



Yuni Franata Sinurat, S.Kom.  
Masrizal, S.Kom., M.Kom.  
Irmayanti, S.Si., M.Pd.



# *Data Mining* Pengelompokan Siswa Berprestasi Menggunakan Metode *Clustering*



*Data Mining* Pengelompokan Siswa Berprestasi  
Menggunakan Metode *Clustering*

Copyright © 2024  
PENERBITANEM

**KUTIPAN PASAL 72:**  
**Ketentuan Pidana Undang-Undang Republik**  
**Indonesia**  
**Nomor 19 Tahun 2002 tentang HAK CIPTA**

1. Barang siapa dengan sengaja dan tanpa hak melakukan perbuatan sebagaimana dimaksud dalam Pasal 2 ayat (1) atau Pasal 49 ayat (1) dan ayat (2) dipidana dengan pidana penjara masing-masing paling singkat 1 (satu) bulan dan/atau denda paling sedikit Rp1.000.000,00 (satu juta rupiah), atau pidana penjara paling lama 7 (tujuh) tahun dan/atau denda paling banyak Rp5.000.000.000,00 (lima miliar rupiah).
2. Barang siapa dengan sengaja menyiarkan, memamerkan, mengedarkan, atau menjual kepada umum suatu Ciptaan atau barang hasil pelanggaran Hak Cipta atau Hak Terkait sebagaimana dimaksud dalam ayat 1, dipidana dengan pidana penjara paling lama 5 (lima) tahun dan/atau denda paling banyak Rp500.000.000,00 (lima ratus juta rupiah).

Yuni Franata Sinurat, S.Kom.  
Masrizal, S.Kom., M.Kom.  
Irmayanti, S.Si., M.Pd.

*Data Mining* Pengelompokan Siswa Berprestasi  
Menggunakan Metode *Clustering*



Pekalongan - Indonesia

# ***Data Mining* Pengelompokan Siswa Berprestasi Menggunakan Metode *Clustering***

Copyright © 2024

## **Penulis:**

Yuni Franata Sinurat, S.Kom.  
Masrizal, S.Kom., M.Kom.  
Irmayanti, S.Si., M.Pd.

## **Penyunting:**

Moh. Nasrudin  
(SK BNSP: No. Reg. KOM.1446.01749 2019)

## **Penata Letak:**

Nurul Diyah Anggraeni

## **Desain Sampul:**

Ahmad Arifin

Diterbitkan oleh:

**PT Nasya Expanding Management**  
(Penerbit NEM - Anggota IKAPI)

Jl. Raya Wangandowo, Bojong  
Pekalongan, Jawa Tengah, Indonesia, 51156  
Telp. (0285) 435833, Mobile: 0853-2521-7257  
[www.penerbitnem.com](http://www.penerbitnem.com) / [penerbitnem@gmail.com](mailto:penerbitnem@gmail.com)

Hak Cipta dilindungi oleh Undang-Undang.  
Dilarang memperbanyak sebagian  
atau seluruh isi buku ini tanpa izin tertulis dari Penerbit

Cetakan ke-1, Juli 2024

**ISBN: 978-623-115-458-3**

## **Prakata**

Segala puji syukur kepada Tuhan YME, berkat-Nya masih memberikan penulis kesehatan sehingga masih berkesempatan untuk menyelesaikan tugas yang mulia yakni menulis buku yang berjudul *“Data Mining Pengelompokan Siswa Berprestasi Menggunakan Metode Clustering”*.

Penulis mengucapkan terima kasih kepada mama tercinta dan terhebatku karena atas perjuangan dan doanyalah penulis mampu menyelesaikan penulisan buku ini. Penulis juga mengucapkan terima kasih kepada Nesty Sridawani Sinurat selaku sebagai kakak penulis, karena berkat dukungannya penulis mampu menyelesaikan buku ini dengan baik.

Tanpa andil dari pihak-pihak tersebut, mustahil buku ini dapat terselesaikan. Penulis juga mengakui jika masih banyak terdapat kekurangan dari buku ini. Oleh karena itu, kritik dan saran yang konstruktif sangat diharapkan penulis. Demikian, semoga buku sederhana ini dapat bermanfaat.

Rantauprapat, Juli 2024

**Yuni Franata Sinurat, S.Kom.**

# Daftar Isi

PRAKATA \_\_ v

DAFTAR ISI \_\_ vi

BAB 1 PENDAHULUAN \_\_ 1

BAB 2 TEORI DATA ANALITIK \_\_ 7

A. Konsep Sistem \_\_ 7

B. Sistem Informasi \_\_ 9

C. *Data Mining* \_\_ 14

D. *Clustering* \_\_ 16

E. Algoritma K-Means \_\_ 22

F. Aplikasi Orange \_\_ 26

G. RapidMiner \_\_ 29

BAB 3 *DATA MINING* PENGELOMPOKAN SISWA  
BERPRESTASI MENGGUNAKAN METODE *CLUSTERING*  
\_\_ 35

BAB 4 PENUTUP \_\_ 53

DAFTAR PUSTAKA \_\_ 55

TENTANG PENULIS



# Bab 1

## PENDAHULUAN

Pada era sekarang perkembangan teknologi khususnya informasi dalam pengolahan data membuat masyarakat dituntut untuk menimba pengetahuan dan meningkatkan kemampuan di bidang teknologi dalam pengolahan data. Kebutuhan informasi menjadi suatu elemen penting dalam pengambilan keputusan. Permasalahan yang sering terjadi adalah informasi yang dibutuhkan harus cepat dan kebutuhan informasi tidak sesuai seperti ketika informasi yang dibutuhkan lambat dan perlu mencari file-file yang tersimpan di lemari arsip yang jumlahnya sangat banyak. Untuk itu dibuatlah suatu sistem dengan pengolahan jumlah data yang besar dengan teknik *data mining* dengan metode k-means.

Siswa atau pelajar merupakan peserta didik yang merupakan komponen masukan dalam proses pendidikan yang bertujuan untuk menghasilkan manusia yang berkualitas sesuai dengan tujuan pendidikan nasional. Dalam rangka meningkatkan dan mendistribusikan kemampuan pembelajaran siswa, saat ini banyak diterapkan model pembelajaran dengan menggunakan pengelompokan siswa atau sering dikenal dengan

kelompok belajar. Kelompok belajar pada dasarnya membagi himpunan siswa ke dalam beberapa kelompok atau himpunan kecil yang mana dalam kelompok tersebut siswa dapat saling bekerja sama dalam proses pembelajaran sehingga kemampuan siswa dapat terdistribusi di antara siswa.

Pengelompokan siswa ke dalam kelompok sering kali mengalami kendala untuk menentukan kelompok yang optimal. Kelompok belajar yang optimal adalah kelompok dimana terdapat siswa dengan kemampuan yang bervariasi sehingga distribusi kemampuan siswa dalam kelompok tersebut dapat terwujud. Sulitnya mengelompokkan siswa ke dalam kelompok-kelompok yang optimal menjadi tantangan tersendiri yang membutuhkan model matematis berdasarkan informasi dari data kemampuan tiap siswa.

Maka dari itu untuk menyelesaikan masalah tersebut, *data mining* sebagai salah satu metode untuk menggali pengetahuan yang dapat digunakan untuk menganalisis data akademik. Salah satu teknik dalam *data mining* untuk mengelompokkan *data yang* memiliki kemiripan karakteristik dengan data lainnya. Dengan menggunakan *clustering*, maka kelompok belajar yang dihasilkan akan beranggotakan siswa-siswi yang memiliki kemiripan dalam nilai suatu mata pelajaran.

Jumlah data murid terus bertambah setiap tahunnya. Hal ini mengakibatkan akumulasi data yang belum diolah secara baik digunakan untuk mengungkap pengetahuan dan informasi baru dengan pattern yang terbentuk sebagai hasil akumulasi data. Jumlah pemrosesan data yang terus meningkat mengharuskan penggunaan strategi dan metode sehingga dapat ditransformasikan menjadi informasi dan pengetahuan yang dapat dimanfaatkan oleh pendidik dalam

proses pembuatan kebijakan. Hal ini menjadi permasalahan bagi guru maupun wali kelas yaitu bagaimana cara menentukan tingkat prestasi murid yang rendah, cukup, dan tinggi serta menemukan *top rank* murid unggulan dalam kelas, agar dapat membentuk kelas yang ideal untuk meningkatkan prestasi maupun memotivasi murid.

Pendidikan merupakan aspek kunci dalam pembentukkan Masyarakat yang berkualitas. Salah satu keberhasilan indikator suatu sistem Pendidikan adalah prestasi siswa. Prestasi siswa mencakup berbagai aspek, termasuk pencapaian akademis, partisipasi dalam kegiatan ekstrakurikuler, dan perkembangan keterampilan individu. Penggunaan *data mining* khususnya klastering dapat memberikan kontribusi yang signifikan dalam pemahaman dan peningkatan prestasi siswa.

Data merupakan salah satu komponen yang penting di era digital, seluruh aktivitas data akan terekam dan disimpan dalam media penyimpanan data atau database, sehingga data ini dapat diproses untuk mendapatkan informasi yang berguna. Dalam era digital pemanfaatan teknologi dalam pengolahan data sudah diterapkan di berbagai sektor seperti aktivitas transaksi, pendidikan, pelayanan masyarakat, perbankan, transportasi, dan masih banyak lagi. Pengolahan data dengan menggunakan komputer sering dikenal dengan nama Electronic Data Processing (EDP). Siklus pengolahan data pada ini mempunyai tiga tahapan yaitu Input, Processing, dan Output. Tahapan tersebut dikembangkan dengan menambahkan tahap Storage dan menjadi model siklus pengolahan data (Data Processing Cycle). Permasalahan yang muncul dalam pengolahan data ialah bagaimana dapat mengetahui informasi yang akan dihasilkan dari basis data yang sangat besar. Knowledge Discovery in Database (KDD)

merupakan sebuah proses untuk menemukan pengetahuan dalam sebuah basis data. KDD dapat dijelaskan sebagai sebuah proses ekstraksi atau identifikasi pola, pengetahuan dan informasi dari sekumpulan data yang besar. Pengetahuan dan informasi yang dihasilkan dari KDD bersifat baru, mudah dimengerti, dan bermanfaat bagi para pengguna informasi.

*Data mining* merupakan sebuah metode dalam bidang ilmu komputer yang digunakan dalam mencari pengetahuan dari data sehingga menjadi sebuah informasi yang bermanfaat. Tahapan dalam proses *data mining* berguna untuk mencari sebuah pola tertentu dari data penilaian yang sangat banyak. Bidang pendidikan pemanfaatan *data mining* dapat sebagai bahan evaluasi dari proses pembelajaran siswa, hasil dari data mining dapat menjadi bahan evaluasi instansi sekolah dalam proses apa yang akan diajarkan serta bagaimana cara mengajar yang baik yang dilakukan oleh guru.

*Prediction* atau fungsi prediksi merupakan salah satu fungsi *data mining*. Maksudnya yaitu dari proses nanti akan menemukan pola tertentu dari suatu data. Pola tersebut dapat diketahui dari variabel yang ada pada data. Pola yang didapat bisa digunakan untuk memprediksi variabel lain yang belum diketahui nilai ataupun jenisnya. Karena itulah fungsi satu ini dikatakan sebagai fungsi prediksi. Nantinya bisa digunakan untuk memprediksi variabel tertentu yang tidak ada dalam suatu data. Hal ini tentunya memudahkan dan menguntungkan bagi mereka pemilik kepentingan yang memerlukan prediksi akurat untuk membuat hal penting tersebut menjadi lebih baik. Pengertian kompetensi adalah suatu kemampuan atau kecakapan yang dimiliki oleh seseorang dalam melaksanakan suatu pekerjaan atau tugas di bidang tertentu, sesuai dengan jabatan yang disandangnya.

Pendapat lain mengatakan arti kompetensi adalah suatu keterampilan, pengetahuan, sikap dasar, dan nilai yang terdapat dalam diri seseorang yang tercermin dari kemampuan berpikir dan bertindak secara konsisten.

*K-Means clustering* merupakan sebuah algoritma *unsupervised learning* yang digunakan dalam pengelompokan data dalam dataset yang tidak memiliki label ke dalam sebuah cluster-cluster yang berbeda. *K-Means clustering* memungkinkan pengguna melakukan pengelompokan data ke dalam cluster berdasarkan variabel-variabel yang ada tanpa harus melalui proses training data terlebih dahulu. Tujuan utama dari algoritma *K-Means clustering* ini untuk meminimalisir jarak antara titik data dengan cluster yang sesuai. Algoritma *k-means* juga sudah banyak digunakan di berbagai bidang. Seperti bidang marketing.

Perbedaan algoritma *clustering* dan klasifikasi adalah *clustering* tidak memiliki *class/target/label*, jadi bisa dikatakan sebagai *unsupervised learning*. *Clustering* bukan klasifikasi karena tidak mengelompokkan kelas data yang telah ada.

*Clustering* adalah teknik untuk membedakan kumpulan data menjadi banyak kelompok dilihat dari kecocokan yang diinginkan. *Clustering* dalam *data mining* adalah kumpulan data atau objek dalam *cluster* (grup) dan menjadikan setiap *cluster* memiliki data yang hampir mirip dengan aslinya serta dapat dibedakan dari objek di *cluster* lainnya. Pengelompokan diperlukan karena data yang belum diolah sangat besar dan tidak mudah untuk dilakukan analisis maupun dipelajari. Tujuan pengelompokan dalam skenario ini adalah untuk lebih memahami data dan menganalisis kualitas data. Metode *clustering* yang paling banyak dimanfaatkan dalam kajian

ilmiah dan industri adalah algoritma *K-Means*. Setiap *cluster* diwakili oleh *mean* dari data dalam *cluster* tersebut.



Copyright © 2024  
PENERBIT NEM



## Bab 2

# TEORI DATA ANALITIK

### A. Konsep Sistem

Sistem biasanya merujuk pada sistem komputer atau sistem informasi, yang terdiri dari perangkat keras, perangkat lunak, dan pengguna yang bekerja bersama untuk mengumpulkan, memproses, menyimpan, dan menyebarkan informasi. Sistem adalah suatu kumpulan atau himpunan dari kegiatan, komponen, unsur, elemen atau variabel yang terorganisir, saling berinteraksi dan saling berhubungan satu sama lain melakukan kerjasama dengan cara-cara tertentu secara harmonis sehingga membentuk kesatuan untuk melaksanakan suatu fungsi untuk mencapai tujuan yang telah ditetapkan.

Adapun karakteristik sistem yaitu:

#### 1. Komponen Sistem

Suatu sistem terdiri dari sejumlah komponen yang saling berinteraksi, yang bekerja sama membentuk satu kesatuan. Komponen-komponen sistem tersebut dapat berupa suatu bentuk subsistem. Setiap subsistem memiliki

sifat-sifat sistem yang menjalankan suatu fungsi tertentu dan mempengaruhi proses sistem secara keseluruhan.

## 2. Batasan Sistem

Ruang lingkup sistem merupakan daerah yang membatasi antara sistem dengan sistem lainnya atau sistem dengan lingkup luarnya. Batasan sistem ini memungkinkan suatu sistem dipandang sebagai satu kesatuan yang tidak dapat dipisah-pisahkan.

## 3. Lingkup Luar Sistem

Bentuk apa pun yang ada di luar ruang lingkup atau batasan sistem yang mempengaruhi operasi sistem tersebut disebut dengan lingkungan luar sistem. Lingkungan luar sistem ini dapat menguntungkan dan dapat juga merugikan sistem tersebut. Lingkungan luar yang menguntungkan merupakan energi bagi sistem tersebut, yang dengan demikian lingkungan luar tersebut harus selalu dijaga dan dipelihara.

## 4. Penghubung Sistem

Media yang menghubungkan sistem dengan subsistem yang lain disebut dengan penghubung sistem atau interface. Penghubung ini memungkinkan sumber-sumber daya mengalir dari satu subsistem ke subsistem yang lain. Keluaran suatu subsistem akan menjadi masukan untuk subsistem yang lain dengan melewati penghubung.

## 5. Masukan Sistem

Energi yang dimasukkan ke dalam sistem disebut masukan sistem, yang dapat berupa pemeliharaan dan

sinyal. Sebagai contoh, di dalam suatu unit sistem komputer, program adalah maintenance input yang digunakan untuk mengoperasikan komputer. Sementara data adalah sinyal input yang akan diolah menjadi informasi.

#### 6. Keluaran Sistem

Hasil energi yang diolah dan diklasifikasikan menjadi keluaran yang berguna. Keluaran ini merupakan masukan bagi subsistem yang lain. Seperti contoh sistem informasi, keluaran yang dihasilkan adalah informasi, di mana informasi ini dapat digunakan sebagai masukan untuk pengambilan keputusan atau hal-hal lain yang merupakan input bagi subsistem lainnya.

#### 7. Pengolahan Sistem

Suatu sistem dapat mempunyai suatu proses yang akan mengubah masukan menjadi keluaran. Sebagai contoh, sistem akuntansi. Sistem ini akan mengolah data transaksi menjadi laporan-laporan yang dibutuhkan oleh pihak manajemen.

#### 8. Sasaran Sistem

Suatu sistem memiliki tujuan dan sasaran yang pasti dan bersifat deterministik. Kalau suatu sistem tidak memiliki sasaran, maka operasi sistem tidak ada gunanya. Suatu sistem dikatakan berhasil bila mengenai sasaran atau tujuan yang telah direncanakan.

### **B. Sistem Informasi**

Sistem informasi adalah kombinasi dari teknologi informasi, orang, dan prosedur yang dirancang untuk

mengumpulkan, memproses, menyimpan, dan mendistribusikan informasi. Tujuan utamanya adalah untuk mendukung operasi, manajemen, dan pengambilan keputusan dalam suatu organisasi. Sistem informasi dapat digunakan dalam berbagai konteks, seperti bisnis, pemerintahan, pendidikan, dan lainnya.

Berikut adalah komponen utama dari sistem informasi:

1. Perangkat Keras

Komputer, server, perangkat jaringan, dan perangkat lain yang digunakan untuk mengumpulkan, menyimpan, memproses, dan mendistribusikan informasi.

2. Perangkat Lunak

Aplikasi dan sistem operasi yang mengendalikan perangkat keras dan melakukan tugas pengolahan data.

3. Data

Informasi mentah yang diolah menjadi bentuk yang berguna untuk pengambilan keputusan.

4. Orang

Pengguna yang berinteraksi dengan sistem informasi, termasuk administrator, pengembang, dan pengguna akhir.

5. Proses

Aturan dan pedoman yang digunakan untuk mengoperasikan sistem dan memproses data.

Sistem informasi dapat dibagi menjadi beberapa jenis berdasarkan fungsinya, antara lain:

1. MIS Sistem Informasi Manajemen

Mendukung fungsi manajemen dalam pengambilan keputusan dengan menyediakan informasi yang relevan dan tepat waktu.

2. Sistem Pendukung Keputusan

Membantu dalam membuat keputusan dengan menyediakan analisis data yang mendalam.

3. Sistem Informasi Eksekutif

Menyediakan informasi yang disesuaikan untuk eksekutif perusahaan.

4. Sistem Informasi Geografis

Mengelola dan menganalisis data geospasial.

5. Sistem Informasi Akuntansi

Mengelola transaksi akuntansi dan keuangan.

Fungsi utama dari sistem informasi meliputi:

1. Pengumpulan Data

Mengumpulkan data dari berbagai sumber untuk diproses lebih lanjut.

2. Pengolahan Data

Mengonversi data mentah menjadi informasi yang bermakna melalui berbagai teknik pengolahan, seperti perhitungan, analisis, dan interpretasi.

### 3. Penyimpanan Data

Menyimpan data dalam basis data atau media penyimpanan lainnya untuk diakses dan digunakan di masa mendatang.

### 4. Distribusi Informasi

Menyebarkan informasi kepada pengguna yang membutuhkan, baik dalam bentuk laporan, dashboard, atau alat lainnya.

Sistem informasi memiliki beberapa sifat atau karakteristik utama yang mempengaruhi bagaimana mereka dirancang, diimplementasikan, dan digunakan. Berikut adalah beberapa sifat penting dari sistem informasi:

#### 1. Kompleksitas

Sistem informasi sering kali kompleks karena melibatkan banyak komponen, termasuk perangkat keras, perangkat lunak, data, dan pengguna. Komponen-komponen ini harus berinteraksi dengan lancar untuk mencapai tujuan sistem.

#### 2. Integrasi

Sistem informasi sering kali harus terintegrasi dengan sistem lain dalam organisasi untuk memungkinkan pertukaran data dan mendukung proses bisnis yang saling terkait.

#### 3. Ketersediaan

Sistem informasi harus tersedia kapan saja dibutuhkan untuk mendukung operasi bisnis yang

berkelanjutan. Ketersediaan yang tinggi sering kali dicapai melalui redundansi dan pemulihan bencana.

#### 4. Keamanan

Sistem informasi harus memiliki mekanisme untuk melindungi data dan informasi dari akses yang tidak sah, kehilangan, atau kerusakan. Ini melibatkan langkah-langkah keamanan fisik dan digital.

#### 5. Keandalan

Sistem informasi harus berfungsi dengan baik dan memberikan hasil yang akurat dan konsisten. Keandalan juga mencakup kemampuan sistem untuk pulih dari kesalahan atau gangguan.

#### 6. Fleksibilitas

Sistem informasi harus dapat beradaptasi dengan perubahan kebutuhan bisnis dan teknologi. Fleksibilitas ini memungkinkan sistem untuk berkembang dan diubah seiring waktu.

#### 7. Efisiensi

Sistem informasi harus dirancang untuk memaksimalkan penggunaan sumber daya dan meminimalkan pemborosan. Efisiensi mencakup penggunaan waktu, tenaga kerja, dan sumber daya lainnya secara optimal.

#### 8. Akurat

Data dan informasi yang dihasilkan oleh sistem harus akurat dan bebas dari kesalahan. Keakuratan ini penting untuk pengambilan keputusan yang tepat.

9. Skalabilitas

Sistem informasi harus dapat menangani peningkatan beban kerja atau jumlah pengguna tanpa mengalami penurunan kinerja. Skalabilitas memungkinkan sistem untuk tumbuh seiring dengan pertumbuhan organisasi.

10. Interaktivitas

Sistem informasi sering kali dirancang untuk interaksi pengguna yang efisien dan efektif, termasuk antarmuka pengguna yang intuitif dan mudah digunakan.

11. Relevansi

Informasi yang dihasilkan oleh sistem harus relevan dengan kebutuhan pengguna dan tujuan bisnis yang ingin dicapai.

Sifat-sifat ini membantu memastikan bahwa sistem informasi dapat mendukung operasi dan strategi bisnis dengan cara yang efektif, efisien, dan aman.

**C. *Data Mining***

*Data mining* merupakan serangkaian proses untuk menggali nilai tambah berupa informasi yang dihasilkan dengan cara mengekstraksi dan mengenali pola yang terdapat dalam basis data. *Data mining* digunakan untuk mencari pengetahuan dalam jumlah data yang besar atau disebut dengan *Knowledge Discovery in Database* (KDD), yang artinya KDD berhubungan dengan pola-pola sejumlah data yang besar dengan teknik integrasi dan penemuan ilmiah, interpretasi dan visualisasi.

Secara skematis proses data mining dibagi beberapa langkah proses pelaksanaan data mining dalam 3 (tiga) aktivitas yaitu:

1. Eksplorasi Data
2. Membuat Model dan Pengujian Validasi Model
3. Penerapan Model

*Data mining* bermaksud menyelesaikan permasalahan dengan membangun model berdasarkan data yang sudah digali. Beberapa metode yang dimiliki data mining dalam menyelesaikan permasalahan tersebut salah satunya clustering.

*Data mining* adalah proses menggunakan pendekatan tertentu untuk menemukan pattern atau informasi tertentu dalam sekumpulan data. Data mining atau bisa disebut *Knowledge Discovery in Database* (KDD) merupakan proses yang melibatkan penghimpunan dan analisis data historis untuk mengungkap sistem, pattern, apa keterkaitan dalam kumpulan data besar. Hasil pemrosesan *data mining* dapat dimanfaatkan untuk meningkatkan kemampuan dalam pengambilan keputusan di waktu mendatang.

Tujuan kajian ini yaitu mengetahui dan membentuk cluster data siswa berdasarkan nilai akademik, nilai sikap, dan nilai disiplin sehingga menjadi sebuah cluster sehingga hasil cluster siswa dapat menjadi acuan dalam meningkatkan nilai siswa dalam proses pembelajaran selanjutnya. Hasil evaluasi dan penilaian terhadap siswa dilakukan oleh tenaga pengajar atau guru dalam melakukan penilaian selama proses pembelajaran.

*Data mining* memiliki banyak sekali fungsi, untuk fungsi utamanya sendiri yaitu ada dua, yaitu fungsi descriptive dan fungsi predictive. Untuk fungsi lainnya akan dibahas di bawah:

1. Fungsi Deskripsi dalam *Data Mining*

Fungsi deskripsi dalam *data mining* adalah sebuah fungsi untuk memahami lebih jauh tentang data yang diamati. Dengan melakukan sebuah proses diharap bisa mengetahui perilaku dari sebuah data tersebut. Dengan menggunakan fungsi *descriptive data mining*, maka nantinya bisa menemukan pola tertentu yang tersembunyi dalam sebuah data. Dengan kata lain jika pola yang berulang dan bernilai itulah karakteristik sebuah data bisa diketahui.

2. Fungsi Prediksi

Fungsi prediksi merupakan sebuah fungsi bagaimana sebuah proses nantinya akan menemukan pola tertentu dari suatu data. Pola-pola tersebut dapat diketahui dari berbagai variabel-variabel yang ada pada data.

**D. *Clustering***

*Clustering* merupakan aktivitas (*task*) yang bertujuan mengelompokkan data yang memiliki kemiripan antara satu data dengan data lainnya ke dalam klaster atau kelompok sehingga data dalam satu klaster memiliki tingkat kemiripan (*similarity*) yang maksimum dan data antar klaster memiliki kemiripan yang minimum. *Clustering* juga dapat diartikan metode segmentasi data yang diimplementasikan dalam beberapa bidang, di antaranya marketing, analisa masalah bisnis segmentasi pasar dan prediksi, pola dalam bidang computer vision, zonasi wilayah hingga identifikasi obyek dan pengolahan citra.

Pada kajian ini dilakukan *clustering* dengan menggunakan algoritma K-Means menggunakan data capaian Ujian Nasional Sekolah Menengah Pertama pada tahun 2018/2019 yang

diperoleh dari website resmi Pusat Penilaian Pendidikan dan Kebudayaan Kementerian Pendidikan dan Kebudayaan Republik Indonesia. Clustering dilakukan untuk mengelompokkan provinsi mana saja di Republik Indonesia yang memiliki nilai Ujian Nasional tingkat Sekolah Menengah Pertama yang rendah, sedang dan tinggi, sehingga nantinya dapat dijadikan bahan evaluasi terhadap hasil ujian di tiap-tiap provinsi.

Klastering (*clustering*) adalah salah satu teknik dalam data mining yang digunakan untuk mengelompokkan data ke dalam kelompok (klaster) berdasarkan kemiripan atau kedekatan tertentu tanpa perlu tahu terlebih dahulu kategorinya.

Berikut adalah beberapa fungsi utama dari klastering:

1. Segmentasi Pasar

Mengelompokkan pelanggan dengan karakteristik atau perilaku yang serupa untuk menargetkan kampanye pemasaran yang lebih efektif dan personalisasi layanan.

2. Analisis Pelanggan

Mengidentifikasi kelompok pelanggan yang memiliki kebiasaan pembelian serupa, sehingga perusahaan dapat memahami kebutuhan dan preferensi mereka dengan lebih baik.

3. Deteksi Anomali

Mengidentifikasi data yang tidak biasa atau mencurigakan dengan mencari data yang tidak cocok dengan klaster manapun. Ini berguna dalam mendeteksi penipuan, kegagalan mesin, atau masalah keamanan.

4. Pengelompokan Dokumen

Mengelompokkan dokumen atau artikel berdasarkan topik atau konten serupa untuk membantu dalam pengelolaan dan pencarian informasi.

5. Riset dan Penemuan Pengetahuan

Membantu peneliti menemukan pola atau struktur yang tersembunyi dalam data yang kompleks, seperti dalam bidang biologi, astronomi, atau ilmu sosial.

6. Segmentasi Citra

Mengelompokkan piksel dalam citra digital ke dalam bagian-bagian yang homogen untuk tujuan pengolahan citra, seperti dalam analisis medis atau pengenalan pola.

7. Pengelompokan Geografis

Mengelompokkan data geografis untuk analisis spasial, seperti mengidentifikasi area dengan karakteristik demografis atau ekonomi yang serupa.

8. Rekomendasi Produk

Mengelompokkan produk berdasarkan fitur atau ulasan yang serupa untuk menyediakan rekomendasi yang lebih tepat kepada pengguna.

9. Optimisasi Rantai Pasokan

Mengelompokkan lokasi gudang atau pusat distribusi berdasarkan pola permintaan untuk mengoptimalkan logistik dan distribusi.

## 10. Penyusunan Kurikulum

Mengelompokkan siswa berdasarkan performa atau minat mereka untuk merancang kurikulum yang lebih sesuai dengan kebutuhan masing-masing kelompok.

## 11. Analisis Jaringan Sosial

Mengidentifikasi komunitas atau kelompok dalam jaringan sosial berdasarkan pola interaksi dan hubungan antar individu.

Fungsi-fungsi tersebut menunjukkan bahwa klustering adalah alat yang sangat berguna untuk memahami struktur data yang kompleks dan membuat keputusan yang lebih informasional di berbagai bidang.

Klustering melibatkan berbagai metode untuk mengelompokkan data berdasarkan kemiripan. Berikut adalah beberapa metode yang umum digunakan dalam klustering:

### 1. K-Means Clustering

Metode yang paling populer dan sederhana di mana data dibagi ke dalam  $k$  jumlah kluster yang telah ditentukan sebelumnya. Setiap titik data ditetapkan ke kluster dengan centroid (titik tengah) terdekat. Centroid diupdate hingga konvergensi tercapai.

### 2. Hierarchical Clustering

Membuat hierarki kluster yang dapat divisualisasikan dalam bentuk dendrogram. Terdapat dua jenis utama:

a. Agglomerative (Bottom-Up)

Mulai dengan setiap titik data sebagai kluster individu dan kemudian secara iteratif menggabungkan kluster yang paling mirip hingga hanya tersisa satu kluster.

b. Divisive (Top-Down)

Mulai dengan satu kluster besar dan secara iteratif memecahnya menjadi kluster yang lebih kecil.

3. DBSCAN (Density-Based Spatial Clustering of Applications with Noise)

Mengidentifikasi kluster berdasarkan kepadatan data. Titik data yang berdekatan dengan cukup banyak titik lain (dalam radius tertentu) akan membentuk kluster. DBSCAN juga mampu mengidentifikasi titik data yang dianggap sebagai noise atau outlier.

4. Mean-Shift Clustering

Teknik non-parametrik yang tidak memerlukan jumlah kluster awal. Metode ini memindahkan setiap titik data ke arah mode distribusi data (puncak kepadatan) hingga konvergensi tercapai, sehingga membentuk kluster berdasarkan kepadatan data.

5. Gaussian Mixture Models (GMM)

Mengasumsikan data berasal dari kombinasi beberapa distribusi Gaussian. Metode ini menggunakan algoritma Expectation-Maximization (EM) untuk menemukan parameter dari distribusi Gaussian tersebut

dan mengelompokkan data berdasarkan probabilitas keanggotaan.

#### 6. Spectral Clustering

Menggunakan spektrum (eigenvalue) dari matriks afinitas (kesamaan) untuk mengurangi dimensi data sebelum menerapkan teknik klustering seperti K-Means. Metode ini berguna untuk data yang tidak dapat dipisahkan secara linier.

#### 7. Agglomerative Nesting (AGNES) dan Divisive Analysis (DIANA)

Agnes adalah pendekatan bottom-up untuk hierarchical clustering, sedangkan Diana adalah pendekatan top-down. Keduanya mengelompokkan data dengan cara membangun hierarki kluster.

#### 8. BIRCH (Balanced Iterative Reducing and Clustering using Hierarchies)

Metode ini dirancang untuk klustering data besar dengan cara membuat pohon clustering feature (CF) untuk mengompres data dan kemudian menerapkan teknik klustering lain pada hasil kompresi.

#### 9. Fuzzy C-Means Clustering

Memperluas K-Means dengan membiarkan setiap titik data memiliki derajat keanggotaan ke beberapa kluster. Setiap titik data dapat menjadi bagian dari beberapa kluster dengan probabilitas tertentu.

#### 10. Self-Organizing Maps (SOM)

Menggunakan jaringan saraf untuk memetakan data tinggi dimensi ke dalam grid dua dimensi, memungkinkan visualisasi dan klastering yang lebih mudah.

#### 11. Affinity Propagation

Menentukan kluster dengan mengirim pesan antara titik data hingga menemukan “exemplar” (contoh) yang paling representatif untuk setiap kluster. Tidak memerlukan jumlah kluster awal sebagai input.

Setiap metode memiliki kelebihan dan kekurangan tergantung pada jenis data, ukuran dataset, dan tujuan analisis. Pemilihan metode yang tepat sangat penting untuk mendapatkan hasil klastering yang akurat dan bermakna.

### E. Algoritma K-Means

K-Means pertama kali dipublikasikan oleh Stuart Lloyd pada tahun 1984 dan merupakan algoritma clustering yang banyak digunakan. K-Means bekerja dengan mensegmentasi objek yang ada ke dalam kelompok atau yang disebut dengan segmen sehingga objek yang berada dalam masing-masing kelompok lebih serupa satu sama lain dibandingkan dengan objek dalam kelompok yang berbeda.

K-Means adalah algoritma klastering yang membagi data menjadi kelompok berdasarkan pusat kluster. Dalam konteks pengelompokan siswa, atribut seperti nilai akademis, tingkat kehadiran, dan partisipasi dalam kegiatan ekstrakurikuler dapat digunakan untuk membentuk kelompok. Clustering adalah algoritma yang bekerja dengan cara menerapkan algoritma yang ada di dalamnya, pada dasarnya K-Means

Clustering merupakan algoritma yang sering kali digunakan dalam proses mengimplementasikan clustering.

Adapun di bawah ini merupakan uraian dari algoritma menggunakan metode K-Means clustering.

1. Tentukan sebanyak kelompok yang kita inginkan.
2. Inisiasi centroid untuk setiap kelompok. Centroid ibarat seperti "leader group", yang mempresentasikan kelompok.
3. Hitung kedekatan antara suatu data dengan centroid, selanjutnya masukkan data tersebut pada kelompok centroid yang memiliki sifat terdekat dengan dirinya.
4. Pilih centroid baru untuk masing-masing kelompok dari anggota kelompok itu sendiri dan lakukan perhitungan kedekatan antara suatu data dengan centroid.
5. Ulangi langkah-langkah sebelumnya hingga tidak ada perubahan anggota untuk semua kelompok.

Clustering atau klasterisasi merupakan metode pengelompokan data. Clustering dapat diartikan sebagai sebuah proses untuk mengelompokan data ke dalam beberapa cluster sehingga data dalam satu cluster tersebut memiliki tingkat kemiripan yang maksimum dan data antar cluster memiliki kemiripan yang minimum.

Langkah-langkah melakukan clustering dengan metode K-Means adalah sebagai berikut:

1. Pilih jumlah cluster K.
2. Inisialisasi K pusat cluster ini bisa dilakukan dengan berbagai cara. Namun yang paling sering dilakukan adalah dengan cara random (acak). Pusat cluster diberi nilai awal dengan angka random.

3. Alokasikan semua data atau objek ke cluster terdekat. Kedekatan dua data atau objek ditentukan berdasarkan jarak kedua data atau objek tersebut. Demikian juga kedekatan suatu data atau objek ke cluster tertentu ditentukan jarak antara data atau objek dengan pusat cluster. Dalam tahap ini perlu dihitung jarak tiap data atau objek ke tiap pusat cluster. Jarak paling antara satu data atau objek dengan satu cluster tertentu akan menentukan suatu data atau objek masuk dalam cluster mana. Untuk menghitung jarak data ke setiap titik pusat cluster dapat menggunakan teori jarak Euclidean yang dirumuskan sebagai berikut:

$$D_{i,j} = \sqrt{(x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2 + \dots + (x_{ki} - x_{kj})^2}$$

Keterangan:

$D(i,j)$  = jarak data ke  $i$  ke pusat *cluster*  $j$

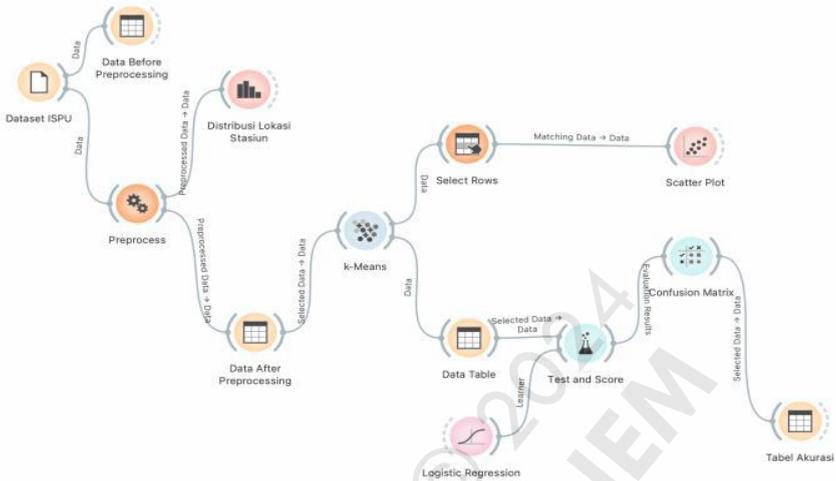
$X(k,i)$  = data ke  $i$  pada atribut data ke  $k$

$X(k,j)$  = titik pusat ke  $j$  pada atribut ke  $k$

4. Hitung kembali pusat cluster dengan keanggotaan cluster yang sekarang. Pusat cluster adalah rata-rata semua dari data atau objek dalam cluster tertentu. Jika dikehendaki bisa juga menggunakan median dari cluster tersebut. Jadi rata-rata (median) bukan satu-satunya ukuran yang bisa dipakai.
5. Tugaskan lagi setiap objek memakai pusat cluster yang baru. Jika pusat cluster tidak berubah lagi maka proses clustering selesai. Atau kembali ke langkah No. 3 sampai pusat cluster tidak berubah lagi.

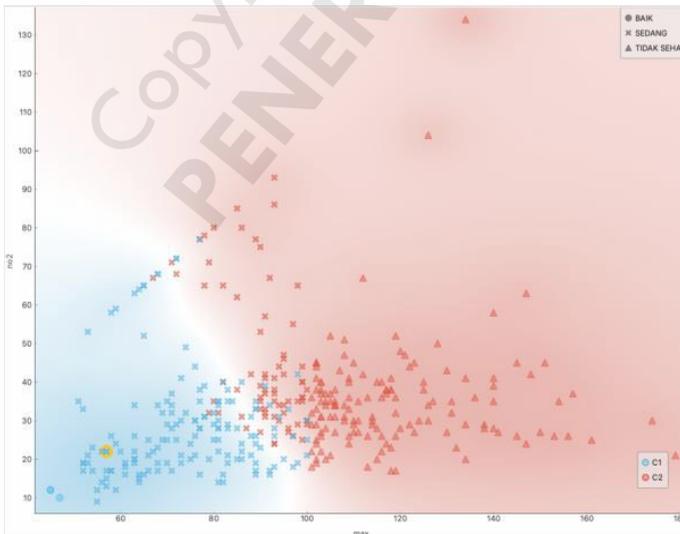
Proses pengujian dilakukan tanpa melakukan perubahan pada aplikasi Orange. Adapun pengujian yang dilakukan yaitu

dengan menggunakan algoritma *k-means*. Berikut proses pengujiannya dilakukan sebagai berikut:



**Gambar 2.1** Pemodelan Clustering K-means

Dengan menggunakan pemodelan *k-means* clustering seperti gambar di atas, dengan inisialisasi jumlah cluster.



**Gambar 2.2** Hasil Scatter Plot Cluster K-Means

## F. Aplikasi Orange

Orange adalah aplikasi perangkat lunak open-source yang dirancang untuk analisis data dan visualisasi. Ia menyediakan antarmuka grafis yang intuitif untuk pengguna yang tidak ahli di bidang pemrograman atau statistik. Orange adalah sebuah platform atau aplikasi yang memungkinkan Anda untuk melakukan data mining meskipun Anda tidak punya pengalaman sama sekali.

Orange menyediakan alat visualisasi yang kuat untuk membantu pengguna menjelajahi dan memahami struktur data mereka. Ini termasuk berbagai jenis plot, grafik, dan alat visualisasi lainnya. Orange menyediakan modul-modul untuk membangun model dan melakukan pembelajaran mesin. Ini termasuk berbagai algoritma klasifikasi, regresi, dan klastering yang dapat digunakan untuk menganalisis data dan membuat prediksi.

Orange mendukung analisis teks dan penambangan teks. Ini memungkinkan pengguna untuk melakukan ekstraksi fitur dari teks, analisis sentimen, dan pemrosesan teks lainnya. Meskipun Orange dirancang untuk digunakan tanpa pengetahuan pemrograman, ia juga memberikan dukungan untuk integrasi dengan bahasa pemrograman seperti Python, memungkinkan pengguna yang terampil dalam pemrograman untuk memperluas fungsionalitasnya.

Orange mendukung ekstensi dan pengaya (add-ons) yang memungkinkan pengguna untuk menambahkan fungsionalitas tambahan sesuai kebutuhan mereka. Orange data mining adalah paket perangkat lunak pemrograman “visual” berbasis komponen untuk visualisasi data machine

learning, data mining, dan analisis data. Pemrograman visual diimplementasikan melalui antarmuka di mana alur kerja dibuat dengan menghubungkan widget yang telah ditentukan sebelumnya atau yang dirancang pengguna. Orange menyediakan lingkungan serbaguna untuk pengembang, peneliti, dan data praktisi pertambangan. Berkat Python, bahasa skrip generasi baru dan lingkungan pemrograman, skrip penambangan data yang mungkin sederhana tetapi hasilnya sangat berkualitas.

Adapun menu-menu yang digunakan pada proses data mining dengan metode algoritma K-means pada aplikasi Orange adalah sebagai berikut:

1. Tampilan Awal Orange



**Gambar 2.3** Tampilan Awal Orange



## G. RapidMiner

RapidMiner adalah platform analitik data dan pembelajaran mesin yang komprehensif yang memungkinkan pengguna untuk merancang, melaksanakan, dan mengotomatiskan proses analitik data tanpa memerlukan keahlian pemrograman tingkat lanjut. Berikut adalah beberapa fitur utama dan kegunaan RapidMiner:

Fitur Utama RapidMiner:

### 1. User-Friendly Interface

Antarmuka drag-and-drop yang intuitif memungkinkan pengguna untuk membuat alur kerja analitik dengan mudah, tanpa perlu menulis kode.

### 2. Komprehensif

Menyediakan berbagai macam alat dan teknik untuk pra-pemrosesan data, analitik prediktif, pembelajaran mesin, dan pemodelan.

### 3. Integrasi Data

Mendukung berbagai format data dan sumber, termasuk basis data SQL, lembar kerja Excel, file teks, dan layanan web.

### 4. Pra-Pemrosesan Data

Berbagai alat untuk pembersihan data, transformasi, pengisian nilai yang hilang, penghapusan duplikasi, dan normalisasi.

5. Algoritma Pembelajaran Mesin

Menyediakan banyak algoritma pembelajaran mesin populer untuk klasifikasi, regresi, klastering, dan deteksi anomali.

6. Visualisasi

Kemampuan untuk membuat visualisasi data yang kaya dan informatif untuk analisis eksploratif dan presentasi hasil.

7. Model Evaluation

Alat untuk mengevaluasi model dengan berbagai metrik kinerja, validasi silang, dan pemilihan model terbaik.

8. Automasi

Fitur untuk mengotomatisasi alur kerja analitik, memungkinkan pengguna untuk menjadwalkan tugas dan proses berulang.

9. Ekstensibilitas

Mendukung integrasi dengan bahasa pemrograman seperti Python dan R untuk analitik lanjutan dan kustomisasi.

10. Deployment

Alat untuk menerapkan model pembelajaran mesin ke lingkungan produksi dan integrasi dengan aplikasi bisnis.

RapidMiner adalah alat yang sangat kuat dan serbaguna untuk data mining. Berikut adalah beberapa kegunaan utama RapidMiner dalam berbagai aspek *data mining*:

## 1. Pengumpulan Data dan Integrasi

### a. Import data

RapidMiner dapat mengimpor data dari berbagai sumber, seperti basis data SQL, lembar kerja Excel, file CSV, dan layanan web.

### b. Konektivitas

Menyediakan konektivitas langsung ke berbagai sumber data seperti Hadoop, NoSQL, dan cloud storage, memungkinkan integrasi data yang lancar.

## 2. Pra-Pemrosesan Data

### a. Pembersihan data

Menawarkan berbagai alat untuk membersihkan data, termasuk pengisian nilai yang hilang, penghapusan duplikasi, dan koreksi kesalahan data.

### b. Transformasi data

Mendukung transformasi data, seperti normalisasi, standarisasi, dan encoding data kategorikal.

### c. Sampling dan reduksi dimensi

Memungkinkan sampling data dan teknik reduksi dimensi seperti PCA (Principal Component Analysis) untuk mengurangi kompleksitas data.

3. Analisis Data Eksploratif

a. Visualisasi

Menyediakan alat visualisasi yang kaya untuk analisis eksploratif, termasuk grafik, diagram, dan plot interaktif.

b. Statistik deskriptif

Menyediakan ringkasan statistik dasar dan analisis distribusi data untuk memahami karakteristik dataset.

4. Pembelajaran Mesin dan Model Prediktif

a. Algoritma pembelajaran mesin

Mendukung berbagai algoritma pembelajaran mesin untuk klasifikasi, regresi, klustering, dan deteksi anomali. Contoh algoritma termasuk K-Means, Random Forest, Support Vector Machines (SVM), dan Neural Networks.

b. Automated Machine Learning (AutoML)

Memungkinkan penggunaan AutoML untuk secara otomatis mencari model terbaik berdasarkan metrik yang ditentukan.

5. Klustering (Clustering)

a. K-Means

Menyediakan implementasi K-Means yang mudah digunakan untuk mengelompokkan data berdasarkan kemiripan.

b. DBSCAN

Mendukung DBSCAN untuk klustering berbasis kepadatan, yang efektif dalam mendeteksi kluster dengan bentuk arbitrer dan menangani noise.

c. Hierarchical clustering

Menawarkan metode klustering hirarkis untuk membangun dendrogram dan mengidentifikasi struktur kluster dalam data.

6. Analisis Asosiasi

a. Apriori algorithm

Mendukung algoritma Apriori untuk menemukan aturan asosiasi dalam data, seperti dalam analisis keranjang pasar.

b. Frequent Pattern Growth (FP-Growth)

Menyediakan algoritma FP-Growth untuk menemukan pola yang sering muncul dalam dataset besar.

7. Evaluasi Model

a. Cross-Validation

Menawarkan teknik validasi silang untuk mengevaluasi kinerja model dan mencegah overfitting.

b. Performance metrics

Menyediakan berbagai metrik evaluasi seperti akurasi, presisi, recall, F1-score, dan AUC-ROC.

8. Deployment dan Integrasi

a. Model deployment

Memungkinkan implementasi model prediktif ke dalam lingkungan produksi untuk digunakan dalam aplikasi bisnis.

b. Web services

Mendukung pembuatan layanan web untuk mengakses model prediktif dari aplikasi eksternal.

9. Automasi dan Orkestrasi

a. Process automation

Memungkinkan otomatisasi alur kerja analitik data, termasuk penjadwalan tugas dan eksekusi batch.

b. Integration with workflows

Mendukung integrasi dengan alur kerja yang lebih luas menggunakan API dan skrip.

10. Kolaborasi dan Versi Kontrol

a. Team collaboration

Mendukung kolaborasi tim dengan fitur seperti versioning, komentar, dan berbagi alur kerja.

b. Version control

Menyediakan kontrol versi untuk melacak perubahan dan pengembangan model secara sistematis.



## Bab 3

# **DATA MINING PENGELOMPOKAN SISWA BERPRESTASI MENGGUNAKAN METODE CLUSTERING**

Berikut ini adalah langkah-langkah dalam pengolahan data, yaitu:

1. Pengelompokan Siswa Berprestasi dengan K-Means

Data siswa ini yang diperoleh ini akan dikelompokkan untuk mencari siswa yang berprestasi dengan kriteria nilai Tugas (A), nilai UH (B), nilai UTS (C) dan nilai UAS (D). Dari semua kriteria di atas, maka akan diperoleh Nilai Rata-Rata (E) siswa. Nilai dari masing-masing kriteria di atas diperoleh dengan melakukan wawancara dan memperoleh data langsung dari pihak sekolah. Terdapat 20 orang siswa yang mengajukan beasiswa prestasi, namun tidak semuanya termasuk siswa yang berprestasi. Maka, diperlukan cara untuk mengelompokkan siswa tersebut, mana yang termasuk

siswa yang berprestasi dan mana yang tidak. Data siswa yang digunakan dapat dilihat pada Tabel 3.1.

Algoritma *K-Means* dimulai dengan menentukan nilai  $k$  yaitu jumlah kluster yang ingin dibentuk. Pada kajian ini akan dibentuk 2 buah kluster yaitu Siswa Berprestasi (C1) dan Siswa Kurang Berprestasi (C2). Langkah berikutnya adalah menentukan titik pusat untuk masing-masing kluster (*centroid*) secara *random*. Untuk C1 diberikan nilai 79 dan untuk C2 diberikan nilai 77. Tahap selanjutnya adalah menghitung jarak masing-masing nilai Rata-rata (E) ke C1 dan C2.

**Tabel 3.1**  
Data Siswa yang Mengajukan Beasiswa

No	NIS	Nilai Tugas	Nilai UH	Nilai UTS	Nilai UAS	Rata-rata
1	1711	82	91	70	85	82,000
2	1715	89	68	89	60	76,500
3	1726	90	85	95	98	92,000
4	1747	87	80	87	95	87,250
5	1751	65	75	65	65	67,500
6	1760	70	75	61	54	65,000
7	1767	77	76	88	85	81,500
8	1768	84	82	65	55	71,500
9	1769	86	87	95	93	90,250
10	1771	78	78	71	70	74,250
11	1773	84	88	80	85	84,250
12	1774	78	77	65	85	76,250
13	1775	77	87	80	78	80,500
14	1780	96	98	78	76	87,000
15	1791	84	81	76	98	84,750
16	1802	82	98	70	95	86,250
17	1804	80	88	63	55	71,500
18	1805	83	71	88	65	76,750
19	1809	87	68	65	84	76,000

**Tabel 3.2**  
Jarak Rata-rata (E) ke C1 dan C2

Jarak Rata-Rata (E) ke C1 dan C2 C1	C2		C1	C2	
E1	3,000	4,000	E11	5,250	6,250
E2	2,500	1,500	E12	2,750	1,750
E3	13,000	14,000	E13	1,500	2,500
E4	8,250	9,250	E14	8,000	9,000
E5	11,500	10,500	E15	5,750	6,750
E6	14,000	13,000	E16	7,250	8,250
E7	2,500	3,500	E17	7,500	6,500
E8	7,500	6,500	E18	2,250	1,250
E9	11,250	12,250	E19	3,000	2,000
E10	4,750	3,750	E20	2,500	1,

Tahapan selanjutnya adalah mengelompokkan nilai Rata-rata (E) berdasarkan jarak terpendek ke C1 dan C2, sehingga dapat diperoleh kelompok nilai Rata-rata (E) sebagai berikut:

Anggota C1:

E1, E3, E4, E7, E9, E11, E13, E14, E15, E16

Anggota C2:

E2, E5, E6, E8, E10, E12, E17, E18, E19, E20

Kemudian tentukan nilai *centroid* baru (C1baru dan C2baru) seperti berikut ini:

$$C1baru = \frac{3,000 + 13,000 + 8,250 + 2,500 + 11,250 + 5,250 + 1,500 + 8,000 + 5,750 + 7,250}{10}$$

$$C1baru = 85,575$$

$$C2baru = \frac{1,500 + 10,500 + 13,000 + 6,500 + 3,750 + 1,750 + 6,500 + 1,250 + 2,000 + 1,500}{10}$$

$$C2baru = 73,175$$

Setelah memperoleh nilai centroid baru, bandingkan nilai centroid baru dengan centroid lama. Centroid baru tidak sama dengan centroid lama, maka dilakukan perhitungan jarak kembali dengan menggunakan centroid baru. Hasil perhitungan dapat dilihat pada Tabel 3.3.

**Tabel 3.3**  
Jarak Rata-Rata (E) ke C1baru dan C2baru

C1	C2		C1	C2	
E1	3,575	8,825	E11	1,325	11,075
E2	9,075	3,325	E12	9,325	3,075
E3	6,425	18,825	E13	5,075	7,325
E4	1,675	14,075	E14	1,425	13,825
E5	18,075	5,675	E15	0,825	11,575
E6	20,575	8,175	E16	0,675	13,075
E7	4,075	8,325	E17	14,075	1,675
E8	14,075	1,675	E18	8,825	3,575
E9	4,675	17,075	E19	9,575	2,825
E10	11,325	1,075	E20	9,075	3,325

Kelompokkan kembali nilai Rata-rata (E) berdasarkan jarak terpendek ke C1baru dan C2baru, sehingga dapat diperoleh kelompok nilai Rata-rata (E) sebagai berikut:

Anggota C1:

E1, E3, E4, E7, E9, E11, E13, E14, E15, E16

Anggota C2:

E2, E5, E6, E8, E10, E12, E17, E18, E19, E20

Karena tidak terjadi perubahan anggota kluster, maka proses *K-Means* dihentikan. Dari hasil perhitungan *K-Means*, terdapat 10 orang siswa yang masuk kategori siswa berprestasi seperti pada Tabel 3.4.

**Tabel 3.4**  
Kelompok Siswa Berprestasi

No	NIS	Nilai Tugas	Nilai UH	Nilai UTS	Nilai UAS	Rata-rata
		(A)	(B)	(C)	(D)	(E)
1	1711	82	91	70	85	82,000
2	1726	90	85	95	98	92,000
3	1747	87	80	87	95	87,250
4	1767	77	76	88	85	81,500
5	1769	86	87	95	93	90,250
6	1773	84	88	80	85	84,250
7	1775	77	87	80	78	80,500
8	1780	96	98	78	76	87,000
9	1791	84	81	76	98	84,750
10	1802	82	98	70	95	86,250

## 2. Aliran Sistem Informasi yang Sudah Ada

Aliran sistem informasi yang sudah ada dalam kajian ini Pengelompokan Siswa Berprestasi dengan cara:

### a. Pengumpulan data

Mulai dengan mengumpulkan data siswa yang relevan, seperti nilai akademis, kehadiran, partisipasi dalam kegiatan ekstrakurikuler, dan atribut lain yang dapat mencerminkan prestasi siswa.

### b. Precocensing data

Lakukan preprocessing data untuk membersihkan data dari nilai yang hilang atau outlier. Proses ini juga mencakup normalisasi data jika diperlukan agar semua atribut memiliki skala yang seragam.

c. Pemilihan atribut

Pilih atribut-atribut yang paling relevan untuk analisis clustering. Ini dapat melibatkan pemilihan fitur yang dianggap berkontribusi besar terhadap prestasi siswa.

d. Transformasi data

Lakukan transformasi data jika diperlukan, seperti mengonversi data kategorikal menjadi bentuk yang sesuai untuk pemrosesan clustering.

e. Pemilihan metode clustering

Pilih metode clustering yang sesuai untuk data dan tujuan analisis. Contoh metode clustering meliputi K-Means, Hierarchical Clustering, atau DBSCAN.

f. Pelaksanaan algoritma clustering

Terapkan algoritma clustering pada data siswa menggunakan metode yang telah dipilih sebelumnya. Proses ini akan menghasilkan kelompok (cluster) siswa berdasarkan kesamaan atribut.

g. Evaluasi dan pemilihan cluster optimal

Evaluasi hasil clustering untuk menentukan jumlah cluster yang optimal. Gunakan metrik evaluasi seperti indeks siluet atau indeks Davies-Bouldin.

h. Interpretasi hasil

Interpretasikan hasil clustering untuk memahami pola dan hubungan di antara kelompok siswa.

Identifikasi ciri-ciri khas dari setiap cluster, termasuk cluster yang mewakili siswa berprestasi tinggi.

i. Visualisasi hasil

Buat visualisasi grafis dari hasil clustering untuk memudahkan pemahaman dan komunikasi. Diagram atau grafik dapat membantu pemangku kepentingan melihat pola-pola secara intuitif.

j. Pengembangan model prediktif

Jika tujuan melibatkan prediksi, pertimbangkan untuk mengembangkan model prediktif menggunakan teknik pembelajaran mesin untuk memprediksi prestasi siswa di masa depan.

k. Implementasi sistem

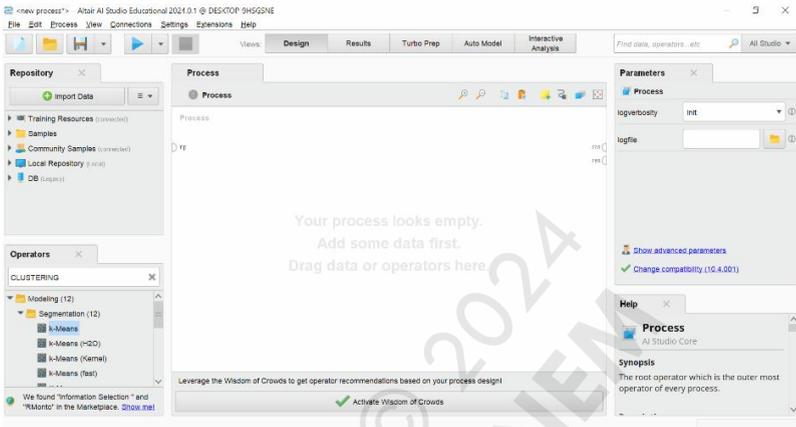
Implementasikan sistem secara keseluruhan, termasuk antarmuka pengguna jika diperlukan. Pastikan bahwa hasil clustering dapat diakses dan dimanfaatkan dengan mudah oleh pemangku kepentingan.

l. Evaluasi dan pemeliharaan

Lakukan evaluasi berkala terhadap sistem untuk memastikan keberlanjutan dan relevansi hasil clustering. Lakukan pemeliharaan dan perbaikan jika diperlukan berdasarkan umpan balik dan perkembangan data.

Berikut ini adalah implementasi dari pengelompokan siswa berprestasi menggunakan metode clustering

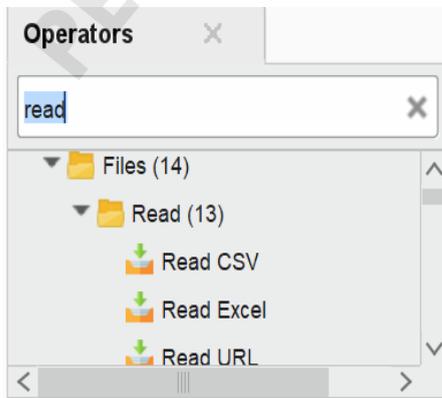
### 1. Tampilan Awal Aplikasi



Gambar 3.1 Tampilan Awal Aplikasi RapidMiner

Gambar di atas merupakan gambar tampilan awal Aplikasi RapidMiner. Aplikasi tersebut yang nantinya akan digunakan untuk melakukan clustering data untuk mengelompokkan data siswa dan siswi yang berprestasi.

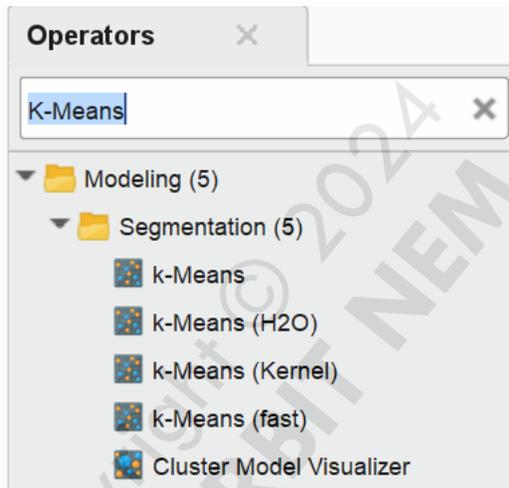
### 2. Widget Data



Gambar 3.2 Widget Data

Pada gambar di atas merupakan widget yang digunakan untuk menambahkan data ataupun file yang akan digunakan pada kajian ini. Penulis menggunakan file dengan format .xlsx, maka untuk widgetnya penulis menggunakan Read Excel.

### 3. Metode K-Means



**Gambar 3.3** Metode K-Means

Pada gambar di atas merupakan metode yang digunakan pada kajian ini yaitu metode K-means.

Berikut ini adalah teknik-teknik dalam pengujian sistem, antara lain yaitu:

#### 1. Analisis Data

Pada tahapan analisis data merupakan tahapan yang dilakukan untuk mengumpulkan data yang akan digunakan pada kajian ini. Dengan adanya data sampel, maka kajian dapat dilakukan. Untuk data sampel yang akan digunakan dapat dilihat pada tabel di bawah ini.

**Tabel 3.5**  
Data yang Digunakan untuk Dikelompokkan  
Menggunakan Metode K-Means Clustering

Nama Lengkap	Nilai Pengetahuan	Nilai Keterampilan	Nilai Sikap
Adrian Sidabalok Sinurat	87	87	87
Aldo Sihombing Simatupang	95	99	95
Alexander Panjaitan Sihombing	97	98	89
Andi Hutagalung Simbolon	95	99	95
Andre Marbun Sinaga	95	99	95
Belinda Situmorang Silalahi	97	98	89
Bella Napitupulu Siagian	95	99	95
Bella Situmorang Siahaan	95	99	95
Berlian Sibarani Sigalingging	87	87	87
Betty Silalahi Simatupang	95	99	95
Carlo Sinaga Silitonga	95	99	95
Carlo Sitanggung Sihite	87	87	87
Carlos Simatupang Sianipar	97	98	89
Christo Sitorus Sihotang	95	99	95
Cindy Saragih Sitorus	97	98	89
Darma Sihombing Simanjuntak	87	82	97
Debora Hutapea Sihotang	95	99	95
Dessy Purba Sitompul	94	94	94
Devi Hutapea Simbolon	95	99	95
Diana Simanjuntak Siregar	97	98	89
Edo Manurung Sitorus	87	82	97
Eko Sihombing Sitorus	95	99	95
Elisa Nababan Simbolon	97	98	89
Endah Sinaga Silalahi	92	87	90
Enrico Gamaliel Silalahi	97	98	89
Evelyn Gultom Sihombing	94	94	94
Fadel Simbolon Sidabutar	95	99	95
Fanny Silalahi Sihombing	87	82	97
Ferdi Tamba Sitinjak	97	98	89
Ferianto Simbolon Sirait	94	94	94
Fina Sitompul Sihotang	95	99	95
Gilbert Sitorus Sihotang	87	82	97
Giska Hutagalung Sihotang	92	92	92
Gita Sinaga Sihotang	97	98	89
Grace Hutabarat Sitorus	94	94	94
Hans Parulian Simatupang	94	94	94
Hendra Pardede Simatupang	98	98	92
Hendrik Marpaung Sitorus	97	98	89
Heni Sihombing Simbolon	87	82	97
Ida Butarbutar Sihombing	97	87	87
Indra Silalahi Sianipar	87	82	97

Irene Hutagalung Sianipar	97	98	89
Irma Hutagalung Siahaan	94	94	94
James Silaban Siagian	97	98	89
Johan Hutapea Sianipar	94	94	94
Johan Situmorang Silalahi	94	94	94
Joni Simbolon Sitorus	87	87	87
Karin Simanungkalit Sinambela	95	99	95
Kevin Sinurat Simbolon	87	87	87
Kristin Hutabarat Sihombing	87	87	87
Kristin Marbun Sinaga	94	94	94
Leo Panggabean Sidabutar	95	99	95
Lina Sihombing Silaban	96	96	96
Linda Pardede Sihotang	94	94	94
Lisa Sinaga Simatupang	87	87	87
Lunggu Nainggolan	87	82	97
Maisie Amandine Saragih	92	87	90
Maria Nainggolan Silitonga	95	99	95
Mario Sitompul Sihotang	97	98	89
Mariska Sidabalok Sihotang	87	87	87
Miko Pasaribu Sihite	94	94	94
Mutiara Friska Hutagalung	95	99	95
Muzy Rosi Saragih	92	92	92
Nanda Sinaga Sitorus	97	98	89
Nando Sitorus Simatupang	87	87	87
Natalia Panjaitan Siringoringo	95	99	95
Niko Siahaan Sitohang	95	99	95
Noah Ben Joachim Siregar	98	98	92
Olin Hutabarat Sihombing	97	98	89
Olivia Hutapea Siahaan	95	99	95
Olivia Sinaga Sihombing	94	94	94
Oscar Hutapea Silaban	95	99	95
Paulus Saragih Simorangkir	95	99	95
Peter Hutagalung Sihotang	94	94	94
Petra Sitompul Sinaga	95	99	95
Poppy Sitompul Silalahi	97	98	89
Regina Nababan Siahaan	95	99	95
Riko Simanjuntak Sitorus	94	94	94
Riko Sinaga Simatupang	97	98	89
Rina Lumban Gaol Siringoringo	87	82	97
Samuel Silalahi Sihotang	95	99	95
Sari Sihombing Sitorus	94	94	94
Sinta Sitanggung Sihotang	97	98	89
Steven Hasibuan Situmeang	87	82	97
Tania Pasaribu Simbolon	94	94	94
Thanta Primasya Ginting	97	87	87
Theresia Samosir Sianturi	87	82	97
Tommy Pasaribu Sidabutar	97	98	89

Trentson Martua Wiliard Siagian	94	94	94
Triana Tampubolon Sinaga	95	99	95
Uli Hutapea Sitorus	97	98	89
Uli Lumbantoruan Simanullang	87	82	97
Untung Hutapea Sihotang	94	94	94
Ursula Manurung Simbolon	95	99	95
Vera Sinaga Sihotang	97	98	89
Victor Pangaribuan Simanjuntak	87	82	97
Vino Hutapea Sitorus	95	99	95
Vivi Sihombing Simatupang	94	94	94
Willy Sibarani Sinurat	97	98	89
Willy Sitorus Sihombing	94	94	94
Wina Sibarani Sihombing	95	99	95
Yana Hutagalung Simbolon	95	99	95
Yehuda Immanuel Saragi	87	87	87
Yessica Chindy Claudya Napitupulu	96	96	96
Yoseph Simatupang Sihotang	95	99	95
Yuli Butarbutar Siagian	87	87	87
Yulia Simanjuntak Sianipar	94	94	94
Zefanya Napitupulu Simatupang	87	87	87
Zefanya Sidabalok Sihotang	94	94	94
Zenni Simanjuntak Sihombing	95	99	95

Pada tabel di atas merupakan data yang digunakan untuk dikelompokkan menggunakan metode k-Means Clustering. Data yang digunakan sebanyak 110 data.

## 2. Model Perancangan



**Gambar 3.4** Model Clustering

Pada gambar di atas merupakan model yang digunakan untuk melakukan clustering data. Untuk teknik pengujiannya menggunakan model pada gambar di atas. Jadi dengan adanya model di atas, maka proses clustering data dapat dilakukan dengan baik.

Berikut ini adalah hasil pengujian menggunakan RapidMiner.

### 1. Hasil Clustering

Untuk hasil clustering yang digunakan oleh penulis yaitu C0, C1, dan C2. Untuk hasilnya dapat dilihat pada tabel di bawah ini.

**Tabel 3.6**  
Hasil Clustering yang Dilakukan dengan Menggunakan Metode K-Means

Nama Lengkap	Nilai Pengetahuan	Nilai Keterampilan	Nilai Sikap	Cluster
Adrian Sidabalok Sinurat	87	87	87	C1
Aldo Sihombing Simatupang	95	99	95	C0
Alexander Panjaitan Sihombing	97	98	89	C0
Andi Hutagalung Simbolon	95	99	95	C0
Andre Marbun Sinaga	95	99	95	C0
Belinda Situmorang Silalahi	97	98	89	C0
Bella Napitupulu Siagian	95	99	95	C0
Bella Situmorang Siahaan	95	99	95	C0
Berlian Sibarani Sigalingging	87	87	87	C1
Betty Silalahi Simatupang	95	99	95	C0
Carlo Sinaga Silitonga	95	99	95	C0
Carlo Sitanggang Sihite	87	87	87	C1
Carlos Simatupang Sianipar	97	98	89	C0
Christo Sitorus Sihotang	95	99	95	C0
Cindy Saragih Sitorus	97	98	89	C0
Darma Sihombing Simanjuntak	87	82	97	C2
Debora Hutapea Sihotang	95	99	95	C0
Dessy Purba Sitompul	94	94	94	C0
Devi Hutapea Simbolon	95	99	95	C0
Diana Simanjuntak Siregar	97	98	89	C0
Edo Manurung Sitorus	87	82	97	C2
Eko Sihombing Sitorus	95	99	95	C0
Elisa Nababan Simbolon	97	98	89	C0
Endah Sinaga Silalahi	92	87	90	C1
Enrico Gamaliel Silalahi	97	98	89	C0
Evelyn Gultom Sihombing	94	94	94	C0
Fadel Simbolon Sidabutar	95	99	95	C0
Fanny Silalahi Sihombing	87	82	97	C2
Ferdi Tamba Sitinjak	97	98	89	C0

Ferianto Simbolon Sirait	94	94	94	C0
Fina Sitompul Sihotang	95	99	95	C0
Gilbert Sitorus Sihotang	87	82	97	C2
Giska Hutagalung Sihotang	92	92	92	C0
Gita Sinaga Sihotang	97	98	89	C0
Grace Hutabarat Sitorus	94	94	94	C0
Hans Parulian Simatupang	94	94	94	C0
Hendra Pardede Simatupang	98	98	92	C0
Hendrik Marpaung Sitorus	97	98	89	C0
Heni Sihombing Simbolon	87	82	97	C2
Ida Butarbutar Sihombing	97	87	87	C1
Indra Silalahi Sianipar	87	82	97	C2
Irene Hutagalung Sianipar	97	98	89	C0
Irma Hutagalung Siahaan	94	94	94	C0
James Silaban Siagian	97	98	89	C0
Johan Hutapea Sianipar	94	94	94	C0
Johan Situmorang Silalahi	94	94	94	C0
Joni Simbolon Sitorus	87	87	87	C1
Karin Simanungkalit Sinambela	95	99	95	C0
Kevin Sinurat Simbolon	87	87	87	C1
Kristin Hutabarat Sihombing	87	87	87	C1
Kristin Marbun Sinaga	94	94	94	C0
Leo Panggabean Sidabutar	95	99	95	C0
Lina Sihombing Silaban	96	96	96	C0
Linda Pardede Sihotang	94	94	94	C0
Lisa Sinaga Simatupang	87	87	87	C1
Lunggu Nainggolan	87	82	97	C2
Maisie Amandine Saragih	92	87	90	C1
Maria Nainggolan Silitonga	95	99	95	C0
Mario Sitompul Sihotang	97	98	89	C0
Mariska Sidabalok Sihotang	87	87	87	C1
Miko Pasaribu Sihite	94	94	94	C0
Mutiara Friska Hutagalung	95	99	95	C0
Muzy Rosi Saragih	92	92	92	C0
Nanda Sinaga Sitorus	97	98	89	C0
Nando Sitorus Simatupang	87	87	87	C1
Natalia Panjaitan Siringoringo	95	99	95	C0
Niko Siahaan Sitohang	95	99	95	C0
Noah Ben Joachim Siregar	98	98	92	C0
Olin Hutabarat Sihombing	97	98	89	C0
Olivia Hutapea Siahaan	95	99	95	C0
Olivia Sinaga Sihombing	94	94	94	C0
Oscar Hutapea Silaban	95	99	95	C0
Paulus Saragih Simorangkir	95	99	95	C0
Peter Hutagalung Sihotang	94	94	94	C0
Petra Sitompul Sinaga	95	99	95	C0
Poppy Sitompul Silalahi	97	98	89	C0

Regina Nababan Siahaan	95	99	95	C0
Riko Simanjuntak Sitorus	94	94	94	C0
Riko Sinaga Simatupang	97	98	89	C0
Rina Lumban Gaol Siringoringo	87	82	97	C2
Samuel Silalahi Sihotang	95	99	95	C0
Sari Sihombing Sitorus	94	94	94	C0
Sinta Sitanggung Sihotang	97	98	89	C0
Steven Hasibuan Situmeang	87	82	97	C2
Tania Pasaribu Simbolon	94	94	94	C0
Thanta Primasya Ginting	97	87	87	C1
Theresia Samosir Sianturi	87	82	97	C2
Tommy Pasaribu Sidabutar	97	98	89	C0
Trentson Martua Wiliard Siagian	94	94	94	C0
Triana Tampubolon Sinaga	95	99	95	C0
Uli Hutapea Sitorus	97	98	89	C0
Uli Lumbantoruan Simanullang	87	82	97	C2
Untung Hutapea Sihotang	94	94	94	C0
Ursula Manurung Simbolon	95	99	95	C0
Vera Sinaga Sihotang	97	98	89	C0
Victor Pangaribuan Simanjuntak	87	82	97	C2
Vino Hutapea Sitorus	95	99	95	C0
Vivi Sihombing Simatupang	94	94	94	C0
Willy Sibarani Sinurat	97	98	89	C0
Willy Sitorus Sihombing	94	94	94	C0
Wina Sibarani Sihombing	95	99	95	C0
Yana Hutagalung Simbolon	95	99	95	C0
Yehuda Immanuel Saragi	87	87	87	C1
Yessica Chindy Claudya Napitupulu	96	96	96	C0
Yoseph Simatupang Sihotang	95	99	95	C0
Yuli Butarbutar Siagian	87	87	87	C1
Yulia Simanjuntak Sianipar	94	94	94	C0
Zefanya Napitupulu Simatupang	87	87	87	C1
Zefanya Sidabalok Sihotang	94	94	94	C0
Zenni Simanjuntak Sihombing	95	99	95	C0

Pada tabel di atas merupakan hasil clustering yang dilakukan dengan menggunakan metode K-Means. Hasil clustering yang dilakukan dengan menggunakan metode K-Means menunjukkan pengelompokan siswa dan siswi berdasarkan prestasi mereka. Dari total 110 data yang dianalisis, sebanyak 82 data masuk ke dalam cluster C0, yang mengindikasikan bahwa mayoritas siswa dan siswi

memiliki tingkat prestasi yang serupa dan mungkin berada pada tingkat prestasi yang umum atau standar. Cluster C1, yang terdiri dari 16 data, mungkin mencerminkan kelompok siswa dan siswi dengan prestasi yang lebih tinggi atau menunjukkan kemampuan akademis yang lebih unggul dibandingkan kelompok mayoritas. Siswa dan siswi dalam cluster ini bisa jadi adalah mereka yang memiliki nilai tertinggi, aktif dalam kegiatan ekstrakurikuler, atau menunjukkan potensi kepemimpinan.

Sementara itu, cluster C2, yang terdiri dari 12 data, kemungkinan menggambarkan kelompok siswa dan siswi dengan prestasi yang paling menonjol atau luar biasa. Mereka mungkin adalah siswa-siswi yang sering memenangkan lomba akademis atau non-akademis, memiliki pencapaian yang luar biasa dalam bidang tertentu, atau secara konsisten menunjukkan performa yang sangat baik di berbagai aspek penilaian sekolah. Dengan demikian, pengelompokan ini membantu dalam mengidentifikasi dan memahami distribusi prestasi di antara siswa dan siswi, serta memungkinkan pihak sekolah untuk merancang program pembinaan yang lebih tepat sasaran untuk setiap kelompok prestasi.

## 2. Hasil Centroid

Attribute	cluster_0	cluster_1	cluster_2
Nilai Pengetahuan	95.293	88.875	87
Nilai Keterampilan	97.122	87	82
Nilai Sikap	93	87.375	97

**Gambar 3.5** Hasil Centroid

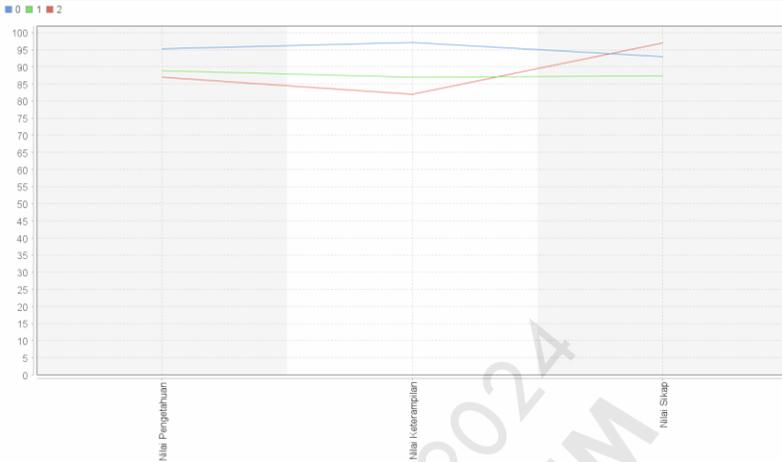
Gambar di atas menunjukkan hasil pengelompokan siswa dan siswi berprestasi berdasarkan tiga atribut: Nilai Pengetahuan, Nilai Keterampilan, dan Nilai Sikap.

Pengelompokan ini dilakukan menggunakan metode K-Means, menghasilkan tiga cluster: cluster\_0, cluster\_1, dan cluster\_2. Cluster\_0, dengan 82 siswa dan siswi, menunjukkan rata-rata Nilai Pengetahuan sebesar 95.293, Nilai Keterampilan sebesar 97.122, dan Nilai Sikap sebesar 93. Siswa dan siswi dalam cluster ini memiliki prestasi yang sangat baik, menunjukkan konsistensi dalam kinerja akademis dan non-akademis yang tinggi.

Cluster\_1 terdiri dari 16 siswa dan siswi dengan rata-rata Nilai Pengetahuan sebesar 88.875, Nilai Keterampilan sebesar 87, dan Nilai Sikap sebesar 87.375. Kelompok ini menunjukkan prestasi yang baik namun sedikit di bawah cluster\_0, mengindikasikan bahwa mereka memiliki kemampuan yang solid tetapi mungkin memerlukan sedikit lebih banyak dorongan atau dukungan untuk mencapai puncak potensi mereka. Cluster\_2, yang terdiri dari 12 siswa dan siswi, memiliki rata-rata Nilai Pengetahuan sebesar 87, Nilai Keterampilan sebesar 82, dan Nilai Sikap sebesar 97. Siswa dan siswi dalam cluster ini menonjol dalam Nilai Sikap, namun prestasi akademis dan keterampilan mereka sedikit lebih rendah dibandingkan dua cluster lainnya. Mereka mungkin menunjukkan keunggulan dalam aspek sikap dan karakter, yang bisa menjadi dasar untuk pengembangan lebih lanjut dalam bidang akademis dan keterampilan.

Pengelompokan ini memberikan wawasan yang berguna bagi pihak sekolah untuk merancang program pembinaan yang sesuai dengan kebutuhan setiap kelompok siswa dan siswi, memastikan bahwa setiap individu mendapatkan dukungan yang tepat untuk mencapai prestasi maksimal.

### 3. Hasil Plot



**Gambar 3.6** Hasil Plot

Pada gambar di atas merupakan hasil plot yang diperoleh dari proses clustering data siswa dan siswi berprestasi menggunakan metode K-Means.

~oOo~



## Bab 4 PENUTUP

Kajian yang dilakukan menggunakan metode K-Means untuk clustering menunjukkan hasil pengelompokan yang informatif berdasarkan data dari 110 siswa dan siswi. Dari hasil analisis, sebanyak 82 data masuk ke dalam cluster\_0, sebanyak 16 data masuk ke dalam cluster\_1, dan sebanyak 12 data masuk ke dalam cluster\_2. Pengelompokan ini membantu dalam mengidentifikasi dan memahami distribusi prestasi siswa dan siswi, serta memberikan gambaran yang jelas mengenai karakteristik masing-masing kelompok. Cluster\_0, yang mencakup mayoritas siswa dan siswi dengan 82 data, menunjukkan rata-rata Nilai Pengetahuan sebesar 95.293, Nilai Keterampilan sebesar 97.122, dan Nilai Sikap sebesar 93. Siswa dan siswi dalam cluster ini memiliki prestasi yang sangat baik secara keseluruhan, dengan kinerja akademis dan non-akademis yang konsisten tinggi. Mereka merupakan kelompok yang menonjol dalam hal pencapaian akademis dan memiliki keterampilan yang baik serta sikap yang positif.

Cluster\_1, yang terdiri dari 16 siswa dan siswi, memiliki rata-rata Nilai Pengetahuan sebesar 88.875, Nilai Keterampilan

sebesar 87, dan Nilai Sikap sebesar 87.375. Meskipun nilai rata-rata mereka sedikit lebih rendah dibandingkan cluster\_0, siswa dan siswi dalam kelompok ini tetap menunjukkan prestasi yang baik. Mereka mungkin memerlukan dorongan atau dukungan tambahan untuk mencapai potensi maksimal mereka, tetapi secara keseluruhan, mereka adalah siswa-siswi yang kompeten dan berbakat. Cluster\_2, yang terdiri dari 12 siswa dan siswi, menunjukkan karakteristik yang berbeda dengan rata-rata Nilai Pengetahuan sebesar 87, Nilai Keterampilan sebesar 82, dan Nilai Sikap sebesar 97. Kelompok ini menonjol dalam aspek sikap dan karakter, meskipun prestasi akademis dan keterampilan mereka sedikit lebih rendah dibandingkan dua cluster lainnya. Siswa dan siswi dalam cluster ini memiliki sikap yang sangat baik, yang bisa menjadi landasan kuat untuk pengembangan lebih lanjut dalam bidang akademis dan keterampilan mereka. Pengelompokan ini memberikan wawasan yang berharga bagi pihak sekolah dalam merancang program pembinaan yang sesuai dengan kebutuhan masing-masing kelompok siswa dan siswi. Dengan memahami karakteristik dan kebutuhan unik dari setiap cluster, sekolah dapat memberikan dukungan yang lebih tepat sasaran, memastikan bahwa setiap individu mendapatkan peluang terbaik untuk berkembang dan mencapai prestasi maksimal.





## DAFTAR PUSTAKA

- [1] A. Aditya, I. Jovian, and B. N. Sari, "Implementasi K-Means Clustering Ujian Nasional Sekolah Menengah Pertama di Indonesia Tahun 2018/2019," *J. MEDIA Inform. BUDIDARMA*, vol. 4, no. 1, p. 51, Jan. 2020.
- [2] F. Pramataning Dewi, P. Siwi Aryni, and Y. Umaidah, "Implementasi Algoritma K-Means Clustering Seleksi Siswa Berprestasi Berdasarkan Keaktifan dalam Proses Pembelajaran," *MEI*, 2011.
- [3] J. Khatib Sulaiman, M. Silalahi, and I. Artikel Abstrak, "Penerapan Data Mining Clustering Dalam Mengelompokkan Buku Dengan Metode K-Means," *Indones. J. Comput. Sci. Attrib.*, vol. 10, no. 1, pp. 2021-77.
- [4] U. Menentukan, K. Bantuan, S. Miskin, and D. Berprestasi, "Clustering Data Siswa SMPN-6 Palangka Raya."
- [5] T. R. Rivanthio and M. Ramdhani, "Penerapan Teknik Clustering Data Mining untuk Memprediksi Kesesuaian Jurusan Siswa (Studi Kasus SMA PGRI 1 Subang)," *Fakt. Exacta*, vol. 13, no. 2, p. 125, Aug. 2020.

- [6] B. Hasmaulina, "Penerapan Data Mining Untuk Membentuk Kelompok Belajar Menggunakan Metode Clustering di SMK Negeri 3 Seluma," 2021.
- [7] A. I. Rizmayanti, N. Hidayati, F. S. Nugraha, and W. Gata, "9~18 Diterima Februari 10," *J. SWABUMI*, vol. 9, no. 1, p. 2021, 2021.
- [8] A. I. Warnilah, "Analisis Algoritma K-Means Clustering untuk Pemetaan Prestasi Siswa Studi Kasus SMP Negeri I Sukahening," *IJCIT (Indonesian J. Comput. Inf. Technol.)*, vol. 1, no. 1, Jun. 2016.
- [9] M. Syukron Nawawi, F. Sembiring, and A. Erfina, "'Program Studi Teknik Informatika-Universitas PGRI Madiun' | 789 Implementasi Algoritma K-Means Clustering Menggunakan Orange Untuk Penentuan Produk Busana Muslim Terlaris."
- [10] 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, Feb. 2023.
- [11] R. Sovia, E. P. W. Mandala, and S. Mardhiah, "Algoritma K-Means dalam Pemilihan Siswa Berprestasi dan Metode SAW untuk Prediksi Penerima Beasiswa Berprestasi," *J. Edukasi dan Penelit. Inform.*, vol. 6, no. 2, p. 181, Aug. 2020.



## Tentang Penulis



**Yuni Franata Sinurat, S.Kom.** lahir di Bagan Batu PT. LTS. ADE (Riau) pada tanggal 04 Mei 2002. Saat ini penulis tinggal di Rantauprapat, Sumatera Utara. Pendidikan tinggi ditempuh pada S1 di Fakultas Sains dan Teknologi (lulus 2024) Universitas Labuhanbatu dengan meraih gelar S.Kom. Penulis baru membuat buku untuk pertama kalinya dengan judul "*Data Mining* Pengelompokan Siswa Berprestasi Menggunakan Metode *Clustering*". Penulis

berharap awal mula buku ini menjadi berkat untuk mencapai karier penulis di hari selanjutnya, sehingga penulis mampu untuk mengembangkan bakat sebagai seorang penulis sesuai dengan harapan dan cita-cita penulis sejak kecil.



**Masrizal, S.Kom., M.Kom.** lahir di Manggung pada tanggal 05 Maret 1994. Proses dalam menempuh pendidikan baru dimulai dari sekolah dasar SD N 13 Manggung, SMP N 7 Pariaman, SMA N 4 Parlaman dan melanjutkan perguruan tinggi swasta dengan Jurusan Sistem Informasi S1 (Sarjana) Universitas Putra Indonesia "YPTK" Padang tahun lulusan 2016 dan melanjutkan Magister 2016 Universitas Putra Indonesia "YPTK" Padang. Tahun lulusan 2018 dengan

Jurusan Sistem Informasi, Fakultas Teknologi Informasi. Pada 2018 saya mengabdikan diri sebagai salah satu dosen di bidang Ilmu Komputer di Fakultas Sains dan Teknologi (FST) Prodi Sistem Informasi di Universitas Labuhanbatu salah satu

menjadi dosen tetap di kampus tersebut. Syukur alhamdulillah sekarang saya mengabdikan diri di bawah Wakil Rektor III di bidang pengembangan, kerja sama dan alumni sebagai Ka. biro bagian pengembangan, kerja sama dan alumni serta divisi tracer studi di Universitas Labuhanbatu Sumatera Utara.



**Irmayanti, S.Si., M.Pd.** lahir di Simonis pada tanggal 24 Agustus 1984. Setelah menamatkan SD Negeri 112324 Aek Kota Batu pada tahun 1996, lalu melanjutkan sekolah Madrasah Tsanawiyah Daar Al-Ulum Asahan Kisaran lulus pada tahun 1999, SMU Al-Azhar Medan lulus pada tahun 2002, S1 Matematika UNIMED lulus pada tahun 2007, dan melanjutkan S2 Pendidikan Matematika lulus pada tahun 2013. Menikah dengan Beni Yunanta Hasibuan pada tahun 2009, mempunyai dua orang anak, satu putri (Ratu Nabila Yunanta Hasibuan) dan satu putra (Rauf Athaya Alby Yunanta Hasibuan). Memulai karier dari guru honor sampai di perguruan tinggi STKIP Labuhanbatu sebagai dosen tetap. Melalui karya ini mudah-mudahan dapat menjadi amal keluarga kami dan terkhusus kedua orang tua (Alm. H. Zulkarnaen Ritonga dan Hj. Normah Dalimunthe) semoga selalu mendapat rahmat dari Allah Swt.



# *Data Mining* Pengelompokan Siswa Berprestasi Menggunakan Metode *Clustering*

**P**restasi adalah pencapaian atau hasil yang diperoleh seseorang atau kelompok sebagai hasil dari usaha, kerja keras, dan keterampilan yang ditunjukkan dalam suatu bidang atau aktivitas. Prestasi bisa berupa berbagai bentuk, seperti penghargaan, pengakuan, pencapaian target, atau kesuksesan dalam kompetisi.

*Data mining* adalah proses mengidentifikasi pola, tren, dan informasi berguna lainnya dari kumpulan data besar. Proses ini melibatkan penggunaan teknik statistik, matematika, dan komputasi untuk menganalisis data dan menemukan hubungan yang mungkin tidak terlihat secara langsung.

Tujuan utama *data mining* adalah untuk mengekstrak informasi yang berguna dari data mentah yang dapat digunakan untuk pengambilan keputusan, peramalan, atau pemecahan masalah. *Data mining* sering digunakan dalam berbagai bidang, seperti bisnis, pemasaran, kesehatan, penelitian ilmiah, dan teknologi informasi.

Buku ini akan membahas data prestasi siswa dari berbagai bentuk prestasi yang dicapai. Selain itu buku ini juga membantu untuk menjadi pedoman bagi siswa yang ingin berprestasi di bidang pendidikan.