Journal of Information System Research (JOSH)
Volume 6, No. 1, Oktober 2024, pp 251−260
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i1.5957

Implementasi CRISP-DM Pada Analisis Pembangunan Pendidikan Prasekolah Menurut Kabupaten/Kota di Indonesia

Putri Chandra Iranti, Dedy Kurniawan*, M. Rudi Sanjaya, Ahmad Rifai, M. Husni Syahbani, Gabriel Ekoputra Hartono Cahyadi, Purwita Sari

Fakultas Ilmu Komputer, Sistem Informasi, Universitas Sriwijaya, Palembang
Jl. Srijaya Negara, Bukit Besar, Kec. Ilir Barat I, , Palembang, Indonesia
Email: [1]09031482326004@student.unsri.ac.id, [2,*]dedykurniawan@ilkom.unsri.ac.id, [3]m.rudi.sjy@ilkom.unsri.ac.id, [4]ahmadrifai@ilkom.unsri.ac.id, [5]husnisyahbani@unsri.ac.id, [6]ekoputra2695@gmail.com, [7]witasari92@gmail.com
Email Penulis Korespondensi: dedykurniawan@ilkom.unsri.ac.id
Submitted: **23/09/2024;** Accepted: **09/10/2024;** Published: **15/10/2024**

Abstrak−Pendidikan prasekolah melalui Taman Kanak-Kanak (TK) penting untuk perkembangan anak di Indonesia, namun akses yang tidak merata menjadi masalah utama. Penelitian ini mengevaluasi kebutuhan pembangunan infrastruktur prasekolah memakai algoritma K-Means'clustering di tools RapidMiner. Pengelompokan wilayah didasarkan pada jumlah peserta didik, jumlah TK, Indeks Pembangunan Manusia (IPM), persentase kemiskinan, jumlah penduduk, dan tingkat pengangguran. Metode CRISP-DM diterapkan melalui tahapan pemahaman, persiapan, pemodelan, evaluasi, dan penerapan. Data dari Badan Pusat Statistik (BPS) dan Dapodik digunakan, dengan normalisasi Z-transformation dan pembersihan data. Hasil clustering menunjukkan tiga cluster utama dengan Davies-Bouldin Index (DBI) terendah pada K=3, yaitu 0.205. Dengan total 514 kabupaten/kota di Indonesia, didapat hasil kebutuhan setiap cluster yaitu Cluster 0 terdiri dari 402 kabupaten/kota membutuhkan peningkatan partisipasi, Cluster 1 mencakup 49 kabupaten/kota membutuhkan fasilitas pendidikan, Cluster 2 yang meliputi 63 kabupaten/kota membutuhkan pembangunan sekolah baru. Penelitian ini memberi wawasan untuk mengatasi ketimpangan akses pendidikan prasekolah.

Kata Kunci: Pendidikan Prasekolah; K-Means Clustering; RapidMiner; Metodologi CRISP-DM; Indeks Davies-Bouldin (DBI)

Abstract−Preschool education through Kindergarten (TK) plays a crucial role in child development in Indonesia, yet unequal access remains a significant issue. This study evaluates the need for preschool infrastructure development using the K-Means clustering algorithm implemented through RapidMiner. Regional clustering is based on the number of students, number of TK schools, Human Development Index (HDI), poverty rate, population size, and unemployment rate. The CRISP-DM methodology is applied, involving stages of understanding, preparation, modeling, evaluation, and deployment. Data from the Central Bureau of Statistics (BPS) and the Ministry of Education's Dapodik system are utilized, incorporating Z-transformation normalization and data cleansing. The clustering results reveal three main clusters with the lowest Davies-Bouldin Index (DBI) at K=3, scoring 0.205. With a total of 514 districts/cities in Indonesia, the results of the needs of each cluster were obtained, namely Cluster 0 consisting of 402 districts/cities requiring increased participation, Cluster 1 covering 49 districts/cities requiring educational facilities, Cluster 2 covering 63 districts/cities requiring the construction of new schools. This study provides valuable insights into addressing disparities in preschool education access and offers guidance for better resource allocation and policy decisions aimed at improving early childhood education infrastructure.

Keywords: Preschool Education; K-Means Clustering; RapidMiner; CRISP-DM Methodology; Davies-Bouldin Index (DBI).

1. PENDAHULUAN

Peningkatan kualitas Sumber Daya Manusia"(SDM) ialah suatu kunci utama dalam menghadapi persaingan di tingkat internasional, terutama bagi Indonesia. Dalam era globalisasi yang semakin kompetitif, kualitas SDM yang unggul dan adaptif berperan signifikan dalam mendorong kemajuan ekonomi, inovasi, dan daya saing bangsa di berbagai sektor [1]. Melalui sistem pendidikan yang efisien, diharapkan dapat terbentuk generasi yang lebih terampil dan berpengetahuan, pada akhirnya akan berkontribusi untuk peningkatan kualitas hidup secara menyeluruh.

Penelitian terdahulu yang telah melakukan upaya membantu peningkatan kualitas pendidikan secara merata dengan menganalisis cluster faktor penunjang pendidikan menggunakan Algoritma K-Means [2]. Pendidikan memiliki peran penting dalam era globalisasi untuk mempersiapkan masyarakat agar mampu secara fisik dan mental dalam menghadapi perubahan yang semakin tak terelakkan di berbagai aspek kehidupan manusia [3]. Pendidikan adalah sebuah strategi untuk meningkatkan kesejahteraan manusia. Selain berfungsi sebagai usaha untuk memperbaiki kualitas hidup, pendidikan juga memainkan peran vital dalam proses pembangunan nasional. Keyakinan ini didukung oleh pandangan bahwa pendidikan memberikan kontribusi signifikan terhadap pembangunan negara, khususnya dalam bidang ekonomi [4].

Pendidikan prasekolah, terutama yang diselenggarakan oleh lembaga Taman Kanak-Kanak (TK), memegang peranan krusial dalam mendukung tahap awal perkembangan anak-anak di Indonesia. TK, sebagai bagian dari sistem pendidikan, memiliki fungsi strategis dalam membangun dasar pendidikan bagi generasi masa depan. Ini adalah tahap awal dari pendidikan terstruktur yang bertujuan untuk membentuk individu menjadi mandiri dan mampu bersaing pada era globalisasi.

Pendidikan anak pada usia dini dianggap sangat penting karena periode 0–5 tahun dikenal sebagai masa keemasan [5]. Pendidikan pada tingkat prasekolah ini merupakan landasan utama yang membantu membentuk

Journal of Information System Research (JOSH)
Volume 6, No. 1, Oktober 2024, pp 251−260
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i1.5957

kemampuan kognitif, sosial, dan emosional anak-anak sebelum memasuki jenjang pendidikan formal. Pada dasarnya, untuk memastikan terpenuhinya hak anak usia dini dalam hal tumbuh kembang, diperlukan berbagai upaya peningkatan di bidang kesehatan, gizi, perawatan, pengasuhan, perlindungan, kesejahteraan, serta stimulasi pendidikan. Semua ini harus dilakukan secara bersamaan, sistematis, menyeluruh, terpadu, dan berkelanjutan [6].

Pendidikan karakter di Taman Kanak-kanak melanjutkan pendidikan karakter yang telah dimulai dalam lingkungan keluarga. Pendidikan karakter di TK dilakukan melalui kebiasaan-kebiasaan yang tidak tertulis, namun dijalankan secara konsisten. Ini menunjukkan bahwa kehidupan anak atau siswa dipengaruhi oleh proses internalisasi nilai-nilai pendidikan karakter Sayangnya, tidak semua wilayah di Indonesia memiliki akses yang setara terhadap fasilitas pendidikan prasekolah ini.

Kesenjangan ini muncul akibat berbagai faktor, salah satunya adalah perbedaan antara jumlah peserta didik dan ketersediaan infrastruktur Taman Kanak-Kanak (TK) di sebagian daerah. Di beberapa wilayah, jumlah TK sangat terbatas, sementara di daerah lain, jumlah peserta didik melebihi kapasitas fasilitas yang ada. Kondisi ini menunjukkan perlunya analisis mendalam untuk mengidentifikasi wilayah-wilayah yang membutuhkan perhatian khusus dalam pengembangan pendidikan prasekolah.

Dengan demikian, penelitian ini bertujuan untuk mengevaluasi pembangunan infrastruktur pendidikan prasekolah di Indonesia dengan memanfaatkan teknologi data mining, khususnya melalui penerapan algoritma K-Means clustering. Teknik tersebut adalah metode mengelompokan data non-hierarkis dengan bertujuan untuk membagi atau mengelompokkan data ke dalam beberapa cluster yang telah ditentukan sebelumnya [7]. Setiap kelompok yang dihasilkan dari algoritma clustering mencerminkan pola atau karakteristik khusus dalam data, yang dapat berguna untuk mengidentifikasi tren, memprediksi perilaku, atau mengelompokkan entitas serupa untuk analisis lanjutan. Namun, pemilihan metode clustering yang tepat dan interpretasi yang akurat dari hasilnya menjadi langkah krusial dalam menjamin keberhasilan penerapan teknik ini di berbagai aplikasi [8]. Dengan menggunakan teknik clustering, sumber data yang dimiliki dengan pola atau karakteristik yang mirip akan dikelompokkan dalam satu cluster.

Hal ini mempermudah dalam menganalisis pola-pola tersebut lebih lanjut [9]. Data mining adalah proses otomatis untuk menganalisis data besar dan kompleks guna menemukan pola dan tren yang signifikan. Proses ini bertujuan untuk mengungkap wawasan yang tersembunyi dan belum dikenal sebelumnya, dengan memanfaatkan teknik statistik dan algoritma pembelajaran mesin seperti penelitian terdahulu melakukan analisis data mining untuk membantu pihak dinas pendidikan terkait dalam mengelompokkan wilayah sekolah [10]. Proses untuk mengidentifikasi pola atau informasi penting pada sumber yang dipilih melibatkan penerapan sumber atau metode tertentu. Beragam cara, sumbere, atau algoritma digunakan dalam data mining [11]. Clustering merupakan cara pembelajaran yang tidak terprediksi karena tidak memerlukan target akhir, dan dikenal sebagai unsupervised learning [12].

Clustering memiliki tujuan utama untuk mengidentifikasi pola atau struktur yang belum terungkap dalam data. Terdapat berbagai algoritma clustering, seperti K-Means dan Hierarchical Clustering [13]. Algoritma-algoritma ini mengelompokkan data ke dalam beberapa cluster berdasarkan perhitungan jarak dan hanya dapat diterapkan pada data numerik, dengan tujuan untuk meminimalkan jarak antara data dan pusat cluster [14]. Tujuan utama dari clustering adalah untuk meningkatkan variasi antar cluster sambil meminimalkan fungsi objektif di dalam setiap cluster [15]. Algoritma K-Means bersifat iteratif, membagi data menjadi K cluster yang telah ditentukan sebelumnya. Secara historis, K-Means merupakan salah satu algoritma yang paling penting dalam data mining [16]. Penelitian ini bertujuan untuk mengidentifikasi kelompok wilayah yang memiliki kebutuhan pembangunan prasekolah yang berbeda, sehingga kebijakan pemerintah dapat lebih tepat sasaran dan efisien.

Penelitian ini menerapkan kerangka kerja CRISP-DM (Cross Industry Standard Process for Data mining), memberikan panduan sistematis dalam analisis data, mulai dari memahami masalah bisnis hingga penerapan hasil analisis. Metode ini dirancang untuk memastikan bahwa setiap tahap dalam proses data mining dilakukan secara terstruktur dan efisien, dari pemahaman awal hingga implementasi solusi [17]. Data dalam studi ini diperoleh dari sumber-sumber resmi, yaitu Badan Pusat Statistik (BPS) dan Data Pokok Pendidikan (Dapodik) yang berasal dari Kementerian Pendidikan dan Kebudayaan (Kemendikbud).

Data tersebut meliputi informasi tentang jumlah peserta didik TK di berbagai kabupaten/kota, jumlah sekolah TK, Indeks Pembangunan Manusia (IPM), populasi, tingkat kemiskinan, dan tingkat pengangguran. Untuk analisis data, perangkat lunak RapidMiner digunakan, yang memungkinkan pengelompokan wilayah berdasarkan karakteristik yang ada. RapidMiner tidak hanya digunakan dalam bisnis, dalam bidang pendidikan, penelitian, pelatihan, dan pengembangan prototipe. Perangkat lunak ini yang mendukung mekanisme pembelajaran mesin melalui visualisasi, pemodelan, persiapan data, serta optimasi [18]. Berdasarkan penelitian sebelumnya, menggunakan RapidMiner dalam Penerapan Algoritma K-Means Clustering pada tingkat penyelesaian pendidikan di provinsi-provinsi Indonesia digunakan sebagai referensi oleh peneliti dalam menganalisis pendidikan prasekolah di berbagai kabupaten/kota di Indonesia. diharapkan menjadi peningkatan pemerataan pendidikansejak awal yaitu dari tingkat Taman Kanak-Kanak [19].

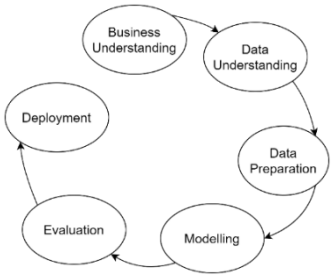
Melalui analisis data ini, penelitian ini diharapkan dapat mengidentifikasi dengan tepat daerah-daerah yang membutuhkan peningkatan infrastruktur pendidikan prasekolah dan memberikan rekomendasi kepada pemerintah untuk mengoptimalkan distribusi pembangunan TK di seluruh Indonesia. Selain daripada itu, output dari penelitian

Journal of Information System Research (JOSH)
Volume 6, No. 1, Oktober 2024, pp 251−260
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i1.5957

ini diharapkan sebagai acuan dalam merancang kebijakan pendidikan yang lebih efektif, sehingga kesenjangan dalam akses pendidikan prasekolah di Indonesia dapat berkurang secara signifikan.

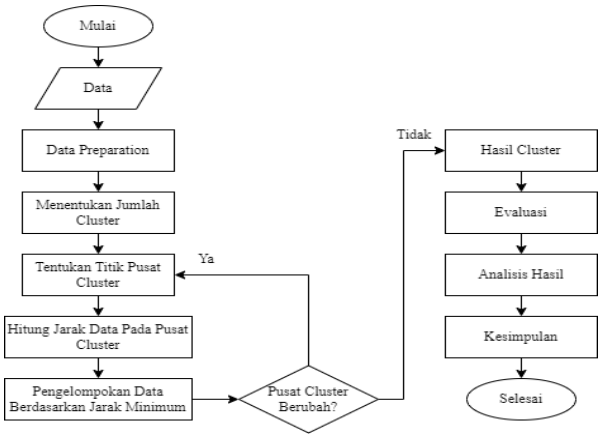
2. METODOLOGI PENELITIAN

Metode yang pakai pada penelitian ini adalah clustering merujuk pada penerapan algoritma K-Means, sebuah teknik yang digunakan untuk mengelompokkan objek berdasarkan kesamaan karakteristiknya. Tujuan dari penggunaan algoritma K-Means pada penelitian ini adalah mengelompokkan wilayah-wilayah di Indonesia sesuai dengan kebutuhan mereka dalam pembangunan pendidikan prasekolah, sehingga analisis ini dapat menawarkan wawasan yang lebih mendalam tentang distribusi infrastruktur pendidikan yang tersedia.



Gambar 1. Metode CRISP-DM

Data mining adalah proses yang mengikuti prosedur yang telah ditetapkan, yaitu CRISP-DM (Cross-Industry Standard Process for Data mining). Prosedur ini yang meliputi seluruh tahapan, dari preprocessing data hingga pembentukan, evaluasi, dan penerapan model. Metode CRISP-DM menawarkan standar untuk data mining sebagai pendekatan untuk menyelesaikan permasalahan yang sering dihadapi oleh bisnis atau unit penelitian [20]. Penelitian ini mengadopsi kerangka kerja CRISP-DM, seperti yang digambarkan pada Gambar 1, yang memberikan panduan menyeluruh dan terstruktur untuk setiap langkah dalam analisis data. Berikut tahapan penelitian berdasarkan kerangka kerja CRISP-DM:



Gambar 2. Tahapan Penelitian

Berdasarkan tahapan penelitian ini pada Gambar 2 meliputi kerangka kerja CRISP-DM, berikut uraian metodologi penelitian yang dilakukan:

2.1 Business Understanding

Pada tahap inilah, fokus penelitian adalah memahami dan menetapkan tujuan utama, yaitu menganalisis serta memetakan pengembangan pendidikan prasekolah di Indonesia, dengan perhatian khusus pada Taman Kanak-Kanak (TK). Ada ketimpangan dalam hal akses dan juga pada kualitas pendidikan di berbagai daerah di Indonesia. Beberapa area menghadapi kekurangan infrastruktur pendidikan, sementara yang lainnya mungkin memiliki jumlah TK yang cukup namun dengan jumlah peserta didik yang rendah.

Penelitian ini bertujuan untuk mengidentifikasi daerah dengan membutuhkan perhatian lebih pada pengembangan taman kanak-kanak, dengan menganalisis data seperti jumlah peserta didik, Indeks Pembangunan Manusia (IPM), persentase warga miskin, serta tingkat pengangguran, dan jumlah sekolah. Metode data mining, khususnya algoritma K-Means clustering, digunakan dengan cara mengelompokkan wilayah-wilayah di Indonesia ke dalam cluster yang berbeda berdasarkan karakteristiknya. Dengan cara ini, ketidakmerataan dalam pendidikan prasekolah dapat dianalisis dengan lebih akurat, memungkinkan pembuatan kebijakan pembangunan yang lebih tepat sasaran oleh pemerintah.

Journal of Information System Research (JOSH)
Volume 6, No. 1, Oktober 2024, pp 251−260
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i1.5957

2.2 Data Understanding

Setelah masalah diidentifikasi, tahap berikutnya adalah pengumpulan dan evaluasi data. Data yang digunakan berasal dari sumber terpercaya seperti Badan Pusat Statistik (BPS) dan Data Pokok Pendidikan (Dapodik) Kementerian Pendidikan dan Kebudayaan. Pada tahap ini, sumber data terkait jumlah peserta didik, jumlah sekolah TK, Indeks Pembangunan Manusia (IPM), populasi masyarakat, persentase kemiskinan, dan tidak ada pekerjaan di setiap kabupaten/kota di Indonesia dievaluasi untuk memastikan relevansi dan kualitasnya.

Populasi penduduk menjadi variabel tambahan yang memberikan gambaran mengenai skala kebutuhan pendidikan prasekolah di setiap wilayah, di mana daerah dengan populasi yang lebih besar cenderung memerlukan lebih banyak infrastruktur. Selain itu, persentase kemiskinan menjadi faktor kunci yang mempengaruhi akses terhadap pendidikan, karena kemiskinan seringkali menjadi penghalang utama bagi partisipasi dalam pendidikan prasekolah. Terakhir, tingkat pengangguran juga dianalisis untuk melihat pengaruh kondisi ekonomi terhadap kemampuan masyarakat mengakses layanan pendidikan, dengan daerah yang memiliki tingkat pengangguran tinggi biasanya mengalami kesulitan lebih besar dalam menyediakan pendidikan yang memadai bagi anak-anak.

Setelah menetapkan tujuan proyek dan mengidentifikasi masalah bisnis, langkah selanjutnya adalah mengumpulkan dan mengevaluasi data demi memastikan relevansi serta kualitas data yang akan digunakan dalam analisis. Data yang digunakan dalam penelitian ini diperoleh dari dua sumber utama, yaitu Badan Pusat Statistik (BPS) dan Data Pokok Pendidikan (Dapodik) dari Kementerian Pendidikan dan Kebudayaan. Data ini mencakup berbagai variabel penting yang memengaruhi pembangunan pendidikan prasekolah di setiap kabupaten/kota di Indonesia.

Variabel utama yang dianalisis meliputi jumlah peserta didik TK, yang mencerminkan total jumlah anak yang telah terdaftar pada Taman Kanak-Kanak (TK) pada tiap kabupaten/kota, menjadi indikator penting untuk menilai kebutuhan infrastruktur pendidikan di suatu daerah. Selain itu, jumlah sekolah TK yang tersedia di setiap kabupaten/kota digunakan untuk mengevaluasi apakah fasilitas yang ada memadai dalam memenuhi kebutuhan pendidikan. Indeks Pembangunan Manusia (IPM) juga menjadi pertimbangan penting, karena indeks ini mengukur kualitas kehidupan manusia berdasarkan pada dimensi pendidikan,'kesehatan,'dan standar hidup dikatakan layak, dengan kabupaten/kota ber-IPM rendah umumnya menghadapi tantangan lebih besar dalam akses pendidikan.

Tahap pemahaman data ini sangat penting untuk mengenali karakteristik dan hubungan antara variabel-variabel tersebut sehingga dapat digunakan dengan tepat dalam model clustering. Sebagai contoh, Tabel 1 berikut menunjukkan variabel-variabel yang digunakan beserta tipe datanya:

Tabel 1. Variabel dan Tipe Data

No	Atribut	Tipe Data
1	Kabupaten/Kota	Polynominal
2	Jumlah Peserta Didik (TK)	Real
3	Jumlah IPM	Real
4	Persentase Kemiskinan	Real
5	Jumlah Penduduk	Real
6	Tingkat Pengangguran	Real
7	Jumlah Sekolah (TK)	Real

Setelah data dikumpulkan, dilakukan data cleaning untuk memastikan tidak ada nilai yang hilang atau duplikat, serta melakukan normalisasi data agar semua atribut berada pada skala yang sebanding. Pada penelitian ini, tidak ditemukan adanya missing value.

Gambar 2. Pengecekan Missing Value

2.3 Data Preparation

Data Preparation ialah bertujuan mengatasi permasalahan dalam data sebelum memasuki fase pemodelan, agar dapat menghasilkan model yang optimal [21]. Pada tahap Data Preparation, dilakukan serangkaian langkah demi

Journal of Information System Research (JOSH)
Volume 6, No. 1, Oktober 2024, pp 251−260
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i1.5957

mempersiapkan elemen yang dibutuhkan dalam analisis clustering. Proses ini meliputi seleksi atribut, normalisasi, pembersihan data, dan transformasi data, sehingga hasil analisis lebih akurat dan dapat diandalkan.

2.3.1 Seleksi Atribut

Sumber data dari Badan Pusat Statistik (BPS) dan juga Data Pokok Pendidikan (Dapodik) Kementerian Pendidikan dan Kebudayaan dimanfaatkan dalam penelitian ini, dengan atribut yang mencakup informasi yaitu Jumlah peserta didik TK di setiap kabupaten/kota, Jumlah sekolah TK, Indeks Pembangunan Manusia (IPM), Persentase kemiskinan, Populasi masyarakat, Tingkat tidak bekerja. Atribut-atribut ini dipilih karena relevansinya yang langsung terhadap infrastruktur dan akses pendidikan prasekolah di Indonesia. Penelitian ini memanfaatkan data dari tahun 2022/2023 yang meliputi semua daerah kabupaten dan kota di Indonesia.

2.3.2 Normalisasi Data

Normalisasi dilakukan untuk memastikan bahwa seluruh atribut berada pada skala sertara, maka tidak ada atribut maka mendominasi hasil clustering karena memiliki rentang nilai yang berbeda. Pada penelitian ini, digunakan metode Z-transformation yang mengubah setiap nilai menjadi distribusi maka rata-rata 0 serta standar deviasi 1. Hal ini memastikan bahwa variabel yang memiliki skala besar, seperti populasi penduduk, tidak akan mendominasi variabel lain seperti jumlah sekolah.

2.3.3 Pembersihan Data

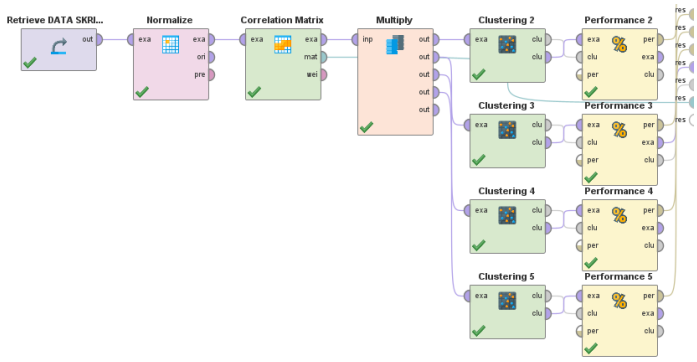
Pembersihan data dilakukan untuk memastikan tidak ada missing values atau data duplikat yang dapat mempengaruhi hasil analisis. Berdasarkan pengecekan awal, tidak ditemukan adanya missing values dalam data ini. Selain itu, dilakukan analisis korelasi menggunakan Correlation matrix untuk memastikan bahwa tidak ada dua atribut yang terlalu erat berkaitan. Dalam analisis korelasi ini, metode yang digunakan adalah Correlation matrix. padaCorrelation'matrix, kita bisa mengetahui indeks korelasi antara berbagai atribut pada dataset. Penggunaan metode ini memungkinkan kita untuk memahami sejauh mana keterkaitan dan interaksi antar atribut yang dianalisis [22].

2.3.4 Transformasi Data

Setelah proses normalisasi dan pembersihan data, langkah selanjutnya adalah mentransformasikan data menggunakan aplikasi RapidMiner. Proses ini mempermudah pembuatan model, khususnya dalam mengelompokkan pencari kerja berdasarkan tingkat pendidikan. Tahap ini memastikan bahwa data berada dalam format yang tepat untuk dianalisis menggunakan algoritma K-Means clustering dengan bantuan alat RapidMiner [23].

2.3.5 Penggunaan RapidMiner

Gambar 3 merupakan penerapan RapidMiner pada penelitian ini, berikut penggunaannya yaitu Retrieve untuk mengambil data yang telah disimpan dalam local repository, Normalize untuk menormalkan nilai atribut untuk memastikan skala yang sama pada semua atribut, Correlation matrix untuk menganalisis korelasi antar-atribut dan mengidentifikasi atribut yang memiliki korelasi tinggi, Clustering untuk menggunakan algoritma K-Means dengan nilai K=2 hingga K=5 untuk menemukan jumlah cluster optimal. Dengan semua tahap persiapan ini, data telah siap untuk dianalisis lebih lanjut, dan hasil dari proses clustering dapat memberikan insight yang lebih akurat mengenai kebutuhan pembangunan pendidikan prasekolah di Indonesia.



Gambar 3. Penggunaan RapidMiner

2.4 Modelling

Pada tahap ini, algoritma K-Means dapat digunakan demi mengelompokkan data menjadi beberapa cluster dari karakteristik yang telah ada. Proses ini dimulai dengan pemilihan algoritma K-Means karena kemampuannya dalam membagi data ke dalam kluster yang berbeda berdasarkan kedekatan atribut yang relevan. K-Means

Journal of Information System Research (JOSH)
Volume 6, No. 1, Oktober 2024, pp 251−260
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i1.5957

berfungsi dengan mengelompokkan data ke dalam sejumlah K kluster, di mana setiap data ditempatkan dalam kluster yang memiliki centroid terdekat.

Untuk menentukan jumlah Untuk menilai kualitas hasil clustering, digunakan Davies-Bouldin Index (DBI) demi menentukan cluster yang maksimal. Indeks ini mengevaluasi seberapa baik suatu cluster terbentuk, sebagaimana nilai DBI begitu rendah memperlihatkan kualitas clustering yang lebih baik. Penelitian ini menguji berbagai nilai K, mulai dari 2 hingga 5, dan menemukan bahwa nilai K yang paling sesuai adalah 3, pada nilai DBI terendah sebesar 0.205.

Setelah menentukan jumlah cluster yang optimal, algoritma K-Means diterapkan dengan presentase yang telah dipersiapkan dan dinormalisasi. Hal ini melibatkan beberapa literasi demi menemukan centroid pada tiap-tiap cluster dan mengkategorikan data ke dalam cluster yang tepat. Evaluasi kembali dari clustering dilakukan menggunakan DBI, di mana nilai DBI terendah pada K=3 menandakan bahwa hasil clustering efektif dalam memisahkan data berdasarkan karakteristik yang ada.

Hasil clustering ini membentuk beberapa cluster yang menunjukkan kelompok wilayah dengan berbagai kebutuhan dalam pembangunan prasekolah. Setiap cluster dianalisis menurut karakteristik utama seperti jumlah peserta didik, jumlah sekolah, IPM, persentase kemiskinan, dan tingkat pengangguran. Analisis lebih lanjut dilakukan untuk menemukan pola dan tren dalam data, dengan bantuan visualisasi dalam bentuk plot dan grafik untuk mempermudah pemahaman distribusi dan karakteristik tiap cluster. Selain itu, centroid dari setiap cluster disediakan untuk memberikan pemahaman yang lebih mendalam mengenai fitur utama yang membedakan masing-masing cluster.

2.5 Evaluation

Setelah model clustering selesai diterapkan, langkah selanjutnya adalah mengevaluasi kualitas hasil clustering tersebut. Pengukuran kualitas ini dilakukan dengan menggunakan indeks Davies-Bouldin (DBI), yang merupakan indikator untuk menilai seberapa baik objek-objek dalam satu cluster memiliki kemiripan dan seberapa berbeda mereka dari objek di cluster lain. Indeks Davies-Bouldin (DBI) tersebut mengukur rasio antara jarak internal cluster (jarak dalam cluster) dan jarak antara cluster (antar cluster). Nilai DBI yang paling sedikit memperlihatkan kualitas clustering untuk lebih baik.

2.6 Deployment

Tahap terakhir adalah penerapan hasil analisis dalam bentuk rekomendasi yang spesifik. Berdasarkan hasil clustering yang telah dilakukan, wilayah-wilayah yang membutuhkan peningkatan infrastruktur pendidikan prasekolah akan diidentifikasi dengan lebih jelas.

Rekomendasi yang disusun diharapkan dapat membantu pihak pemerintah atau pemangku kebijakan dalam merencanakan strategi pembangunan pendidikan yang lebih terfokus dan efektif. Dengan memanfaatkan hasil clustering, pihak berwenang untuk mengalokasikan sumber daya dengan lebih efisien dan menetapkan prioritas pembangunan berdasarkan kebutuhan spesifik setiap wilayah.

Untuk menerapkan metode clustering dan algoritma K-Means melalui tahapan CRISP-DM, penelitian ini diharapkan memberikan gambaran yang komprehensif mengenai kebutuhan pembangunan prasekolah pada wilayah Indonesia. Maka output penelitian ini dapat diharapkan bersinergi untuk mendukung pengambilan keputusan yang lebih baik, meratakan akses pendidikan prasekolah, serta meningkatkan mutu pendidikan di seluruh daerah Indonesia.

3. HASIL DAN PEMBAHASAN

Dari analisis korelasi menggunakan Correlation matrix terdapat dua atribut yang terlalu erat berkaitan, bagian ini bertujuan dalam pembersihan data. Berikut hasil nilai korelasi untuk mengidentifikasi hubungan antar-atribut:

Attributes	Jumlah Peserta Didik (TK)	Jumlah IPM	Persentase Kemiskinan	Jumlah Penduduk	Tingkat Pengangguran	Jumlah Sekolah (TK)
Jumlah Peserta Didik (TK)	1	-0.156	0.186	-0.173	-0.031	-0.302
Jumlah IPM	-0.156	1	-0.702	0.254	0.461	0.280
Persentase Kemiskinan	0.186	-0.702	1	-0.199	-0.342	-0.228
Jumlah Penduduk	-0.173	0.254	-0.199	1	0.351	0.710
Tingkat Pengangguran	-0.031	0.461	-0.342	0.351	1	0.222
Jumlah Sekolah (TK)	-0.302	0.280	-0.228	0.710	0.222	1

Gambar 4. Nilai Korelasi

Tabel 2. Nilai Korelasi Tertinggi

Atribut 1	Atribut 2	Nilai Korelasi
Jumlah Penduduk	Jumlah Sekolah (TK)	0,710

Gambar 4 menunjukkan semua hasil korelasi yang ada pada setiap atribut dan Tabel 2 menunjukkan denga terdapat korelasi yang begitu kuat antara jumlah penduduk dan jumlah sekolah TK di suatu wilayah dengan nilai

Journal of Information System Research (JOSH)
Volume 6, No. 1, Oktober 2024, pp 251−260
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i1.5957

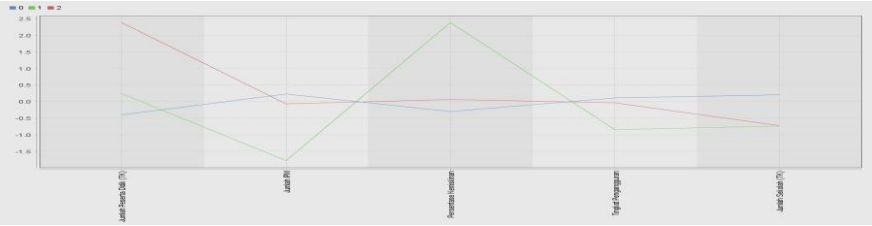
korelasi sebesar 0.710. Ini mengindikasikan bahwa wilayah dengan jumlah penduduk yang lebih besar cenderung memiliki jumlah sekolah TK yang lebih banyak, meskipun jumlah peserta didik yang terdaftar bisa jadi tidak sesuai dengan kapasitas sekolah yang ada. Maka, atribut yang dihapus yaitu jumlah penduduk.

Tabel 3. Nilai K

K	Hasil DBI	Measure Types	Divergence
K=2	0,278	Bregman Divergences	Squared'Euclidean Distance
K=3	0,205	Bregman Divergences	Squared Euclidean Distance
K=4	0,233	Bregman Divergences	Squared Euclidean Distance
K=5	0,212	Bregman Divergences	Squared Euclidean Distance

Dalam penelitian ini, DBI dihitung untuk setiap nilai K yang diuji, dengan rentang nilai K dari 2 hingga 5. Berdasarkan Tabel 3, hasil perhitungan DBI, nilai K yang optimal adalah yang memberikan nilai DBI terendah. Dalam penelitian ini, nilai DBI terendah diperoleh pada K=3 dengan nilai DBI sebesar 0.205. hal ini memperlihatkan bahwa clustering dengan K=3 adalah yang paling optimal, di mana objek-objek dalam cluster memiliki kemiripan yang lebih tinggi dan perbedaan antar cluster lebih jelas dibandingkan dengan nilai K lainnya. Dengan nilai DBI yang lebih rendah pada K=3, dapat disimpulkan bahwa hasil clustering pada K=3 memiliki kualitas yang lebih baik. Ini berarti bahwa cluster yang terbentuk memiliki konsistensi internal yang lebih tinggi dan perbedaan yang jelas antar cluster, sesuai dengan karakteristik data yang dianalisis.

Hasil clustering mengidentifikasi tiga cluster utama dengan karakteristik yang berbeda. Berdasarkan hasil tersebut, visualisasi hasil clustering dapat melakukan pemahaman distribusi data dan karakteristik setiap cluster dengan lebih jelas. Berikut plot yang menunjukkan visualisasi cluster:



Gambar 5. Visualisasi Cluster

Pada Gambar 5 dapat dilihat bahwa Cluster 0 (C0) memiliki wilayah dengan jumlah peserta didik yang cukup rendah dan jumlah sekolah yang tinggi, Cluster 1 (C1) memiliki wilayah dengan jumlah peserta didik yang standar namun jumlah sekolah yang rendah, dengan tingkat kemiskinan tinggi dan IPM rendah, dan Cluster 2 (C2) memiliki wilayah dengan jumlah peserta didik yang sangat tinggi dan jumlah sekolah yang rendah, meskipun IPM, kemiskinan, dan tingkat pengangguran berada pada standar. Berikut Gambar 6 merupakan kabupaten/kota di Indonesia pada setiap cluster yaitu cluster 0 memiliki 402 kabupaten/kota, cluster 1 memiliki total 49 kabupaten/kota, dan cluster 2 memiliki 63 kabupaten/kota, dengan total sampel 514 data pada kabupaten atau kota yang ada di Indonesia:

Cluster	Kabupaten/Kota						
2	Kota Sabang	Kab. Empat Lawang	Kota Banjar	Kab.Maluku Tenggara	Kota Gunungsitoli	Kab.Simalungun	Kab.Barru
	Kab. Batubara	Kab. Lahat	Kota Tegal	Kota Tual	Kota Padang Sidimpuan	Kab.Tapanuli Tengah	Kab.Kepulauan Selayar
	Kab. Dairi	Kab. Musi Rawas Utara	Kab. Belu	Kab.Kepulauan Morotai	Kota Pematangsiantar	Kab.Tapanuli Utara	Kab.Soppeng
	Kab. Humbang Hasudutan	Kab. Ogan Ilir	Kab. Malaka	Kab.Biak Numfor	Kota Tanjung Balai	Kota Binjai	Kab.Tana Toraja
	Kab. Karo	Kab. Ogan Komering Ulu	Kab. Manggarai Barat	Kab.Kepulauan Yapen	Kab.Lingga	Kota Pagar Alam	Kab.Toraja Utara
	Kab. Labuhan Batu	Kab. Ogan Komering Ulu Selatan	Kab. Timor Tengah Utara	Kab.Sarmi	Kab.Kepulauan Seribu	Kota Prabumulih	Kab.Enrekang
	Kab. Labuhan Batu Utara	Kab. Penukal Abab Lematang Ilir	Kab. Mahakam Ulu	Kota Palopo	Kota Cilegon	Kab.Bengkulu Tengah	Kab.Bone
	Kab. Mandailing Natal	Kab. Padang Lawas utara	Kab. Kepulauan Siau Tagulandang Biaro	Kab.Nias Selatan	Kota Sukabumi	Kota Lubuk Linggau	Kab.Nias Selatan
1	Kab. Nias Barat	Kab. Rote-Ndao	Kab. Supiori	Kab.Deiyai	Kab.Lembata	Kab.Manokwari Selatan	Kab.Sorong
	Kab. Nias Utara	Kab. Sabu Raijua	Kab. Waropen	Kab.Dogiyai	Kab.Manggarai	Kab.Pegunungan Arfak	Kab.Tambrauw
	Kab. Kepulauan Meranti	Kab. Sumba Barat	Kab. Membramo Raya	Kab.Puncak Jaya	Kab.Manggarai Timur	Kab.Puncak	Kab.Maybrat
	Kab. Sampang	Kab. Sumba Timur	Kab. Mappi	Kab.Pegunungan Bintang	Kab.Sumba Barat Daya	Kab.Yahukimo	Kab.Mamberamo Tengah
	Kab. Lombok Utara	Kab. Timor Tengah Selatan	Kab. Asmat	Kab.Yalimo	Kab.Sumba Tengah	Kab.Jayawijaya	Kab.Teluk Wondama
	Kab. Alor	Kab. Kepulauan Tanimbar	Kab. Paniai	Kab.Nduga	Kab.Seram Bagian Timur	Kab.Nabire	Kab.Kupang
	Kab. Ende						
0	Kab. Aceh Barat	Kab. Rokan Hilir	Kota Jakarta Timur	Kota Pekalongan	Kota Tidore Kepulauan	Kab.Temanggung	Kab.Karimun
	Kab. Aceh Barat Daya	Kab. Rokan Hulu	Kota Jakarta Barat	Kota Salatiga	Kota Jayapura	Kab.Wonogiri	Kab.Natuna
	Kab. Aceh Besar	Kab. Siak	Kota Jakarta Selatan	Kota Semarang	Kab.Jayapura	Kab.Wonosobo	Kota Batam
	Kab. Aceh Jaya	Kota Dumai	Kota Jakarta Utara	Kota Surakarta	Kab.Keerom	Kota Magelang	Kota Tanjungpinang
	Kab. Aceh Selatan	Kota Pekanbaru	Kota Jakarta Pusat	Kab.Sleman	Kab.Mandobwari	Kab.Bulukumba	Kab.Hulu Sungai Selatan
	Kab. Aceh Singkil	Kab. Batang Hari	Kab. Bogor	Kab.Bantul	Kab.Kaimana	Kab.Enrekang	Kab.Hulu Sungai Tengah
	Kab. Aceh Tamiang	Kab. Bungo	Kab. Bandung	Kab.Gunung Kidul	Kab.Merauke	Kab.Gowa	Kab.Hulu Sungai Utara
	Kab. Aceh Tengah	Kab. Kerinci	Kab. Bekasi	Kota Yogyakarta	Kab.Mimika	Kab.Jeneponto	Kab.Kotabaru
	Kab. Aceh Tenggara	Kab. Merangin	Kab. Garut	Kab.Kulon Progo	Kota Sorong	Kab.Luwu	Kab.Tabalong
	Kab. Aceh Timur	Kab. Muaro Jambi	Kab. Sukabumi	Kota Surabaya	Kab.Halmahera Timur	Kab.Luwu Timur	Kab.Tanah Bumbu
	Kab. Aceh Utara	Kab. Sarolangun	Kab. Ciamjur	Kab.Bangkalan	Kab.halmahera Utara	Kab.Luwu Utara	Kab.Tanah Laut
	Kab. Bener Meriah	Kab. Tanjung Jabung Barat	Kota Bekasi	Kab.Banyuwangi	Kab.Kepulauan Sula	Kab.Maros	Kab.Tapin
	Kab. Bireuen	Kab. Tanjung Jabung Timur	Kab. Karawang	Kab.Blitar	Kab.Pangkajene Kepulauan	Kab.Pinrang	Kota Banjarbaru
	Kab. Gayo Lues	Kab. Tebo	Kota Bandung	Kab.Bojonegoro	Kab.Sigi	Kab.Pinrang	Kota Banjarmasin
	Kab. Nagan Raya	Kota Jambi	Kab. Cirebon	Kab.Bondowoso	Kab.Toyo Una-Una	Kab.Sidenreng Rappang	Kab.Berau
	Kab. Pidie	Kota Sungai Penuh	Kota Depok	Kab.Gresik	Kab.Tolitoli	Kab.Sinjai	Kab.Kutai Barat
	Kab. Pidie Jaya	Kab. Banyuasin	Kab. Indramayu	Kab.Jember	Kota Palu	Kab.Takalar	Kab.Kutai Kartanegara
	Kab. Simeulue	Kab. Muara Enim	Kab. Tasikmalaya	Kab.Jombang	Kab.Bantaeng	Kab.Wajo	Kab.Kutai Timur
	Kota Banda Aceh	Kab. Musi Banyuasin	Kab. Subang	Kab.Kediri	Kab.Bone	Kota Makassar	Kab.Paser
	Kota Langsa	Kab. Musi Rawas	Kab. Bandung Barat	Kab.Lamongan	Kab.Balangan	Kab.Bombana	Kab.Penajam Paser Utara
	Kota Lhokseumawe	Kab. Ogan Komering Ilir	Kab. Majalengka	Kab.Lumajang	Kab.Banjar	Kab.Buton	Kota Balikpapan
	Kota Subulussalam	Kab. Ogan Komering Ulu Timur	Kab. Kuningan	Kab.Madiun	Kab.Barito Kuala	Kab.Buton Selatan	Kota Bontang
	Kab. Asahan	Kota Palembang	Kab. Sumedang	Kab.Magetan	Kota Padang Panjang	Kab.Buton Tengah	Kota Samarinda
	Kab. Deli Serdang	Kab. Bengkulu Selatan	Kota Bogor	Kab.Malang	Kota Pariaman	Kab.Buton Utara	Kota Tarakan
	Kab. Labuhan Batu Selatan	Kab. Bengkulu Utara	Kab. Purwakarta	Kab.Mojokerto	Kota Payakumbuh	Kab.Kolaka	Kab.Nunukan
	Kab. Langkat	Kab. Kaur	Kab. Ciamis	Kab.Nganjuk	Kota Solok	Kab.Kolaka Timur	Kab.Bulungan
	Kab. Nias	Kab. Kepahiang	Kota Tasikmalaya	Kab.Ngawi	Kab.Bengkalis	Kab.Kolaka Utara	Kab.Malinau
	Kab. Padang Lawas	Kab. Lebong	Kab. Pangandaran	Kab.Pacitan	Kab.Indragiri Hili	Kab.Konawe	Kab.Tana Tidung
	Kab. Pakpak Bharat	Kab. Muko-muko	Kota Cirebon	Kab.Pamekasan	Kab.Bangka Selatan	Kab.Konawe Selatan	Kab.Bolaang Mongondow
	Kab. Samosir	Kab. Rejang Lebong	Kab. Banjarnegara	Kab.Pasuruan	Kab.Bangka Tengah	Kab.Konawe Utara	Kab.Bolaang Mongondow Se
	Kab. Serdang Bedagai	Kab. Seluma	Kab. Banyumas	Kab.Ponorogo	Kab.Belitung	Kab.Muna	Kab.Bolaang Mongondow Ti
	Kab. Tapanuli Selatan	Kota Bengkulu	Kab. Batang	Kab.Probolinggo	Kab.Belitung Timur	Kab.Muna Barat	Kab.Bolaang Mongondow Ut
	Kab. Toba	Kab. Lampung Barat	Kab. Blora	Kab.Sidoarjo	Kab.Pangkalpinang	Kab.Wakatobi	Kota Kep. Sangihe
	Kota Medan	Kab. Lampung Selatan	Kab. Boyolali	Kab.Situbondo	Kab.Bintan	Kota Baubau	Kab.Kepulauan Talaud
	Kota Sibolga	Kab. Lampung Tengah	Kab. Brebes	Kab.Sumenep	Kab.Purwoejo	Kota Kendari	Kab.Minahasa
	Kab. Agam	Kab. Lampung Timur	Kab. Cilacap	Kab.Trenggalek	Kab.Rembang	Kab.Boalemo	Kab.Minahasa Selatan
	Kab. Dharmasraya	Kab. Lampung Utara	Kab. Demak	Kab.Tuban	Kab.Semarang	Kab.Bone Bolango	Kab.Minahasa Tenggara
	Kab. Kepulauan Mentawai	Kab. Mesuji	Kab. Grobogan	Kab.Tulungagung	Kab.Sragen	Kab.Gorontalo	Kab.Minahasa Utara
	Kab. Lima Puluh Koto	Kab. Pesawaran	Kab. Jepara	Kota Batu	Kab.Sukoharjo	Kab.Gorontalo Utara	Kota Bitung
	Kab. Padang Pariaman	Kab. Pesisir Barat	Kab. Karanganyar	Kota Blitar	Kab.Tegal	Kab.Pohuwato	Kota Kotamobagu
	Kab. Pasaman	Kab. Pringsewu	Kab. Kebumen	Kota Kediri	Kab.Tangerang	Kota Gorontalo	Kota Manado
	Kab. Pasaman Barat	Kab. Tanggamus	Kab. Kendal	Kota Madiun	Kota Cilegon	Kab.Polewali Mandar	Kota Tomohon
	Kab. Pesisir Selatan	Kab. Tulang Bawang	Kab. Klaten	Kota Malang	Kota Serang	Kab.Mamuju	Kab.Banggai
	Kab. Sijunjung	Kab. Tulang Bawang Barat	Kab. Kudus	Kota Mojokerto	Kota Tangerang	Kab.Majene	Kab.Banggai Kepulauan
	Kab. Solok	Kab. Way Kanan	Kab. Magelang	Kota Pasuruan	Kota Tangerang Selatan	Kab.Mamasa	Kab.Banggai Laut
	Kab. Solok Selatan	Kota Bandar Lampung	Kab. Pati	Kota Probolinggo	Kota Ternate	Kab.Pasangkayu	Kab.Buol
	Kab. Tanah Datar	Kota Metro	Kab. Pekalongan	Kab.Lebak	Kab.Halmahera Barat	Kab.Mamuju Tengah	Kab.Donggala
	Kota Bukittinggi	Kab. Bangka	Kab. Pemalang	Kab.Pandeglang	Kab.Halmahera Selatan	Kab.Buru	Kab.Konawaringin Timur
	Kota Padang	Kab. Bangka Barat	Kab. Purbalingga	Kab.Serang	Kab.Halmahera Tengah	Kab.Buru Selatan	Kab.Lamandau
	Kab. Pulang Pisau	Kab. Morowali	Kota Ambon	Kab.Parigi Moutong	Kab.Sukamara	Kota Palangka Raya	Kab.Poso
	Kab. Sengan	Kab. Morowali Utara					

Gambar 6. Kabupaten/Kota Disetiap Cluster

Journal of Information System Research (JOSH)
Volume 6, No. 1, Oktober 2024, pp 251−260
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i1.5957

Berdasarkan Gambar 6 yang telah menjabarkan kabupaten/kota pada tiga cluster tersebut, berikut adalah karakteristik dan analisis dari setiap cluster:

3.1 Cluster 0 (Wilayah dengan peserta didik rendah, jumlah sekolah cukup tinggi)

Cluster 0 terdiri dari 402 kabupaten/kota yang memiliki karakteristik peserta didik TK yang relatif rendah, sementara jumlah sekolah yang tersedia cukup tinggi. Meski infrastruktur TK cukup memadai, rendahnya partisipasi peserta didik menandakan perlunya inisiatif untuk meningkatkan kesadaran dan partisipasi masyarakat dalam pendidikan prasekolah.

Beberapa daerah di cluster ini juga memiliki tingkat pengangguran yang cukup tinggi, yang dapat berdampak pada kemampuan keluarga untuk menyekolahkan anak mereka. Selain itu, data menunjukkan bahwa cluster ini memiliki aksesibilitas yang baik terhadap fasilitas pendidikan yang ada, tetapi tetap menghadapi masalah dalam hal partisipasi. Fokus utama pada wilayah-wilayah ini adalah meningkatkan partisipasi peserta didik melalui kampanye kesadaran tentang pentingnya pendidikan prasekolah, pemberian bantuan pendidikan bagi keluarga kurang mampu, serta program-program yang mendukung aksesibilitas. Selain itu, upaya untuk mengurangi tingkat pengangguran juga dapat membantu meningkatkan kemampuan keluarga dalam mendukung pendidikan anak.

3.2 Cluster 1 (Wilayah dengan peserta didik standar, jumlah sekolah rendah)

Cluster 1 mencakup 49 kabupaten/kota yang memiliki jumlah peserta didik yang tergolong standar, namun infrastruktur sekolah TK masih kurang memadai. Wilayah-wilayah ini umumnya memiliki kondisi sosial-ekonomi yang lebih lemah, dengan tingkatan kemiskinan yang begitu tinggi dan Indeks Pembangunan Manusia (IPM) yang dibawah, meskipun tingkat pengangguran relatif lebih baik.

Data menunjukkan bahwa di cluster ini terdapat kekurangan infrastruktur yang signifikan dalam hal jumlah sekolah TK yang tersedia dibandingkan dengan jumlah peserta didik. Selain itu, adanya ketidakmerataan dalam distribusi sumber daya pendidikan dapat memperburuk kondisi pendidikan di wilayah ini. Daerah-daerah dalam cluster ini memerlukan pembangunan sekolah baru untuk memenuhi kebutuhan infrastruktur pendidikan prasekolah. Pembangunan sekolah juga harus mempertimbangkan kemampuan ekonomi masyarakat lokal, sehingga biaya pendidikan bisa dijangkau seluruh lapisan masyarakat. maka, program pelatihan bagi tenaga pengajar juga penting untuk memastikan kualitas pendidikan yang diterima peserta didik.

3.3 Cluster 2 (Wilayah dengan peserta didik sangat tinggi, jumlah sekolah rendah)

Cluster 2 terdiri dari 63 kabupaten/kota yang memiliki jumlah peserta didik TK yang sangat tinggi, tetapi jumlah sekolah masih sangat terbatas. Hal ini mengindikasikan adanya kebutuhan mendesak akan pembangunan sekolah TK untuk menampung peserta didik yang ada. Meskipun IPM dan tingkat kemiskinan di wilayah ini tergolong standar, kapasitas infrastruktur yang rendah mempengaruhi kualitas pendidikan prasekolah yang dapat diberikan.

Data menunjukkan bahwa kekurangan jumlah sekolah TK di cluster ini menyebabkan kepadatan peserta didik yang tinggi di setiap sekolah, yang dapat mempengaruhi kualitas pendidikan dan perhatian individual yang diterima peserta didik. Infrastruktur yang terbatas ini berpotensi menjadi kendala besar dalam pencapaian tujuan pendidikan di wilayah ini. Pembangunan sekolah TK harus menjadi prioritas utama di wilayah-wilayah ini, mengingat tingginya jumlah peserta didik yang tidak dapat tertampung oleh infrastruktur yang ada. Penambahan sekolah di daerah dengan populasi peserta didik yang tinggi sangat dibutuhkan untuk memenuhi permintaan yang ada. Pengalokasian anggaran untuk pembangunan sekolah baru dan peningkatan fasilitas pendidikan juga harus diperhatikan. Selain itu, program-program bantuan untuk keluarga kurang mampu dapat membantu meningkatkan akses pendidikan.

Setiap cluster memiliki karakteristik masing-masing dan rekomendasi pembangunan prasekolah. Setelah melakukan analisis berdasarkan hasil yang didapat, maka dari itu berikut rekomendasi disetiap cluster yang menjadi hal utama dari setiap cluster:

Tabel 4. Rekomendasi Cluster

Cluster	Peserta Didik	Pembangunan Sekolah	Saran
0	Rendah	Tinggi	Tingkatkan partisipasi
1	Standar	Rendah	Tambah sekolah
2	Sangat Tinggi	Sangat Rendah	Prioritas pembangunan

Tabel 4 memberikan ringkasan informasi utama untuk setiap cluster. Cluster 0 memiliki peserta didik TK yang rendah tetapi jumlah sekolah tinggi. Rekomendasi adalah untuk meningkatkan partisipasi peserta didik melalui kampanye kesadaran dan bantuan pendidikan. Cluster 1 memiliki jumlah peserta didik standar dan jumlah sekolah rendah. Rekomendasi adalah untuk membangun lebih banyak sekolah TK. Cluster 2 memiliki jumlah peserta didik sangat tinggi tetapi jumlah sekolah sangat rendah. Pembangunan sekolah TK harus menjadi prioritas utama di wilayah ini.

Journal of Information System Research (JOSH)
Volume 6, No. 1, Oktober 2024, pp 251−260
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i1.5957

4. KESIMPULAN

Berdasarkan hasil penelitian yang memanfaatkan algoritma pengelompokan K-Means serta menggunakan indeks Davies-Bouldin (DBI) sebesar 0.205, ditemukan tiga kelompok wilayah dengan karakteristik yang berbeda. Cluster 0 terdiri dari 402 kabupaten/kota yang menunjukkan jumlah peserta didik yang rendah namun memiliki jumlah sekolah yang memadai. Wilayah ini perlu fokus pada peningkatan partisipasi peserta didik, dengan perhatian khusus pada masalah sosial seperti tingginya tingkat pengangguran. Sebaliknya, Cluster 1 mencakup 49 kabupaten/kota dengan jumlah peserta didik yang standar tetapi jumlah sekolah yang masih kurang, sehingga memerlukan penambahan fasilitas pendidikan untuk memenuhi kebutuhan yang ada. Cluster 2, yang meliputi 63 kabupaten/kota, menghadapi masalah serius dengan jumlah peserta didik yang sangat tinggi dan fasilitas sekolah yang sangat terbatas. Daerah-daerah dalam cluster ini memerlukan perhatian segera untuk membangun lebih banyak sekolah agar dapat menampung jumlah peserta didik yang terus bertambah. Hasil keseluruhan menunjukkan perlunya distribusi pembangunan TK yang lebih seimbang di seluruh Indonesia, dengan Cluster 1 dan Cluster 2 sebagai prioritas utama untuk pengembangan infrastruktur baru.

REFERENCES

[1] R. Adhitama, A. Burhanuddin, and R. Ananda, "Penentuan Jumlah Cluster Ideal Smk Di Jawa Tengah Dengan Metode X-Means Clustering Dan K-Means Clustering Determining Vocational Ideal Cluster Number in Central Java With X-Means Clustering and K-Means Clustering Methods," J. Inform. dan Komputer) Akreditasi KEMENRISTEKDIKTI, vol. 3, no. 1, pp. 1–5, 2020, doi: 10.33387/jiko.

[2] Abdussalam Amrullah, Intam Purnamasari, Betha Nurina Sari, Garno, and Apriade Voutama, "Analisis Cluster Faktor Penunjang Pendidikan Menggunakan Algoritma K-Means (Studi Kasus: Kabupaten Karawang)," J. Inform. dan Rekayasa Elektron., vol. 5, no. 2, pp. 244–252, 2022, doi: 10.36595/jire.v5i2.701.

[3] N. D. Wahyuni, "Pengaruh Tingkat Pendidikan, Pekerjaan, Jumlah Tanggungan Anak, dan Pendapatan Orangtua terhadap Kemampuan Memenuhi Kebutuhan Pendidikan Anak di Indonesia," J. Pendidik. dan Ekon., vol. 9, no. 3, pp. 204–212, 2020.

[4] E. Yanty, T. Putri, and I. Kamila, "Pendukung Pendidikan Dengan Jumlah Sekolah Dan Jumlah Guru Menggunakan Algoritma K-Means," J. Ilm. Mat., vol. 2, no. 1, pp. 1–12, 2022.

[5] M. M. Astriani and M. A. Alfahnum, "Peningkatan Kompetensi Guru PAUD dalam Mengembangkan Media Pembelajaran Inovatif," J. PkM Pengabdi. Kpd. Masy., vol. 3, no. 4, p. 366, 2020, doi: 10.30998/jurnalpkm.v3i4.8151.

[6] E. Sugian, F. Fahruddin, and A. H. Witono, "Implementasi Program Pengembangan PAUD "Holistik Integratif"di PAUD LSM Ampenan Kota Mataram," J. Ilm. Mandala Educ., vol. 7, no. 3, pp. 675–685, 2021, doi: 10.58258/jime.v7i3.2342.

[7] Ramadhana, Islamiyah, and A. P. A. Masa, "Penerapan Data Mining Menggunakan Metode K-Means Clustering Pada Data Ekspor Batubara," Adopsi Teknol. dan Sist. Inf., vol. 2, no. 1, pp. 35–42, 2023, doi: 10.30872/atasi.v2i1.595.

[8] N. Hendrastuty, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa," J. Ilm. Inform. Dan Ilmu Komput., vol. 3, no. 1, pp. 46–56, 2024, [Online]. Available: https://doi.org/10.58602/jima-ilkom.v3i1.26

[9] M. Rachman Mulyandi et al., "Implementasi Algoritma K-Means Clustering Dalam," J. Student Res., vol. 1, no. 3, pp. 101–114, 2023.

[10] K. S. Purba, D. Hartama, and S. Suhada, "Analisis Data Mining Pesebaran Siswa Smp Di Pematangsiantar Dengan Metode Algoritma K-Means Clustering," Kesatria J. Penerapan Sist. Inf. (Komputer dan Manajemen), vol. 3, no. 1, pp. 1–8, 2022, doi: 10.30645/kesatria.v3i1.91.

[11] F. Martha and D. Anggraini, "Data Mining Untuk Pemeliharaan Prediktif Mesin Produksi berdasarkan Database Kerusakan Mesin menggunakan Naïve Bayes Classifier," J. Ilm. Komputasi, vol. 20, no. 2, pp. 143–154, 2021, doi: 10.32409/jikstik.20.2.368.

[12] Z. Muttaqin, "Implementasi Unsupervised Learning Pada Nilai Jasmani Kesamaptaan Sekolah Polisi Negara Dengan Metode Clustering Analysis," PROSISKO J. Pengemb. Ris. dan Obs. Sist. Komput., vol. 10, no. 1, pp. 18–23, 2023, doi: 10.30656/prosisko.v10i1.6269.

[13] M. Rafi Nahjan, Nono Heryana, and Apriade Voutama, "Implementasi Rapidminer Dengan Metode Clustering K-Means Untuk Analisa Penjualan Pada Toko Oj Cell," JATI (Jurnal Mhs. Tek. Inform., vol. 7, no. 1, pp. 101–104, 2023, doi: 10.36040/jati.v7i1.6094.

[14] M. Yamin Nurzaman and B. Nurina Sari, "Implementasi K-Means Clustering Dalam Pengelompokkan Banyaknya Jumlah Petani Berdasarkan Kecamatan Di Provinsi Jawa Barat," J. Tek. Inform. dan Sist. Inf., vol. 10, no. 3, pp. 131–144, 2023, [Online]. Available: http://jurnal.mdp.ac.id

[15] I. Ramadhani and M. Megawati, "Implementasi Algoritma K-Means Untuk Klustering Data Produktivitas Kelapa Sawit," Indones. J. Inform. Res. Softw. Eng., vol. 3, no. 1, pp. 56–64, 2023, doi: 10.57152/ijirse.v3i1.488.

[16] D. Cahya, M. Hidayat, and F. Asnawi, "Journal of Engineering and Informatic Implementasi Algoritma K-Means Clustering untuk Menentukan Calon," vol. 1, no. 1, pp. 28–34, 2022, doi: 10.56854/jei.v1i1.16.

[17] E. Muningsih, N. Hasan, and G. B. Sulistyo, "Penerapan Metode Principle Component Analysis (PCA) untuk Clustering Data Kunjungan Wisatawan Mancanegara ke Indonesia," Bianglala Inform., vol. 8, no. 1, pp. 58–62, 2020, doi: 10.31294/bi.v8i1.8470.

[18] J. Nasir, "Penerapan Data Mining Clustering Dalam Mengelompokan Buku Dengan Metode K-Means," Simetris J. Tek. Mesin, Elektro dan Ilmu Komput., vol. 11, no. 2, pp. 690–703, 2021, doi: 10.24176/simet.v11i2.5482.

[19] A. T. Basalamah and R. Setyadi, "Penerapan Algoritma K-Means Clustering Pada Tingkat Penyelesaian Pendidikan Di Provinsi Indonesia," J. Inform. dan Teknol. Komput., vol. 4, no. 2, pp. 114–121, 2023, [Online]. Available: https://ejurnalunsam.id/index.php/jicom/

[20] Sutisna and N. M. Yuniar, "Klasifikasi Kualitas Air Bersih Menggunakan Metode Naïve baiyes," J. Sains dan Teknol.,

Journal of Information System Research (JOSH)
Volume 6, No. 1, Oktober 2024, pp 251−260
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i1.5957

vol. 5, no. 1, pp. 243–246, 2023, [Online]. Available: https://doi.org/10.55338/saintek.v5i1.1383

[21] Y. Suhanda, I. Kurniati, and S. Norma, "Penerapan Metode Crisp-DM Dengan Algoritma K-Means Clustering Untuk Segmentasi Mahasiswa Berdasarkan Kualitas Akademik," J. Teknol. Inform. dan Komput., vol. 6, no. 2, pp. 12–20, 2020, doi: 10.37012/jtik.v6i2.299.

[22] E. N. R. Khakim, A. Hermawan, and D. Avianto, "Implementasi Correlation Matrix Pada Klasifikasi Dataset Wine," JIKO (Jurnal Inform. dan Komputer), vol. 7, no. 1, p. 158, 2023, doi: 10.26798/jiko.v7i1.771.

[23] S. R. Hani, "Clustering Data Pencari Kerja Menurut Tingkat Pendidikan Menggunakan Algoritma K-Means," J. Minfo Polgan, vol. 12, no. 1, pp. 1–14, 2023, doi: 10.33395/jmp.v12i1.12217.