATULLAH	1992	87	90	86	87	90	94	89	82		92	
2	AHMAD TRI JULIANSYAH	1993	82	80	81	83	90			89	84	86	
3	Ahmat Ikbal	1994	82	78	81	83	90			86	84	86	
4	Andi Arifiyanto .S	2003	80	76	79	81	90			85	80	90	
5	Aura Nurhaliza Balqis	2015	80	78	82	85	90	80	75			91	82
6	BAGAS KURNIAWAN	2018	86	82	85	84	90				87	90	85
7	CINTA CLAURA HELMALIA	2022	86	80	81	86	90			80	87	91	83
8	Ferdy Apriadi Yudistira	2038	82	75	84	82	90			82	87	91	
9	Fina Febriana	2039	85	80	91	85	90	85		84	86	88	
10	Ibnul Haidar Utomo	2048	83	80	89	83	90				85	88	85
11	Ikbal	2138	84	82	84	83	90	88				92	80
12	Jhonatan Rea	2054	88	86	89	83	90	89			88	93	
13	Jumedia	2058	85	80	86	85	91			81		88	80
14	KEYSHA DHIYA SASIKIRANA	2061	86	88	84	84	90	84		82	85	89	
15	KHAIRINA HASYA NABILA	2062	86	82	86	85	90	82			87	93	85
16	LUSY MARSELA	2064	85	80	80	86	90	85			87	91	80
17	M. Aditya Tarwi	2065	88	80	89	85	91		85			91	85
18	M. ASRIL RAHMADAN	2066	88	84	83	85	90		87			90	86
19	MISNAWATI HARDIA VICHA ARYANI	2073	86	86	82	86	90				86	91	80
20	muhammad oktananda	2078	82	76	79	82	90	77		87		88	

Gambar 3. Data setelah diseleksi

Gambar 3 menunjukkan hasil seleksi data secara manual menggunakan excel yang akhirnya menyisakan kolom No, Nama Siswa, NIS, B.Indo, B.Ing, MTK(U), Sej, Bio, Fis, Kim, Geog, Sos, Eko, dan TIK. Seleksi data secara manual ini bertujuan agar memudahkan program python saat akan melakukan pre-processing data.

3.2 Pre-Processing Data

Sebelum data dimodelkan, harus dilakukan *pre-processing* dahulu untuk menghilangkan nilai-nilai yang tidak sesuai, seperti yang ditunjukkan pada Gambar 4 masih ada *missing value* atau nilai-nilai yang kosong.

	B.Indo	B.Ing	MTK(U)	Sej	Bio	Fis	Kim	Geog	Sos	Eko	Tik
0	87	90	86	87	90	94.0	89.0	82.0	NaN	92	NaN
1	82	80	81	83	90	NaN	NaN	89.0	84.0	86	NaN
2	82	78	81	83	90	NaN	NaN	86.0	84.0	86	NaN
3	80	76	79	81	90	NaN	NaN	85.0	80.0	90	NaN
4	80	78	82	85	90	80.0	75.0	NaN	NaN	91	82.0
5	86	82	85	84	90	NaN	NaN	NaN	87.0	90	85.0
6	86	80	81	86	90	NaN	NaN	80.0	87.0	91	83.0
7	82	75	84	82	90	NaN	NaN	82.0	87.0	91	NaN
8	85	80	91	85	90	85.0	NaN	84.0	86.0	88	NaN
9	83	80	89	83	90	NaN	NaN	NaN	85.0	88	85.0
10	84	82	84	83	90	88.0	NaN	NaN	NaN	92	80.0

Gambar 4. Sebelum penanganan *missing value*

3.2.1 Missing Value

Pada bagian ini, dilakukan penanganan terhadap *missing value* dan kolom yang tidak diperlukan pada data frame karena tidak relevan dengan penelitian. Kemudian missing value diisi dengan nilai rata-rata seperti yang ditunjukkan pada Gambar 5.

```
Data setelah dibersihkan:
   B.Indo  B.Ing  MTK(U)  Sej  Bio        Fis        Kim   Geog        Sos  Eko        Tik
0      87     90      86   87   90  94.000000  89.000000  82.00  85.727273   92  82.434783
1      82     80      81   83   90  84.529412  84.777778  89.00  84.000000   86  82.434783
2      82     78      81   83   90  84.529412  84.777778  86.00  84.000000   86  82.434783
3      80     76      79   81   90  84.529412  84.777778  85.00  80.000000   90  82.434783
4      80     78      82   85   90  80.000000  75.000000  83.75  85.727273   91  82.000000
```

Gambar 5. Hasil data cleaning

Gambar 5 menunjukkan bahwa nilai kosong(NaN) diubah menjadi nilai rata-rata yang diperoleh dari menghitung total semua nilai pada kolom tertentu dengan mengabaikan nilai NaN

JURTI, Vol.9 No.2, Juni 2025, ISSN: 2579-8790

dan membaginya dengan jumlah nilai NaN pada kolom tersebut. Maka hasil dari rata-rata akan digunakan untuk mengganti semua nilai NaN pada kolom.

3.3 Transformasi Data

Pada tahap ini, digunakan dua metode normalisasi dengan peruntukkan yang berbeda. Pertama, menggunakan Min-Max Scaler pada semua kolom kecuali kolom Biologi (Bio) untuk mengubah nilai dalam rentang 0 hingga 1. Kedua, menggunakan Standard Scaler hanya untuk kolom Biologi (Bio) karena kolom ini memiliki rentang nilai yang sempit, hanya 90 hingga 91, tidak seluas kolom nilai lain. Hasil transformasi ditunjukkan pada Gambar 6.

Data setelah normalisasi dan standarisasi:

	B.Indo	B.Ing	MTK(U)	Sej	Bio	Fis	Kim	Geog	Sos	Eko	Tik
0	0.875	1.000000	0.583333	0.857143	-0.301511	1.000000	1.000000	0.222222	0.715909	0.857143	0.405797
1	0.250	0.333333	0.166667	0.285714	-0.301511	0.442907	0.698413	1.000000	0.500000	0.000000	0.405797
2	0.250	0.200000	0.166667	0.285714	-0.301511	0.442907	0.698413	0.666667	0.500000	0.000000	0.405797
3	0.000	0.066667	0.000000	0.000000	-0.301511	0.442907	0.698413	0.555556	0.000000	0.571429	0.405797
4	0.000	0.200000	0.250000	0.571429	-0.301511	0.176471	0.000000	0.416667	0.715909	0.714286	0.333333

Gambar 6. Bentuk data setelah di transformasi

3.4 Data Modelling

Setelah melakukan pengolahan data *mining*, maka selanjutnya melakukan pemodelan *dataset* dari data nilai siswa menggunakan K-Means. Algoritma *K-means* dipilih untuk menentukan jumlah *cluster* yang optimal pada *dataset* nilai siswa. Percobaan dilakukan dengan menggunakan k = 2 sebagai nilai uji untuk mencari jumlah *cluster* yang optimal dari algoritma ini dan memperoleh hasil seperti yang ditunjukkan pada Gambar 7.

	B.Indo	B.Ing	MTK(U)	Sej	Bio	Fis	Kim	Geog	Sos	Eko	Tik	Cluster
0	0.250	0.384615	0.166667	0.285714	-0.306186	0.578125	0.770833	1.000000	0.500000	0.000000	0.405797	0
1	0.250	0.230769	0.166667	0.285714	-0.306186	0.578125	0.770833	0.666667	0.500000	0.000000	0.405797	0
2	0.000	0.076923	0.000000	0.000000	-0.306186	0.578125	0.770833	0.555556	0.000000	0.571429	0.405797	0
3	0.000	0.230769	0.250000	0.571429	-0.306186	0.250000	0.000000	0.429630	0.715909	0.714286	0.333333	0
4	0.750	0.538462	0.500000	0.428571	-0.306186	0.578125	0.770833	0.429630	0.875000	0.571429	0.833333	2
5	0.750	0.384615	0.166667	0.714286	-0.306186	0.578125	0.770833	0.000000	0.875000	0.714286	0.500000	2
6	0.250	0.000000	0.416667	0.142857	-0.306186	0.578125	0.770833	0.222222	0.875000	0.714286	0.405797	0
7	0.625	0.384615	1.000000	0.571429	-0.306186	0.666667	0.770833	0.444444	0.750000	0.285714	0.405797	2
8	0.375	0.384615	0.833333	0.285714	-0.306186	0.578125	0.770833	0.429630	0.625000	0.285714	0.833333	0
9	0.500	0.538462	0.416667	0.285714	-0.306186	0.916667	0.770833	0.429630	0.715909	0.857143	0.000000	2

Gambar 7. Output *clustering* dengan *K-means*

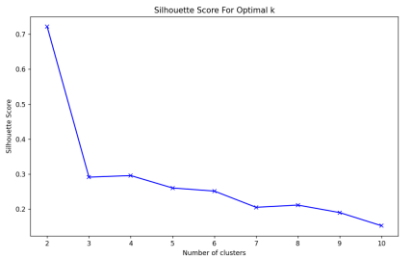
3.5 Uji Validasi

Setelah melakukan proses pemodelan dengan algoritma *K-means*, tahap selanjutnya adalah uji validasi untuk menilai kualitas dan efektivitas hasil pengelompokkan data. Uji validasi diperlukan agar dapat memastikan bahwa jumlah *cluster* yang dipilih sudah tepat, serta hasil *clustering* yang benar dalam merepresentasikan pola pada data. Untuk menguji kebenaran nilai k pada *K-means* dipilih metode validasi yang terbaik yaitu *Silhouette Score* dan *Davies Bouldin Index* (DBI).

3.5.1 Validasi Jumlah Cluster

Setelah menguji nilai k yang optimal, maka perlu adanya validasi terlebih dahulu agar mengetahui apakah nilai k = 2, k =3, atau k = 4 merupakan pilihan yang optimal untuk algoritma *K-means*. Jika menerapkan metode *Silhouette Score* dan DBI dengan python maka hasil visualisasi dari metode tersebut akan ditampilkan di layar perangkat, seperti yang ditunjukkan pada Gambar 8 dan Tabel 1.

JURTI, Vol.9 No.2, Juni 2025, ISSN: 2579-8790



Gambar 8. Visualisasi *Silhouette Score*

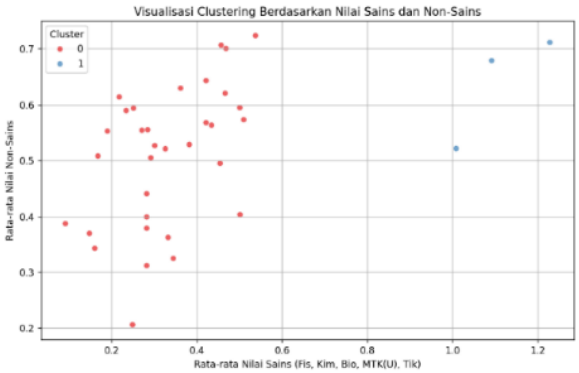
Tabel 1. Perbandingan Silhouette Score dan DBI

K	Silhouette Score	Davies Bouldin Index (DBI)
2	0.7380	0.3383
3	0.6302	0.4807
4	0.5823	0.5728

Dari hasil validasi kedua metode berdasarkan Gambar 8 dan Tabel 1 diperoleh bahwa jumlah *cluster* lebih ideal di angka 4 jika berdasarkan *Silhouette Score* dengan nilai 0.7380499034900105 atau 0.74 yang artinya nilai sudah bagus dalam pemisahan antara *cluster*. Selanjutnya, dari metode *Davies Bouldin Index* diperoleh nilai sebesar 0.3383745697588879 atau jika dibulatkan menjadi 0.34 yang artinya nilai ini bagus, karena semakin kecil nilainya atau n<1 berarti semakin baik kualitas *clustering*. Hal ini menandakan bahwa jumlah *cluster* yang ideal sebanyak dua *cluster*.

3.5.2 Visualisasi dengan Scatter Plot

Setelah diperoleh hasil dari *Silhouette Score* dan *Davies Bouldin Index* bahwa dua merupakan jumlah *cluster* yang optimal, selanjutnya adalah menerapkannya dengan kode dan disimpan dalam variabel k-means dan memberikan label untuk kolom *cluster*. Kemudian menampilkan hasil dari *clustering* dengan algoritma *K-means* menggunakan *Scatter Plot* yang dimana nanti terdapat nilai x dan y. Nilai x atau sains yang menyimpan pola nilai siswa seperti fisika, kimia, biologi, matematika (u), dan TIK untuk mengetahui perbedaan antar individu. Lalu, nilai y atau non-sains menyimpan pola nilai-nilai yang bukan termasuk mata pelajaran bidang sains. Hasil visualisasi ditunjukkan pada Gambar 9.



Gambar 9. Visualisasi Scatter Plot hasil *clustering*

Dari hasil visualisasi pada Gambar 9 diperoleh informasi bahwa *cluster* 0 sebanyak 33 siswa merupakan siswa yang memiliki minat pada bidang keilmuan non-sains (soshum dan bahasa), *cluster* 1 sebanyak 3 anggota adalah siswa yang minat pada bidang keilmuan sains (sains dan teknologi).

JURTI, Vol.9 No.2, Juni 2025, ISSN: 2579-8790

4. KESIMPULAN

Evaluasi performa *Silhouette Score* menunjukkan nilai sebesar 0.73 untuk k = 2 yang artinya sudah cukup tinggi dalam pemisahan *cluster*. *Davies Bouldin Index* menunjukkan nilai sebesar 0.34 untuk k = 2 yang artinya *clustering* semakin baik. Setelah visualisasi dengan *Scatter Plot* maka diperoleh cluster 0 sebanyak 33 yang merupakan siswa dengan minat pada bidang keilmuan soshum, sedangkan cluster 1 sebanyak 3 siswa merupakan mereka dengan minat pada bidang keilmuan saintek. Hasil clustering nilai yang diperoleh dapat digunakan untuk mempermudah proses penentuan minat bidang akademik siswa sebelum terjun ke perguruan tinggi. Normalisasi yang digunakan lebih dari satu sehingga nilai tidak terlihat konsisten, ini dikarenakan terdapat nilai biologi yang dinilai 90 sampai 91 untuk semua siswa yang menyebabkan nilai tersebut menjadi 0 jika menggunakan Min-Max Scaler. Sedangkan jika nilai berentang sempit seperti itu lebih cocok menggunakan Standard Scaler.

5. SARAN

Fitur non-akademik seperti ekstrakurikuler yang diikuti, minat dan tujuan kuliah perlu ditambahkan untuk mencegah nilai-nilai yang dipukul rata seperti pada nilai biologi. Kemudian data tahunan dapat ditambahkan agar dapat mengetahui pola minat siswa untuk tetap konsisten atau tidak.

UCAPAN TERIMA KASIH

Terima kasih banyak kepada Ibu Iin Fauzia, S.pd selaku guru sekolah yang telah bersedia memberikan data akademik siswa siswi XI 3 SMA N 1 Loa Janan, sehingga penelitian ini dapat terlaksana dengan baik.

DAFTAR PUSTAKA

[1] Tasya Alifia Izzani, Selva Octaria, and Linda Linda, “Perkembangan Masa Remaja,” JISPENDIORA Jurnal Ilmu Sosial Pendidikan Dan Humaniora, vol. 3, no. 2, pp. 259–273, Jul. 2024, doi: 10.56910/jispendiora.v3i2.1578.

[2] A. Febi, A. Zannah, A. I. Rindhiyani, S. S. Mutia, and L. Faizah, “Krisis Jati Diri, Eksistensi, dan Konflik Perbatinan yang Terjadi pada Remaja di Era Sekarang,” 2022. [Online]. Available: http://jurnalilmiah.org/journal/index.php/majemuk

[3] Cahyo Wulandari, Heni Prasetyaningrum, & Salma Salsabila Firdausa. (2024). Meningkatkan Potensi, Minat, dan Self-Awareness Siswa SMA Memasuki Perguruan Tinggi Melalui Program Faculty Fair KKN-PPM UGM. Jurnal Pengabdian, Riset, Kreativitas, Inovasi, Dan Teknologi Tepat Guna, 2(2), 314–324. https://doi.org/10.22146/parikesit.v2i2.17403

[4] F. Leonadus Lopo, W. Olok, and Y. Taebenu, “Pengaruh Pergaulan Terhadap Pemilihan Karir Siswa SMA Negeri 1 Amarasi,” Jurnal Penelitian dan Pengembangan Pendidikan, vol. 5, pp. 60–69, Jul. 2022.

[5] Nadi, N., & Agustini, N. (2020). PERAN MOTIVASI BERPRESTASI DAN STATUS SOSIAL EKONOMI TERHADAP MINAT MELANJUTKAN STUDI KE PERGURUAN TINGGI PADA SISWA PEREMPUAN DI TABANAN. PSIKOLOGI KONSELING, 17, 766–777. https://doi.org/10.24114/konseling.v17i2.2208

[6] A. A. Nabilah, C. A. Pravesti, K. Kunci, : Konformitas, T. Sebaya, and M. Karir, “Hubungan Konformitas Teman Sebaya Terhadap Minat Karir Siswa SMA Negeri 4 Sidoarjo The Relationship between Peer Conformity and Career Interests of Sidoarjo State High School 4 Students,” Hubungan

JURTI, Vol.9 No.2, Juni 2025, ISSN: 2579-8790

Konformitas Teman Sebaya Terhadap Minat Karir Siswa SMA Negeri 4 Sidoarjo Jurnal Guiding World, vol. 7, no. 2, 2024.

[7] Afrillia, S., Fadhli, T., Aini, Z., Konseling, B., & Jabar Ghafur, U. (2024). PENGARUH KONFORMITAS TEMAN SEBAYA TERHADAP MINAT SISWA MELANJUTKAN PENDIDIKAN KE PENGURUAN TINGGI. JURNAL PSIKO-KONSELING, 2(2), 9–18.

[8] H. Fahrunnisa and A. Murad, "Tabularasa: Jurnal Ilmiah Magister Psikologi Hubungan Empati dan Dukungan Sosial Teman Sebaya dengan Komunikasi Interpersonal pada Siswa Madrasah Aliyah Negeri (MAN) Binjai The Relationship between Empathy and Peer Social Support with Interpersonal Communication in Binjai State Aliyah Madrasa (MAN) Students," Jurnal Ilmiah Magister Psikologi, vol. 5, no. 1, pp. 11–20, 2023, doi: 10.31289/tabularasa.v5i1.

[9] Mrs. Listiowatty, "Keterlibatan Orang Tua pada Pemilihan Jurusan Perguruan Tinggi Siswa SMA," JIRA: Jurnal Inovasi dan Riset Akademik, vol. 2, no. 8, pp. 1226–1240, Aug. 2021, doi: 10.47387/jira.v2i8.205.

[10] Femas Aji Saputra, Adityawarman Adityawarman, and Salwa Rahman Nursyabani, "Analisis Dampak Kesalahan Pemilihan Jurusan terhadap Prestasi Akademik dan Kesejahteraan Psikologis Mahasiswa," Corona: Jurnal Ilmu Kesehatan Umum, Psikolog, Keperawatan dan Kebidanan, vol. 2, no. 2, pp. 180–192, May 2024, doi: 10.61132/corona.v2i2.418.

[11] W. Kokoh Andriyan, D. Palembang, J. A. Jenderal Yani No, and P. Sumatera Selatan, "Penerapan Data Mining Dengan Menggunakan Metode K-Means Clustering Dalam Pengelompokan Data Nilai Pada SMA YKPP PENDOPO Untuk Menentukan Jurusan Ipa Dan Ips,"Jurnal Jupiter, vol. 15, pp.452-461, 2023.

[12] S. Butsianto and N. Saepudin, "Penerapan Data Mining Terhadap Minat Siswa Dalam Mata Pelajaran Matematika Dengan Metode K-Means," Jurnal Nasional Komputasi dan Teknologi Informasi (JNKTI), vol. 3, no. 1, pp. 51–59, Apr. 2020, doi: 10.32672/jnkti.v3i1.2008.

[13] A. F. Zabidi, "Penerapan Algoritma K-Means untuk Pengelompokan Koleksi Perpustakaan dengan Data Mining," Media Jurnal Informatika, vol. 16, no. 2, p. 233, Dec. 2024, doi: 10.35194/mji.v16i2.4814.

[14] Susilowati, D., & Wicaksono, Y. (2024). Penerapan Data Mining Untuk Clustering Data Pencari Kerja dengan Menggunakan Algoritma K-Means. Pseudocode, 11(2), 54–58. https://doi.org/10.33369/pseudocode.11.2.54-58

[15] Awad, F. H., & Hamad, M. M. (2022). Improved k-Means Clustering Algorithm for Big Data Based on Distributed SmartphoneNeural Engine Processor. Electronics (Switzerland), 11(6). https://doi.org/10.3390/electronics11060883