

## **Penerapan Algoritma *K-Means Clustering* Pada Tingkat Penyelesaian Pendidikan Di Provinsi Indonesia**

Akhmad Thoriq Basalamah<sup>1</sup>, Resad Setyadi<sup>2</sup>

<sup>1</sup>Sistem Informasi, Informatika, Institut Teknologi Telkom Purwokerto

<sup>2</sup>Sistem Informasi, Informatika, Institut Teknologi Telkom Purwokerto

<sup>1</sup>18103034@ittelkom-pwt.ac.id, <sup>2,\*</sup>resad@ittelkom-pwt.ac.id

### **Abstract**

*The level of educational attainment plays a crucial role in the progress of a country, as higher levels of education achieved have a positive impact on the development of the nation. The data used in this research is obtained from the Indonesian Central Bureau of Statistics (BPS), which is a reliable and authoritative source of statistical information in the country. The objective of this study is to conduct data mining analysis and examine the educational attainment levels across 34 provinces in Indonesia based on education levels and provinces from 2020 to 2022. The analysis process is performed using RapidMiner Data Mining software. The findings of this study include the mapping of education levels and provinces into specific clusters. The method applied is K-Means, a data mining technique used for classifying data with the aim of identifying patterns or groups within it. In this research, cluster mapping is divided into three groups: high cluster (K0), medium cluster (K1), and low cluster (K2). For elementary school (SD) level, there are 29 high clusters, 4 medium clusters, and 1 low cluster. For junior high school (SMP) level, there is 1 high cluster province, 20 medium clusters, and 13 low clusters. The mapping results indicate that as the educational level increases, there is a decrease in the percentage of educational attainment, with the SD level having a good rate, SMP level being fairly good, and high school (SMA) level having a relatively poor rate. Additionally, the province of Papua is classified under the low cluster. It is hoped that the information derived from the research findings has the potential to provide a deeper understanding and valuable insights as a comprehensive overview of the educational attainment levels in Indonesian provinces over the past few years, aiming to enhance the quality of education in Indonesia.*

**Keywords:** Data Mining, BPS, RapidMiner, K-Means, Education completion rate.

### **Abstrak**

Tingkat penyelesaian pendidikan memegang peranan krusial dalam kemajuan suatu negara, karena semakin tinggi taraf pendidikan yang berhasil diraih, semakin positif pula perkembangan negara tersebut. Data yang digunakan dalam riset ini termuat di Badan Pusat Statistik (BPS) Indonesia, yang merupakan sumber yang terpercaya dan otoritatif dalam hal informasi statistik di negara ini. Tujuan dari penelitian ini adalah untuk melakukan analisis data mining dan menganalisis tingkat pencapaian pendidikan di 34 provinsi di Indonesia berdasarkan jenjang pendidikan dan provinsi dari tahun 2020 hingga 2022. Proses analisis dilakukan dengan menggunakan perangkat lunak RapidMiner Data Mining. Temuan dari studi ini adalah pemetaan yang mengelompokkan jenjang pendidikan dan provinsi ke dalam klaster-klaster tertentu. Metode yang diterapkan adalah *K-Means*, sebuah teknik dalam data mining yang digunakan pengklasifikasian data dengan tujuan mengidentifikasi pola atau kelompok yang ada di dalamnya. Dalam penelitian ini, pemetaan klaster dibagi tiga kelompok, yaitu klaster tinggi (K0), klaster sedang (K1), dan klaster rendah (K2). Pada jenjang SD terdapat 29 klaster tinggi, 4 klaster sedang, dan 1 klaster rendah. Pada jenjang SMP terdapat 1 provinsi klaster tinggi, 20 klaster sedang, dan 13 klaster rendah. Hasil pemetaan menunjukkan bahwa setiap naik jenjang pendidikan mengalami penurunan dalam persentase tingkat penyelesaian pendidikan, dengan jenjang SD memiliki tingkat yang baik, SMP cukup baik, dan SMA memiliki tingkat yang cukup buruk. Selain itu, provinsi Papua tergabung dalam klaster rendah. Harapannya, Informasi yang dimuat dari hasil penelitian ini memiliki potensi untuk memberikan pemahaman yang lebih dalam dan wawasan yang berharga sebagai gambaran secara luas tentang tingkat pencapaian pendidikan di provinsi-provinsi Indonesia dalam kurun waktu beberapa tahun terakhir untuk meningkatkan mutu pendidikan Indonesia.

**Kata kunci:** Data Mining, BPS, RapidMiner, *K-Means*, Tingkat penyelesaian pendidikan.

Diterima Redaksi : 13-06-2023 | Selesai Revisi : 14-08-2023 | Diterbitkan Online : 01-10-2023

## 1. Pendahuluan

Pendidikan adalah salah satu aspek kunci dalam pembangunan suatu negara. Meningkatnya tingkat pendidikan di suatu negara berdampak langsung pada perkembangan ekonomi, kualitas sumber daya manusia, dan kualitas mutu hidup masyarakat. Oleh karena itu, penting untuk mempelajari serta menganalisis tingkat penyelesaian pendidikan di berbagai jenjang pendidikan. Tingkat penyelesaian pendidikan dihitung berdasarkan tingkat penyelesaian pendidikan seharusnya dengan kisaran umur 1-3 tahun pada umumnya, SD (Sekolah Dasar) memiliki durasi pembelajaran selama 6 tahun dengan kisaran umur 12-15 tahun, SMP (Sekolah Menengah Pertama) memiliki durasi pembelajaran selama 3 tahun dengan kisaran umur 15-18 tahun, dan SMA (Sekolah Menengah Atas) memiliki durasi pembelajaran selama 3 tahun dengan kisaran umur 18-20 tahun [1]. Tingkat penyelesaian pendidikan dipengaruhi oleh berbagai faktor diantaranya siswa tinggal kelas, putus sekolah, angka partisipasi kasar dan faktor ekonomi. Penelitian ini fokus pada eksplorasi metode *K-Means* cluster untuk mengidentifikasi dan memvalidasi tingkat penyelesaian pendidikan di masa depan.

Pada penelitian ini, dilakukan penggunaan RapidMiner dan pendekatan *K-Means* sebagai metode yang diterapkan. *RapidMiner* adalah sebuah platform analitik yang kuat yang memungkinkan peneliti untuk mengakses, memproses, dan menganalisis data dengan mudah. Alat ini memberikan kemampuan untuk melakukan berbagai tugas analisis data, termasuk pemrosesan data, pemodelan prediktif, dan visualisasi hasil. Metode *K-Means* merupakan sebuah pendekatan yang sangat berguna dalam proses pengklasifikasian data, yang memungkinkan kita untuk mengelompokkan entitas-entitas dalam sekumpulan data berdasarkan kesamaan karakteristik mereka, *K-Means* juga populer dan efektif. Metode ini membagi data menjadi beberapa kelompok berdasarkan kesamaan karakteristik. provinsi-provinsi berdasarkan tingkat penyelesaian pendidikan pada jenjang SD, SMP, dan SMA. Hasil analisis akan disajikan dalam bentuk visualisasi yang informatif.

Berdasarkan beberapa penelitian sebelumnya, penelitian yang dilakukan oleh Nanda Try Luchia dan Mustakim pada tahun 2022, clusterisasi dengan metode Davies Bouldin Index (DBI) lebih baik ketimbang tidak melakukannya [2]. Penelitian yang dilakukan Agustin dan Sirait pada 2021, *K-Means* telah terbukti menjadi pilihan utama yang sangat efektif berdasarkan pengukuran variansi. [3]. Penelitian oleh Tonny Tendean, Windania Purba, penelitian tentang cluster bahan pangan dengan *K-Means* yang membagi jadi 3 Cluster dengan penentuan secara acak [4]. Penelitian oleh Sena Wijayanto, M Yoka Fathoni, penelitian ini mengelompokkan hasil pertumbuhan tanaman padi di Jawa Tengah menggunakan *K-Means* dan perhitungan jarak Euclidean untuk klusterisasi

[5]. Penelitian oleh Abdul Aziz, Amril Mutoi Siregar, Candra Zonyfar penelitian ini mengelompokkan produktifitas padi di Jawa Barat dengan mengimplementasikan *K-Means* dan fuzzy C-Means untuk klusterisasi [6]. Penelitian oleh Agung Jaelani, penelitian ini melakukan analisis produk domestik regional bruto menggunakan *K-Means* dan membagi jadi 7 cluster [7]. Penelitian oleh Nur Ahlina Febriyati, Achmad Daengs GS, Anjar Wanto, penelitian ini Penelitian ini mengimplementasikan algoritma *K-Means* untuk mengklasifikasikan laju pertumbuhan Produk Domestik Regional Bruto di Kota Surabaya yang terdiri dari 3 cluster [8]. Berdasarkan penelitian sebelumnya, menunjukan jika metode *K-Means* merupakan metode yang sangat baik untuk Data Mining.

Penelitian ini menganalisis tingkat penyelesaian pendidikan di jenjang SD, SMP, dan SMA di 34 provinsi Indonesia selama periode 2020-2022. Tujuannya adalah untuk memahami secara mendalam tingkat penyelesaian pendidikan di jenjang-jenjang tersebut di provinsi-provinsi Indonesia. Melalui analisis data penyelesaian pendidikan selama periode tersebut, peneliti dapat mengidentifikasi pola dan tren dalam tingkat penyelesaian pendidikan di setiap jenjang dan provinsi.

Penelitian ini juga memberikan kontribusi penting dalam rangka memperbaiki standar pendidikan di Indonesia. Dengan mengetahui tingkat penyelesaian pendidikan di setiap provinsi, pemerintah dan pemangku kepentingan lainnya dapat mengidentifikasi daerah yang membutuhkan perhatian khusus dan mengalokasikan sumber daya yang tepat guna untuk meningkatkan tingkat penyelesaian pendidikan di daerah tersebut. Selain itu, Penelitian ini memberikan pemahaman yang lebih mendalam tentang faktor-faktor yang berpengaruh terhadap tingkat keberhasilan dalam menyelesaikan pendidikan, seperti tingkat siswa tinggal kelas, putus sekolah, dan faktor ekonomi. Selanjutnya, penelitian ini diharapkan dapat memberikan rekomendasi kebijakan yang relevan untuk meningkatkan tingkat penyelesaian pendidikan di provinsi-provinsi Indonesia. Dengan pemahaman yang lebih mendalam dan komprehensif tentang faktor-faktor yang memengaruhi penyelesaian pendidikan, pemerintah dapat merancang kebijakan yang lebih efektif dalam meningkatkan akses, kualitas, dan kesetaraan pendidikan di seluruh provinsi. Penelitian ini memiliki dampak signifikan dalam pembangunan pendidikan dan pembentukan kebijakan pendidikan di Indonesia. Dengan pemahaman yang lebih baik tentang tingkat penyelesaian pendidikan dan faktor-faktor yang memengaruhi, langkah-langkah strategis dapat diambil untuk memperbaiki sistem pendidikan, mengurangi kesenjangan antar wilayah, dan meningkatkan standar mutu sumber daya manusia.

Dalam rangka menyelesaikan tujuan penelitian ini, penulis akan menggunakan sumber data resmi dari Badan Pusat Statistik Indonesia dan Susenas. Selain itu,

penulis juga akan mengacu pada literatur terkait dan sumber informasi lainnya yang relevan untuk memperkuat argumentasi dan memberikan dasar teoretis yang kuat bagi penelitian ini. Penelitian ini juga akan mempertimbangkan batasan dan kendala yang mungkin terjadi dalam pengumpulan dan analisis data, serta melakukan validasi hasil penelitian melalui metode yang relevan.

## 2. Metode Penelitian

### 2.1 Tahapan Penelitian

Penelitian ini memiliki tahapan yang di mulai dari pengumpulan data, proses data, perhitungan *K-Means* dengan *RapidMiner*, lalu diakhiri hasil dari perhitungan *K-Means*.



Gambar 1. Tahapan Penelitian

Pada gambar 2. *K-Means* akan diproses dengan *RapidMiner* dan hasilnya akan mengclusterisifikasi tingkat penyelesaian pendidikan pada jenjang SD, SMP, dan SMA.

### 2.2 Pengumpulan Data

Sumber data penelitian ini di peroleh dari Badan Pusat Statistik (BPS) Indonesia, selama 3 tahun dari 2020-2022. Dalam penelitian ini data yang digunakan adalah data tingkat penyelesaian pendidikan SD, SMP, dan SMA pada tahun 2020-2022.

Tabel 1. Tingkat penyelesaian pendidikan menurut jenjang dan provinsi pada jenjang SD

No.	Provinsi	SD / Sederajat		
		2020	2021	2022
1	ACEH	98.33	99.44	99.45
2	SUMATERA UTARA	96.48	98.57	98.74
3	SUMATERA BARAT	93.26	95.29	97.87
4	RIAU	95.18	96.91	98.20
5	JAMBI	95.34	98.54	97.62

6	SUMATERA SELATAN	95.93	97.82	97.53
7	BENGKULU	95.16	98.16	97.65
8	LAMPUNG	97.69	98.33	98.41
9	KEP. BANGKA BELITUNG	96.05	96.61	96.45
10	KEP. RIAU	95.25	98.16	98.38
11	DKI JAKARTA	97.80	99.26	98.58
12	JAWA BARAT	97.76	98.45	99.08
13	JAWA TENGAH	97.20	98.06	98.01
14	DI YOGYAKARTA	96.43	98.48	98.91
15	JAWA TIMUR	96.81	97.76	98.71
16	BANTEN	96.03	98.82	97.24
17	BALI	96.66	97.02	97.55
18	NUSA TENGGARA BARAT	97.29	98.71	98.47
19	NUSA TENGGARA TIMUR	86.92	91.84	92.35
20	KALIMANTAN BARAT	91.79	94.29	95.13
21	KALIMANTAN TENGAH	95.45	97.45	98.51
22	KALIMANTAN SELATAN	94.10	95.67	94.68
23	KALIMANTAN TIMUR	95.84	96.82	99.18
24	KALIMANTAN UTARA	96.38	95.77	96.94
25	SULAWESI UTARA	95.58	96.10	96.74
26	SULAWESI TENGAH	95.05	96.19	97.19
27	SULAWESI SELATAN	96.13	97.30	98.05
28	SULAWESI TENGGARA	95.28	95.58	97.24
29	GORONTALO	91.90	93.44	95.12
30	SULAWESI BARAT	94.70	95.93	97.15
31	MALUKU	97.47	98.50	98.98
32	MALUKU UTARA	95.59	96.97	97.72
33	PAPUA BARAT	89.25	91.81	93.94
34	PAPUA	78.40	78.43	81.99

Tabel 2. Tingkat penyelesaian pendidikan menurut jenjang dan provinsi pada jenjang SMP

No.	Provinsi	SMP / Sederajat		
		2020	2021	2022
1	ACEH	90.92	93.43	97.63
2	SUMATERA UTARA	91.63	91.35	92.84
3	SUMATERA BARAT	87.12	89.49	88.83
4	RIAU	86.84	87.11	88.53
5	JAMBI	86.31	89.00	86.57
6	SUMATERA SELATAN	85.02	87.68	88.41
7	BENGKULU	87.68	89.94	90.81

8	LAMPUNG	86.66	89.46	90.99	9	KEP. BANGKA BELITUNG	56.74	63.98	66.87
9	KEP. BANGKA BELITUNG	81.91	80.99	84.72	10	KEP. RIAU	78.65	81.07	73.93
10	KEP. RIAU	94.80	92.71	95.72	11	DKI JAKARTA	85.67	84.98	87.71
11	DKI JAKARTA	94.57	95.00	95.40	12	JAWA BARAT	63.56	64.89	67.05
12	JAWA BARAT	88.17	88.18	89.29	13	JAWA TENGAH	55.82	59.90	58.75
13	JAWA TENGAH	89.00	88.44	90.02	14	DI YOGYAKARTA	87.99	90.12	87.92
14	DI YOGYAKARTA	92.36	94.94	97.06	15	JAWA TIMUR	63.53	66.33	66.87
15	JAWA TIMUR	89.98	90.30	90.47	16	BANTEN	64.24	66.90	66.02
16	BANTEN	86.70	90.63	92.65	17	BALI	74.88	75.86	76.59
17	BALI	93.21	94.26	94.15	18	NUSA TENGGARA BARAT	64.66	65.71	61.00
18	NUSA TENGGARA BARAT	89.91	92.19	95.39	19	NUSA TENGGARA TIMUR	50.65	44.88	38.47
19	NUSA TENGGARA TIMUR	78.81	78.83	83.25	20	KALIMANTAN BARAT	55.23	54.27	58.40
20	KALIMANTAN BARAT	80.88	79.65	81.82	21	KALIMANTAN TENGAH	60.77	61.04	61.88
21	KALIMANTAN TENGAH	84.83	89.76	87.79	22	KALIMANTAN SELATAN	63.05	63.59	67.81
22	KALIMANTAN SELATAN	82.48	84.06	87.95	23	KALIMANTAN TIMUR	71.63	74.26	74.00
23	KALIMANTAN TIMUR	92.40	95.34	95.32	24	KALIMANTAN UTARA	67.77	62.30	54.80
24	KALIMANTAN UTARA	87.14	90.14	90.55	25	SULAWESI UTARA	73.79	68.56	66.66
25	SULAWESI UTARA	86.69	91.05	91.98	26	SULAWESI TENGAH	57.68	61.16	53.73
26	SULAWESI TENGAH	82.79	85.42	88.90	27	SULAWESI SELATAN	66.22	69.43	68.32
27	SULAWESI SELATAN	85.08	88.18	90.55	28	SULAWESI TENGGARA	68.28	70.65	65.97
28	SULAWESI TENGGARA	89.18	90.88	91.19	29	GORONTALO	55.35	53.73	45.12
29	GORONTALO	80.10	81.22	80.56	30	SULAWESI BARAT	56.60	56.22	55.18
30	SULAWESI BARAT	83.08	86.09	84.14	31	MALUKU	70.55	68.12	72.08
31	MALUKU	88.04	93.08	93.65	32	MALUKU UTARA	66.52	66.95	67.10
32	MALUKU UTARA	87.41	92.93	94.92	33	PAPUA BARAT	61.49	59.08	57.07
33	PAPUA BARAT	83.47	85.18	87.03	34	PAPUA	30.92	32.95	39.01
34	PAPUA	65.75	66.06	66.16					

Tabel 3. Tingkat penyelesaian pendidikan menurut jenjang dan provinsi pada jenjang SMA

No.	Provinsi	SMA / Sederajat		
		2020	2021	2022
1	ACEH	70.07	74.36	70.67
2	SUMATERA UTARA	70.39	72.81	77.16
3	SUMATERA BARAT	67.11	70.06	65.96
4	RIAU	66.62	68.94	66.91
5	JAMBI	63.66	64.51	65.85
6	SUMATERA SELATAN	65.42	67.20	67.07
7	BENGKULU	62.73	62.46	64.88
8	LAMPUNG	57.59	60.09	62.42

Tabel 1-3 menunjukan data yang didapat dari BPS berupa tingkat penyelesaian pendidikan, atribut yang dipakai adalah 34 Provinsi Indonesia.

### 2.3 Data Mining

Data mining atau penambangan data adalah serangkaian proses yang digunakan untuk mengekstraksi, menganalisis, dan menggali informasi berharga dari sejumlah besar data. Tujuan utamanya adalah untuk mengidentifikasi pola, tren, atau hubungan yang tersembunyi atau tidak terdeteksi sebelumnya [9]–[11].

Data mining atau penambangan data merupakan konseptualisasi yang sering digunakan secara bergantian dengan penggalian informasi dari lapisan dalam basis data yang melimpah. Ia adalah sebuah prosesi yang melibatkan pengumpulan data dalam proporsi monumental, yang kemudian ditransformasikan menjadi inti pengetahuan yang berharga [9], [12].

Data mining mengungkap semangat penemuan epistemik baru yang muncul dari kelompok data besar, yang berperan sebagai penuntun dalam pengambilan keputusan yang bijak. Ia menentang pendekatan mendasar dalam menggali inti data dan mengekstraksi informasi bernilai yang terperangkap dalam kegelapan, melalui penggunaan teknik dan instrumen analitis yang beragam untuk menemukan keterkaitan dan pola-pola yang terselubung [9], [13].

#### 2.4 RapidMiner

*RapidMiner* merupakan sebuah perangkat lunak yang secara khusus diciptakan untuk memfasilitasi pengguna dalam mengoperasikan software ini dengan lebih mudah. Aplikasi ini menawarkan lebih dari 500 operator yang beragam, termasuk operator untuk input, output, pemrosesan data, dan visualisasi. [9], [14].

*RapidMiner* menggunakan berbagai teknik deskriptif dan prediktif. Software ini ditulis dalam bahasa pemrograman Java, sehingga dapat berjalan secara baik di berbagai sistem operasi [9], [14]. *RapidMiner* juga memungkinkan hasil analisis ditampilkan secara visual melalui grafik, membuatnya menjadi salah satu pilihan utama dalam ekstraksi data menggunakan metode-metode data mining [9], [15].

#### 2.5 K-Means

*K-Means* adalah sebuah metode klustering yang tergolong dalam pembelajaran tanpa pengawasan, yang aplikasikan untuk mengelompokkan data ke dalam beberapa kelompok melalui proses partisi. Dengan algoritma ini, data dapat diolah tanpa memerlukan label kelas. Inilah perbedaan *K-Means* dengan metode *K-Nearest Neighbor (KNN)*, *Hierarchical Clustering* serta algoritma pembelajaran unsupervised learning lainnya yang membutuhkan vektor sebagai masukan [16]. Algoritma ini akan mengklasifikasikan data atau objek ke dalam kelompok yang telah ditentukan. Setiap kelompok akan memiliki titik pusat yang disebut *centroid* yang merepresentasikan kelompok tersebut. Secara sederhana, algoritma *K-Means* adalah algoritma data mining yang digunakan untuk memecahkan masalah pengelompokan (*clustering*) [17], [18].

Dalam proses pengolahan data menggunakan algoritma *K-Means Clustering*, langkah awalnya adalah menentukan *centroid* awal secara acak untuk setiap kelompok, kemudian dilakukan perhitungan berulang agar posisi *centroid* optimal [16], [18]. *K-Means* akan diproses dengan *RapidMiner* dengan perhitungan *Euclidean Distance*.

### 3. Hasil dan Pembahasan

#### 3.1 Dataset

Penelitian ini menggunakan dataset yang berisi data tingkat penyelesaian pendidikan menurut jenjang pendidikan dan provinsi. Dataset ini termuat di Badan

Pusat Statistik (BPS) Indonesia, yang merupakan sumber yang terpercaya dan otoritatif dalam hal informasi statistik di negara ini dan akan digunakan untuk melakukan analisis perjenjang pendidikan. Dataset tersebut terstruktur dalam bentuk tabel, yang kemudian dibagi menjadi tiga kategori berdasarkan jenjang pendidikan, yaitu SD, SMP, dan SMA.

Metode yang akan digunakan dalam penelitian ini adalah *clustering* dengan menggunakan metode *K-Means*. *Clustering* akan dilakukan dengan menggunakan teknik *Random Clustering*, di mana *centroid* awal untuk setiap cluster ditentukan secara acak. Penelitian ini akan membagi data menjadi tiga cluster berdasarkan karakteristik penyelesaian pendidikan.

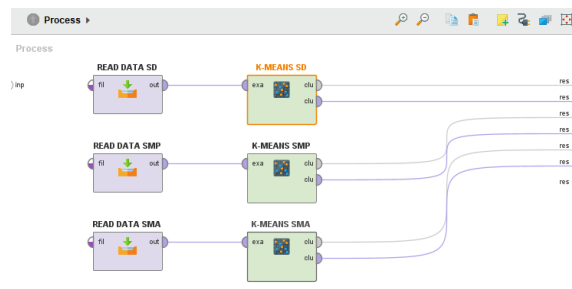
Untuk melakukan analisis dan pengujian, alat yang digunakan adalah *RapidMiner*. *RapidMiner* adalah perangkat lunak yang memudahkan analisis data dan penerapan algoritma data mining. Dengan menggunakan *RapidMiner*, peneliti dapat melakukan pengolahan data, pengujian algoritma *clustering*, dan visualisasi hasil secara efisien.

Dengan menggunakan metode *K-Means* dan alat *RapidMiner*, penelitian ini bertujuan untuk mengidentifikasi pola atau kelompok dalam data tingkat penyelesaian pendidikan berdasarkan jenjang dan provinsi. Hasil dari penelitian ini diharapkan dapat memberikan pemahaman yang lebih dalam tentang karakteristik penyelesaian pendidikan di setiap jenjang dan provinsi yang diteliti.

Dalam penelitian ini kluster akan membagi nilai kluster,  $K=0$  adalah kluster tinggi,  $K=1$  kluster sedang dan  $K=2$  kluster rendah. Kluster tinggi berarti provinsi tersebut memiliki tingkat penyelesaian pendidikan yang tinggi, kluster sedang memiliki nilai sedang dan kluster rendah memiliki nilai yang rendah.

#### 3.2 Penerapan algoritman *K-Means* pada *RapidMiner*

Attribute yang digunakan untuk mengolah data adalah jenjang pendidikan dan provinsi. Attribute class data yang digunakan adalah persentase penyelesaian pendidikan. Dataset di masukan ke excel dg format .CSV dan diimpor ke dalam aplikasi, Lalu memasukan data setelah itu memasukan operator data dan operator *K-Means* dalam jendela process setelah itu disambungkan Selanjutnya, operator data dan operator *K-Means* digunakan dalam jendela proses untuk menerapkan algoritma *K-Means* pada dataset. Pada operator *K-Means*, pengaturan  $K=3$  digunakan untuk membagi data menjadi tiga kluster yang berbeda. Selain itu, perhitungan jarak menggunakan *Euclidean Distance* dipilih untuk mengukur jarak antara data dan *centroid*. Hasilnya seperti gambar 3.



Gambar 2. Proses pada RapidMiner

Pada gambar 2. Dijelaskan jika Dataset dibagi jadi 3 kategori yaitu Data Sd, Data SMP, dan Data SMA. Operator *K-Means* sudah diatur untuk 3 Cluster dengan menggunakan Euclidean Distance.

#### a. Analisis pemetaan pada jenjang SD

Berikut adalah hasil proses *K-Means* pada data Tingkat Penyelesaian pendidikan jenjang SD dengan jumlah Kluster 3 menggunakan alat RapidMiner seperti pada gambar 2.

### Cluster Model

```
Cluster 0: 29 items
Cluster 1: 4 items
Cluster 2: 1 items
Total number of items: 34
```

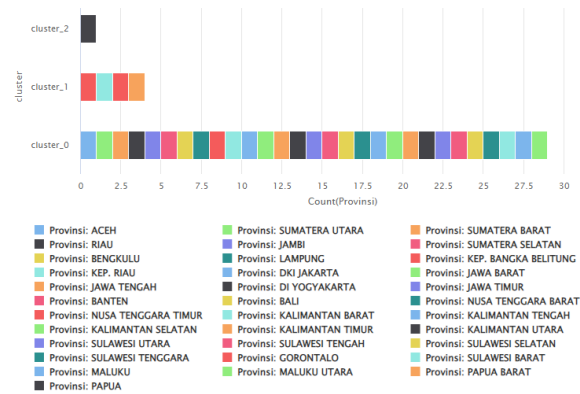
Gambar 3. Hasil Clustering jenjang SD

Pada gambar 3. Hasil output cluster terbagi jadi 3 kluster pada jenjang SD. Pada K=0 terdapat 29 provinsi diantaranya adalah Aceh, Lampung, DKI Jakarta, Banten serta Bali. Lalu pada K=1 ada 4 provinsi yaitu Nusa Tenggara Timur, Gorontalo, Kalimantan Barat, dan Papua Barat. Pada K=2 hanya ada satu yaitu Provinsi Papua.

Attribute	cluster_0	cluster_1	cluster_2
2020	96.077	89.965	78.400
2021	97.471	92.845	78.430
2022	97.904	94.135	81.990

Gambar 4. Nilai centroid pada jenjang SD

Pada gambar 4 menunjukkan penentuan nilai klaster tinggi, sedang, dan rendah berdasarkan pada nilai *centroid* dimana nilai *centroid* akhir dari pemetaan jenjang SD yakni K=0 {96.077 | 97.471 | 97.904}, K=1 {89.965 | 92.845 | 94.135}, dan K=2 {78.400 | 78.430 | 81.990}.



Gambar 5. Visualisasi Clustering pada jenjang SD

Pada gambar 5 merupakan visualisasi pemetaan jenjang SD dengan chart Bar, hasil visualisasi menunjukkan 3 chart cluster dengan pembagian warna setiap provinsi.

#### b. Analisis pemetaan pada jenjang SMP

Berikut adalah hasil proses *K-Means* pada data Tingkat Penyelesaian pendidikan jenjang SMP dengan jumlah Kluster 3 menggunakan alat RapidMiner seperti pada gambar 6.

### Cluster Model

```
Cluster 0: 20 items
Cluster 1: 13 items
Cluster 2: 1 items
Total number of items: 34
```

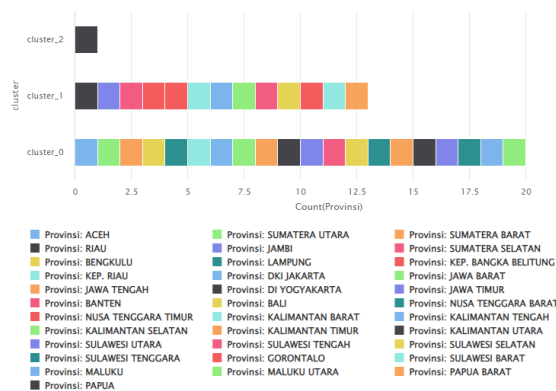
Gambar 6. Hasil Clustering jenjang SMP

Pada gambar 6 hasil kluster pada provinsi Indonesia jenjang SMP terbagi jadi 3. K=0 terdiri dari 20 provinsi yaitu Aceh, Bengkulu, dan Lampung. Pada cluster K=1 terdapat 13 provinsi diantaranya Sumatra Selatan, Jambi dan Kepulauan Riau. Pada cluster K=2 terbagi jadi 1 yaitu Papua.

Attribute	cluster_0	cluster_1	cluster_2
2020	89.679	83.200	65.750
2021	91.687	84.859	66.060
2022	92.943	86.171	66.160

Gambar 7. Nilai centroid pada jenjang SMP

Pada gambar 7 menunjukkan penentuan nilai klaster tinggi, sedang, dan rendah berdasarkan pada nilai *centroid* dimana nilai *centroid* akhir dari pemetaan jenjang SMP yakni K=0 {89.679 | 91.687 | 92.943}, K=1 {83.200 | 84.859 | 85.171}, dan K=2 {65.750 | 66.060 | 66.160}.



Gambar 8. Visualisasi clustering pada jenjang SMP

Pada gambar 8 merupakan visualisasi pemetaan jenjang SMP dengan chart Bar, hasil visualisasi menunjukkan 3 chart cluster dengan pembagian warna setiap provinsi.

### c. Analisis pemetaan pada jenjang SMA

Berikut adalah hasil proses *K-Means* pada data Tingkat Penyelesaian pendidikan jenjang SMA dengan jumlah Kluster 3 menggunakan alat RapidMiner seperti pada gambar 8.

## Cluster Model

```
Cluster 0: 7 items
Cluster 1: 24 items
Cluster 2: 3 items
Total number of items: 34
```

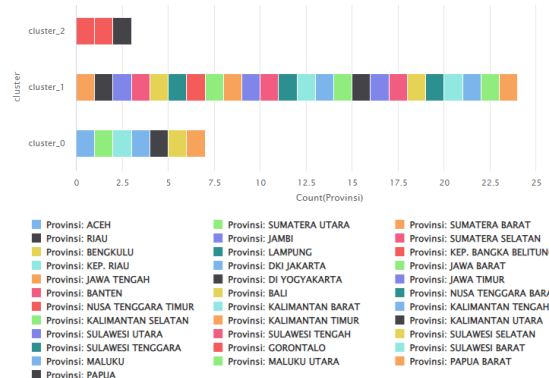
Gambar 9. Hasil clustering jenjang SMA

Pada gambar 9 hasil kluster pada provinsi Indonesia jenjang SMA terbagi jadi 3. K=0 terdiri dari 7 provinsi yaitu Aceh, Bali, dan DKI Jakarta. Pada cluster K=1 terdapat 24 provinsi diantaranya Riau, Jambi, dan Bengkulu. Pada cluster K=2 terbagi jadi 3 diantaranya Nusa Tenggara Timur, Gorontalo, dan Papua.

Attribute	cluster_0	cluster_1	cluster_2
2020	77.040	63.318	45.640
2021	79.066	64.264	43.853
2022	78.283	63.694	40.867

Gambar 10. Nilai centroid pada jenjang SMA

Pada gambar 10 menunjukkan penentuan nilai klaster tinggi, sedang, dan rendah berdasarkan pada nilai *centroid* dimana nilai *centroid* akhir dari pemetaan jenjang SMA yakni K=0 {77.040 | 79.066 | 78.263}, K=1 {63.318 | 64.264 | 63.694}, dan K=2 {45.640 | 43.853 | 40.867}.



Gambar 11. Visualisasi clustering pada jenjang SMA

Pada gambar 9 merupakan visualisasi pemetaan jenjang SMA dengan chart Bar, hasil visualisasi menunjukkan 3 chart cluster dengan pembagian warna setiap provinsi.

### 3.3 Pembahasan

Pada penelitian sebelumnya ada yang menggunakan 2 metode untuk *clustering* data, ada yang menggunakan Davies Bouldin Index untuk mencari berapa cluster terbaik dan ada yang menggunakan perhitungan manual untuk mencari *centroid*. Pada penelitian ini hanya menggunakan metode *K-Means* saja dengan random *clustering*. Penelitian ini juga menggunakan 3 dataset yang berbeda.

## 4. Kesimpulan

Berdasarkan hasil penelitian ini, terdapat beberapa kesimpulan yang dapat diambil, hasil penelitian menunjukkan bahwa tingkat penyelesaian pendidikan menurun secara signifikan setiap naik ke jenjang pendidikan yang lebih tinggi, terutama pada jenjang SMA. Hal ini mengindikasikan adanya tantangan yang perlu ditangani dalam meningkatkan tingkat penyelesaian pendidikan pada jenjang SMA. Lalu pada provinsi bagian Indonesia Timur seperti provinsi Papua, Gorontalo, dan Nusa Tenggara Timur perlu mendapatkan perhatian lebih dalam upaya meningkatkan tingkat penyelesaian pendidikan. Data menunjukkan bahwa provinsi-provinsi ini memiliki tingkat penyelesaian pendidikan yang relatif rendah, sehingga perlu dilakukan upaya khusus untuk mengatasi permasalahan pendidikan di daerah-daerah tersebut. Kesimpulan ini menunjukkan pentingnya perhatian dan tindakan yang lebih intensif dalam meningkatkan tingkat penyelesaian pendidikan di provinsi-provinsi yang teridentifikasi sebagai area dengan tingkat penyelesaian pendidikan yang rendah. Dengan memperbaiki pengaturan kluster, mengatasi tantangan pada jenjang SMA, dan memberikan perhatian khusus pada provinsi-provinsi yang membutuhkan, diharapkan dapat meningkatkan kesempatan pendidikan dan taraf hidup masyarakat di Indonesia secara keseluruhan.



## Daftar Rujukan

- [1] Badan Pusat Statistik, “Tingkat Penyelesaian Pendidikan Menurut Jenjang Pendidikan dan Jenis Kelamin,” 2017. <https://www.bps.go.id/linkTableDinamis/view/id/960>.
- [2] N. T. Luchia and Mustakim, “Perbandingan Algoritma K-Means dan K-Medoids pada Pengelompokan Humidity, Temperature, dan Voltage di Data Center Perawang,” *J. Inf. Syst. Res.*, vol. 4, no. 1, pp. 184–190, 2022, doi: 10.47065/josh.v4i1.2385.
- [3] V. AGUSTIN, “Perbandingan Analisis Clustering K-Means Dan K-Medoids Pada Data Penyakit Di Indonesia Tahun 2019,” *Repository.Unri.Ac.Id*, no. Icmi, pp. 1–45, 2021, [Online]. Available: <https://repository.unri.ac.id/handle/123456789/10436>
- [4] T. Tendean and W. Purba, “Analisis Cluster Provinsi Indonesia Berdasarkan Produksi Bahan Pangan Menggunakan Algoritma K-Means,” *Saintek (Jurnal Sains dan Teknol.*, vol. 1, no. 2, pp. 5–11, 2020.
- [5] S. Wijayanto and Y. Fathoni, M., “Pengelompokan Produktivitas Tanaman Padi di Jawa Tengah Menggunakan Metode Clustering K-Means,” *J. JUPITER*, vol. 13, no. 2, pp. 212–219, 2021.
- [6] A. Aziz, A. Siregar, and C. Zonyfar, “Penerapan Algoritma K-Means dan Fuzzy C-Means untuk Pengelompokan Kabupaten Kota Berdasarkan Produksi Padi di Provinsi Jawa Barat,” ... *Student J.* ..., vol. III, pp. 1–8, 2022, [Online]. Available: <https://journal.ubpkarawang.ac.id/mahasiswa/index.php/ssj/article/view/411>
- [7] A. Jaelanie, “Analisis Klaster Kabupaten/Kota Berdasarkan Indikator Produk Domestik Regional Bruto (PDRB) Menggunakan Metode Unsupervised Learning-Cluster Analysis Of ...,” 2016, [Online]. Available: <https://repository.its.ac.id/72395/>
- [8] N. A. Febriyati, A. Daengs Gs, and A. Wanto, “GRDP Growth Rate Clustering in Surabaya City uses the K-Means Algorithm,” *Int. J. Inf. Syst. Technol. Akreditasi*, vol. 3, no. 36, pp. 276–283, 2020.
- [9] Aprilla Dennis, *Belajar Data Mining dengan RapidMiner*, vol. 5, no. 4, 2013. [Online]. Available: [http://esjournals.org/journaloftechnology/archive/vol1no6/vol1no6\\_6.pdf%5Cnhttp://www.airccse.org/journal/nsa/5413nsa02.pdf](http://esjournals.org/journaloftechnology/archive/vol1no6/vol1no6_6.pdf%5Cnhttp://www.airccse.org/journal/nsa/5413nsa02.pdf)
- [10] J. Sumarah and A. T. Wulandari, “Pemanfaatan Algoritma K-Means untuk Pengelompokan Angka Partisipasi Sekolah di Jawa Tengah,” *J. Media Inform. Budidarma*, vol. 5, no. 4, p. 1507, 2021, doi: 10.30865/mib.v5i4.3277.
- [11] J. Ha, M. Kambe, and J. Pe, *Data Mining: Concepts and Techniques*. 2011. doi: 10.1016/C2009-0-61819-5.
- [12] G. Aprilianur and E. L. Hadisaputro, “Penerapan Data Mining Menggunakan Metode K-Means Clustering Untuk Analisa Penjualan Toko Myam Hijab Penajam,” *J. JUPITER*, vol. 14, no. 1, pp. 161–170, 2022.
- [13] K. Rahayu, L. Novianti, and M. Kusnandar, “Implementation Data Mining with K-Means Algorithm for Clustering Distribution Rabies Case Area in Palembang City,” *J. Phys. Conf. Ser.*, vol. 1500, no. 1, 2020, doi: 10.1088/1742-6596/1500/1/012121.
- [14] s wuni, a m Siregar, and ..., “K-Means Clustering Untuk Mengelompokan Tingkat Putus Sekolah Jenjang Smp Di Indonesia,” *Sci. Student J. Information, Technol. Sci.*, vol. 1, pp. 47–54, 2020, [Online]. Available: <http://journal.ubpkarawang.ac.id/mahasiswa/index.php/ssj/article/download/9/8>
- [15] M. D. Chandra, E. Irawan, I. S. Saragih, A. P. Windarto, and D. Suhendro, “Penerapan Algoritma K-Means dalam Mengelompokkan Balita yang Mengalami Gizi Buruk Menurut Provinsi,” *BIOS J. Teknol. Inf. dan Rekayasa Komput.*, vol. 2, no. 1, pp. 30–38, 2021, doi: 10.37148/bios.v2i1.19.
- [16] F. Simanjorang, R. Winanjaya, and F. Rizki, “Penerapan Algoritma K-Means Dalam Pengelompokan Rasio Angka Partisipasi Kasar di Tingkat Pendidikan Perguruan Tinggi Menurut Provinsi,” *TIN Terap. Inform. Nusantara*, vol. 2, no. 7, pp. 454–459, 2021, [Online]. Available: <http://ejurnal.seminar-id.com/index.php/tin/article/view/997>
- [17] N. Purba, P. Poningsih, and H. S. Tambunan, “Penerapan Algoritma K-Means Clustering Pada Penyebaran Penyakit Infeksi Saluran Pernapasan Akut (ISPA) di Provinsi Riau,” *J. Inf. Syst. Res.*, vol. 2, no. 3, pp. 220–226, 2021, [Online]. Available: <http://ejurnal.seminar-id.com/index.php/josh/article/view/736>
- [18] N. Nurahman, A. Purwanto, and S. Mulyanto, “Klasterisasi Sekolah Menggunakan Algoritma K-Means berdasarkan Fasilitas, Pendidik, dan Tenaga Pendidik,” *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 21, no. 2, pp. 337–350, 2022, doi: 10.30812/matrik.v21i2.1411.