

## **ANALISIS BIG DATA APS SLTA DAN STRATEGI PENDIDIKAN MENGGUNAKAN K-MEANS BERBASIS RAPIDMINER MENUJU INDONESIA EMAS**

**Khoirun Nisa<sup>✉</sup>, Nor Anisa**

Fakultas Sains dan Teknologi, Universitas Sari Mulia, Banjarmasin, Indonesia

Email: [khoirun.nisa@student.unism.ac.id](mailto:khoirun.nisa@student.unism.ac.id)

### **ABSTRACT**

*The Golden Indonesia Vision 2045 places education as the main pillar in creating superior human resources. School Participation Rate (APS) data is an important indicator to evaluate student access and participation in education. This study utilizes the K-Means Clustering method to analyze APS big data to identify patterns of education participation in Indonesia. The results of the analysis show significant participation clusters based on demographic, socio-economic, and geographical factors, and reveal gaps and potential for education improvement in various regions. In this study, RapidMiner is used as an analysis tool to process and visualize APS data. The results of clustering show a striking difference between areas with good access to education and areas with poor access to education. Factors such as income levels, educational infrastructure, and geographical location were discovered to significantly affect student participation rates. Strategic recommendations include increasing access to education in disadvantaged areas through equitable distribution of education facilities, infrastructure development, and flexible data-based policies. In addition, scholarship programs in vulnerable areas are also proposed as a solution. This research supports strategic efforts towards the vision of Golden Indonesia 2045 by providing a strong foundation for policies that focus on the sustainability of national education.*

**Keyword:** *Big Data, School Participation Rate, K-Means Clustering, Education Strategy,*

*Golden Indonesia 2045.*

### **ABSTRAK**

*Visi Indonesia Emas 2045 menempatkan pendidikan sebagai pilar utama dalam menciptakan sumber daya manusia unggul. Data Angka Partisipasi Sekolah (APS) menjadi indikator penting untuk mengevaluasi akses dan partisipasi siswa dalam pendidikan. Penelitian ini memanfaatkan metode K-Means Clustering untuk menganalisis big data APS guna mengidentifikasi pola partisipasi pendidikan di Indonesia. Hasil analisis menunjukkan klaster partisipasi yang signifikan berdasarkan faktor demografi, sosial-ekonomi, dan geografis, serta mengungkap kesenjangan dan potensi peningkatan pendidikan di berbagai wilayah. Dalam penelitian ini, RapidMiner digunakan sebagai alat analisis untuk mengolah dan memvisualisasikan data APS. Hasil clusterisasi menunjukkan perbedaan mencolok antara wilayah dengan akses Pendidikan yang baik dan wilayah yang tertinggal. Faktor seperti Tingkat pendapatan, infrastruktur pendidikan, dan lokasi geografis ditemukan memiliki dampak besar terhadap tingkat partisipasi siswa. Rekomendasi strategis mencangkup peningkatan akses Pendidikan di wilayah tertinggal melalui pemerataan fasilitas Pendidikan, Pembangunan infrastruktur, dan kebijakan berbasis data yang fleksibel. Selain itu, program beasiswa di daerah rentan juga di usulkan sebagai Solusi. Penelitian ini mendukung Upaya strategis menuju visi Indonesia Emas 2045 dengan memberikan dasar kuat bagi kebijakan yang berfokus pada keberlanjutan Pendidikan nasional.*

**Kata Kunci:** *Big Data, Angka Partisipasi Sekolah, K-Means Clustering, Strategi Pendidikan,*

*Indonesia Emas 2045.*

### **PENDAHULUAN**

Indonesia ditengah ambisi untuk mewujudkan visi Indonesia emas 2045. Pilar utama untuk mencapai visi tersebut adalah peningkatan kualitas Pendidikan di Indonesia. Di Era digital ini, data menjadi asset berharga yang dapat dimanfaatkan untuk meningkatkan efektivitas dan efisiensi dari berbagai sektor, salah satunya pendidikan.

Di era teknologi informasi saat ini, tidak diragukan lagi ada begitu banyak informasi dari segala bidang, dan dunia pendidikan adalah salah satu sumber, dimana sejumlah besar data terus-menerus dihasilkan (Solihin & Wibisono, 2017). Salah satu sumber data yang signifikan dalam dunia pendidikan adalah Angka Partisipasi Sekolah (APS). Data APS menyediakan informasi penting mengenai capaian pembelajaran siswa, termasuk tingkat akses, keikutsertaan, dan

keberlanjutan siswa dalam proses pendidikan formal. Namun, pengelolaan dan analisis data ini masih menghadapi berbagai tantangan, seperti volume data yang besar, pola yang kompleks, serta keterbatasan pemanfaatan teknologi analitik modern.

Teknik pengelompokan K-Means menjadi sebuah Solusi utama guna diterapkan dalam menganalisis big data APS. Dengan teknik ini, data APS dapat dikelompokkan berdasarkan karakteristik tertentu sehingga menghasilkan wawasan yang mendalam. Wawasan ini tidak hanya membantu memahami pola-pola partisipasi pendidikan, tetapi juga bisa dipakai dalam merumuskan pendekatan Pendidikan yang lebih efektif dan inklusif. Seperti studi kasus yang serupa menunjukkan bahwa penggunaan k-means clustering dapat membantu dalam menentukan daerah dengan akses pendidikan yang rendah dan membuat intervensi kebijakan yang lebih tepat sasaran (Hanifah & Primandari, 2023). Pemilihan Teknik *K-Means Clustering* dalam penelitian ini didasarkan pada kemampuannya mengelompokkan data siswa menggunakan kriteria tertentu yang dapat menjadi panduan untuk memahami bagaimana Tingkat partisipasi siswa dalam minat Pendidikan terutama di Tingkat SLTA dari tahun ke tahun di berbagai wilayah Kalimantan selatan.

*K-Means Clustering* dapat menggabungkan data dengan maksimalkan kesamaan antar kelompok dan mengurangi kesamaan antar kelompok, pengukuran kesamaan cluster didasarkan pada jangka terkecil antara data dalam cluster dan titik pusat (Nur et al., 2018). Dengan Metode *k-means Clustering*, Informasi yang sudah dikumpulkan dapat dikelompokkan ke dalam beberapa cluster yang didasarkan pada Tingkat kemiripannya. Sehingga data yang mempunyai kesamaan atribut akan dikumpulkan dalam satu kelompok dan data dengan atribut beda akan digolongkan ke dalam cluster lain dengan atribut yang sama (Dewi et al., 2022).

Penelitian ini bertujuan untuk memanfaatkan analisis big data berbasis K-Means Clustering guna memperbaiki mutu Pendidikan di Indonesia, terkhusus di daerah Kalimantan Selatan. Melalui analisis data dari badan pusat statistic Kalimantan Selatan (BPS Kalsel) mengenai angka partisipasi sekolah SLTA di kalimantan selatan, penelitian ini berupaya memberikan kontribusi nyata dalam merancang strategi pendidikan yang mendukung pemerataan akses dan kualitas pendidikan. Dengan demikian, langkah-langkah ini diharapkan dapat menjadi bagian dari upaya strategis menuju Indonesia Emas 2045.

Adapun urgensi penelitian ini terletak pada potensi pemanfaatan teknologi analitik modern untuk

menjawab tantangan pendidikan yang kompleks di Indonesia. Dengan mengintegrasikan pendekatan berbasis big data dan strategi pendidikan yang adaptif, Studi ini diharapkan dapat memberikan dampak yang signifikan dan positif terhadap kemajuan pendidikan nasional.

## **TINJAUAN PUSTAKA**

### **Angka Partisipasi Sekolah (APS)**

APS adalah indikator penting yang harus terus ditingkatkan untuk mencapai kualitas manusia yang lebih baik dan berpengertahuan luas. Angka Partisipasi Sekolah (APS) menunjukkan tingginya Tingkat penerimaan Pendidikan formal pada empat jenjang Pendidikan, yang dibagi berdasarkan kelompok usia dan presentasi populasi tertentu yang sedang bersekolah, sehingga memberikan Gambaran sejauh mana Pendidikan dapat di akses oleh Masyarakat di suatu wilayah. Pada sekolah dasar mayoritas siswa berumur 7 - 17 tahun, siswa, Sekolah Menengah Pertama Sebagian besar siswa berumur 13-15 tahun dan Sekolah Menengah Keatas Sebagian besar siswa berusia 16-18 tahun (Virdam et al., 2023).

### **Big Data**

Di era digital saat ini, big data telah menjadi asset berharga yang mempengaruhi berbagai sektor, salah satunya Pendidikan. Big data dalam konteks Pendidikan mencangkup Kumpulan data yang besar, cepat berkembang dan beragam, seperti data kehadiran siswa, data penduduk sampai angka partisipasi sekolah (APS).

Dari Pemanfaatan big data, dengan mudah dapat memahami pola dan tren Pendidikan diberbagai wilayah, sehingga bisa mengidentifikasi kesenjangan dan membuat strategi yang lebih efektif.

### **K-Means Clustering**

Data *Clustering* adalah sebuah teknik data mining yang tidak di awasi. Dua jenis pengelompokan data yang paling umum dalam proses *Clustering* data adalah hirarkis dan non-hirarkis. Nilai K merupakan Metode clustering data non-hirarkis yang bertujuan membagi data saat ini ke satu atau lebih banyak cluster. Teknik ini menggabungkan data ke dalam cluster sebagai karakteristik yang mempunyai kesamaan dan kelompok dengan karakteristik yang berbeda (Agusta, 2007).

Tujuan clustering data adalah sebagai upaya meningkatkan keberagaman antar cluster serta menurunkan fungsi tujuan yang dikumpulkan melalui sebuah cluster (Asroni et al., 2018).

## RapidMiner

RapidMiner adalah perangkat lunak pengolahan data sekaligus platform analisis data yang memungkinkan pengguna untuk melakukan proses analitik tanpa harus menulis code secara manual. RapidMiner mengintegrasikan teknik statistic, kecerdasan buatan dan basis data untuk menemukan pola dalam Kumpulan dataset yang begitu banyak, dengan menerapkan prinsip dan algoritma penambangan data.

Dengan memanfaatkan perangkat RapidMiner, pengguna akan lebih mudah dalam mengerjakan perhitungan pada data dengan jumlah besar. Perangkat ini berfungsi untuk mengubah data (Fajriani & Pasrun, 2017).

## METODE PENELITIAN

### Identifikasi masalah

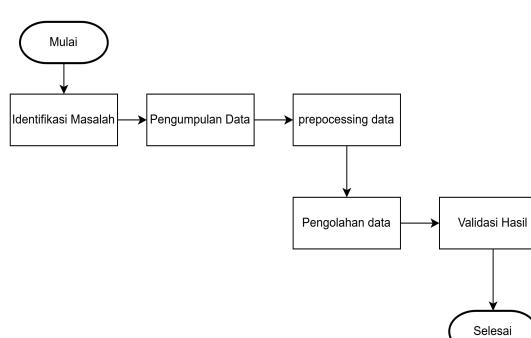
Penelitian ini di mulai dengan mengidentifikasi permasalahan rendahnya angka partisipasi sekolah Tingkat SLTA pada beberapa wilayah di Indonesia khususnya Kalimantan Selatan. Permasalahan ini mempengaruhi pemerataan akses Pendidikan, yang menjadi tantangan utama dalam mencapai visi Indonesia Emas 2045.

Tujuan Penelitian ini adalah untuk menemukan pola partisipasi Pendidikan dan faktor yang berpengaruh melalui pendekatan berbasis data menggunakan Teknik *k-means clustering*.

### Pengumpulan data

Studi ini memanfaatkan data APS Tingkat SLTA yang di ambil dari Badan Pusat Statistik (BPS) Kalimantan Selatan. Data ini mencangkup informasi terkait Tingkat partisipasi siswa dalam Pendidikan, kondisi sosial ekonomi, infrakstruktur pendidikan dan Lokasi geografis.

Data dikumpulkan dalam format excel dengan metode Power Query untuk mempermudah pengolahan menggunakan RapidMiner. Data yang di ambil mulai periode tahun 2011 sampai 2022.



Gambar 1. Flowchart Penelitian

Tabel 1. Dataset APS SLTA Kalimantan Selatan

Kabupaten	Tahun	Jenis Kelamin	Aps
Barito Kuala	2011	Total	53,61
Tapin	2011	Total	41,28
Banjarmasin	2011	Total	61,05
Banjar Baru	2011	Total	78,72
.....	.....	.....	.....
Hulu Sungai Selatan	2022	Total	69,97
Hulu Sungai Tengah	2022	Total	66,19
Hulu Sungai Utara	2022	Total	67,9
Tabalong	2022	Total	79,69

### Preprocessing data

Data Preprocessing terdiri dari beberapa langkah yaitu pengambilan data, pembersihan data, integrasi data dan modifikasi data. Preprocessing data adalah proses mengolah data mentah menjadi data yang memiliki kualitas. Tahapan ini dikerjakan untuk menyelesaikan masalah pada data mentah yang memiliki noise, atau komponen data tidak lengkap.

Data pre-processing ini dilakukan karena dalam data real pada sumber data tertentu sering kali data tidak lengkap dan tidak konsisten, sehingga dapat mengurangi hasil analisis (Wardhani et al., 2023).

Id kabupaten	Polynomial	0	Least TAPIN (36)	Most BALANGAN (36)
Cluster cluster	Nominal	0	Least cluster_0 (127)	Most cluster_1 (237)
Tahun	Integer	0	Min 2011	Max 2022
APS( Angka Partisipasi Sekolah )	Real	0	Min 28.710	Max 93.800

Gambar 2. Pengecekan Missing Value

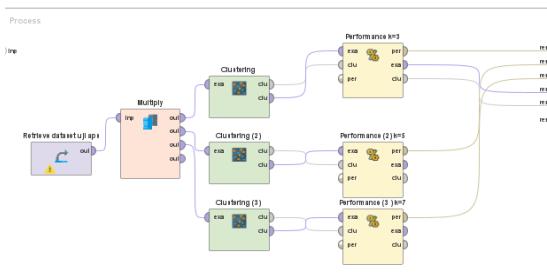
Dapat dilihat bahwa atribut-atribut ini tidak memiliki missing value dan proses bisa di lanjutkan ke tahap berikutnya.

### Pengolahan data

Data yang Selesai pada tahap Preprocessing akan lanjut diproses menggunakan algoritma *K-means Clustering* pada Repidminer. Langkah-langkahnya meliputi:

1. **Retrieve dataset** adalah Langkah awal mengambil dataset dari file excel yang sudah di sediakan dan akan digunakan untuk proses analisis. Dataset ini berisi data real dari Angka partisipasi sekolah (APS).
2. **Multiply Data** operator ini dilakukan untuk memungkinkan pengujian beberapa sekenario cluster yang berbeda secara pararel.

3. **Clustering** menggunakan 3 operator untuk menunjukkan penggunaan metode k-means dalam membentuk tiga cluster. Didalam setiap operator ini mewakili data yang diolah untuk menemukan pola berdasarkan atribut tertentu.
4. **Performance Evaluation** digunakan setelah memasukkan operator clustering dan proses selesai, untuk mengevaluasi hasil klasterisasi. Evaluasi ini mengukur kinerja model cluster berdasarkan metrik. Dipenelitian kali ini menggunakan 3 proses dalam menentukan jumlah cluster untuk memberikan hasil klasterisasi terbaik sesuai dengan struktur data, yakni  $k=3$ ,  $k=5$ , dan  $k=7$ . Proses ini membutuhkan evaluasi, maka fungsi dari operator performance sebagai pembanding kualitas klasterisasi yang dihasilkan dari berbagai nilai  $k$ .



Gambar 3. Proses Pengolahan Dataset

## HASIL DAN PEMBAHASAN

### Hasil Pengolahan Data dengan K-Means Clustering menggunakan Rapidminer

Penelitian ini menggunakan data Angka Partisipasi Sekolah (APS) pada Tingkat SLTA dari tahun 2011 sampai 2022 di wilayah Kalimantan Selatan. Penelitian ini mencoba tiga scenario Analisis jumlah *cluster* ( $k=3$ ,  $k=5$ ,  $k=7$ ) untuk mendapat hasil terbaik. Analisis Jumlah *cluster* sangat penting dilakukan karena tujuannya adalah untuk menggabungkan objek sedemikian rupa sehingga setiap objek memiliki kesamaan tertinggi dengan objek lainnya akan masuk dalam klaster yang sama (Sulistianingsih & Perdana, 2016).

Tahapan preprocessing dan clustering di kerjakan melalui pemanfaatan algoritma *K-means Clustering* dengan RapidMiner.

Tabel 2. Cluster Model

Model Klaster	
Klaster 0	127 item
Klaster 1	237 item
Klaster 2	140 item
Jumlah Total Item	504

### Performance Vector (Performance k=3)

Result not stored in repository.

```

PerformanceVector:
Avg. within centroid distance: 28.370
Avg. within centroid distance_cluster_0: 25.669
Avg. within centroid distance_cluster_1: 19.351
Avg. within centroid distance_cluster_2: 46.085
Davies Bouldin: 0.745
  
```

Gambar 4. Evaluasi Performance k=3

### Performance Vector (Performance (2) k=5)

Result not stored in repository.

```

PerformanceVector:
Avg. within centroid distance: 15.814
Avg. within centroid distance_cluster_0: 13.978
Avg. within centroid distance_cluster_1: 17.460
Avg. within centroid distance_cluster_2: 19.024
Avg. within centroid distance_cluster_3: 12.227
Avg. within centroid distance_cluster_4: 23.397
Davies Bouldin: 0.852
  
```

Gambar 5. Evaluasi Performance k=5

### Performance Vector (Performance (3 ) k=7)

Result not stored in repository.

```

PerformanceVector:
Avg. within centroid distance: 11.177
Avg. within centroid distance_cluster_0: 11.822
Avg. within centroid distance_cluster_1: 10.307
Avg. within centroid distance_cluster_2: 18.934
Avg. within centroid distance_cluster_3: 16.995
Avg. within centroid distance_cluster_4: 7.848
Avg. within centroid distance_cluster_5: 13.498
Avg. within centroid distance_cluster_6: 6.416
Davies Bouldin: 0.839
  
```

Gambar 6. Evaluasi Performance k=7

Berdasarkan evaluasi performance klasterisasi dari nilai  $k=3$ ,  $k=5$ , dan  $k=7$ , dapat disimpulkan bahwa jumlah *cluster*  $k=3$  adalah hasil paling optimal yang didukung oleh nilai Davies Bouldin Index (0.745) yang merupakan nilai paling rendah dibandingkan dengan nilai  $k=5$  ( 0.852 ) dan  $k=7$  ( 0.839 ).

Hal ini menunjukkan bahwa *cluster* yang terbentuk pada  $k=3$  memiliki pemisah lebih baik antar klaster, serta kepadatan yang lebih besar didalam masing- masing klaster.

### Karakteristik Klaster

Hasil Clustering menunjukkan ada tiga klaster dengan masing-masing karakteristik.

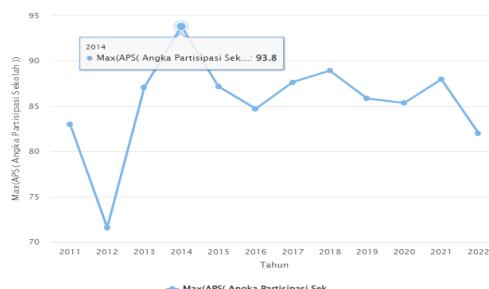
1. Klaster 0 dengan APS tinggi, Infrastruktur Pendidikan yang baik, dan ekonomi Masyarakat yang kuat.
2. Klaster 1 dengan APS sedang, Memiliki akses Pendidikan yang cukup, tetapi masih memiliki masalah pada infrasturnya.
3. Klaster 2 dengan APS rendah, Sering terletak di daerah terpencil dengan fasilitas Pendidikan terbatas.

## Analisis Faktor yang Berpengaruh

Tingkat Pendidikan, akses teknologi dan infrastruktur Pendidikan di masing-masing wilayah adalah faktor utama yang mempengaruhi hasil klasterisasi. *Cluster 0* menunjukkan daerah dengan fasilitas Pendidikan yang lebih tinggi, *Cluster 1* mencerminkan daerah dengan APS sedang, sedangkan *Cluster 2* menunjukkan daerah dengan keterbatasan akses Pendidikan. Faktor-faktor ini merupakan perbedaan utama antar klaster dan dapat digunakan dalam menentukan strategi baru untuk meningkatkan akses Pendidikan.

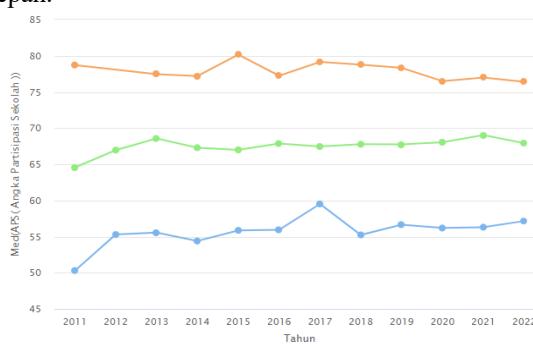
## Visualisasi Data

Grafik tren Angka Partisipasi Sekolah(APS) dalam penelitian ini akan di analisis maksimum dan median dari hasil visualisasi datanya. APS maksimum tahun 2011 sampai 2022 menunjukkan variasi yang mencerminkan perubahan dalam Tingkat partisipasi Pendidikan. Di tahun 2014, APS maksimum mencapai titik tertinggi dengan angka lebih dari 90. Akan Tetapi, tahun 2012 justru terjadi penurunan yang signifikan diikuti dengan fluktuasi stabil sampai ditahun 2022.



Gambar 7. Visualisasi Data Max

Grafik ini menunjukkan dinamika parisipasi sekolah dipengaruhi banyak faktor. Seperti, kebijakan Pendidikan, kondisi sosial ekonomi dan aksesibilitas Pendidikan. Visual ini akan membantu menganalisis lanjut faktor yang mempengaruhi Tingkat partisipasi sekolah dan rencana peningkatakan lebih baik dimasa depan.



Gambar 8. Visualisasi Data Median

Dibandingkan dengan fungsi maksimum, hasil visualisasi data dengan agregasi median menunjukkan tren Angka Partisipasi Sekolah (APS) menghasilkan kondisi yang lebih stabil. Nilai Tengah distribusi APS setiap tahunnya ditampilkan pada grafik median, memberikan Gambaran yang lebih akurat terkait kondisi umum partisipasi sekolah. Nilai median meningkat dan menurun secara bertahap dari tahun 2011 sampai 2022 tanpa lonjakan signifikan. Hal ini tentu menunjukkan fluktuasi nilai maksimum APS mungkin dipengaruhi oleh anomaly dari beberapa daerah tertentu. Sementara median lebih tepat untuk memberikan Gambaran mayoritas populasi.

## Rekomendasi Strategi

Setalah menganalisis data hasil klasterisasi, ada beberapa rekomendasi strategi yang disarankan untuk meningkatkan Angka Partisipasi Sekolah dan meminimalkan kesenjangan di berbagai daerah.

1. Wilayah Klaster 0 dengan nilai APS tertinggi menunjukkan bahwa lebih banyak anak usia sekolah benar-benar bersekolah pada cluster ini.
2. Wilayah klaster 2 dengan nilai rendah menunjukkan bahwa angka partisipasi sekolah rendah ini menggambarkan adanya kendala lebih besar dalam akses Pendidikan didaerah atau kelompok dalam cluster.
3. Wilayah klaster 1 dengan nilai sedang menunjukkan kelompok ini dapat dianggap sebagai peralihan dengan Tingkat partisipasi sekolah tidak terlalu tinggi atau terlalu rendah.

Dari perbedaan kualitas di berbagai cluster ini kemungkinan terdapat perbedaan kualitas Pendidikan dimasing-masing cluster. Cluster paling tinggi mungkin memiliki fasilitas lebih bagus, pengajar lebih berkualitas dan lingkungan belajar yang lebih kondusif atau tersruktur dengan baik.

Hal ini Perlu dilakukan analisis lanjutan dalam mengidentifikasi faktor apa saja yang menyebabkan perbedaan APS di tiap cluster. Faktor yang mungkin mempengaruhi bisa jadi faktor ekonomi, faktor geografis, faktor sosial dan budaya serta dukungan fasilitas belajar yang memadai.

Dari analisis ini dapat diberikan rencangan strategi lebih spesifik dalam meningkatkan APS pada masing masing cluster. Pada APS rendah dan sedang perlu difokuskan untuk Upaya peningkatan akses dan kualitas Pendidikan, serta memberikan dorongan bagi Masyarakat untuk memasukkan anak-anak bersekolah.

## KESIMPULAN

Penelitian ini menganalisis Angka Partisipasi Sekolah (APS) tingkat SLTA di Kalimantan Selatan

menggunakan metode K-means Clustering untuk memahami pola partisipasi Pendidikan. Hasil klasterisasi menunjukkan tiga wilayah cluster. Cluster 0 dengan akses Pendidikan lebih baik, cluster 1 dengan APS sedang dan cluster 2 dengan akses pendidikan terbatas. Faktor yang mempengaruhi klasterisasi adalah Tingkat pendapatan, infrastruktur Pendidikan dan Lokasi geografis.

Dengan nilai Davies Bouldin Index (0.745), klaster k=3 memberikan hasil yang lebih optimal. Dari visualisasi tren APS menunjukkan fluktuasi perubahan stabil dari tahun 2011 sampai 2022. Adapun strategi peningkatan APS meliputi pemerataan fasilitas Pendidikan, Pembangunan infrakstruktur dan kebijakan strategi dalam focus pada klaster rendah. Penelitian ini memberikan rekomendasi strategis untuk mendukung visi Indonesia Emas 2045 melalui peningkatan akses dan kualitas Pendidikan.

## DAFTAR PUSTAKA

- Agusta, Y. (2007). K-Means-Penerapan, Permasalahan dan Metode Terkait. In *Jurnal Sistem dan Informatika* (Vol. 3).
- Asroni, A., Fitri, H., & Prasetyo, E. (2018). Penerapan Metode Clustering dengan Algoritma K-Means pada Pengelompokan Data Calon Mahasiswa Baru di Universitas Muhammadiyah Yogyakarta (Studi Kasus: Fakultas Kedokteran dan Ilmu Kesehatan, dan Fakultas Ilmu Sosial dan Ilmu Politik). *Semesta Teknika*, 21(1). <https://doi.org/10.18196/st.211211>
- Dewi, N. L. P. P., Purnama, I. N., & Utami, N. W. (2022). Penerapan Data Mining Untuk Clustering Penilaian Kinerja Dosen Menggunakan Algoritma K-Means (Studi Kasus: STMIK Primakara). *Jurnal Ilmiah Teknologi Informasi Asia*, 16(2), 105. <https://doi.org/10.32815/jitika.v16i2.761>
- Fajriani, N., & Pasrun, Y. P. (2017). Implementasi K-Means Clustering Pada Rapidminer Untuk Analisis Daerah Rawan Kecelakaan. <https://www.researchgate.net/publication/316661764>
- Hanifah, S., & Primandari, A. H. (2023). Implementasi Metode K-Means Clustering Dalam Pengelompokan Kabupaten/Kota Di Provinsi Ntb Berdasarkan Indikator Pendidikan. *Emerging Statistics and Data Science Journal*, 1(3).
- Nur, F., Faizan, R., Aziz, J., Darma Setiawan, B., & Arwani, I. (2018). Implementasi Algoritma K-Means untuk Klasterisasi Kinerja Akademik Mahasiswa (Vol. 2, Issue 6). <http://j-ptiik.ub.ac.id>
- Solihin, I. P., & Wibisono, M. B. (2017). Desain Kebijakan Sistem Informasi dan Teknologi Informasi Dengan Framework Zachman Pada Perguruan Tinggi Negeri Baru di Jakarta. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 1(3), 266–276. <https://doi.org/10.29207/resti.v1i3.83>
- Sulistianingsih, E., & Perdana, H. (2016). Penentuan Jumlah Cluster Optimal Pada Median Linkage Dengan Indeks Validitas Silhouette. In *Buletin Ilmiah Math. Stat. dan Terapannya (Bimaster)* (Vol. 05, Issue 2).
- Virdam, F., Bernadette, M., & Ariani, N. (2023). Analisis Faktor-Faktor Yang Mempengaruhi Angka Partisipasi Sekolah Pada Provinsi Di Pulau Sulawesi. *Jurnal Of Development Economic And Digitalization*, 2(1), 20–35.
- Wardhani, D., Chyan, P., Mallu, S., Suseno, A. T., Fauzi, A., Gede, I., Sanjaya, W., Rahmah, S. A., Zidny, H., Indra, J., Wicaksana, Y. E., Sofyan, M., Thayf, S., Wayan, I., Nirawana, S., Rikson, R., Sijabat, M., & Al Mudzakir, T. (2023). *Pengantar Data Mining PT. Mifandi Mandiri DigitaL*.