

Segmentasi Wilayah Berdasarkan Komposisi Tenaga Pendidik Menggunakan Algoritma K-Means (Studi Kasus: Data Kemendikdasmen SD-MI di Indonesia Tahun 2022 - 2024)

Fransiskus Xaverius Prasetyo Satriatama

Teknik Informatika S-2, Program Pascasarjana, Universitas Pamulang, Kota Tangerang Selatan, Banten

e-mail: riosatriatama0801@gmail.com

Abstrak—Penelitian ini bertujuan untuk melakukan Segmentasi Wilayah yang ada di Indonesia berdasarkan komposisi tenaga pendidik pada jenjang pendidikan dasar (SD-MI) menggunakan Algoritma K-Means. Permasalahan pemerataan Tenaga Pendidik masih menjadi isu yang sangat penting, baik dari sisi jumlah maupun karakteristik guru antarwilayah. Data yang digunakan dalam penelitian ini bersumber dari Kementerian Pendidikan Dasar dan Menengah (Kemendikdasmen) periode tahun 2022 sampai dengan 2024, yang mencakup Provinsi, Jumlah PNS, Bukan PNS dan Total Guru. Metode K-Means diterapkan untuk mengelompokkan wilayah ke dalam beberapa klaster yang memiliki kemiripan karakteristik tenaga pendidik. Hasil Pengelompokan menunjukkan adanya perbedaan yang cukup signifikan antarwilayah, terutama antara daerah perkotaan besar dengan daerah terpencil atau kota kecil. Temuan ini sangat diharapkan dapat memberikan gambaran yang lebih jelas mengenai kondisi distribusi Tenaga Pendidik di Indonesia serta menjadi bahan pertimbangan bagi para pengambil kebijakan dalam merumuskan strategi pemerataan dan peningkatan kualitas pendidikan dasar secara berkelanjutan.

Kata Kunci—Segmentasi Wilayah, Kmeans Clustering, Tenaga Pendidik, Pendidikan Dasar.

I. PENDAHULUAN

Pendidikan merupakan usaha untuk mewujudkan suasana belajar dan proses pembelajaran agar peserta didik secara aktif mengembangkan potensi dirinya untuk memiliki kekuatan spiritual keagamaan, pengendalian diri, kepribadian, kecerdasan, akhlak mulia, serta keterampilan yang diperlukan dirinya, masyarakat, bangsa dan negara (PMPK KEMDIKBUD, 2020). Dalam pembukaan Undang-Undang Dasar Negara Republik Indonesia tahun 1945 diamanatkan bahwa salah satu tujuan Pemerintah Negara Indonesia adalah untuk mencerdaskan kehidupan bangsa. Pendidikan menjadi landasan kuat yang diperlukan untuk meraih kemajuan bangsa di masa depan. Selain itu, pendidikan penting sebagai bekal dalam menghadapi era global dan persaingan ketat antarbangsa. Dengan demikian, pendidikan menjadi sangat penting untuk dipenuhi karena merupakan faktor bagi suatu bangsa untuk bisa bertahan dan unggul dalam kompetisi global. Oleh karena itu, Pemerintah Indonesia perlu untuk mengupayakan pemerataan pendidikan. Pemerataan pendidikan ini berarti bahwa persamaan kesempatan bagi seluruh warga untuk mendapatkan hak pendidikan dan keadilan dalam memperoleh pendidikan yang sama dalam kehidupan masyarakat.

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 (Badan Pusat Statistik, 2017). 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 persebaran atau komposisi tenaga pendidik tiap provinsinya di masa depan.

Pada penelitian ini, dilakukan penggunaan Google Colab dan pendekatan K-Means sebagai metode yang diterapkan. Google Colab adalah sebuah IDE untuk pemrograman Python dimana pemrosesan akan dilakukan oleh server Google yang memiliki perangkat keras dengan performa yang tinggi (R. T. Handayanto and H. Herlawati, 2020). Dari sisi perangkat lunak, Google Colab telah menyediakan hampir sebagian besar pustaka (library) yang dibutuhkan. Dalam penelitian ini pustaka yang dibutuhkan adalah Keras, TensorFlow, NumPy, Pandas, dan pendukung lainnya, misalnya untuk pembuatan grafik lewat Matplotlib. Bahkan seluruh versi disediakan, misalnya TensorFlow versi 1.x maupun 2.x tersedia, begitu juga versi Python yang mulai dari versi 2.x hingga 3.x. Dari sisi perangkat keras, Google Colab menyediakan layanan berupa media penyimpanan yang terintegrasi dengan Google Drive, prosesor yang berupa CPU, GPU, dan TPU, serta RAM. Dengan jaminan kemampuan servernya yang stabil hampir keseluruhan pemrosesan tidak menemukan kendala dengan Google Colab selama koneksi jaringan internet lancar. 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, jumlah PNS dan Jumlah Non PNS dari seluruh Provinsi yang ada di Indonesia. Hasil analisis akan disajikan dalam bentuk visualisasi yang informatif.

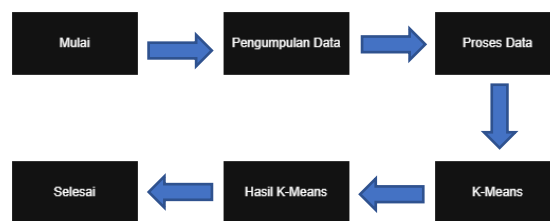
Penelitian ini menganalisis tingkat persebaran guru atau tenaga pendidik pada jenjang SD di 38 provinsi Indonesia selama periode 2022-2024. Tujuannya adalah untuk memahami secara mendalam tingkat penyelesaian pendidikan di jenjang-jenjang tersebut di provinsi-provinsi Indonesia. Melalui analisis data penyelesaian pendidikan selama periode tersebut, peneliti dapat mengidentifikasi pola dan tren dalam tingkat penyelesaian pendidikan di setiap jenjang dan provinsi.

Dalam rangka menyelesaikan tujuan penelitian ini, penulis akan menggunakan sumber data resmi dari Kemendikdasmen. Selain itu, 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.

II. METODE PENELITIAN

2.1 Tahapan Penelitian

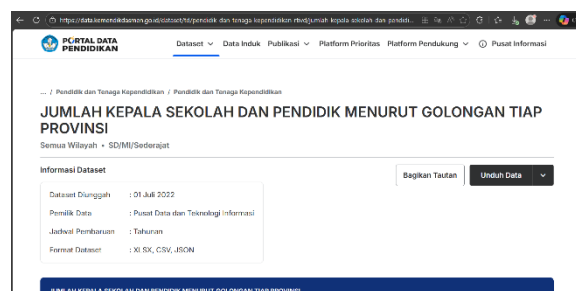
Penelitian ini memiliki tahapan yang di mulai dari pengumpulan data, proses data, perhitungan K-Means dengan Google Colab, lalu diakhiri hasil dari perhitungan K-Means.



Gambar 1. Tahapan Penelitian

2.2 Pengumpulan Data

Sumber data penelitian ini di peroleh dari Kementerian Pendidikan Dasar dan Menengah (Kemendikdasmen), selama 3 tahun dari 2022 - 2024. Dalam penelitian ini data yang digunakan adalah data jumlah kepala sekolah dan pendidik menurut golongan tiap provinsi di Indonesia dari tahun 2022-2024.



Gambar 2. Pengumpulan Data

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 (J. Sumarah and A. T. Wulandari, 2021).

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 (J. Ha, M. Kambe, and J. Pe, 2011).

2.4 Google Colaboratory

Google Colaboratory, juga dikenal sebagai Colab, adalah platform komputasi awan gratis yang menyediakan lingkungan notebook tanpa server interaktif untuk menulis, berbagi, dan mengeksekusi kode (F. M. Götz, R. Maertens, S. Loomba, and S. van der Linden, 2023). Colab beroperasi pada notebook Jupyter, yang kompatibel dengan bahasa pemrograman populer seperti Python, R, dan Julia (F. M. Götz, R. Maertens, S. Loomba, and S. van der Linden, 2023). Colab menawarkan sumber daya komputasi yang kuat, termasuk akses ke mesin virtual jarak jauh berbasis Linux dengan unit pemrosesan grafis (GPU) yang canggih hingga 12 jam (F. M. Götz, R. Maertens, S. Loomba, and S. van der Linden, 2023). Hal ini menjadikannya platform yang nyaman untuk berbagai aplikasi, termasuk tugas pembelajaran mendalam dan pembelajaran mesin (M. L. Hilton, J. M. Goessling, L. M. Knezevich, and J. M. Downer, 2022). Peneliti dapat menggunakan Colab untuk memadukan penjelasan naratif, kode perangkat lunak, hasil komputasi, dan sumber daya multimedia dalam satu dokumen alur kerja. Colab telah digunakan di berbagai bidang seperti biologi, ilmu komputer, dan kedokteran (M. L. Hilton, J. M. Goessling, L. M. Knezevich, and J. M. Downer, 2022). Colab juga telah digunakan untuk pengkodean kolaboratif dan kerja sama tim dalam lingkungan pendidikan. Secara umum, Google Colaboratory menyediakan platform yang mudah digunakan dan dapat diakses oleh para peneliti dan mahasiswa untuk melakukan tugas komputasi dan berbagi kode.

2.5 K-Means

K-Means adalah sebuah metode klustering yang tergolong dalam pembelajaran tanpa pengawasan, yang mengaplikasikan 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 (F. Simanjourang, R. Winanjaya, and F. Rizki, 2021). 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) (N. Nurahman, A. Purwanto, and S. Mulyanto, 2022).

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 (F. Simanjourang, R. Winanjaya, and F. Rizki, 2021). K-Means akan diproses dengan Google Colab dengan perhitungan Euclidean Distance.

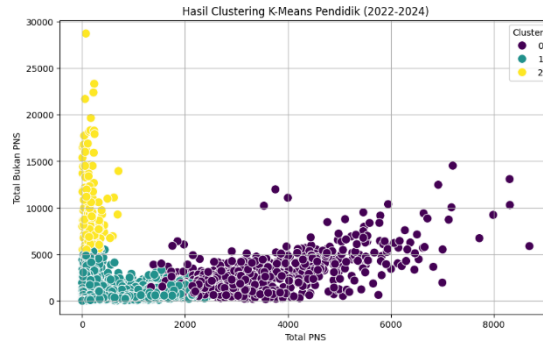
III. HASIL DAN DISKUSI

3.1 Dataset

Penelitian ini menggunakan dataset dari Kementerian Pendidikan Dasar dan Menengah periode 2022 hingga 2024 dengan total 3.031 entri data (wilayah/status sekolah). Fitur yang digunakan untuk pengelompokan meliputi jumlah PNS Golongan II, III, IV, dan jumlah tenaga Bukan PNS.

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 tenaga pendidik.

3.2 Hasil Clustering K-Means Pendidik



Gambar 3. Hasil Clustering K-Means

1. Sumbu dan Titik Data

- Sumbu X (Total PNS)
Menunjukkan jumlah guru PNS di suatu wilayah.
- Sumbu Y (Total Bukan PNS)
Menunjukkan jumlah guru non-PNS (honorer, kontrak, dll).
- Setiap titik merepresentasikan satu wilayah/daerah (misalnya kabupaten/kota).

2. Clustering K-Means

Data dikelompokkan menjadi 3 klaster ($k = 3$), ditandai dengan warna berbeda:

a) Cluster 0 (ungu tua)

- Wilayah dengan jumlah PNS tinggi dan jumlah non-PNS sedang hingga tinggi
- Umumnya daerah dengan kapasitas tenaga pendidik besar
- Biasanya wilayah maju atau berpenduduk besar

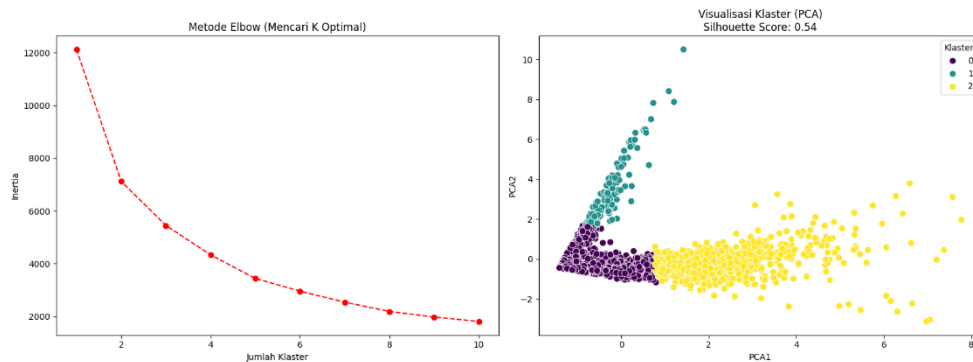
b) Cluster 1 (hijau kebiruan)

- Wilayah dengan jumlah PNS rendah–menengah dan non-PNS rendah
- Menunjukkan daerah dengan jumlah guru relatif sedikit
- Bisa merupakan daerah kecil atau terpencil

c) Cluster 2 (kuning)

- Wilayah dengan jumlah PNS sangat rendah, tetapi non-PNS tinggi
- Mengindikasikan ketergantungan tinggi pada guru non-PNS
- Menjadi indikator potensi ketimpangan distribusi guru PNS

3.4 Validasi Model



Gambar 4. Hasil Metode Elbow dan Silhouette Score

- Silhouette Score yang diperoleh sebesar 0,54. Nilai ini menunjukkan bahwa hasil pengelompokan data tergolong cukup baik, di mana setiap objek dalam satu kluster memiliki tingkat kemiripan yang relatif tinggi serta pemisahan antar kluster yang sudah cukup jelas.
- Berdasarkan visualisasi PCA yang sudah dilakukan, dapat diamati bahwa data membentuk pola pengelompokan yang cukup terstruktur. Namun demikian, masih terdapat beberapa outlier, khususnya pada kluster yang memiliki jumlah tenaga non-PNS dengan nilai yang sangat ekstrem.

IV. KESIMPULAN

1. Mayoritas wilayah dan satuan pendidikan di Indonesia saat ini berada pada skala kecil hingga menengah, namun secara konsisten menunjukkan bahwa jumlah tenaga pendidik non-PNS sudah melampaui jumlah PNS di hampir semua kelompok.
2. Terdapat konsentrasi beban pendidik yang sangat masif di segelintir wilayah tertentu (Cluster 1) yang mengelola tenaga pengajar hingga sepuluh kali lipat lebih banyak dibanding wilayah biasa, namun ironisnya wilayah-wilayah besar ini justru yang paling minim memiliki tenaga PNS.
3. Hasil clustering membuktikan bahwa kebijakan pendidikan tidak bisa dipukul rata secara nasional, melainkan memerlukan penanganan berbeda antara wilayah yang kekurangan guru secara total dengan wilayah yang hanya kekurangan guru berstatus PNS.

DAFTAR PUSTAKA

- [1] PMPK KEMDIKBUD. (2020). Direktorat Pendidikan Masyarakat dan Pendidikan Khusus, Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi, Republik Indonesia. Retrieved November 26, 2022, from <https://pmpk.kemdikbud.go.id/>
- [2] 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.
- [3] R. T. Handayanto and H. Herlawati, "Prediksi Kelas Jamak dengan Deep Learning Berbasis Graphics Processing Units," *Jurnal Kajian Ilmiah*, vol. 20, no. 1, pp. 67–76, Jan. 2020, doi: 10.31599/jki.v20i1.71.
- [4] F. M. Götz, R. Maertens, S. Loomba, and S. van der Linden, "Let the algorithm speak: How to use neural networks for automatic item generation in psychological scale development.," *Psychol. Methods*, 2023.
- [5] M. L. Hilton, J. M. Goessling, L. M. Knezevich, and J. M. Downer, "Utility of machine learning for segmenting camera trap time-lapse recordings," *Wildl. Soc. Bull.*, vol. 46, no. 4, p. e1342, 2022.
- [6] 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.
- [7] J. Ha, M. Kambe, and J. Pe, *Data Mining: Concepts and Techniques*. 2011. doi: 10.1016/C2009-0-61819-5.
- [8] 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>