ANALISIS KLASTER K-MEDOID UNTUK PENGELOMPOKAN DAN PEMETAAN PROVINSI DI INDONESIA BERDASARKAN NILAI UJIAN NASIONAL

Anita Puri[1], Dodi Solihudin[2], Saeful Anwar [3], Denni Pratama [4], Edi Wahyudin [5]
[1,2,3] Teknik Informatika, STMIK IKMI Cirebon
[4,5] Komputerisasi Akuntansi, STMIK IKMI Cirebon
Jalan Perjuangan No.10B, Karyamulya, Kec. Kesambi, Kota Cirebon, Jawa Barat 45135, Indonesia
Anitapuri3809@gmail.com

ABSTRAK

Ujian Nasional (UN) adalah suatu metode standar mengevaluasi tahap pendidikan awal yang membantu menentukan apakah siswa lulus atau tidak. Sejalan dengan Penilaian hasil belajar ini bertujuan unntuk menilai kinerja siswa pada tingkat nasional. Ujian nasional juga mempunyai dampak penting dalam melanjutkan pendidikan pada jenjang selanjutnya dan dapat meningkatkan mutu pendidikan di sekolah. Penelitian ini bertujuan mengidentifikasi pola atau struktur dalam Algoritma *K-Medoid Clustering* yang digunakan untuk mengelompokkan nilai rata-rata siswa berdasarkan hasil ujian nasional tingkat provinsi di Indonesia. Dalam mengelompokkan pencapaian nilai ujian siswa berdasarkan Nama Provinsi, Ujian Nasional (UN) tingkat (SMP), UN (SMA) (IPA), UN SMA Program Ilmu Pengetahuan Sosial (IPS), UN SMA Program Bahasa, UN Sekolah Menengah Kejuruan (SMK), Ujian Institusi (IIUN) tingkat SMP, IIUN tingkat SMA (IPA), IIUN SMA (IPS), IIUN SMA Program Bahasa, dan IIUN SMK. Algoritma *K-Medoid* ini juga dikenal sebagai Algoritma Partitioning Around Medoid atau PAM, dari data yang sudah diolah menggunakan teknik clustering hasil yang di dapat dalam mengelompokan nilai rata-rata di provinsi Indonesia berdasarkan *Davies Bouldin Index*, dapat diharapkan memudahkan dan memahami algoritma K-Medoid dalam pengelompokan nilai ujian siswa di Indonesia menggunakan teknik clustering untuk mendapatkan jumlah kelompok terbaik pada data hasil ujian nasional di Indonesia menggunakan algoritma *k-medoid*. Dari hasil penerapan tersebut nilai K terbaik diperoleh pada nilai iterasi K9 dengan jumlah 1.016 *berdasarkan Davies Bouldin Index*, Berdasarkan hasil pengukuran nilai max runs dari 1 sampai 10, menunjukkan nilai DBI yang konsisten, yakni 1.016 pada K9 menjadi nilai terbaik, pada cluster ini, terdapat 25 items dalam cluster 0, dan 5 items dalam cluster 1. Stabilitas nilai DBI pada K9 menunjukkan kualitas pengelompokan yang baik.

Kata kunci: *Algoritma K-medoid, Nilai Ujian Nasional, Metode clustering, Data Mining*

1. PENDAHULUAN

Menjadi salah satu pendorong utama kemajuan dalam sistem data. Ujian Nasional (UN) dianggap sebagai suatu Mengevaluasi standar pendidikan inti yang mendukung penyelesaian sekolah. Evaluasi keberhasilan pembelajaran akan terapkan sesuai dengan peraturan yang berlaku di Indonesia. dilaksanakan dengan maksud menilai pencapaian siswa yang telah menyelesaikan pendidikan nasional. UN tidak hanya berfungsi sebagai kriteria kelulusan, tetapi juga memainkan peran penting dalam penentuan kelanjutan pendidikan ke tingkat berikutnya, serta dapat berkontribusi pada peningkatan nilai pendidikan tersebut. Peneliti menggunakan Nilai rata-rata di provinsi indonesia pada website resmi Kemendikbud, Informasi disajikan dalam format Excel.

Penelitian bertujuan untuk mengidentifikasi pola atau struktur dalam Algoritma *K-Medoid Clustering* yang diterapkan pada pengelompokkan nilai rata-rata pendidikan siswa SMP, SMA, dan SMK berdasarkan hasil ujian nasional tingkat provinsi di Indonesia pada tahun 2022. Pendekatan menggunakan algoritma K-Medoid, yang bertujuan untuk memahami pola nilai ujian siswa tersebut dengan harapan dapat memberikan wawasan yang lebih mendalam terkait karakteristik kelompok prestasi yang berbeda. Metodologi ini mencakup tahapan pengolahan data, seperti normalisasi nilai menggunakan algoritma k-medoid clustering, penentuan cluster optimum, dan pengukuran nilai dbi dari hasil pengelompokan nilai prestasi siswa di Indonesia.

Peneliti sebelumnya telah menjalankan studi untuk algoritma K-Means dan K-Medoid dalam menghasilkan cluster. Salah satu penelitian yang dilakukan pada tahun 2019, menyimpulkan bahwa perbandingan antara kedua algoritma tersebut tidak menunjukkan perbedaan yang sangat signifikan dalam proses pengelompokkan data. Hasil penelitian juga menunjukkan bahwa baik algoritma K-Means maupun K-Medoid memiliki tingkat akurasi yang memadai dan fleksibilitas yang cukup, karena peneliti dapat dengan mudah menentukan jumlah cluster yang diinginkan.[1]

Analisis pemahaman pola nilai dan segmentasi. Sebaliknya, penilaian nilai ujian nasional siswa di Indonesia memiliki peran kunci dalam perbaikan sistem pendidikan, di mana pengelompokan siswa berdasarkan kinerja dapat membantu penyelenggara pendidikan merancang strategi pembelajaran yang lebih efektif. Dalam konteks tersebut, penerapan algoritma *K-Medoid Clustering* menjadi relevan sebagai metode analisis yang dapat mengungkap pola tersembunyi dan menghasilkan kelompok atau cluster dengan karakteristik serupa. Penelitian ini bertujuan untuk membandingkan penggunaan algoritma K-

Medoid Clustering pada dua domain yang berbeda ini, mengeksplorasi efektivitas dan generalitas algoritma dalam memberikan wawasan yang berguna dalam konteks analisis nilai ujian siswa di Indonesia. Dengan memahami perbedaan dan kesamaan dalam penerapan algoritma k-medoid, dari memahami pengembangan metode analisis yang dapat diadopsi dalam berbagai konteks data nilai.[2]

Permasalahan penelitian ada pada data yang bersifat acak, bukan disusun berdasarkan kategori atau nilai atribut tertentu. Akibatnya, informasi yang disajikan kurang terstruktur, dan beberapa atribut masih belum dikelompokkan. Untuk mengatasi kendala dilakukan teknik klaster.

Hasil dari penelitian ini akan memberikan kontribusi dalam mempermudah pemahaman tentang penggunaan algoritma *K-Medoid* untuk mengelompokkan nilai rata-rata ujian siswa di seluruh provinsi Indonesia. Metode clustering akan digunakan untuk menentukan jumlah kelompok terbaik dalam data hasil ujian nasional Indonesia. Penelitian ini akan fokus pada penerapan teknik *clustering* untuk mempertimbangkan nilai (DBI) dan hasil yang optimal.

2. TINJAUAN PUSTAKA

2.1. *Data Mining*

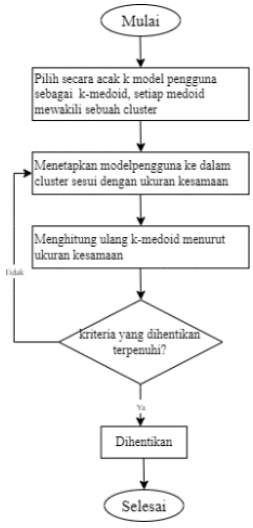
Proses penerapan penggunaan teknik statistik, matematika, kecerdasan buatan, serta pembelajaran mesin untuk mengekstrak dan mengidentifikasi informasi. berharga menggabungkan pengetahuan sumber data yang besar. Definisi tersebut, dijelaskan sebagai pencarian pola dan penerapan metode matematika dan statistik, kecerdasan buatan, dan *Machine learning* untuk menarik dan mengenali informasi terdapat dalam data yang telah diolah.[3][4]

2.2. *Clustering*

Pengelompokan data dalam suatu atribut ke beberapa klaster, dimana tingkat persamaan antar data dalam satu *cluster* lebih besar dibandingkan dengan kesamaannya terhadap data dalam *cluster* lainnya. Clustering merupakan aspek krusial dalam pembelajaran tanpa pengawasan, mirip dengan masalah sejenis lainnya yang berkaitan dengan pengidentifikasian pola atau struktur dalam kumpulan data yang tidak memiliki label. Secara umum, clustering dapat didefinisikan sebagai proses mengelompokkan objek-objek sehingga anggota dalam satu kelompok dimiliki kesamaan tertentu dalam beberapa aspek. *Cluster* sekumpulan entitas kemiripan atau kesamaan tertentu di dalam suatu dataset. Grup ini dibentuk berdasarkan karakteristik atau atribut tertentu objek yang termasuk dalam *cluste*r lainnya.[3][5]

2.3. *K-Medoid*

Merupakan komponen yang menentukan pemisahan antara kelompok medoid dalam *cluster* terkecil serta pemisahan antara titik individu dan *cluster* besar. Data akan diurutkan ke dalam *cluster* terdekat sesuai persamaan *Euclidian*. Titik pusat *cluster* di dalam suatu grup ditemukan menggunakan pendekatan *K-Medoid.* Misalnya *K-Medoid* menggunakan kuadrat hasil jarak *Euclidean* pada data objek, namun *K-Means* menggunakan k sebagai representasi objek yang dikurangi dengan hasil ketidaksamaan objek data. Inilah mengapa *K-Medoid* lebih sukses dibandingkan *K-Means*. Selain itu, metrik jarak ini menurunkan noise dan outlier data.[3] Menurut[6] beberapa peneliti pernah mengadakan penelitian yang membandingkan efektivitas algoritma *K-Means* dan *K-Medoid* dalam menghasilkan cluster. Kedua algoritma ini dianggap cukup tepat dan mudah disesuaikan karena peneliti memiliki keleluasaan untuk menentukan jumlah cluster yang diinginkan.[7] Penelitian tersebut menciptakan kelompok terbaik. Dalam pendekatan ini, kita pilih objek aktual sebagai representasi kelompok objek ada kelompok titik referensi. Proses-proses yang terlibat dalam metode K-Medoids melibatkan:



Gambar 1 *Flowchart K-medoid*

Pada Gambar 1 merupakan *flowchart* algoritma K-Medoid.

1. Diberikan k
2. Pilih secara acak k instance sebagai medoid awal
3. Tetapkan setiap instance ke medoid x terdekat
4. Hitung fungsi tujuan jumlah ketidaksamaan semua instance dengan medoid terdekatnya
5. Pilih secara acak sebuah instance y
6. Tukar x dengan y jika penukaran tersebut mengurangi fungsi tujuan
7. Ulangi (3-6) hingga tidak ada perubahan.

2.4. *Davies-Bouldin Index (DBI)*

Cluster dievaluasi menggunakan indeks Davies-Bouldin. Evaluasi DBI mencakup skema evaluasi klaster internal di mana kualitas klaster ditentukan berdasarkan jumlah dan kedekatan hasil klaster. DBI merupakan suatu metode untuk mengukur validitas hasil *cluster* dengan metode pengelompokan. Kohesi

didefinisikan sebagai banyaknya hubungan data antara pusat suatu *cluster* dengan cluster berikutnya. Sedangkan pemisahannya didasarkan pada jarak antara pusat klaster dengan klaster lainnya. Pengukuran DBI bertujuan untuk memaksimalkan jarak klaster (Ci dan Cj) sekaligus meminimalkan jarak antar titik dalam suatu cluster. Ketika jarak antar *cluster* semakin maksimal, berarti nilai antar *cluster* semakin berkurang kemiripannya dan perbedaan antar cluster semakin terlihat. Jarak minimum dalam sebuah cluster berarti bahwa nilai setiap objek dalam cluster sangat mirip. [8][9][10]
DBI dapat dihitung menggunakan persamaan:

$$R_i = \frac{\max}{j=1\ldots k, i\neq j} R_{ij}$$

$$Var(x) = \frac{1}{N-1}\sum_{i-1}^{n} x_I - \bar{x})^2$$

$$\frac{R_{ij}}{i \neq j}\frac{var(c_i) + var(c_j)}{||c_{i-} c_j||}$$

$$DBI = \frac{1}{k}\sum_{i=1}^{n} R_i$$

Keterangan:

R	: (jarak antar cluster),
Var	: (varians dari data),
x	: (data ke-i), dan
$\bar{x}$	: (rata-rata dari tiap cluster) digunakan.
DBI	: (*Validasi Davies Bouldin Index*)

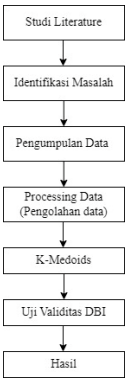
Dengan menerapkan *Davies Bouldin Index*, suatu *cluster* pengelompokan dianggap optimal jika memiliki nilai *Davies Bouldin Index* yang minimal.[8]

2.5. *Rapid Miner*

RapidMiner dirancang untuk menyederhanakan penggunaan perangkat lunak ini, memberikan kemudahan bagi penggunanya. Hasil yang ditampilkan di Rapid Miner juga dapat direpresentasikan secara visual dalam grafik, menjadikan RapidMiner sebagai alat untuk mempelajari kinerja. Perangkat lunak yang dapat digunakan untuk mengekstraksi data menggunakan berbagai metode data mining.[9] Rapid Miner ialah suatu platform perangkat lunak sumber terbuka yang dimanfaatkan untuk menganalisis prediktif, machine learning, data mining, serta analisis bisnis. Platform ini diciptakan untuk membantu para ahli di berbagai sektor dalam mengeksplorasi wawasan berharga dari data yang tersedia dan mengambil keputusan yang lebih baik berdasarkan pemahaman yang mendalam terhadap pola penelitian.

3. METODE PENELITIAN

Metode penelitian ini melibatkan pendekatan numerik dengan penerapan metode kuantitatif. Penelitian ini menggunakan metode data mining dan algoritma k-medoid, yang menghasilkan beberapa kelompok atau kluster sesuai dengan aturan clustering dan pendekatan *Knowledge Discovery in Database (KDD).*



Gambar 2 Alur Metode Penelitian

1. Studi literatur dalam penelitian ini merupakan proses pengumpulan dan analisis sumber daya literatur yang terkait dengan Algoritma K-Medoids Clustering, serta permasalahan yang terkait dengan nilai ujian nasional di Indonesia.
2. Identifikasi masalah dalam penelitian ini melibatkan kurangnya pemahaman terhadap nilai kluster yang optimal untuk beberapa atribut kelompok dalam prestasi siswa. Permasalahan ini dilihat dari data yang diperoleh dari berbagai sumber.
3. Setelah mengidentifikasi masalah, langkah selanjutnya melibatkan teknik pengumpulan data melalui metode penelusuran online research, yang mencakup eksplorasi data melalui internet dengan mengakses situs web seperti kemendikbud.
4. Pada tahap berikutnya, setelah mendapatkan data yang relevan, peneliti melakukan pengolahan data untuk menentukan atribut mana yang dapat dikelompokkan menggunakan metode clustering. Proses pengolahan data bertujuan untuk mengelompokkan nilai ujian siswa ke dalam kelompok-kelompok (cluster) yang serupa berdasarkan karakteristik tertentu.
5. Algoritma *K-Medoid* kemudian diaplikasikan secara iteratif, memastikan bahwa setiap objek yang mewakili benar-benar menjadi medoid dalam kelompok. Proses clustering ini berguna untuk mengidentifikasi pola atau struktur yang mungkin tersembunyi dalam data, di mana objek dalam satu kelompok memiliki kesamaan tinggi, sementara objek antar kelompok memiliki perbedaan yang signifikan.
6. Validasi digunakan sebagai metode untuk mengukur kualitas hasil clustering, dengan Davies Bouldin Index (DBI) sebagai ukuran validitas yang dipilih. DBI mengukur keunikan cluster dengan mempertimbangkan kekompakan dan pemisahan antar cluster.
7. Hasil dari proses validasi ini menjadi dasar untuk evaluasi dan analisis lebih lanjut setelah

penelitian selesai, dengan tujuan untuk memastikan keberhasilan penelitian ini.

3.1. Sumber Data

Data yang diterapkan dalam penelitian ini merupakan data sekunder yang diambil dari situs web resmi Kementerian Pendidikan dan Kebudayaan Indonesia. Peneliti menggunakan informasi Nilai Ujian Nasional yang tersedia di website tersebut, sebagaimana disampaikan oleh [2] dan rekan-rekan pada tahun 2023. Data ini telah dipublikasikan dalam format file Microsoft Excel (.xls) pada tahun 2022. Variabel yang digunakan Nama Provinsi, Nilai Ujian Nasional SMP, Nilai Ujian Nasional SMA IPA, Nilai Ujian Nasional SMA IPS, Nilai Ujian Nasional SMA Bahasa, Nilai Ujian Nasional SMK, Indeks Integrasi Ujian Nasional SMP, Indeks Integrasi Ujian Nasional IPA, Indeks Integrasi Ujian Nasional IPS, Indeks Integrasi Ujian Nasional SMA Bahasa, dan Indeks Integrasi Ujian Nasional SMK. Selanjutnya, langkah-langkah yang ditempuh dalam penelitian ini akan diuraikan:

1. Persiapkan data yang akan diproses, seperti data hasil ujian nasional di Indonesia pada tahun 2022, dan lakukan analisis deskriptif pada data tersebut.
2. Lakukan standardisasi pada data.
3. Terapkan analisis clustering menggunakan algoritma K-Medoids dengan menggunakan Tools Rapid Miner.
4. Evaluasikan performa algoritma dengan merujuk pada nilai DBI dalam persamaan.

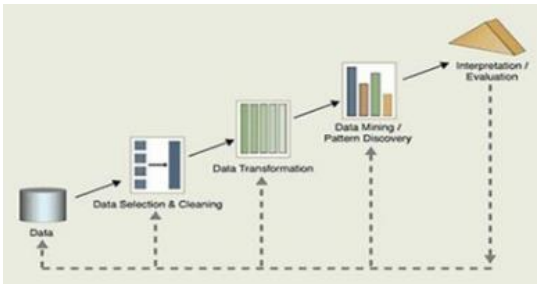
3.2. Teknik Pengumpulan Data

Teknik Pengumpulan mengumpulkan data melibatkan penelitian literatur, dengan fokus pada analisis dan tinjauan karya ilmiah yang relevan dengan topik yang ditentukan. Metode pengumpulan data ini terkait dengan teknik penelitian online, yang melibatkan pencarian informasi melalui internet dengan mengakses situs web. Data yang dianalisis berasal dari ujian nasional di Indonesia, dengan total 500 data dalam dataset.

Data yang digunakan adalah data sekunder, yang merupakan informasi yang telah dikumpulkan sebelumnya oleh peneliti dari ringkasan yang sudah ada. Sumber data sekunder ini mencakup berbagai sumber seperti buku, publikasi, jurnal ilmiah, dan sumber online, yang diperoleh dari website.

3.3. Teknik Analisis Data

Dalam penelitian ini, metode analisis data yang digunakan adalah pendekatan KKD (*Knowledge Discovery in Database*), yang merupakan suatu proses untuk mengidentifikasi pola atau informasi yang bermanfaat dari sejumlah besar data. Proses ini melibatkan beberapa langkah, dan teknik analisis data adalah salah satu aspek penting dalam proses KDD. Adapun Langkah-langka dari teknik ini mengacuh pada seleksi data, processing, transformasi, data mining, evaluasi, interpretasi hasil, merupakan suatu proses untuk mendapatkan informasi dari sebuah data yang tersembunyi dan belum terpecahkan menjadi pola karakteristik dari data tersebut.



Gambar 3 Langkah Knowledge Discovery in Database U(KDD) (Sumber : [10])

1. *Selection Data*

Tahap ini melibatkan pemilihan data yang relevan untuk analisis. Data yang dipilih harus sesuai dengan tujuan analisis.

2. *Processing Data*

Langkah pra-pemrosesan melibatkan pembersihkan data yang hilang atau memiliki nilai yang tidak konsisten pada bagian proses pembersihan. Sebelum melakukan proses ini perlu di laukan analisis untuk di lakukan apakah dataset mengandung nilai yang hilang dan sejauh mana nilai konsisten yang ada.

3. *Data Transformation*

Transformasi data pada Langkah pra-prossesing adalah upaya mengubah data untuk menyesuaikan dengan langkah-langkah data mining, pada fase pengolahan data bertipe polynominal di ubah menjadi bentuk numerik, memungkinkan pengolahan data berdasarkan jarak. Proses ini dapat di lakukan dengan menggunakan operator “nominal tonumerical” untuk mengonfersi data polynominal menjadi format numerik.

4. *Data Mining*

Langkah analisis data yang dilakukan dalam skala besar dengan maksud untuk menemukan pola atau informasi tertentu. Pemilihan metode teknik, atau algoritma yang paling sesuai sangat bergantung pada tujuan dan tahapan keseluruhan dalam proses KDD (*Knowledge Discovery in Database*). Proses KDD dilakukan untuk menganalisis data yang sudah dibersihkan.

5. *Evaluation*

Diperlukan penerjemahan hasil alur data mining ke dalam format yang mudah dipahami. Langkah ini merupakan bagian dari proses KDD (*Knowledge Discovery in Database*), yang mencakup pengecekan kesesuaian atau ketidaksesuaian alur informasi dengan fakta atau hipotesis yang telah ada sebelumnya.

4. HASIL DAN PEMBAHASAN

4.1. *Selection Data*

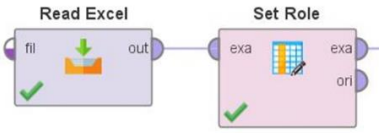
Data yang digunakan dalam penelitian ini berasal dari kumpulan nilai ujian siswa di Indonesia. Data tersebut kemudian dianalisis untuk mendapatkan hasil

nilai rata-rata pada setiap provinsi dan disusun ke dalam kelompok-kelompok tertentu. Pengumpulan data dilakukan melalui metode penelitian online, yaitu dengan melakukan pencarian data melalui internet dengan mengakses berbagai situs web.

Tabel 1 Dataset

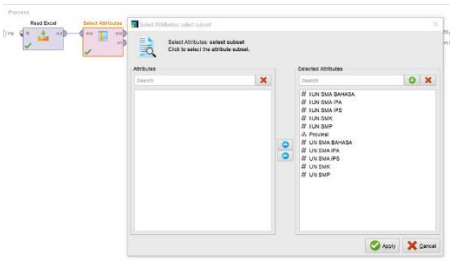
Provinsi	UN SMP	UN SMA IPA	UN SMA IPS	….	IIUN SMK
Jawa barat	56,8	59,8	54,2	….	100
Jawa Tengah	58,5	61,5	56,7	….	100
……	……	…..	…..	…..	
Papua Barat	54,0	47,5	43,5	78,0	78,0
Kalimantan Utara	49,6	50,1	45,5	83,3	83,3

Pada tabel 4.1 Dataset nilai ujian siswa di indonesia ada tipe data yang di perbaharui yaitu integer dan nominal. Pada tabel tersebut mengalami pembaruan secara rutin, dan beberapa atribut tersedia untuk keperluan analisis. Beberapa atribut yang digunakan melibatkan informasi seperti Nama Provinsi, Nilai Ujian Nasional SMP, Nilai Ujian Nasional SMA IPA, Nilai Ujian Nasional SMA IPS, Nilai Ujian Nasional SMA Bahasa, Nilai Ujian Nasional SMK, Indeks Integrasi Ujian Nasional SMP, Indeks Integrasi Ujian Nasional IPA, Indeks Integrasi Ujian Nasional IPS, Indeks Integrasi Ujian Nasional SMA Bahasa, dan Indeks Integrasi Ujian Nasional SMK.



Gambar 4 Operator Read Excel dan Set Rol

Berdasarkan proses operator diatas berfungsi untuk mengakses data dari file Excel, dan menentukan peran atau jenis atribut (variabel) dalam suatu dataset. Hasilnya dapat terlihat pada gambar di bawah ini.



Gambar 5 *Selection* Data

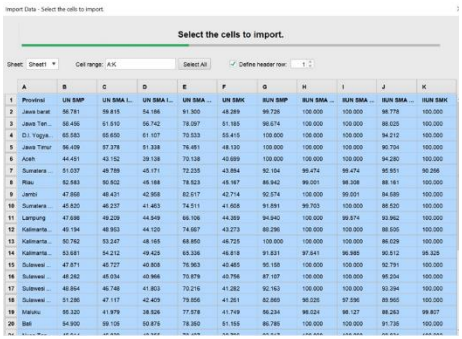
Dalam penelitian ini, terdapat 11 atribut yang digunakan sebagai data. Proses penyiapan dataset melibatkan pemilihan atribut yang akan diolah, beserta parameter yang digunakan dalam proses pengolahan data.

Tabel 2 Parameter Set Role

No	Parameter	Isi	
1.	Attribute name	no	
2.	Nama Provinsi	id	
	Set adittional rules	Attribute name no	Target role id

4.2. *Processing Data*

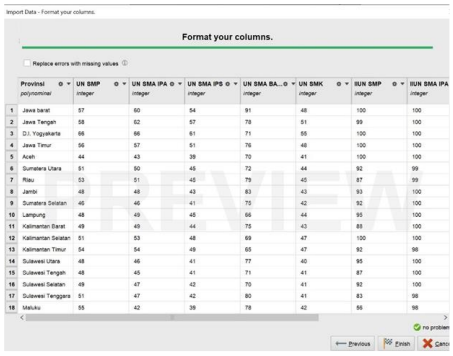
Pada tahapan processing atau pembersihan data, kemudian data dimasukkan dan dipilih berdasarkan atribut yang akan digunakan, kemudian dilakukan pre-processing untuk memastikan tidak ada entri yang sama, tidak ada nilai yang hilang, dan untuk mengurangi potensi kesalahan dalam dataset baru yang disajikan dalam format excel. Pembersihan atau penyiapan data saat ini dilakukan untuk memastikan bahwa data mining dapat dilakukan menggunakan data yang relevan dan tidak mengalami perubahan, seperti yang ditunjukkan dalam gambar di bawah ini.



Gambar 6 Data Preprocessing

4.3. *Transformation Data*

Pada tahap ini dilakukan langkah transformasi untuk memasukkan data ke dalam format yang sesuai untuk diproses dengan teknik data mining. Hal ini dilakukan dengan tujuan untuk menyelaraskan data yang akan diproses oleh algoritma dan perangkat yang digunakan dalam penelitian, yakni RapidMiner. Pada tahap transformasi, atribut nomor diabaikan sebagai pengganti fungsi atribut periode.

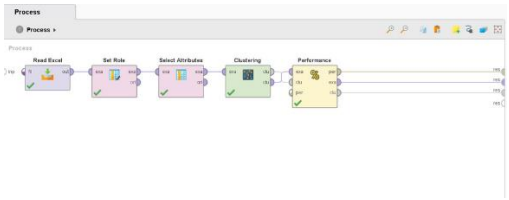


Gambar 7 Transformation Data

4.4. *Data Mining*

Penelitian ini memanfaatkan teknik data mining *Clustering K-Medoid* sebagai metode analisis. Dalam upaya untuk menyederhanakan proses pengolahan

data, digunakanlah perangkat lunak RapidMiner versi 10.3. Dalam implementasinya, algoritma *k-medoid* diterapkan dengan menggunakan *operator read excel* untuk mengakses dataset dalam format xls. Langkah selanjutnya melibatkan operator select atribut untuk memilih fitur pada satu layer data. Proses berlanjut dengan penerapan *clustering k-medoid* untuk mengelompokkan dan memodelkan data yang telah disediakan. Terakhir, kinerja pengelompokan diuji menggunakan *cluster distance performance*.



Gambar 8 Proses *Clustering K-Medoid*

Berikut proses *k-medoid clustering* dengan aplikasi RapidMiner memasukan Data set yang telah melalui preprocessing, diolah dengan menggunaka algoritma *K-medoid* dan menambahkan fitur Performance untuk mengetahui hasil. *Clustering* berperan sebagai operator utama dalam RapidMiner untuk menerapkan algoritma k-medoid, yang bertujuan untuk mengelompokkan data ke dalam kategori yang memiliki ciri-ciri serupa, sehingga dapat mengidentifikasi pola atau struktur dalam data tersebut.

4.5. Uji Validitas

Melalui proses Uji Validitas menggunakan *Davies Bouldin Index* (DBI) dalam konteks *clustering* dilakukan untuk mengevaluasi kualitas hasil *clustering* yang telah dihasilkan. Validitas ini memberikan pemahaman tentang sejauh mana *cluster* yang dihasilkan relevan dan memiliki struktur yang baik. Dengan cara menghasilkan pengelompokan nilai clustering, selanjutnya dilakukan pengitungan jarak antara cluster, menghitung keunikan cluster, kemudian melakukan perhitungan DBI, selanjutnya mengevaluasi dan menginterpretasi hasil cluster, yang terakhir mengetahui nilai iterasi terbaik.

Berikut merupakan hasil Uji Validitas pembacaan operator *K-Medoid Clustering* berdasarkan *Davies Bouldin Index.*

Tabel 3 Nilai DBI

Nilai DBI	
k2	1.448
k3	1.425
k4	1.220
k5	1.153
k6	1.119
k7	1.435
k8	1.320
k9	1.016

Hasil dari uji validitas berdasarkan nilai Davies Bouldin Index berada pada K9 dengan jumlah 1.016 adalah nilai terbaik.

5. KESIMPULAN DAN SARAN

Berdasarkan temuan dan analisis yang dilakukan, dapat disimpulkan sebagai berikut: Hasil penerapan nilai ujian nasional maka nilai K terbaik diperoleh pada nilai iterasi K9 dengan jumlah 1.016 berdasarkan *Davies Bouldin Index* kelompok cluster pada atribut data Nama Provinsi, Ujian Nasional untuk Sekolah Menengah Pertama (SMP), Ujian Nasional untuk Sekolah Menengah Atas (SMA) jurusan IPA, Ujian Nasional untuk SMA jurusan IPS, Ujian Nasional untuk SMA jurusan Bahasa, Ujian Nasional untuk Sekolah Menengah Kejuruan (SMK), IIUN SMP, IIUN SMA jurusan IPA, IIUN SMA jurusan IPS, IIUN SMA jurusan Bahasa, dan IIUN SMK. Berdasarkan hasil Evaluasi pengukuran nilai max runs dari 1 sampai 10, menggunakan menunjukkan nilai DBI yang konsisten, yakni 1.016 pada K9 menjadi nilai terbaik, pada *cluster* ini, terdapat 25 items dalam cluster 0, dan 5 items dalam cluster 1. Stabilitas nilai DBI pada K9 menunjukkan kualitas pengelompokan yang baik. Dengan demikian, hasil Uji Validitas memberikan indikasi bahwa *Mixedmeasure* efektif dalam membentuk kelompok yang signifikan. Kesimpulan ini memperkuat keandalan metode pengukuran numerik yang digunakan dalam analisis clustering ini, yang dapat menjadi dasar bagi Langkah-langkah selanjutnya dalam pemrosesan dan interpretasi data lebih lanjut. Hasil karakteristik penelitian ini mencari nilai K terbaik untuk dataset yang relevan dengan rentang K9 hingga K10. Hasil evaluasi menunjukkan bahwa K9 menunjukkan kinerja optimal dengan nilai mendekati 0, yaitu 1.016, sementara K2 memiliki nilai 1.448. Dalam penelitian ini, terdapat dua cluster yang dianggap optimal atau mendekati 0. Temuan ini mencerminkan keberhasilan algoritma K-Medoid dalam mengelompokkan data pada K9, dengan *Cluster Distance Performance* berdasarkan *Davies Bouldin Index* sebesar 1.016 sebagai yang terbaik.

Saran untuk penelitian selanjutnya adalah melakukan identifikasi lebih lanjut terhadap data hasil pengelompokan nilai rata-rata ujian nasional di provinsi Indonesia dengan menggunakan algoritma k-medoid guna mengevaluasi Davies Bouldin Index (DBI). Proses evaluasi melibatkan suatu skema evaluasi klaster internal di mana kualitas klaster ditentukan berdasarkan jumlah dan kedekatan hasil klaster. Terdapat beberapa aspek yang perlu mendapatkan perhatian lebih lanjut agar proses ini dapat ditingkatkan di masa mendatang. Hasil penelitian menunjukkan bahwa setiap klaster yang dihasilkan memiliki jarak yang berbeda dari klaster lainnya, dan hasil klaster dapat diukur berdasarkan jarak antar atribut klaster. Hal ini memungkinkan penentuan atribut mana yang perlu menjadi prioritas evaluasi guna meningkatkan kinerja siswa. Selain itu, diperlukan perbaikan signifikan pada tahap

preprocessing teks untuk menghindari ambiguitas atau duplikasi data, sehingga hasil yang diperoleh dapat menjadi lebih optimal. Saran lainnya adalah melakukan analisis lebih mendalam terhadap penggunaan algoritma clustering dan metode klasifikasi sistem lainnya sebagai pembanding. Hal ini bertujuan untuk memperoleh hasil yang lebih akurat dan baik, sehingga dapat memberikan kontribusi positif dalam pengembangan evaluasi dan peningkatan performa siswa secara keseluruhan.

DAFTAR PUSTAKA

[1] D. Sartika dan J. Jumadi, "Clustering Penilaian Kinerja Dosen Menggunakan Algoritma K-Means (Studi Kasus: Universitas Dehasen Bengkulu)," *Semin. Nas. Teknol. Komput. ...*, 2019, [Daring]. Tersedia pada: http://seminar-id.com/prosiding/index.php/sainteks/article/view/218

[2] A. R. Lashiyanti, I. Rasyid Munthe, F. A. Nasution, dan E. P. Korespondensi, "Optimisasi Klasterisasi Nilai Ujian Nasional dengan Pendekatan Algoritma K-Means, Elbow, dan Silhouette," *J. Ilmu Komput. dan Sist. Inf. (JIKOMSI*, vol. 6, no. 1, hal. 14–20, 2023.

[3] M. Herviany, P. Delima, T. Nurhidayah, dan K. Kasini, "Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokkan Daerah Rawan Tanah Longsor Pada Provinsi Jawa Barat," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 1, no. 1, hal. 34–40, 2021, doi: 10.57152/malcom.v1i1.60.

[4] S. Nurani, Y. Syahra, dan A. Calam, "Penerapan Data Mining Dalam Clustering Pencapaian Target Penjualan Menggunakan Algoritma K-Means," *J. Sist. Inf. ...*, 2023, [Daring]. Tersedia pada: http://ojs.trigunadharma.ac.id/index.php/jsi/article/view/6552

[5] S. F. Intan, W. Elvira, S. Rahayu, dan ..., "Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Pengeluaran Mahasiswa: Comparison of the K-Means and K-Medoids Algorithms for …," *... Nas. Penelit. dan ...*, hal. 35–40, 2023, [Daring]. Tersedia pada: https://journal.irpi.or.id/index.php/sentimas/article/view/543%0Ahttps://journal.irpi.or.id/index.php/sentimas/article/download/543/335

[6] I. Kamila, U. Khairunnisa, dan M. Mustakim, "Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau," *J. Ilm. Rekayasa dan Manaj. Sist. Inf.*, vol. 5, no. 1, hal. 119, 2019, doi: 10.24014/rmsi.v5i1.7381.

[7] M. Azmi, A. A. Putra, D. Vionanda, dan A. Salma, "Comparison the Performance of K-Means and K-Medoids Algorithms in Grouping Regencies / Cities in Sumatera Based on Poverty Indicators," vol. 1, hal. 59–66, 2023, doi: 10.24036/ujsds/vol1-iss2/25.

[8] M. Mughnyanti, S. Efendi, dan M. Zarlis, "Analysis of determining centroid clustering x-means algorithm with davies-bouldin index evaluation," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 725, no. 1, 2020, doi: 10.1088/1757-899X/725/1/012128.

[9] M. D. Chandra, E. Irawan, I. S. Saragih, A. P. Windarto, dan D. Suhendro, "Penerapan Algoritma K-Means dalam Mengelompokkan Balita yang Mengalami Gizi Buruk Menurut Provinsi," *BIOS J. Teknol. Inf. dan Rekayasa Komput.*, vol. 2, no. 1, hal. 30–38, 2021, doi: 10.37148/bios.v2i1.19.

[10] N. Agung, O. Nurdiawan, dan G. Dwilestari, "Penerapan Data Mining Metode K-Means Clustering Untuk Analisa Penjualan Pada Toko Yana Sport," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 6, no. 2, hal. 1–7, 2022.