

BAB II

LANDASAN TEORI

2.1 Minat dan Bakat

Minat adalah suatu sikap yang berlangsung terus menerus yang mampu membuat pola perhatian seseorang sehingga membuat dirinya jadi selektif terhadap objek minatnya. Sedangkan bakat adalah suatu kondisi pada seseorang yang memungkinkan untuk mencapai suatu kecakapan, pengetahuan, dan keterampilan dengan suatu latihan khusus [1]

Dalam Kamus Psikologi J.P. Chaplin menjelaskan bahwa minat adalah suatu sikap yang berlangsung terus menerus yang mampu membuat pola perhatian seseorang sehingga membuat dirinya jadi selektif terhadap objek minatnya. Sedangkan bakat menurut KBBI adalah dasar seperti kepandaian, sifat, dan pembawaan yang dibawa sejak lahir. Bakat dan minat seseorang dapat diketahui dengan berbagai cara, salah satunya adalah melakukan tes minat dan bakat.

Pelaksanaan tes minat dan bakat yang dilakukan seseorang membentuk data individu. Data dari dilaksanakan tes tersebut akan terus bertambah seiring berjalannya waktu dan hanya digunakan sebagai arsip atau sekedar laporan oleh lembaga penyelenggara. Dari kumpulan data tersebut dapat digali informasi baru melalui data mining yang dapat berguna bagi lembaga-lembaga tertentu.

2.1.1 Faktor-Faktor yang Mempengaruhi Minat dan Bakat Siswa

Faktor-faktor yang mempengaruhi minat dan bakat siswa meliputi sejumlah aspek internal dan eksternal yang memengaruhi preferensi, kecenderungan, dan kemampuan siswa dalam bidang tertentu. Berikut adalah faktor-faktor utama tersebut:

Faktor Internal

1. Kepribadian: Tipe kepribadian (misalnya introvert vs. ekstrovert) mempengaruhi minat siswa terhadap bidang tertentu (misalnya seni vs. sains).
2. Kecerdasan/kognisi: Tipe dan tingkat kecerdasan (logis, verbal, spasial, musikal, kinestetik, dll.) turut menentukan kecenderungan bakat.
3. Minat pribadi: Hal-hal yang menarik perhatian dan memberikan kesenangan secara subjektif pada siswa.
4. Bakat alami: Kemampuan atau potensi bawaan yang terlihat sejak dini dalam bidang tertentu.
5. Motivasi dan nilai pribadi: Apa yang dianggap penting atau memotivasi siswa, seperti prestasi, kreativitas, atau membantu orang lain.

Faktor Eksternal

1. Lingkungan keluarga: Dukungan, harapan, dan latar belakang keluarga sangat mempengaruhi pilihan minat dan bakat.
2. Pendidikan dan kurikulum: Akses terhadap berbagai mata pelajaran dan ekstrakurikuler dapat memperluas atau mempersempit minat siswa.
3. Guru dan bimbingan konseling: Peran guru dalam mengenali dan mengarahkan minat serta bakat siswa sangat signifikan.

4. Lingkungan sosial dan teman sebaya: Tekanan sosial atau ketertarikan terhadap apa yang populer di kalangan teman dapat memengaruhi minat.
5. Akses terhadap media dan teknologi: Informasi yang diperoleh dari media dapat mempengaruhi persepsi siswa tentang profesi atau bidang yang menarik.

2.2 Klaster (*Clustering*)

Klastering dapat diartikan sebagai metode untuk melakukan pengelompokan dengan kriteria tertentu berdasarkan kemiripan karakteristik dari setiap objek yang akan dikelompokkan. Sehingga objek yang memiliki karakteristik yang hampir sama akan terkumpul dalam satu kelompok. Sementara itu, Sifat objek antar kelompok memiliki kesamaan yang rendah. Dengan kata lain, keragaman (variasi) dalam suatu kelompok kecil sedangkan keragaman antar kelompok tinggi [2]. *Clustering* merupakan suatu metode untuk mencari dan mengelompokkan data yang memiliki kemiripan karakteristik (*similarity*) antara satu data dengan data yang lain [3].

2.3 Metode *K-Means*

Metode ini dilakukan pengelompokan data (*clustering*) dengan system partisi dan pemodelannya tanpa supervisi. Dalam metode ini pengelompokan data dilakukan menjadi beberapa kelompok, dan tiap kelompok memiliki kesamaan karakteristik dan memiliki karakteristik dengan kelompok lainnya [4].

2.3.1 *Knowledge Discovery in Database (KDD)*

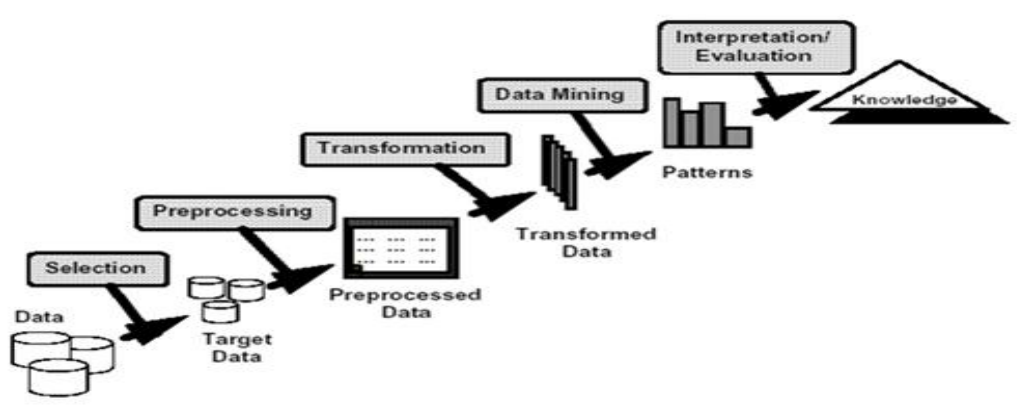
Knowledge Discovery in Database (KDD) merupakan metode yang digunakan untuk mencari pengetahuan atau informasi yang belum diketahui dari sebuah database. *Knowledge Discovery in Database (KDD)* merupakan keseluruhan proses non-trivial untuk mencari dan mengidentifikasi pola (pattern) dalam data, dimana pola yang ditemukan bersifat sah, baru, dapat bermanfaat dan dapat dimengerti.

Teknologi dan Informasi adalah suatu hal yang lumrah bagi masyarakat. Pesatnya perkembangan teknologi dan informasi tentunya memberikan pengaruh besar bagi masyarakat dan juga pastinya bagi pengusaha atau pebisnis dalam kalangsungan bisnisnya, terutama yang berhubungan dengan data. Tidak sedikit segala sesuatu yang berkaitan dengan bisnis, teknologi dan informasi tentunya selalu menggunakan data sebagai sumber informasinya.

Data mining adalah kegiatan untuk mengidentifikasi beragam jenis data. Dari database yang tersimpan, data-data akan diidentifikasi baik dicari kemungkinan adanya pola ataupun lainnya yang dianggap berpotensi untuk menghasilkan sesuatu yang bias dipakai oleh organisasi atau perusahaan yang memiliki data tersebut. *Knowledge Discovery In Database (KDD)* adalah salah satu metode yang bisa digunakan dalam melakukan data mining.

Knowledge Discovery In Database (KDD) merupakan nama lain dari Data Mining walaupun sesungguhnya kedua istilah tersebut memiliki konsep yang tidak sama, namun berkaitan satu sama lain, dan salah satu dari tahapan proses

keseluruhan *Knowledge Discovery In Database* (KDD) merupakan Data Mining yang menjadi inti dari proses KDD.



Gambar 2.1 Ilustrasi Proses Knowledge Discovery In Database

KDD adalah sebuah proses untuk mencari dan mengidentifikasi pattern dalam sebuah database, pada sebuah *Knowledge Discovery In Database* atau KDD memiliki beberapa tahapan di antaranya:

1. Seleksi Data dilakukan dengan memilih data yang relevan dengan tugas menganalisis dari database, menciptakan himpunan data target, atau memfokuskan pada contoh data dimana discovery akan dilakukan dan hasil dari seleksi disimpan dalam suatu berkas terpisah dari database operasional.
2. Pemrosesan dan Pembersihan Data Pada dilakukan menghilangkan noise dan data yang inkonsisten. Sebelum proses Data Mining dikerjakan, perlu dilakukn proses cleaning pada data yang akan menjadi fokus KDD dan proses cleaning mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data. Dilakukan proses enrichment, yaitu proses “memperkaya” data yang sudah ada dengan data lain yang relevan untuk keperluan KDD.

3. Transformasi adalah proses transformasi pada data yang dipilih, sehingga data tersebut sesuai untuk proses Data Mining. Proses ini merupakan proses kreatif dan sangat tergantung pada jenis atau pattern informasi yang akan dicari pada database.
4. Data Mining Pada tahapan ini merupakan proses mencari pattern atau pola dan informasi dari sebuah database dengan menggunakan teknik atau metode. Pada proses Data Mining terdapat banyak teknik, metode atau algoritma yang dapat digunakan dan sangat bervariasi dan untuk menentukan pemilihan metode yang akan digunakan tergantung pada tujuan dan proses KDD secara keseluruhan.
5. Evaluasi merupakan bagian dari proses KDD yang mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesa yang ada sebelumnya. Proses ini merupakan proses penerjemahan pola-pola yang didapatkan dari Data Mining. Pola informasi yang dihasilkan dari proses Data Mining perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan.

2.3.2 Data Mining

Data Mining merupakan sebuah inti dari proses KDD, meliputi dugaan algoritma yang mengeksplor data, membangun model dan menemukan pola yang belum diketahui. KDD bersifat otomatis, dapat didefinisikan sebagai pengorganisasian proses untuk pengidentifikasian yang benar, berguna dan penemuan pola dari kumpulan data yang besar dan kompleks. Data mining adalah suatu proses pengumpulan informasi dan data yang penting dalam jumlah yang besar.

Dalam proses ini sering kali memanfaatkan beberapa metode, seperti matematika, statistika dan pemanfaatan teknologi. Pengertian data mining ini juga dikenal dengan istilah lain yakni *Knowledge Discovery In Database* (KDD) dan Data Analysis.

Data Mining merupakan salah satu teknik untuk menemukan, mencari, atau menggali informasi atau pengetahuan baru dari sekumpulan data yang sangat besar, dengan integrasi atau penggabungan dengan disiplin ilmu lain seperti statistika, kecerdasan buatan, serta machine learning, menjadikan Data Mining sebagai salah satu alat bantu untuk menganalisa data yang kemudian menghasilkan informasi yang berguna. Dari penjelasan tersebut, Data Mining dapat diartikan sebagai sebuah proses dari sekumpulan data yang memiliki jumlah besar untuk mendapatkan data yang hasilnya tidak hanya sekedar informasi melainkan merupakan sebuah pengetahuan atau knowledge yang tersembunyi dari data tersebut. Knowledge yang diperoleh dari hasil mining terhadap sekumpulan data tersebut dapat membantu dalam pengambilan sebuah keputusan.

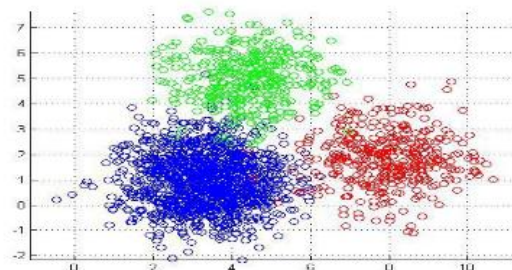
2.3.3 *K-Means*

K-Means sendiri merupakan suatu metode penganalisaan data atau metode Data Mining yang melakukan proses pemodelan tanpa supervisi (unsupervised) dan merupakan salah satu metode yang melakukan pengelompokan data dengan sistem partisi. Metode *K-Means* berusaha mengelompokkan data yang ada ke dalam beberapa kelompok, dimana data dalam satu kelompok mempunyai karakteristik yang sama satu sama lainnya dan mempunyai karakteristik yang berbeda dengan data yang ada didalam kelompok yang lain [5].

2.4 Algoritma K Means

Salah satu metode dalam klastering yang digunakan untuk mengelompokkan data ke dalam beberapa grup berdasarkan kesamaan atau kedekatannya. Algoritma ini termasuk dalam kategori *Unsupervised Learning* atau pembelajaran tanpa pengawasan, di mana tidak ada label atau kategori yang diberikan pada data yang dianalisis.

Tahapan tahapan yang akan di lakukan dalam algoritma *K-Means*:



K-Means Clustering

Gambar 2.2 K-Means Clustering

Nilai pada metode ini difokuskan untuk data bertipe numerik. Pada *clustering* terdapat istilah anomali. Anomali merupakan data yang telah di uji pada *clustering* dan tidak menemukan kelas yang cocok.

Metode *clustering* memiliki konsep apabila kedua objek yang memiliki nilai kemiripan yang tinggi maka akan menghasilkan nilai kesamaan yang tinggi [6]. Pada dasarnya kualitas hasil dari data *clustering* bergantung pada metode yang digunakan. Ada banyak cara untuk melakukan perhitungan *clustering* salah satunya Euclidean Distance. Euclidean distance merupakan rumus yang digunakan dalam menghitung jarak pada dua buah objek dengan menggunakan

nilai dari masing- masing objek. Berikut adalah rumus perhitungan dan pengukuran pada Euclidean distance:

1. Pertama, tentukan berapa banyak jumlah k (*cluster*)
2. Kedua, secara acak tentukan record yang menjadi lokasi pusat *cluster*.
3. Ketiga, temukan pusat *cluster* terdekat untuk setiap record.

Adapun persamaan yang sering digunakan dalam pemecahan masalah dalam menentukan jarak terdekat adalah persamaan Euclidean berikut:

$$d_{Euclidean}(x, y) = \sqrt{\sum_i (x_i - y_i)^2}$$

Dimana $x=x_1, x_2, x_3, \dots, x_m$ dan $y=y_1, y_2, y_3, \dots, y_m$, sementara m menyatakan banyaknya nilai atribut dari 2 buah record.

4. Keempat, tentukan *cluster* terdekat untuk setiap data dengan membandingkan nilai jarak terdekat, lalu perbaharui nilai pusat *clusternya*.

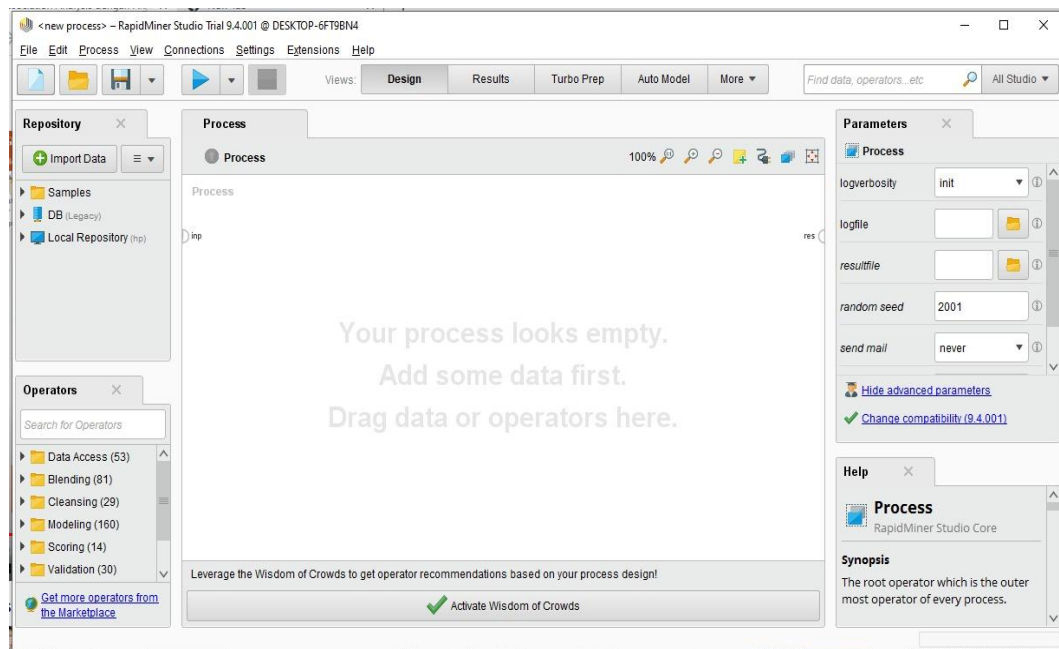
$$ClusterCenter = \sum \frac{a_i}{n}$$

5. Kelima, ulangi langkah 3 sampai 5 hingga tidak ada record yang berpindah *cluster* atau convergen.

2.5 Rapidminer

Rapid Miner merupakan perangkat lunak yang bersifat terbuka (open source). *Rapid Miner* adalah solusi untuk melakukan analisis terhadap *Data Mining*, text mining dan analisis pengelompokan. *RapidMiner* menggunakan

berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik.



Gambar 2.3 Lembar KerjaRapidminer

2.6 Alur Penelitian

Alur sistem penelitian adalah serangkaian konsep yang disusun secara sistematis untuk memastikan penelitian berjalan sesuai dengan tujuan yang telah ditetapkan. Kerangka penelitian ini disusun sebelum proses penelitian dimulai, dengan tujuan untuk menghubungkan berbagai variabel yang akan dianalisis. Dengan adanya kerangka kerja yang jelas, penelitian dapat dilaksanakan secara lebih terarah dan diterima oleh berbagai pihak terkait.

Berdasarkan diagram alur sistem penelitian, langkah awal yang dilakukan adalah mengidentifikasi permasalahan yang akan diteliti. Selanjutnya, peneliti menjelaskan urgensi masalah tersebut serta konteks penelitian yang relevan. Setelah permasalahan diidentifikasi, data yang diperlukan dikumpulkan dan

dianalisis untuk mendapatkan hasil yang valid. Dengan demikian, penelitian dapat memberikan solusi yang tepat terhadap permasalahan yang dihadapi.

Tahapan Penelitian:

1. Pengumpulan Data:

Menyusun dan menyebarkan kuesioner minat dan bakat dan mengumpulkan data dari sekolah jika tersedia.

2. Pra-pemrosesan Data:

Pembersihan data (handling missing value, normalisasi) dan transformasi skor menjadi fitur numerik.

3. Penentuan Nilai K:

Menggunakan metode Elbow untuk menentukan jumlah kluster optimal.

4. Proses *Clustering*:

Menerapkan algoritma *K-Means* dan mengelompokkan siswa berdasarkan hasil *clustering*.

5. Analisis Hasil:

Menginterpretasi masing-masing kluster (misalnya: kluster olahraga, seni, akademik) dan visualisasi hasil (misalnya: scatter plot, tabel distribusi siswa).

6. Penyusunan Rekomendasi:

Memberikan saran program/kegiatan yang sesuai dengan masing-masing kluster.

2.7 Penelitian Terdahulu

Tabel 2.1 Penelitian Terdahulu

No.	Tahun	Judul	Peneliti	Tujuan Penelitian	Kesimpulan
1	2024	Penerapan Metode <i>K-Means</i> untuk Klasterisasi Minat dan Bakat Siswa terhadap Ekstrakurikuler Sekolah	Subagio, Ridho Taufiq, and Viar Dwi Kartika.	Penelitian ini bertujuan untuk menerapkan metode <i>K-Means</i> dalam mengklasterisasi minat dan bakat siswa terhadap ekstrakurikuler di SMPN 2 Sedong untuk membantu menghemat waktu dalam menyeleksi siswa berdasarkan minat dan bakatnya.	Proses klasterisasi dilakukan menggunakan algoritma <i>K-Means</i> berdasarkan hasil kuesioner pada siswa kelas 7 tahun 2023/2024 untuk menentukan <i>cluster</i> berdasarkan data kuesioner masuk yang telah di-input oleh siswa serta sistem akan menghasilkan rekomendasi ekstrakurikuler.
2	2022	Klasterisasi Konsentrasi Keahlian Siswa SMK Berdasarkan Kurikulum Merdeka	Sukmayadi, Firman, Alamsyah Firdaus, and Christina Juliane.	Penelitian ini bertujuan untuk mengelompokkan minat peserta didik dalam memilih konsentrasi keahlian di SMK YPC Tasikmalaya berdasarkan Kurikulum Merdeka	Hasil penelitian menunjukkan bahwa pengelompokan minat peserta didik dalam memilih konsentrasi keahlian terbentuk menjadi 4 <i>cluster</i> .
3	2023	Penerapan Data Mining Untuk Mengetahui Minat Siswa Pada Pelajaran IPA Menggunakan	Mardiansa, Mardiansa, Herlina Latipa Sari, and Prahasti Prahasti.	Tujuan dari penelitian ini adalah membangun aplikasi data mining untuk mengetahui minat siswa pada mata pelajaran IPA di SMPN 18 Kota	Hasil dari penelitian ini terdapat 3 kelompok minat siswa dalam mata pelajaran IPA yaitu siswa yang berminat tinggi berjumlah 12 orang, siswa yang minat sedang berjumlah 17 orang dan siswa yang

		Metode <i>K-Means Clustering</i>		Bengkulu.	minat rendah berjumlah 1 orang di SMPN 18 Kota Bengkulu.
4	2024	Implementasi kurikulum merdeka dalam meningkatkan minat bakat siswa.	Saiful, Febby Eka Putri, Milatul Hasanah, and Amir Bandar Abdul Majid	Tujuan kegiatan ini adalah untuk memberikan pendampingan dan mereview Kurikulum Merdeka dengan melihat minat siswa.	Hasil kegiatan pengabdian menunjukkan guru-guru telah memahami terkait kurikulum merdeka dan telah mampu mengimplementasikan kurikulum merdeka dalam sebuah perencanaan pembelajaran dalam modul ajar sesuai dengan panduan pembelajaran dan asesmen pendidikan usia dini, pendidikan dasar, dan menengah.
5	2022	<i>Clustering</i> Data Rekomendasi Penjurusan Ke Perguruan Tinggi Menggunakan Metode <i>K-Means</i> Studi Kasus Sma Negeri 2 Palembang.	Purboyo, R., and Ahmad Dicky Syarief.	Memberikan solusi untuk penentuan jurusan ke perguruan tinggi menggunakan metode <i>k-means clustering</i> dengan metode ini nantinya siswa-siswi dapat mengetahui jurusan apa yang cocok bagi mereka setelah mereka lulus sekolah dan ingin melanjutkan pendidikan ke perguruan tinggi.	Dalam penelitian ini siswa-siswi dikelompokkan dengan menerapkan metode data mining. Kelompok didasarkan pada atribut jurusan, minat, sifat, hobi, bakat, dan nilai rata-rata mata pelajaran jurusan ipa dan jurusan ips. untuk melakukan <i>clustering</i> data digunakan metode <i>K-Means</i> dan metode pengukuran jarak Euclidean dan dianalisis menggunakan perhitungan manual menggunakan Microsoft excel dan tools <i>RapidMiner</i> .
6	2024	Implementasi Algoritma <i>K-Means Clustering</i> Dengan	Palevi, Muhammad Rheza, and Zulfahmi	penelitian ini dilakukan untuk membantu siswa dalam menentukan jurusan ke perguruan tinggi	<i>K-Means Clustering</i> dengan penerapan Active Learning digunakan untuk menentukan jurusan yang tepat dengan menggunakan data kuesioner dan nilai

		Pendekatan Active Learning Pada Siswa SMA Untuk Menentukan Jurusan Ke Perguruan Tinggi.	Indra	menggunakan algoritma <i>K-Means Clustering</i> dengan penerapan Active Learning in Machine Learning.	rapot yang didapat dari pihak sekolah. Pada penelitian ini, didapat hasil bahwa 113 siswa yang menjadi responden terbagi kedalam 12 kategori atau klaster pada bidang ilmu yang berbeda-beda.
7	2019	Penerapan data mining penentu minat Dan bakat siswa Smk dengan metode C4. 5	Merawati, Dita, and Rino Rino.	Penelitian ini dilakukan untuk mengetahui tingkat keakuratan yang didapatkan berdasarkan penerapan metode yang dilakukan.	Hasil dari penerapan metode C4.5 menunjukkan bahwa metode C4.5 dapat diterapkan pada pengolahan data tes minat dan bakat tersebut dengan hasil akurasi yaitu 82,65% pada tes bakat dan 90,23 pada tes minat. Penerapan metode C4.5 menghasilkan pohon keputusan.
8	2023	Pengelompokan Data Calon Siswa Baru Di Sekolah Menengah Kejuruan menggunakan Algoritma <i>K-Means</i>	Purnamasari, Ningsi Multi, Achmad Syauqi, and Danar Ardian Pramana	Penelitian ini bertujuan untuk mengelompokan data calon siswa baru berdasarkan jurusan dengan menggunakan variabel nilai dan minat siswa, dengan menerapkan metode <i>K-Means</i> .	Hasil dari penelitian ini diharapkan dapat membantu dalam pengolahan data untuk penentuan jurusan.

9	2024	<i>Clustering the K-means Algorithm with the Approach to Student Interpersonal Communication Patterns in Selecting Secondary Schools.</i>	Wulan, Rayung, Themotia Titi Widaningsih, and Fit Yanuar	This research aims to understand students' communication patterns in choosing secondary schools by identifying existing group patterns, and understanding the factors that influence students' decisions in choosing secondary schools.	The results of tests carried out with 4 <i>clusters</i> show an accuracy value close to 0, with <i>cluster</i> 2 having a value of -0.150. The results show that students who choose a secondary school based on their psychological test results and interpersonal communication between parents, homeroom teachers and the school are the dominant values in the continuity of selecting a senior secondary school
10	2020	Perbandingan Hasil Analisis Klasterisasi Data Bimbingan Konseling Peserta Didik Tingkat Menengah Pertama sebagai Identifikasi Kinerja Akademik.	Shanty, Ratna Nur Tiara, Rizki Nurhidayat, and Dwi Cahyono.	Pada penelitian ini dilakukan perbandingan hasil analisis pola kinerja siswa menggunakan metode <i>K-Means</i> dan Hierarchical Single Linkage, Average Linkage dan Complete Linkage.	Hasil pada penelitian ini disimpulkan bahwa metode klasterisasi menghasilkan kelompok-kelompok siswa dengan kecenderungan nilai fitur yang berbeda, dan metode Hierarchical Single Linkage memiliki performa klasterisasi terbaik dengan rasio sebesar 0.004435.