



JURNAL ILMIAH MULTIDISIPLIN ILMU

Halaman Jurnal: <https://journal.smartpublisher.id/index.php/jimi>

Halaman UTAMA Jurnal : <https://journal.smartpublisher.id/>



DOI: <https://doi.org/10.69714/bncwkr16>

PENERAPAN ALGORITMA K-MEANS PADA SMA PROVINSI DKI JAKARTA UNTUK MENENTUKAN SEKOLAH TERBAIK BERDASARKAN NILAI UN

Muhammad Hasan ^{a*}, Zaehol Fatah ^b

^a Sistem Informasi, hasan12052022@gmail.com, Universitas Ibrahimy, Situbondo Jawa Timur

^b Sistem Informasi, zaeholfatah@gmail.com, Universitas Ibrahimy, Situbondo Jawa Timur

* korespondensi

ABSTRACT

Education is an important aspect in human resource development in Indonesia. One indicator to measure the quality of education is the National Examination (UN) score. However, parents or students often have difficulty choosing the best school in DKI Jakarta Province because of the many choices available. This research aims to apply the K-Means algorithm to group high school schools in DKI Jakarta Province based on National Examination scores. By using the clustering method, it is hoped that groups of schools with the best achievements can be found, making it easier to select schools based on these criteria. In this research, the data used are high school National Examination scores in DKI Jakarta obtained from the Ministry of Education and Culture. The results of this research show that the K-Means algorithm can be effectively used to group schools based on National Examination scores, thereby providing a clearer picture for the public in determining quality schools.

Keywords: K-Means Algorithm, Best Schools, National Examination Scores, Classification, Clustering, Education, DKI Jakarta.

Abstrak

Pendidikan merupakan salah satu aspek penting dalam pembangunan sumber daya manusia di Indonesia. Salah satu indikator untuk mengukur kualitas pendidikan adalah nilai Ujian Nasional (UN). Namun, sering kali orangtua atau siswa kesulitan dalam memilih sekolah terbaik di Provinsi DKI Jakarta karena banyaknya pilihan yang ada. Penelitian ini bertujuan untuk menerapkan algoritma K-Means dalam mengelompokkan sekolah-sekolah SMA di Provinsi DKI Jakarta berdasarkan nilai UN. Dengan menggunakan metode clustering, diharapkan dapat ditemukan kelompok sekolah yang memiliki prestasi terbaik, sehingga memudahkan dalam pemilihan sekolah berdasarkan kriteria tersebut. Dalam penelitian ini, data yang digunakan adalah nilai UN SMA di DKI Jakarta yang diperoleh dari Kementerian Pendidikan dan Kebudayaan. Hasil dari penelitian ini menunjukkan bahwa algoritma K-Means dapat efektif digunakan untuk mengelompokkan sekolah berdasarkan nilai UN, sehingga memberikan gambaran yang lebih jelas bagi masyarakat dalam menentukan sekolah yang berkualitas.

Kata Kunci: Algoritma K-Means, Sekolah Terbaik, Nilai UN, Klasifikasi, Clustering, Pendidikan, DKI Jakarta.

1. PENDAHULUAN

Kemajuan teknologi kedokteran telah membuka peluang besar untuk mengatasi berbagai tantangan dalam mendiagnosis dan mengkategorikan kondisi kesehatan, termasuk gangguan tidur. Gangguan tidur, seperti insomnia, sleep apnea, dan lainnya merupakan kondisi medis yang memengaruhi kemampuan seseorang untuk mendapatkan kualitas atau kuantitas tidur yang cukup. Gangguan tersebut dapat ditandai dengan berbagai gejala, seperti sulit tidur, sering terbangun di malam hari, merasa lelah meskipun sudah cukup tidur, atau pola tidur tidak teratur sehingga memengaruhi fungsi tubuh sehari-hari. Menurut survei yang

dilakukan oleh National Sleep Foundation (NSF) di Amerika, orang-orang yang mengaku mengalami kesulitan tidur dilaporkan mengalami lebih banyak kesulitan berkonsentrasi, kesulitan menyelesaikan sejumlah tugas, serta sedikit emosional atau mudah tersinggung.

Algoritma K-means adalah proses pengelompokan sekumpulan objek data ke dalam beberapa himpunan bagian yang disebut cluster. Objek-objek dalam satu cluster memiliki karakteristik yang serupa, sedangkan objek-objek di cluster yang berbeda memiliki karakteristik yang berbeda satu sama lain[3]. Tujuan K-means adalah untuk menemukan struktur atau pola tersembunyi dalam data[4] sekolah-sekolah di DKI Jakarta berdasarkan nilai UN yang diperoleh oleh siswa-siswa di setiap sekolah.

Penelitian ini bertujuan untuk menggunakan algoritma K-Means dalam mengelompokkan SMA di DKI Jakarta berdasarkan hasil Ujian Nasional (UN), serta menyusun rekomendasi sekolah-sekolah unggulan yang dapat menjadi referensi bagi siswa dan orang tua dalam memilih sekolah.

2. TINJAUAN PUSTAKA

2.1. Data Mining

Data Mining adalah ekstraksi pengetahuan dari database. Data mining adalah proses mengekstraksi dan mengidentifikasi informasi berguna dan pengetahuan terikat dari database besar menggunakan teknik statistik, matematika, kecerdasan buatan, dan pembelajaran mesin. Berdasarkan definisi data mining yang dijelaskan diatas, data mining adalah pengetahuan yang disembunyikan dalam database, dan digunakan untuk mengekstrak dan mengubah informasi pengetahuan dari database dengan menemukan pola dan teknik dalam statistik matematika, kecerdasan buatan, dan pembelajaran mesin.[5]

2.2. Algoritma K-means

Algoritma K-means adalah metode pengelompokan data yang bersifat partisional, di mana data dipisahkan ke dalam kelompok-kelompok berbeda. Dalam pelaksanaannya, algoritma K-means mengharuskan setiap data terkelompok ke dalam salah satu cluster pada setiap tahapan proses. Namun, pada iterasi berikutnya, data tersebut dapat berpindah ke cluster lain jika ditemukan bahwa jaraknya lebih dekat ke centroid cluster yang baru. Proses ini akan terus berulang hingga tidak ada perubahan signifikan dalam pembagian cluster atau sampai kriteria tertentu terpenuhi.

Secara umum, keberhasilan penggunaan algoritma K-means dalam proses clustering sangat bergantung pada karakteristik data yang dianalisis serta tujuan akhir yang ingin dicapai.[6]

2.3. Nilai UN sebagai Indikator Kualitas Pendidikan

Ujian Nasional (UN/UNAS) merupakan mekanisme penilaian standar untuk pendidikan dasar dan menengah di seluruh Indonesia. Sistem ini bertujuan untuk memastikan kesetaraan kualitas pendidikan di berbagai daerah. Pelaksanaannya dikoordinasikan oleh Pusat Penilaian Pendidikan, Kementerian Pendidikan Nasional, sesuai dengan amanat Undang-Undang Republik Indonesia Nomor 20 Tahun 2003. Undang-undang tersebut menyatakan bahwa pengendalian mutu pendidikan secara nasional dilakukan melalui evaluasi sebagai bentuk tanggung jawab penyelenggara pendidikan kepada berbagai pihak yang terkait. Evaluasi ini dilakukan oleh lembaga independen secara berkala, menyeluruh, transparan, dan sistematis, guna menilai pencapaian standar nasional pendidikan. Proses pemantauan evaluasi juga harus berlangsung secara berkelanjutan untuk menjamin keberlanjutan mutu pendidikan.[7]

2.4. RapidMiner

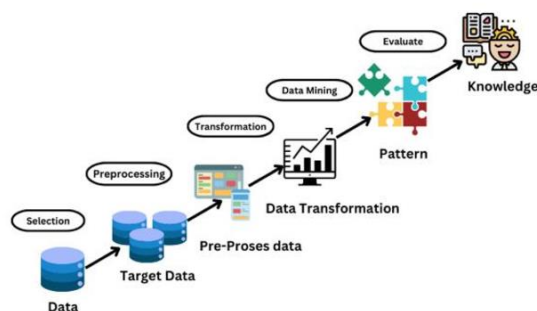
RapidMiner merupakan perangkat lunak yang digunakan untuk pengolahan data mining. Tugas utama RapidMiner dalam text mining meliputi analisis teks, menemukan pola-pola dari kumpulan data yang besar, dan mengintegrasikannya dengan teknik statistik, kecerdasan buatan, serta basis data.[8]

3. METODOLOGI PENELITIAN

3.1. Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah data nilai UN SMA di Provinsi DKI Jakarta yang diperoleh dari Kementerian Pendidikan dan Kebudayaan Indonesia. Data yang digunakan mencakup nilai UN dari berbagai mata pelajaran seperti Matematika, Bahasa Indonesia, Bahasa Inggris, dan IPA.

Dalam penelitian ini, metode K-Means diterapkan menggunakan alat bantu seperti RapidMiner, yang memudahkan dalam proses visualisasi, analisis, dan eksekusi algoritma clustering.[9]



Gambar 1. Tahapan Knowledge Discovery in Database

3.2. Prakonsepsi K-Means

Pada tahap awal, dilakukan praproses data dengan normalisasi nilai-nilai UN untuk memastikan bahwa skala nilai tidak mempengaruhi hasil clustering. Setelah itu, dilakukan seleksi jumlah karakteristik data yang akan digunakan dan memprioritaskan data untuk temuan yang disimpan dalam dokumen terpisah dari basis data. Memilih data dan mengambil dataset dalam format Excel pada RapidMiner. Proses memilih subbagian atau bagian tertentu dari data yang berkaitan.

3.3. Proses K-Means

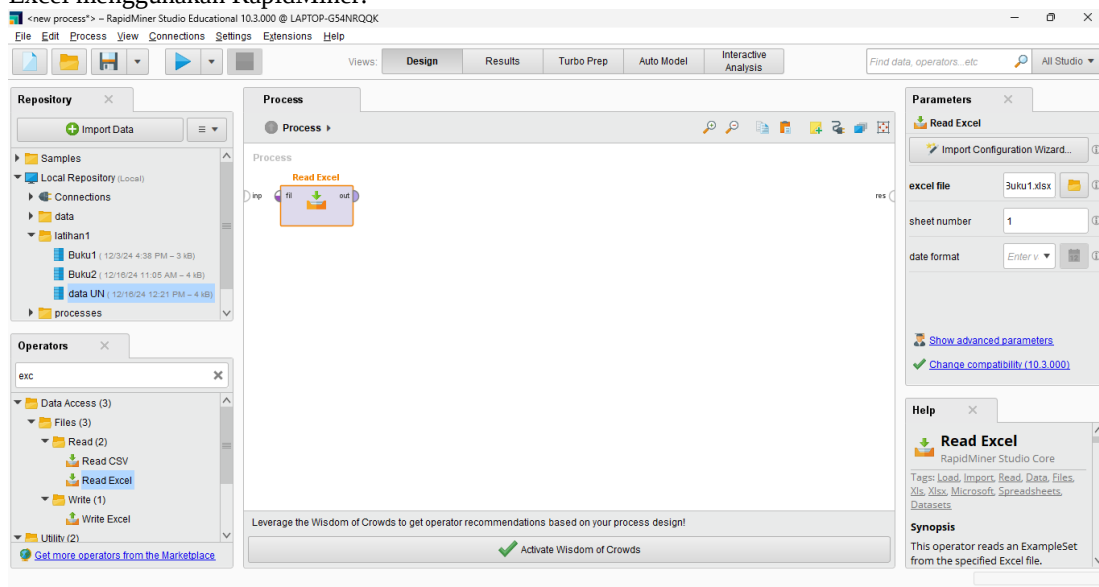
Setelah praproses data, algoritma K-Means dijalankan dengan mengelompokkan SMA berdasarkan nilai UN mereka menggunakan platform RapidMiner. Setiap sekolah akan dikelompokkan ke dalam cluster. Proses ini dilakukan dengan beberapa iterasi hingga konvergensi tercapai, yaitu ketika posisi pusat cluster tidak berubah secara signifikan.[10].

4. HASIL DAN PEMBAHASAN

Data penelitian ini diambil dari file Excel yang berisi lebih dari 40 catatan nilai hasil Ujian Nasional selama periode April hingga Juli. Hasil analisis studi tersebut menyajikan pemahaman mengenai cara data nilai ujian dapat diorganisasi dan diklasifikasikan melalui penerapan algoritma K-means. Proses pengelompokan tersebut dilakukan menggunakan teknik pembelajaran mesin.

4.1 Pengujian Data

Setelah proses impor data selesai, sebuah operator baru bernama Read Excel akan muncul di proses utama. Operator akan memasukkan file dalam format. Xlsx yang diimpor langsung dari file Excel, sehingga data siap digunakan untuk pengujian. Kemudian data akan diimpor dan dipilih dari kumpulan data dalam format Excel menggunakan RapidMiner.



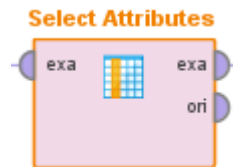
Gambar 2. Read Excel

4.2 Data Selection

Seleksi data merupakan proses untuk pemilihan dan pengumpulan subset atau bagian data tertentu yang memiliki relevansi terhadap tujuan spesifik. Tujuan utama dari proses pemilihan data adalah untuk memusatkan perhatian pada informasi yang paling penting dan relevan, mengurangi tingkat kompleksitas, dan mempercepat langkah analisis. Proses pada operator RapidMiner terdiri dari 40 item, 1 atribut spesial, dan 4 atribut reguler, meliputi nilai MTK, B. Indonesia, B. Inggris, dan IPA.

4.3 Data Preprocessing

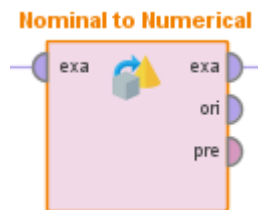
Data preprocessing (pra-pemrosesan data) adalah rangkaian langkah atau teknik yang diterapkan pada data mentah sebelum data tersebut digunakan untuk analisis atau pemodelan. Penelitian ini memanfaatkan data nilai hasil Ujian Nasional yang terdiri atas 40 record dengan 4 atribut.



Gambar 3. Select Attributes

4.4 Data Transformation

Data Transformation adalah proses mengubah atau mengelola data dari format aslinya ke format yang lebih sesuai atau berguna untuk analisis maupun pemodelan. Transformasi data dilakukan untuk mengonversi data nominal menjadi data numerik dengan memanfaatkan operator Nominal ke Numerik. Proses ini melibatkan pengaturan jenis filter atribut sebagai subset, pemilihan atribut (seperti nilai B. Indonesia) yang akan diubah menjadi numerik, serta penyesuaian atribut. Jenis pengkodean ditentukan sebagai bilangan bulat unik, dan operator Nominal-ke-Numerik digunakan untuk mengubah data nominal menjadi data numerik.



Gambar 4. Nominal to Numerical

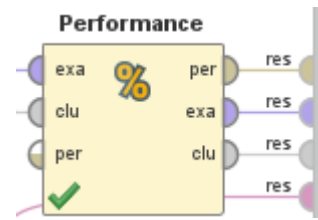
4.5 Data Mining

Data mining merupakan proses ekstraksi pola dan wawasan yang bermanfaat dari kumpulan data yang berukuran besar. Tujuan utama dari data mining adalah mengidentifikasi pola, hubungan, atau informasi tersembunyi yang dapat mendukung proses pengambilan keputusan.



Gambar 5. K-Means Clustering

Sesuaikan parameter clustering yang diperlukan dengan menetapkan nilai k menjadi 2 dan amati sekolah mana yang mendominasi di setiap kelompok. Proses dijalankan hingga maksimal 10 kali, memungkinkan loop untuk mengamati objek secara lebih rinci, sambil memberi kesempatan sistem untuk memeriksa objek lain dengan properti serupa hingga 10 kali agar hasil proses lebih akurat. Jenis pengukuran yang digunakan adalah perhitungan jarak numerik dengan metode Euclidean distance. Metode ini dipilih karena saat mempertimbangkan jarak terpendek antara suatu objek dengan objek lainnya, perhitungan menggunakan jarak Euclidean lebih mudah diterapkan.



Gambar 6. Cluster Distance Performance

Cluster Model

Cluster 0: 21 items
 Cluster 1: 19 items
 Total number of items: 40

Gambar 7. Hasil Cluster Model

Dua cluster dibentuk dengan mengintegrasikan data hasil nilai Ujian Nasional yang telah diproses dan algoritma K-Means ke dalam model cluster yang terbagi menjadi 2 cluster. Cluster 0 terdiri atas 21 item, sedangkan cluster 1 mencakup 19 item, dengan total keseluruhan 40 item.

Tabel 1. Hasil Cluster Optimal

Clustering.k	Davies Bouldin
2	-126.857
3	-162.840
4	-179.633

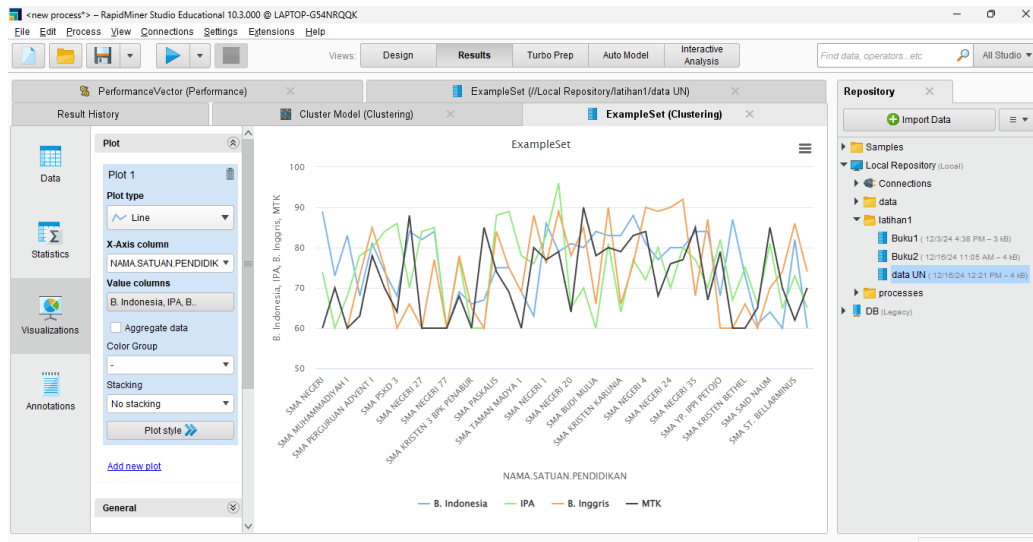
Dari hasil clustering, dari sini dapat menyimpulkan Cluster 0 lebih banyak nilai hasil Ujian Nasional, cenderung lebih berkualitas. Oleh karena itu, disarankan untuk mengoptimalkan persediaan terkait penerimaan calon siswa baru pada sekolah dalam Cluster_0 guna memenuhi penerimaan siswa baru. Langkah-langkah seperti peningkatan pendaftaran, perencanaan persediaan yang efisien, dan manajemen yang baik dapat membantu memastikan ketersediaan sekolah yang memadai.

Sementara itu, Cluster_1 mungkin termasuk dalam kategori sekolah yang kurang berkualitas. Untuk meningkatkan kualitas dalam Cluster_1 diperlukan evaluasi, hal ini dapat meningkatkan kualitas sekolah.

Hasil menunjukkan Kesimpulan akhirnya Ada 21 produk yang dikategorikan sekolah berkualitas, 19 sekolah dikategorikan kurang berkualitas. Melalui analisis clustering Dari temuan tersebut, disimpulkan bahwa sekolah yang tergolong berkualitas disarankan untuk mengoptimalkan persediaan terkait penerimaan calon siswa baru pada sekolah guna memenuhi penerimaan siswa baru. Sekolah yang dikategorikan kurang berkualitas memerlukan perbaikan.

4.6 Visualization Hasil Clustering

Chart merupakan tampilan grafik hasil pengelompokan atau cluster data nilai pada setiap sekolah dengan jumlah 2 cluster. Berikut ini tampilan visualization gambar Tampilan Bar (column) pada chart dan Tampilan Pie pada gambar 8 berikut:



Gambar 8. Tampilan Bar (column) pada Chart

5. KESIMPULAN DAN SARAN

Penelitian ini berhasil menerapkan algoritma K-Means untuk mengelompokkan SMA di DKI Jakarta berdasarkan nilai UN. Hasil clustering menunjukkan adanya perbedaan yang signifikan antara kelompok sekolah dengan kualitas pendidikan yang baik, menengah, dan rendah. Dengan menggunakan teknik clustering ini, calon siswa dan orangtua dapat lebih mudah dalam memilih sekolah yang sesuai dengan kriteria yang diinginkan, khususnya dalam hal nilai UN.

Penelitian lebih lanjut dapat dilakukan dengan mempertimbangkan faktor lain selain nilai UN, seperti fasilitas sekolah, kehadiran siswa, atau tingkat kelulusan untuk memperoleh gambaran yang lebih komprehensif mengenai kualitas sebuah sekolah.

Ucapan Terima Kasih

Dengan penuh rasa syukur, kami mengucapkan terima kasih kepada semua pihak yang telah memberikan dukungan dan kontribusi dalam penyelesaian penelitian ini.

Kami juga menghaturkan terima kasih kepada pembimbing dan pengajar yang telah memberikan arahan, masukan, dan motivasi dalam setiap tahapan penelitian. Tidak lupa, kepada keluarga dan rekan-rekan yang telah memberikan doa, semangat, serta dukungan moral yang tiada henti.

DAFTAR PUSTAKA

- [1] N. sukmadinata Syaodih, "ILMU & APLIKASI PENDIDIKAN," 2007, [Online]. Available: https://www.google.co.id/books/edition/Illmu_dan_aplikasi_pendidikan/B8cfnF69lOEC?hl=id&gbpv=1&dq=pendidikan&pg=PA98&printsec=frontcover
- [2] A. M. Dama yanti S.Si ; Mira Juangsih,S. Si. Asep sukanar, Spd; Lingga kartini, "RINGKASAN MATERI & LATIHAN SOAL-SOAL UN SMP 2011," 2010, [Online]. Available: https://www.google.co.id/books/edition/Ringkasan_Materi_Latihan_Soal_soal_UN_SM/4V5xBAAQBAJ?hl=id&gbpv=1&dq=mengetahui+bagusnya+sekolah+dari+nilai+un&pg=PR3&printsec=frontcover
- [3] M.-Y.-A. M.-N. Ahmad-Dkk, "DATA MINING & Aplikasinya," 2021, [Online]. Available: https://books.google.co.id/books?id=53FXEAAAQBAJ&newbks=0&printsec=frontcover&dq=data+mining&hl=id&source=newbks_fb&redir_esc=y#v=onepage&q=data+mining&f=false
- [4] M. F. Edy Irawansyah, *ADVANCED CLUSTERING Teori dan Aplikasi*, 1st ed. 2015. [Online]. Available: https://www.google.co.id/books/edition/Advanced_Clustering/8y80BgAAQBAJ?hl=id&gbpv=1&dq=algoritma+k-means+clustering+adalah&printsec=frontcover
- [5] P. M. A. K-means, S. Nurani, Y. Syahra, and A. Calam, "Penerapan Data Mining Dalam Clustering Pencapaian Target," vol. 2, pp. 355–363, 2023.
- [6] M. S. Deny Jollyta, Prihandoko, Alyauma Hajjah, Elin Haerani, *ALGORITMA KLASIFIKASI*

- UNTUK PEMULA. DEEPUBLISH DIGITAL, 2023. [Online]. Available: https://books.google.co.id/books?id=y84TEQAAQBAJ&newbks=0&printsec=frontcover&pg=PA138&dq=Rapidminer+bahasa+indonesia&hl=id&source=newbks_fb&redir_esc=y#v=onepage&q&f=true
- [7] ensiklopedia bebas Wikipedia bahasa Indonesia, “Ujian Nasional,” *Wikipedia bahasa Indonesia, ensiklopedia bebas*, 2024.
 - [8] M. Faid, “Perbandingan Kinerja Tool Data Mining Weka dan Rapidminer Dalam Algoritma Klasifikasi,” vol. 8, 2019, doi: 10.34148/teknika.v8i1.95.
 - [9] A. perdana windarto Kiki Fatmawati, “DATA MINING PENERAPAN RAPIDMINER DENGAN K-MEANS CLUSTER PADA DAERAH TERJANGKIT DEMAM BERDAARAH DENGUE (DBD) BERDASARKAN PROVINSI,” vol. 3, 2018.
 - [10] Y. R. Sari *et al.*, “PENERAPAN ALGORITMA K- MEANS UNTUK CLUSTERING DATA KEMISKINAN PROVINSI BANTEN MENGGUNAKAN RAPIDMINER,” vol. 5, no. 2, pp. 192–198, 2020.