

## Pemodelan K-Means untuk Klasifikasi Siswa Berprestasi di Sekolah Menengah Kejuruan (SMK) Migas Cepu

Muksan Junaidi

Program Studi Informatika, Sekolah Tinggi Teknologi Ronggolawe Cepu, Indonesia

Korespondensi penulis: [muchsan.djunaedi@gmail.com](mailto:muchsan.djunaedi@gmail.com)

**Abstract.** *The determination of outstanding students is based on different criteria, depending on the type of achievement that is to be measured. At SMK Migas Cepu, this assessment is typically based on the highest academic score from the class promotion exam. However, this method is considered less accurate and problematic in terms of grouping students. To address this issue, a clustering method using the K-Means algorithm can be applied. The purpose of this research is to build a K-Means model to determine outstanding students. The data used in this study comes from the report card ledger of class XI Machine A and B for the year 2022, which includes 71 students at SMK Migas Cepu. The RapidMiner tool was used to build the K-Means model and cluster the data based on student characteristics. The first test conducted using Excel resulted in two clusters: 35 outstanding students and 36 non-outstanding students. Meanwhile, the second test using the RapidMiner model produced two clusters with a distribution of 26 outstanding students and 45 non-outstanding students.*

**Keywords:** *Excel Application, K-Means, Leger Value Reports, Outstanding students, Rapid Miner*

**Abstrak.** Penentuan siswa berprestasi di SMK Migas Cepu bergantung pada berbagai kriteria yang disesuaikan dengan jenis prestasi yang ingin diukur. Umumnya, penilaian dilakukan berdasarkan nilai tertinggi dari ujian kenaikan kelas. Namun, pendekatan ini dirasakan kurang efektif dan menyulitkan dalam proses pengelompokan siswa. Sebagai solusi, metode clustering menggunakan algoritma K-Means dapat diterapkan. Penelitian ini bertujuan untuk mengembangkan model K-Means yang dapat digunakan dalam menilai siswa berprestasi. Data yang digunakan berasal dari ledger nilai rapor kelas XI Mesin A dan B tahun 2022, yang mencakup 71 siswa di SMK Migas Cepu. Untuk membangun model K-Means dan mengelompokkan data berdasarkan karakteristik siswa, aplikasi RapidMiner digunakan. Pada pengujian pertama yang dilakukan dengan aplikasi Excel, diperoleh dua cluster: 35 siswa berprestasi dan 36 siswa tidak berprestasi. Sedangkan pada pengujian kedua dengan menggunakan model RapidMiner, diperoleh dua cluster dengan distribusi yang berbeda: 26 siswa berprestasi dan 45 siswa tidak berprestasi.

**Kata kunci :** Aplikasi Excel, K-Means, Leger Nilai Rapor, Siswa berprestasi, Rapid Miner

### 1. PENDAHULUAN

SMK Migas Cepu merupakan institusi pendidikan yang memiliki jumlah siswa yang sangat besar. Pada tahun 2021, tercatat sebanyak 408 siswa berhasil menyelesaikan studi mereka di sekolah ini. Siswa-siswa tersebut berasal dari berbagai jurusan, antara lain: 71 siswa dari jurusan Teknik Pemesinan, 27 siswa dari jurusan Teknik Instalasi Tenaga Listrik, 31 siswa dari jurusan Teknik Elektronika Industri, 96 siswa dari jurusan Teknik Kendaraan Ringan Otomotif, 85 siswa dari jurusan Teknik Pengolahan Minyak, Gas, dan Petrokimia, 42 siswa dari jurusan Pemboran Migas, serta 56 siswa dari jurusan Teknik Produksi Migas.

Penentuan siswa berprestasi di SMK ini didasarkan pada nilai tertinggi yang diperoleh dari ujian kenaikan kelas. Hal ini dimaksudkan agar penilaian prestasi akademik dapat mencakup kategori yang seimbang, yakni prestasi yang tinggi, sedang, dan kurang. Namun, permasalahan muncul berdasarkan hasil pengamatan dan evaluasi dari para guru yang

mengajar. Penentuan siswa berprestasi di sekolah ini sering kali terlihat kurang akurat atau sulit untuk ditentukan.

Pada jurusan Teknik Pemesinan, banyak siswa yang tidak menunjukkan prestasi memadai, meskipun ada beberapa siswa yang berhasil berprestasi. Hal ini menunjukkan bahwa proses pembentukan siswa berprestasi di tingkat menengah (sedang) belum sepenuhnya efektif. Berdasarkan data tahun 2021, jumlah siswa di jurusan Teknik Pemesinan yang berprestasi adalah 71 siswa.

Berdasarkan pembahasan mengenai tantangan dalam penentuan siswa berprestasi dan beberapa temuan dari penelitian sebelumnya, penelitian ini akan mengintegrasikan berbagai data relevan, seperti kegiatan ekstrakurikuler yang diikuti, nilai tugas, UTS, UAS, absensi, serta nilai sikap. Dengan demikian, proses analisis menjadi lebih menyeluruh dan komprehensif. Hasil penelitian ini diharapkan dapat mempermudah proses penentuan siswa berprestasi secara lebih objektif dan terukur..

Tujuan penelitian ini adalah untuk mengembangkan model sistem penilaian siswa berprestasi di SMK Migas Cepu menggunakan algoritma K-Means, yang diharapkan dapat memudahkan pemilihan siswa berprestasi berdasarkan nilai efektif.

Manfaat dari penelitian ini adalah sebagai pengembangan sistem untuk mengelompokkan prestasi akademik siswa, sehingga dapat mempermudah pekerjaan para wali siswa. Salah satu cara pengelompokan data yang digunakan dalam penelitian ini adalah dengan metode *clustering*.

## 2. KERANGKA TEORI

### Prestasi Akademik Siswa

Prestasi akademik siswa menggambarkan pencapaian luar biasa yang diraih oleh individu, baik dalam aspek akademik maupun non-akademik di sekolah, yang layak untuk mendapatkan pengakuan atau penghargaan. Siswa, dalam hal ini, adalah pelajar yang tengah menempuh pendidikan, baik di tingkat Sekolah Dasar (SD), Sekolah Menengah Pertama (SMP), hingga Sekolah Menengah Atas (SMA). Mereka berusaha keras untuk mendapatkan pengetahuan serta pemahaman yang mendalam tentang berbagai bidang ilmu, guna mempersiapkan diri untuk masa depan yang lebih cerah.

Secara umum, prestasi dapat diartikan sebagai hasil yang diperoleh dari usaha atau kegiatan tertentu. Setiap pencapaian dalam berbagai jenis aktivitas yang berhasil diselesaikan dapat dianggap sebagai prestasi. Prestasi menggambarkan keterampilan atau pencapaian nyata

yang diraih dalam jangka waktu tertentu. Dalam konteks penelitian ini, prestasi siswa dipahami sebagai hasil yang diperoleh siswa selama proses pembelajaran berlangsung.

### **Sekolah Menengah Kejuruan (SMK) Migas Cepu**

SMK Migas Cepu, yang terletak di Kompleks Mentul, Cepu, Provinsi Jawa Tengah, telah dikenal luas di kalangan masyarakat. Sekolah ini merupakan lembaga pendidikan formal dengan standar mutu nasional yang berperan penting dalam pembangunan nasional, bertujuan menciptakan individu yang memiliki keterampilan untuk siap bekerja dan mandiri. Untuk mewujudkan lulusan yang siap menghadapi dunia kerja, manajemen yang efisien dan perhatian terhadap lingkungan sekitar sangat diperlukan. Informasi lebih lanjut dapat ditemukan melalui situs web resmi SMK Migas Cepu (<http://www.smkmigas.com>).



**Gambar 1** Tampilan Profil SmK Migas Cepu

Jumlah siswa yang terdaftar di setiap jurusan di SMK Migas Cepu adalah sebagai berikut: jurusan Teknik Pemesinan memiliki 71 siswa, jurusan Teknik Instalasi Tenaga Listrik terdiri dari 27 siswa, jurusan Teknik Elektronika Industri memiliki 31 siswa, jurusan Teknik Kendaraan Ringan Otomotif mencatatkan 96 siswa, jurusan Teknik Pengolahan Minyak, Gas, dan Petrokimia berjumlah 85 siswa, jurusan Pemboran Minyak dan Gas memiliki 42 siswa, serta jurusan Teknik Produksi Minyak dan Gas terdapat 56 siswa.

### **Data Mining**

Data mining merupakan teknik pengolahan data yang digunakan untuk mengidentifikasi pola-pola tersembunyi dalam sejumlah besar data. Temuan dari proses ini dapat digunakan sebagai dasar dalam pengambilan keputusan di masa depan. Dalam konteks pengolahan data berskala besar, data mining memainkan peran krusial di berbagai sektor, seperti industri, keuangan, meteorologi, serta dalam bidang ilmu pengetahuan dan teknologi (Haryati et al., 2015).

Metode ini dapat digunakan pada berbagai jenis basis data dan sistem penyimpanan informasi, dengan pola yang ditemukan bergantung pada berbagai teknik data mining, seperti deskripsi kelas/konsep, asosiasi, analisis korelasi, klasifikasi, prediksi, serta analisis kluster,

dan lainnya.. Penerapan algoritma data mining bertujuan untuk mengekstraksi pola-pola tertentu dari data. Proses ini secara otomatis mengidentifikasi pola-pola sederhana dalam kumpulan data besar dengan memanfaatkan teknik analisis yang tepat.

### **Clustering**

Clustering merupakan proses pengelompokan objek-objek yang memiliki kesamaan ke dalam grup-grup yang berbeda. Tujuannya adalah membagi dataset menjadi subset-subset, di mana data dalam setiap subset memiliki relevansi yang berguna. Setiap cluster berisi objek-objek yang serupa satu sama lain dan berbeda dengan objek-objek yang ada pada cluster lainnya..

Salah satu metode populer dalam model clustering adalah K-Means, yang digunakan untuk mengelompokkan data berdasarkan kesamaan karakteristik tertentu. Dengan algoritma ini, data akan dibagi menjadi beberapa cluster, di mana setiap cluster berisi objek yang memiliki kedekatan atau kesamaan yang lebih tinggi dengan anggota lainnya, serta perbedaan yang jelas dengan cluster lain. Proses clustering dengan metode K-Means terdiri dari beberapa langkah utama, yaitu :

1. Menentukan jumlah cluster (k) yang ingin dibentuk dari data yang ada.
2. Menginisialisasi titik pusat cluster (centroid) sebanyak k. Titik ini dapat dipilih secara acak dari data yang tersedia, salah satu cara yang sering digunakan adalah dengan pemilihan acak (random).
3. Menghitung jarak antara setiap data dengan masing-masing centroid menggunakan rumus jarak Euclidean (Euclidean Distance), sehingga dapat diketahui data mana yang memiliki jarak terdekat dengan centroid tertentu, berdasarkan hasil perhitungan tersebut

$$dy = \sum_{k=1}^p \{x_{ik} - x_{jk}\}^2 \dots \dots \dots (2.1)$$

dij = Jarak objek antara objek i dan j

P = Dimensi data

Xik = Koordinat dari obyek i pada dimensi k

Xjk = Koordinat dari obyek j pada dimensi

### **Algoritma K-Means**

Berbeda dari algoritma lain, K-Means dimulai dengan menentukan jumlah kelompok (cluster) yang diinginkan dan mendefinisikan nilai centroid awalnya (Dhuhita, 2015). Proses ini menggunakan iterasi berulang untuk mencapai pembentukan cluster yang paling optimal, di mana jumlah cluster yang diinginkan akan menghasilkan titik centroid akhir sebagai output.

Pada pemilihan pola k sebagai titik centroid awal dilakukan secara random. Proses iterasi/pengulangan yang diperlukan dalam menemukan centroid yang tepat sangat bergantung pada pemilihan awal ini. Oleh karena itu, penting untuk memilih centroid dengan mempertimbangkan kepadatan data yang tinggi agar menghasilkan kinerja yang lebih optimal dalam proses clustering (Sumadikarta & Abeiza, 2014).

Sebagai algoritma clustering berbasis partisi (Partitioning method), K-Means menggunakan titik pusat (centroid) untuk membagi data menjadi kelompok-kelompok yang lebih terstruktur. Han, J.W., Kamber, M., dan Pei (2012) juga menyatakan bahwa ada syarat minimal dan tantangan yang harus dipenuhi agar algoritma clustering ini dapat berfungsi dengan baik.

### ***Rapid Miner***

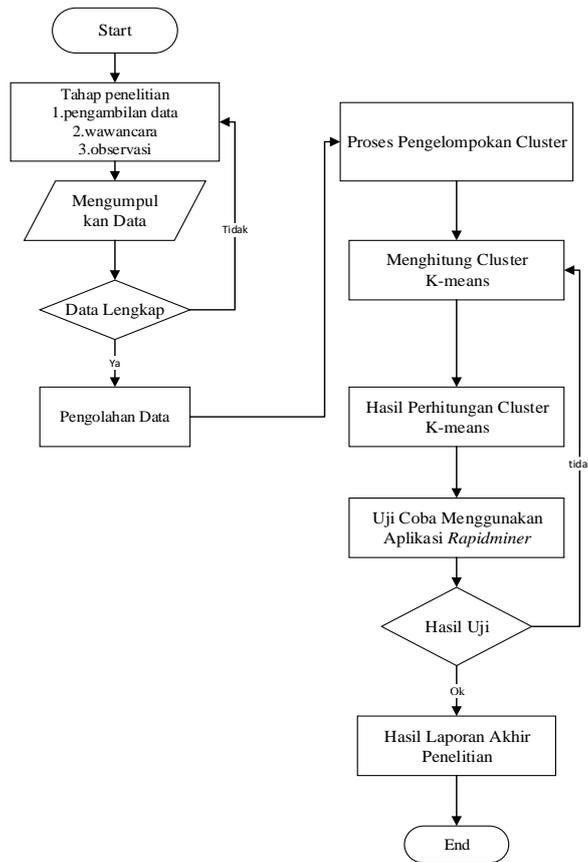
RapidMiner merupakan perangkat lunak sumber terbuka yang menyediakan solusi komprehensif untuk analisis data. Dengan memanfaatkan berbagai metode deskriptif dan prediktif, RapidMiner memberikan pemahaman mendalam bagi penggunanya, yang pada gilirannya memungkinkan mereka untuk mengambil keputusan yang lebih akurat dan efektif.

Aplikasi ini dilengkapi dengan operator untuk input, output, preprocessing data, dan visualisasi, RapidMiner menawarkan fleksibilitas yang luas dalam pengolahan data. Sebagai software mandiri untuk analisis data, RapidMiner juga dapat diintegrasikan dengan produk lainnya, menjadikannya mesin data mining yang kuat dan serbaguna. Ditulis menggunakan bahasa pemrograman Java, RapidMiner dapat berjalan di berbagai sistem operasi, menjadikannya solusi yang dapat diakses secara universal.

## **3. METODOLOGI**

### **Desain Penelitian**

Diagram alir penelitian sangat penting dalam merencanakan, menganalisis, dan mendokumentasikan setiap tahapan proses penelitian. Diagram ini berfungsi sebagai panduan yang jelas untuk menjalankan penelitian dengan sistematis. Gambar 3.1 merupakan diagram alir penelitian yang digunakan dalam studi ini.



**Gambar 2** Diagram Alir Penelitian

## Teknik Dan Analisis Data

### 1. Pengolahan Data siswa

Pengolahan data adalah serangkaian proses yang bertujuan untuk mengubah data mentah menjadi informasi atau pengetahuan yang berguna. Setelah diprogram, proses ini dapat dilakukan secara otomatis oleh komputer, di mana data yang telah diproses membentuk sebuah sistem informasi. Terdapat tiga metode utama dalam pengolahan data, yaitu:

- Pengolahan data manual pada proses pengumpulan, penyaringan, penyortiran, perhitungan, serta operasi logika lainnya, semuanya dikerjakan langsung oleh manusia tanpa bantuan perangkat lunak.
- Pengolahan data mekanis melibatkan pemanfaatan perangkat dan mesin untuk memproses data, yang sangat efektif untuk mempercepat aktivitas dalam proyek-proyek berskala kecil, sehingga meningkatkan efisiensi dan produktivitas.
- Pengolahan data elektronik adalah metode paling mutakhir dan tercepat, di mana data diproses menggunakan teknologi canggih yang didukung oleh perangkat lunak dan program pengolahan data modern, memungkinkan hasil yang lebih efisien dan akurat.

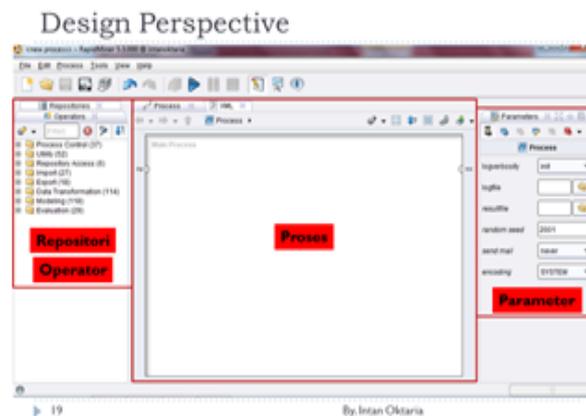
## 2. Proses pengelompokan *Cluster*

Pengelompokan data dengan teknik K-Means adalah salah satu metode dalam data mining yang membagi data ke dalam beberapa kelompok. Setiap cluster akan diisi oleh beberapa anggota yang memiliki kesamaan. Berikut adalah langkah-langkah dalam proses perhitungan pengelompokan data menggunakan K-Means:

1. Tentukan jumlah cluster yang diinginkan.
2. Tentukan nilai pusat (centroid) untuk setiap cluster.
3. Hitung jarak antara data dan pusat cluster.
4. Kelompokkan data ke dalam cluster yang sesuai.
5. Kembali ke langkah 2 dan ulangi proses hingga nilai centroid stabil dan anggota cluster tidak berpindah ke cluster lainnya.

## Uji Model Klasifikasi Dengan Rapid Miner

Proses pengolahan data set akan dilakukan menggunakan RapidMiner dengan penerapan algoritma K-Means, bertujuan untuk mencari pola yang lebih efisien. Algoritma ini akan digunakan untuk mengeksplorasi berbagai data serta nilai siswa yang ada. Selama proses tersebut, data siswa akan dipersiapkan dan dikategorikan ke dalam kelompok berdasarkan tingkat kemampuan, yakni tinggi, sedang, dan rendah. Hasil klasifikasi akan dipakai untuk menentukan penilaian dan seleksi siswa berprestasi di sekolah.



**Gambar 3.** Tampilan Aplikasi Rapidminer

## 4. HASIL DAN PEMBAHASAN

### Hasil Pengumpulan Data

Sumber utama data dalam penelitian ini berasal dari database leger nilai rapor kelas XI jurusan Teknik Permesinan dengan total 71 siswa angkatan 2021/2022. Data ini diterima dalam format tabel melalui aplikasi Microsoft Excel, yang memudahkan proses penentuan siswa berprestasi. Dataset tersebut mencakup atribut dari nilai rapor, yang disederhanakan menjadi

beberapa mata pelajaran sebagai variabel, di antaranya: Pendidikan Agama Islam, Pendidikan Pancasila dan Kewarganegaraan, Bahasa Indonesia, Matematika, Bahasa Inggris, Pendidikan Jasmani Olahraga dan Kesehatan, Gambar Teknik, Teknik Pemesinan Bubut, Teknik Pemesinan Frais, Teknik Pemesinan NC/CNC dan CAM, Produk Kreatif Kewirausahaan, serta Bahasa Jawa.

Data ini awalnya diserahkan dalam bentuk lembaran kertas leger nilai rapor oleh wakil kurikulum di SMK Migas Cepu. Selanjutnya, data tersebut diperbarui dan diorganisir dalam Microsoft Excel untuk mempermudah proses perhitungan menggunakan metode K-Means. Setelah diperbarui, data dihitung menggunakan metode K-Means untuk mengelompokkan siswa ke dalam cluster yang telah ditentukan.

**Tabel 1** Data Sampel Nilai Leger Raport Siswa

PAI	PKN	BHS IND	MTK	BHS ING	PENJAS	G.Teknik	T.BUBUt	T.Frais	T.NC/CNC	Produk Kreatif	BHS Jawa
79	79	84	80	78	82	50	80	82	89	79	64
79	79	84	79	79	82	42	78	80	83	79	64
79	82	83	79	78	86	21	80	81	83	80	83
79	78	84	75	78	79	21	78	78	84	79	81
79	82	84	65	79	84	56	78	78	85	80	80
85	85	84	80	78	84	85	82	90	93	85	87
79	79	84	79	78	81	47	80	80	87	81	79
79	79	84	79	78	79	42	78	78	84	79	80
79	79	84	79	78	84	47	78	78	87	80	79
87	89	84	82	80	90	88	82	85	92	86	89
83	81	84	80	78	84	47	80	90	83	85	83
84	91	85	80	78	90	85	80	87	91	86	86
79	78	84	80	78	81	47	80	78	85	80	80
79	81	84	77	78	79	13	78	78	83	79	84
79	80	84	80	78	82	47	80	80	85	80	85
79	81	83	80	78	84	47	80	80	84	81	79
79	85	84	78	78	90	81	80	80	85	82	80

Setelah proses pengelompokan selesai, nilai-nilai yang ada pada leger rapor dihitung dan disesuaikan dengan cluster yang terbentuk. Hasil perhitungan ini kemudian diuji menggunakan aplikasi RapidMiner untuk memperoleh hasil yang lebih valid.

### Perhitungan K Means

#### 1. Perhitungan K-Means literasi 1

Penentuan jumlah cluster adalah langkah awal dalam proses pengelompokan, yang menentukan seberapa banyak kelompok yang akan terbentuk. Dalam penelitian ini, jumlah cluster yang dipilih adalah dua cluster. Selanjutnya, proses menentukan centroid awal dilakukan secara acak. Centroid awal ini berfungsi sebagai titik pusat untuk cluster pertama. Dalam penelitian ini, centroid awal yang digunakan adalah:

$$C1 = (87,89,84,82,80,90,88,82,85,92,86,89)$$

$$C2 = (86,85,84,80,80,82,85,80,90,85,83,83)$$

Langkah berikutnya adalah menghitung jarak setiap data terhadap masing-masing pusat cluster. Proses perhitungan ini dilakukan dengan menggunakan rumus Euclidean Distance, yang terdiri dari dua langkah utama:

- a) Menghitung jarak antara data siswa pertama dan pusat cluster pertama.

$$C1 \sqrt{\begin{array}{l} (79 - 87)^2 + (79 - 89)^2 + (84 - 84)^2 + (80 - 82)^2 \\ + (78 - 80)^2 + (82 - 90)^2 + (50 - 88)^2 + (80 - 82)^2 \\ + (82 - 85)^2 + (89 - 92)^2 + (79 - 86)^2 + (64 - 89)^2 \\ = 48,7442 \end{array}}$$

- b) Menghitung jarak antara data siswa pertama dan pusat cluster kedua.

$$C2 \sqrt{\begin{array}{l} (79 - 86)^2 + (79 - 85)^2 + (84 - 84)^2 + (80 - 80)^2 \\ + (78 - 80)^2 + (82 - 82)^2 + (50 - 85)^2 + (80 - 80)^2 \\ + (82 - 90)^2 + (89 - 85)^2 + (79 - 83)^2 + (64 - 83)^2 \\ = 42,0833 \end{array}}$$

Setelah jarak dihitung, tahap selanjutnya adalah pengelompokan data berdasarkan hasil perhitungan jarak ke centroid, dimana data akan dikelompokkan ke dalam cluster yang paling sesuai.

**Tabel 2** hasil pengelompokan data

C 1	C 2	JARAK TERPENDEK	KELOMPOK
48,7442	42,0833	42,0833	cluster2
56,1249	49,2037	49,2037	cluster2
69,3181	65,3452	65,3452	cluster2
71,3442	66,3400	66,3400	cluster2
41,1582	35,9861	35,9861	cluster2
10,1980	9,8489	9,8489	cluster2
46,0217	40,7308	40,7308	cluster2
51,7011	46,0652	46,0652	cluster2
46,0435	41,4126	41,4126	cluster2
0,0000	14,7309	0,0000	cluster1

Hasil dari pengelompokan

C1 = 3 data (10,12,38)

C2 = 68 data

(1,2,3,4,5,6,7,8,9,,10,11,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,39,40,41,42,43,44,45,46,47,48,49,50,51,52,53,54,55,56,57,58,59,60,61,62,,63,64,65,66,67,68,69,70,71).

### Menentukan Cluster Baru pada literasi 2

Setelah semua data dikelompokkan ke dalam cluster yang sesuai, langkah selanjutnya adalah menghitung kembali pusat cluster yang baru. Proses ini dilakukan dengan mencari rata-rata nilai dari semua anggota dalam cluster tersebut. Untuk menentukan pusat cluster yang baru, nilai setiap anggota dalam cluster dijumlahkan dan kemudian dibagi dengan jumlah anggota dalam cluster tersebut, seperti yang ditunjukkan pada contoh berikut:

$$C1 \sqrt{(87 + 84 + 83) \div 3} = 85$$

$$C2 \sqrt{\begin{matrix} (79 + 79 + 79 + 79 + 79 + 85 + 79 + 79 \\ +83 + 84 + 79 + 79 + 79 + 79 + 79 + 79 + 79 + 79 + 79 \\ +79 + 79 + 79 + 79 + 79 + 79 + 91 + 78 + 79 + 78 + 80 \\ +79 + 85 + 91 + 79 + 82 + 86 + 79 + 79 + 79 + 79 \\ +79 + 79 + 79 + 79 + 79 + 79 + 86 + 79 + 79 + 79 + 79 \\ +79 + 79 + 79 + 79 + 79 + 79 + 79 + 79 + 79 + 79 \\ +79 + 79 + 79 + 79 + 79 + 79 + 79 + 79 + 79 + 79 \end{matrix} \div 68} = 80$$

Setelah diperoleh titik pusat cluster yang baru untuk setiap cluster, langkah selanjutnya adalah menghitung ulang data dengan pusat cluster yang telah diperbarui. Proses ini diulang hingga diperoleh pola yang stabil, di mana hasil pengelompokan tidak berubah dan data tidak berpindah antar cluster. Dalam penelitian ini, perhitungan dilakukan hingga iterasi ke-7, di mana setiap cluster tetap konstan dan tidak ada data yang berpindah dari satu cluster ke cluster lainnya.

### Hasil Evaluasi

1. Hasil perbandingan pengolahan data menggunakan aplikasi Microsoft Excel dengan metode K-Means menghasilkan dua kelompok cluster. Cluster pertama terdiri dari siswa berprestasi, yang mencakup 35 data. Dalam cluster ini, ditemukan bahwa nilai keseluruhan tertinggi ada pada cluster 1, yang menunjukkan bahwa siswa berprestasi dapat ditentukan berdasarkan nilai tertinggi yang diperoleh. Pada proses tersebut, Microsoft Excel menghasilkan dua tabel cluster, di mana siswa berprestasi tercatat pada cluster pertama. Tabel tersebut menunjukkan bahwa siswa dengan nilai tertinggi adalah Arfika Irawan dengan total nilai 1034, diikuti oleh Bayu Guntur dengan total nilai 1023, dan Amirulanam Mulyansyah dengan total nilai 1022.

**Tabel 3** siswa berprestasi pada cluster 1

NO	NAMA SISWA	PAI	PKN	IND	MTK	ING	PENJAS	G.Teknik	T.BUBUK	T.Frais	T.NQ/CVC	Produk Kreatif	BHS Jawa	JUMLAH NILAI
1	ARFIKAIRAWAN	87	89	84	82	80	90	88	82	85	92	86	89	1034
2	BAYU GUNTUR	84	91	85	80	78	90	85	80	87	91	86	86	1023
3	AMIRULANAM MULYANSYAH	83	89	85	80	80	86	87	82	86	87	88	89	1022

2. Pengujian metode K-Means menggunakan aplikasi RapidMiner menghasilkan bahwa cluster pertama mencakup 26 data siswa berprestasi dari total 71 data yang dianalisis. Dalam proses ini, aplikasi RapidMiner menghasilkan dua tabel cluster, dengan cluster 0 berisi siswa berprestasi. Di dalam tabel tersebut, siswa dengan nilai tertinggi adalah Aditya Wahana dengan total nilai 1018, diikuti oleh Arfika Irawan dengan total nilai





2. Hasil pengujian menggunakan aplikasi RapidMiner menunjukkan pola yang serupa dengan Microsoft Excel, yaitu dua cluster yang terbentuk dengan karakter yang sama: siswa berprestasi dan siswa yang tidak berprestasi. Pada pengujian ini, cluster pertama terdiri dari 26 data siswa berprestasi, dan cluster kedua terdiri dari 45 data siswa tidak berprestasi, dengan total keseluruhan data sebanyak 71.

## **DAFTAR PUSTAKA**

### **Artikel Jurnal**

- Dhuhita, W. M. P. (2015). Clustering menggunakan metode K-Means untuk menentukan status gizi balita. *Jurnal Informatika Darmajaya*, 15(2), 160–174.
- Han, J. W., Kamber, M., & Pei, J. (2012). Findings seminal papers using data mining techniques. *Open Journal of Social Sciences*. <https://doi.org/10.4236/JSS.2020.89023>
- Haryati, S., Sudarsono, A., & Suryana, E. (2015). Implementasi data mining untuk memprediksi masa studi mahasiswa menggunakan algoritma C4.5 (Studi kasus: Universitas Dehasen Bengkulu). *Jurnal Media Infotama*, 11.
- Sumadikarta, I., & Abeiza, E. (2014). Penerapan algoritma K-Means pada data mining untuk memilih produk dan pelanggan potensial (Studi kasus: PT Mega Arvia Utama). *Jurnal Satya Informatika*, 1, 1–12.

### **Buku Kompilasi (Edited Book)**

- Ediyanto, Muhlasah Novitasari Mara, & Neva Satyahadewi. (2013). Pengklasifikasian karakteristik dengan metode K-Means cluster analysis. *BIMASTER*, 2(02).

### **Internet**

- SMK Migas Cepu. (n.d.). Retrieved August 24, 2022, from <http://www.smkmigas.com/>