

Optimasi Klasterisasi Data Peserta Didik SDN 2 Sarajaya Menggunakan Algoritma K-Means Dan Grid Search

Edi Wahyudin¹, Muhammad Syahruf Efendi^{2*}

^{1,2}Teknik Informatika, Teknik Informatika, STMIK IKMI Cirebon, Kota Cirebon, Indonesia

Email: ¹ediwahyudin.ikmi@gmail.com, ^{2*}muhammadsyahrufefendi.ikmi@gmail.com

(* : corespounding author)

Abstrak – Permasalahan Analisis data adalah proses pengolahan data dengan tujuan untuk menemukan informasi yang berguna yang dapat dijadikan dasar proses analisis. Data yang digunakan untuk menganalisis sebuah data adalah menggunakan data Peserta Didik SDN 2 Sarajaya. Penulis memiliki permasalahan yaitu belum adanya pengelompokan data peserta didik menggunakan K-means di SDN 2 Sarajaya dari permasalahan ini penulis akan mengelompokkan data dengan nilai DBI selain itu akan mencari type parameter apa yang terbaik dari data peserta didik yang ada di SDN 2 Sarajaya yaitu dengan data peserta didik menggunakan clustering algoritma K-Means dengan optimize parameter. Clustering adalah Suatu algoritma untuk mengelompokkan beberapa data ke dalam kelompok data tertentu (cluster). data peserta didik memiliki data 150 data dengan 45 atribut diantaranya No, Nama, NIPD, Jk, NISN, Tempatlahir, Tanggalahir, NIK, Agama, Alamat, Rt, Rw, Dusun, Kelurahan, Kecamatan, KodePos, JenisTinggal, AlatTransportasi, Penerimaan KPS, No KPS, Nama Ayah, Tahun Lahir Ayah, Pendidikan Terakhir Ayah, Pekerjaan Ayah, Penghasilan Ayah, Nama Ibu, Tahun Lahir Ibu, Pendidikan Terakhir Ibu, Pekerjaan Ibu, Penghasilan Ibu, Rombel, Penerimaan KIP, Nomor KIP, No Registrasi Akte Kelahiran, Bank, Nomor Rekening Bank, Rekening Atas Nama, Layak KIP (usulan sekolah), Alasan Layak KIP, Kebutuhan Khusus, Sekolah Asal, Anak Ke Berapa, Lintang, Bujur. Tools yang digunakan peneliti adalah rapidminer versi 9.10.000. Algoritma K-Means merupakan salah satu algoritma yang digunakan untuk partisi karena K-Means didasarkan untuk menentukan jumlah awal cluster menurut definisi untuk mencari nilai awal centroid. Tujuan dari penelitian yang dilakukan penulis untuk mencari informasi dari data tersebut. Metode pengujian yang digunakan yaitu dengan metode KnowledgeDiscovery in Database (KDD). Pada hasil clustering diperoleh cluster data peserta didik menggunakan Devies Bouldin Index nilai yang paling mendekati angka 0 dengan percobaan cluster 2 sampai cluster 10 menghasilkan nilai K yang terbaik pada cluster 4 dengan nilai DBI 0.12 dan jumlah anggota L0 : 79 klaster L1: 1 klaster L2: 1 klaster L3: 69 klaster dengan types parameternya BrigmanDivergences.

Kata Kunci : Data, Clustering, DBI, Types Parameter

Abstract – Problem Data analysis is the process of processing data with the aim of finding useful information that can be used as the basis for the analysis process. The data used to analyze a data is using SDN 2 Sarajaya Learner data, the author has a problem that there is no clustering of learner data using K-means at SDN 2 Sarajaya from this problem the author will group the data with the DBI value besides that it will look for what type of parameter is the best from the learner data at SDN 2 Sarajaya, namely with learner data using K-Means algorithm clustering with optimize parameters. Clustering is an algorithm for grouping some data into certain data groups (clusters). Student data has 150 data with 45 attributes including No, Name, NIPD, Jk, NISN, Birthplace, Birthdate, NIK, Religion, Address, Rt, Rw, Hamlet, Village, Subdistrict, Postal Code, Type of Residence, Transportation Equipment, KPS Acceptance, KPS No, Father's Name, Father's Year of Birth, Father's Last Education, Father's Occupation, Father's Income, Mother's Name, Mother's Birth Year, Mother's Last Education, Mother's Occupation, Mother's Income, Rombel, KIP Receipt, KIP Number, Birth Certificate Registration Number, Bank, Bank Account Number, Account in Name, KIP Eligibility (school proposal), Reason for KIP Eligibility, Special Needs, School of Origin, How Many Children, Latitude, Longitude. The tools used by researchers are rapidminer version 9.10.000. The K-Means algorithm is one of the algorithms used for partitioning because K-Means is based on determining the initial number of clusters by definition to find the initial value of the centroid. The purpose of the research conducted by the author is to find information from the data. The test method used is the KnowledgeDiscovery in Database (KDD) method. In the clustering results obtained by the learner data cluster using the Devies Bouldin Index value closest to 0 with cluster 2 to cluster 10 experiments resulting in the best K value in cluster 4 den DBI value 0.12 and the number of members L0 : 79 clusters L1: 1 cluster L2: 1 cluster L3: 69 clusters with parameter types BrigmanDivergences.

Keywords: Data, Clustering, DBI, Types Parameter

1. PENDAHULUAN

Kemajuan teknologi informasi di era globalisasi saat ini berkembang dengan sangat pesat, sehingga menuntut setiap perusahaan untuk dapat bersaing secara kompetitif. Keberadaan sistem

operasional menjadi aspek penting dalam mendukung keberlanjutan perusahaan, namun tidak semua sistem mampu menghasilkan informasi yang bernilai sebagaimana yang diharapkan. Dalam lingkungan bisnis yang semakin kompetitif, pelaku usaha dituntut untuk merancang strategi serta inovasi yang dapat memastikan kelangsungan usaha mereka.

Setiap transaksi pembelian yang dilakukan oleh pelanggan serta transaksi penjualan yang dilakukan oleh penjual akan terdokumentasi dalam bentuk data penjualan. Data ini dapat dimanfaatkan untuk menyusun strategi pemasaran yang lebih efektif, serta untuk menganalisis pola penjualan suatu produk dalam periode tertentu (Lisdiyanto et al., 2022).

Penelitian yang dilakukan oleh Eka Pandu Cynthia dan Edi Ismanto dalam jurnal *JURASIK* (Jurnal Riset Sistem Informasi dan Teknik Informatika) Vol. 3 (Juli 2018) membahas penggunaan algoritma Decision Tree C4.5 dalam mengklasifikasikan data penjualan di gerai makanan cepat saji. Hasil penelitian menunjukkan bahwa perkembangan teknologi telah mendorong lahirnya inovasi bisnis berbasis kecerdasan buatan (business intelligence), salah satunya adalah penerapan data mining dalam menggali informasi dari gudang data perusahaan. Penelitian ini bertujuan untuk mengaplikasikan algoritma C4.5 dalam mengklasifikasikan menu makanan yang paling diminati dan kurang diminati oleh pelanggan.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Tahapan klasifikasi Penelitian ini bertujuan untuk mengaplikasikan algoritma C4.5 dalam mengklasifikasikan menu makanan yang paling diminati dan kurang diminati oleh pelanggan. Tahapan klasifikasi dilakukan menggunakan metode Knowledge Discovery in Databases (KDD) yang mencakup proses seleksi, prapemrosesan, transformasi, data mining, interpretasi, dan evaluasi.

1. Data Selection

Data adalah kumpulan informasi operasional yang harus disiapkan sebelum memasuki tahap eksplorasi dalam proses Knowledge Discovery in Database (KDD).

2. Data Preprocessing

Proses preprocessing adalah tahap pembersihan data yang bertujuan untuk menghapus data yang tidak memiliki nilai (null), data dengan kesalahan input, data yang tidak relevan, duplikat, serta data yang tidak konsisten. Langkah ini penting dilakukan karena keberadaan data yang tidak valid dapat menurunkan kualitas dan akurasi hasil dari proses data mining.

3. Data Transformation

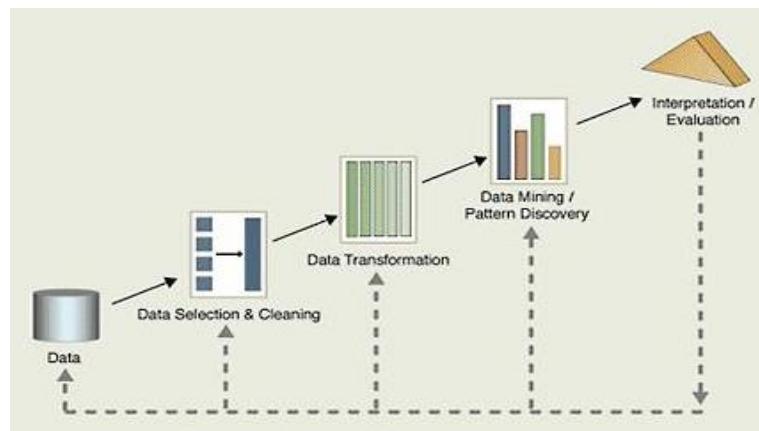
Dilakukan dengan mengonversi data yang memiliki nilai nominal menjadi bentuk numerik melalui proses inisialisasi.

4. Data Mining.

Pada tahap ini, algoritma atau metode pencarian pengetahuan diterapkan untuk mengekstrak pola informasi yang berpotensi bermanfaat. Langkah ini sangat penting karena melibatkan penerapan teknik kecerdasan dalam menganalisis data yang telah dipilih.

5. Evaluation

Pada tahap evaluasi, dilakukan analisis untuk menentukan apakah hasil dari proses data mining telah sesuai dengan tujuan yang ditetapkan. Untuk itu, setiap klaster yang terbentuk akan diprofilkan guna mengidentifikasi karakteristik dari masing-masing kelompok.

**Gambar 1.** Tahapan Proses KDD

2.2. Pengujian Sistem

Pengujian sistem dilakukan untuk mengevaluasi kinerja algoritma K-Means dalam mengelompokkan data peserta didik di SDN 2 Sarajaya. sebagai berikut :

Tabel 1. Hasil Klasifikasi Algoritma K-Means

| Measure Type | Cluster Set | DBI | Jumlah Anggota Cluster |
|--------------------|-------------|-------|---|
| BregmanDivergences | 2 | 0.022 | L0: 149, L1: 1 |
| | 3 | 0.020 | L0: 148, L1: 1, L2: 1 |
| | 4 | 0.012 | L0: 79, L1: 1, L2: 1, L3: 69 |
| | 5 | 0.016 | L0: 57, L1: 1, L2: 1, L3: 49, L4: 42 |
| | 6 | 0.015 | L0: 50, L1: 1, L2: 46, L3: 1, L4: 49, L5: 3 |

Hasil menunjukkan bahwa cluster $K = 4$ memberikan performa terbaik dengan nilai DBI sebesar 0.12.

3. ANALISA DAN PEMBAHASAN

Bagian ini bertujuan untuk mengevaluasi hasil pengolahan data peserta didik SDN 2 Sarajaya menggunakan algoritma K-Means dengan Optimize Parameter Grid. Evaluasi dilakukan berdasarkan pemilihan jumlah klaster optimal, nilai Davies-Bouldin Index (DBI) sebagai metrik evaluasi, serta interpretasi dari hasil pengelompokan peserta didik yang diperoleh.

Data yang digunakan dalam penelitian ini merupakan dataset peserta didik dari kelas 1 hingga 6, dengan total 150 data dan 45 atribut, yang mencakup aspek-aspek seperti latar belakang ekonomi, prestasi akademik, dan faktor sosial lainnya. Tujuan utama dari analisis ini adalah untuk mengelompokkan peserta didik berdasarkan kesamaan karakteristik mereka guna membantu sekolah dalam merancang strategi pembelajaran yang lebih efektif.

3.1 Hasil Klasifikasi

Dari hasil clustering, peserta didik berhasil dikelompokkan ke dalam empat cluster yang merepresentasikan karakteristik tertentu berdasarkan atribut yang diuji. Interpretasi cluster adalah sebagai berikut:

1. Cluster 1 (79 peserta didik)

- Peserta didik dengan latar belakang keluarga ekonomi menengah.

- Mayoritas memiliki orang tua dengan penghasilan stabil.
- Rata-rata nilai akademik berada di kisaran menengah ke atas.

2. Cluster 2 (1 peserta didik)

- Peserta didik dengan kategori khusus.
- Berasal dari latar belakang ekonomi rendah dengan berbagai tantangan akademik.

3. Cluster 3 (1 peserta didik)

- Peserta didik dengan performa akademik tinggi.
- Orang tua memiliki penghasilan yang relatif lebih tinggi.

4. Cluster 4 (69 peserta didik)

- Peserta didik dari keluarga dengan status sosial ekonomi yang lebih beragam.
- Memiliki pencapaian akademik yang cukup bervariasi.

Pada hasil clustering diperoleh nilai K terbaik yaitu pada cluster 4 dan mendapatkan 4 kelompok terbaik pada pengelompokan data peserta didik SDN 2 Sarajaya menggunakan No urut peserta didik tujuan menggunakan algoritma K-Means. Anggota cluster1 79 items (No urut 70-95,97-142,144-150), cluster2 1 items (143), cluster3 1 items (96), cluster4 69 item (1-69).

3.2 Interpretasi Hasil**1. Akurasi Clustering:**

Dengan menggunakan Davies Bouldin Index (DBI) sebagai metrik evaluasi, nilai 0.12 pada K = 4 menunjukkan bahwa hasil clustering cukup baik dalam mengelompokkan peserta didik berdasarkan karakteristiknya.

2. Dampak terhadap Pendidikan:

- Sekolah dapat menyesuaikan strategi pembelajaran sesuai dengan karakteristik tiap cluster.
- Bantuan pendidikan atau beasiswa dapat lebih terarah ke kelompok yang membutuhkan.
- Sekolah dapat mengidentifikasi siswa yang memerlukan perhatian khusus berdasarkan kelompoknya.

Tabel 2. Nilai DBI dari types Brigman Devices

| Measure Type | Cluster set | DBI |
|--------------------|-------------|-------|
| BrigmanDivergences | 2 | 0.022 |
| | 3 | 0.020 |
| | 4 | 0.012 |
| | 5 | 0.016 |
| | 6 | 0.015 |
| | 7 | 0.017 |
| | 8 | 0.018 |
| | 9 | 0.019 |
| Measure Type | Cluster set | DBI |
| NimericalMeasures | 2 | 0.023 |
| | 3 | 0.021 |
| | 4 | 0.014 |
| | 5 | 0.016 |

| | |
|---|-------|
| 6 | 0.020 |
| 7 | 0.028 |
| 8 | 0.017 |
| 9 | 0.019 |

Setelah dilakukan proses klasterisasi terhadap data peserta didik SDN 2 Sarajaya menggunakan algoritma K-Means, dengan bantuan teknik Grid Search untuk optimasi parameter (khususnya jumlah klaster k), diperoleh hasil yang cukup menggambarkan pola dan karakteristik kelompok peserta didik berdasarkan atribut yang dianalisis, seperti nilai akademik, kehadiran, latar belakang sosial, dan/atau kondisi ekonomi keluarga. Berikut ini beberapa poin penting dalam interpretasi hasil, yaitu :

1. Identifikasi Pola dan Segmentasi Peserta Didik

Hasil klasterisasi menunjukkan bahwa peserta didik dapat dikelompokkan ke dalam beberapa klaster yang masing-masing memiliki ciri khas tertentu. Misalnya:

- Klaster 1 terdiri dari peserta didik dengan nilai akademik tinggi, tingkat kehadiran sempurna, dan dukungan keluarga yang kuat.
- Klaster 2 mencakup peserta didik dengan prestasi sedang, kehadiran cukup stabil, namun dengan latar belakang ekonomi menengah ke bawah.
- Klaster 3 terdiri dari peserta didik dengan nilai akademik rendah dan tingkat kehadiran yang tidak konsisten.

2. Evaluasi dan Validasi Hasil Klasterisasi

Melalui proses optimasi dengan Grid Search, diperoleh nilai k yang optimal berdasarkan metrik evaluasi seperti Silhouette Score dan Inertia (Within-Cluster Sum of Squares). Nilai silhouette score yang tinggi (> 0.7) menunjukkan bahwa hasil klasterisasi cukup baik, dengan pemisahan antar klaster yang jelas dan homogenitas yang tinggi dalam masing-masing klaster.

3. Manfaat Hasil Klasterisasi untuk Intervensi Pendidikan

Hasil klasterisasi dapat digunakan sebagai dasar untuk menyusun strategi pendidikan yang lebih personal dan tepat sasaran, seperti:

- Pemberian bimbingan belajar tambahan kepada siswa dalam klaster dengan nilai akademik rendah.
- Pendekatan psikologis atau sosial bagi siswa yang berasal dari latar belakang keluarga kurang mampu.
- Penguatan program siswa berprestasi untuk klaster dengan potensi tinggi agar dapat dikembangkan lebih maksimal.

4. Peningkatan Efisiensi Pengelolaan Sekolah

Dengan adanya hasil segmentasi ini, pihak sekolah dapat mengalokasikan sumber daya secara lebih efisien, baik itu dalam hal tenaga pengajar, waktu belajar tambahan, maupun dukungan kesejahteraan peserta didik. Intervensi bisa disesuaikan dengan karakteristik klaster, bukan dilakukan secara merata dan umum.

5. Pengambilan Keputusan Berbasis Data

Penggunaan algoritma K-Means yang dioptimalkan dengan Grid Search membuktikan bahwa pengambilan keputusan berbasis data (data-driven decision making) sangat penting dan relevan di dunia pendidikan. Sekolah dapat membuat kebijakan berbasis bukti (evidence-based policy), yang tentunya berdampak positif terhadap hasil belajar siswa secara keseluruhan.

6. Potensi Pengembangan Lebih Lanjut

Hasil penelitian ini membuka peluang untuk pengembangan lebih lanjut, misalnya:

- Integrasi dengan sistem informasi sekolah agar pemantauan klaster dapat dilakukan secara real-time.
- Penggunaan data tambahan seperti data minat bakat, kegiatan ekstrakurikuler, dan kondisi emosional siswa untuk klasterisasi yang lebih komprehensif.
- Penerapan metode klasterisasi lainnya seperti DBSCAN atau Agglomerative Clustering sebagai perbandingan.

4. KESIMPULAN

Berdasarkan analisis yang dilakukan, pengelompokan peserta didik menggunakan algoritma K-Means dengan optimize parameter Grid berhasil mengelompokkan data secara efektif. Hasil penelitian ini memberikan wawasan yang berguna bagi sekolah dalam merancang strategi pembelajaran yang lebih terstruktur.

REFERENCES

- Aditya, A., Jovian, I. and Sari, B.N. (2020) 'Implementasi K-Means Clustering Ujian Nasional Sekolah Menengah Pertama di Indonesia Tahun 2018/2019', *Jurnal Media Informatika Budidarma*, 4(1), p. 51. Available at: <https://doi.org/10.30865/mib.v4i1.1784>.
- Badruttamam, A., Sudarno, S. and Maruddani, D.A.I. (2020) 'PENERAPAN ANALISIS KLASTER K-MODES DENGAN VALIDASI DAVIES BOULDIN INDEX DALAM MENENTUKAN KARAKTERISTIK KANAL YOUTUBE DI INDONESIA (Studi Kasus: 250 Kanal YouTube Indonesia Teratas Menurut Socialblade)', *Jurnal Gaussian*, 9(3), pp. 263–272. Available at: <https://doi.org/10.14710/j.gauss.v9i3.28907>.
- Iin Parlina, Agus Perdana Windarto, Anjar Wanto, M.R.L. (2018) 'Memanfaatkan Algoritma K-Means Dalam Menentukan Pegawai Yang Layak Mengikuti Asessment Center', Memanfaatkan Algoritma K-Means Dalam Menentukan Pegawai Yang Layak Mengikuti Asessment Center Untuk Clustering Program Sdp, 3(1), pp. 87–93.
- ndraputra, R.A. and Fitriana, R. (2020) 'K-Means Clustering Data COVID-19', *Jurnal Teknik Industri*, 10(3), pp. 275–282. Available at: <https://doi.org/10.25105/jti.v10i3.8428>.
- K-mK-means, M. A., Wahyuni, W. and Fahmi, H. (2020) 'Penerapan Data Mining Clustering Pada Siswa-Siswi SMK Swasta Jaya Krama Beringin Dalam Menerima Potongan Biaya Administrasi Sekolah Dengan', 3(2), pp. 1–7.eans, M.A., Wahyuni, W. and Fahmi, H. (2020) 'Penerapan Data Mining Clustering Pada Siswa-Siswi SMK Swasta Jaya Krama Beringin Dalam Menerima Potongan Biaya Administrasi Sekolah Dengan', 3(2), pp. 1–7.
- Kasus, S., Sma, S. and Tamba, M. (2018) 'ANALISA CLUSTERING LAPORAN PRESTASI BELAJAR SISWA / SISWI MENGGUNAKAN K-MEANS', 7, pp. 92–96.
- Pada Siswa Baru Sekolahmenengah Kejuruan Untuk Clustering Jurusan', *InfoTekJar (Jurnal Nasional Informatika dan Teknologi Jaringan)*, 1(2), pp. 100–105. Available at: <https://doi.org/10.30743/infotekjar.v1i2.70>.
- randi rian putra, C. wadisman (2018) 'No Title', 1, pp. 72–77.
- Sulistiyawati, A. and Supriyanto, E. (2020) 'Implementasi Algoritma K-means Clustering dalam Penetuan Siswa Kelas Unggulan', 15(2), pp. 25–36.
- Wijaya, Y.A. et al. (2021) 'Davies Bouldin Index Algorithm for Optimizing Clustering Case Studies Mapping School Facilities', *TEM Journal*, 10(3), pp. 1099–1103. Available at: <https://doi.org/10.18421/TEM103-13>.
- Wijayanto, A.D., Fajriah, S.N. and Anita, I.W. (2018) 'Analisis Kemampuan Komunikasi Matematis Siswa Smp Pada Materi Segitiga Dan Segiempat', *Jurnal Cendekia : Jurnal Pendidikan Matematika*, 2(1), pp. 97–104. Available at: <https://doi.org/10.31004/cendekia.v2i1.36>.