IMPLEMENTASI ALGORITMA *K-MEANS* DENGAN *OPTIMIZE PARAMETER GRID* PADA DATA KECELAKAAN LALU LINTAS DI KOTA CIREBON

Tedy Maulana [1], Rini Astuti [2], Yudhistira Arie Wijaya [3]
[1] Program Studi Teknik Informatika, STMIK IKMI Cirebon
[2,3] Program Studi Sistem Informasi, STMIK LIKMI Bandung
[3]Program Studi Sistem Informasi, STMIK IKMI Cirebon
Jl. Perjuangan No. 10 B Majasem, Kec. Kesambi, Kota Cirebon, Jawa Barat
maulanatedy1234@gmail.com

ABSTRAK

Kecelakaan lalu lintas merupakan salah satu peristiwa yang terjadi dijalan raya. Banyak faktor yang menyebabkan terjadinya kecelakaan lalu lintas, namun faktor manusia itu sendiri *(human error)* menjadi masalah utama. Kecelakaan lalu lintas menjadi salah satu permasalahan yang ada di Kota Cirebon. Penelitian ini bertujuan untuk melakukan pengelompokkan data kecelakaan lalu lintas di Kota Cirebon dengan maksud ingin mengetahui informasi karakteristik yang ada dalam dataset kecelakaan lalu lintas di Kota Cirebon dengan mengetahui pengaruh parameter yang ada pada *Numerical Measure type* dan pengaruh parameter atribut waktu (jam atau menit) agar mendapatkan k optimal berdasarkan *Davies Bouldin Index* (DBI). Penelitian ini menerapkan *data mining* dengan algoritma *K-Means Clustering*. Algoritma *K-Means* dipilih karena mudah diimplementasikan dan efisiensinya. Data yang digunakan yaitu data kecelakaan lalu lintas di Kota Cirebon dari tahun 2020-2023 dengan total 406 data record. Metode analisis data dilakukan menggunakan teknik *Knowledge Discovery in Database(KDD)* dengan *Optimize Parameter Grid*. Hasil analisis diperoleh parameter dengan semua jenis nya pada *Numerical Measure type* dan parameter *hour*(jam) yang berpengaruh dalam menghasilkan nilai k optimal yaitu pada *cluster* k=2 dengan nilai DBI sebesar 0,762. Hasil *cluster* k=2 dengan nilai DBI sebesar 0,762 merupakan hasil yang sama untuk semua jenis parameter yang ada pada *Numerical Measure type*.

Kata kunci : *Data Mining, K-Means Clustering, Kecelakaan*

1. PENDAHULUAN

Kecelakaan lalu lintas adalah suatu peristiwa dan penyebab kematian terbesar di Indonesia yang tidak dapat diprediksi dengan pasti. Kecelakaan lalu lintas terjadi ketika satu atau lebih kendaraan bertabrakan dengan objek lain yang mengakibatkan kerusakan. Kejadian ini cukup memberikan dampak yang besar dalam berbagai aspek, seperti kerugian materi dan berakibat pada korban jiwa. Permasalahan terkait kecelakaan lalu lintas ini menjadi tantangan untuk dipecahkan dalam upaya mengurangi angka kecelakaan lalu lintas. Kejadian kecelakaan lalu lintas di Kota Cirebon menjadi salah satu permasalahan yang ada. Oleh karena itu, penelitian terkait kecelakaan lalu lintas coba dilakukan dengan menerapkan *data mining*. Analisis dilakukan dengan cara mengelompokkan data kecelakaan lalu lintas di Kota Cirebon sehingga nantinya bisa memperoleh informasi terkait karakteristik apa saja yang ada dalam data kecelakaan lalu lintas tersebut.

Berdasarkan penelitian sebelumnya yang dilakukan oleh Kurniawan dan Jajuli, penelitian ini melakukan analisis terkait peramasalahan kecelakaan lalu lintas di Kecamatan Cileungsi denggan menggunakan algoritma *K-Means.* Analisis dilakukan untuk mengetahui faktor penyebab dari kecelakaan lalu lintas di Kecamatan Cileungsi. Metode analisa data yang digunakan yaitu *Knowledge Discovery in Database (*KDD). Penelitian ini menghasilkan *cluster* dengan k=3 dengan kategori dari setiap cluster yaitu lokasi tidak rawan, rawan, dan sangat rawan. Hasil evaluasi terhadap *cluster* yang dihasilkan dengan evaluasi menggunakan *silhouette coefficient* yaitu *cluster 1* didapat hasil dengan nilai 0.35, *cluster* 2 didapat dengan nilai 0.22, dan *cluster* 3 didapat dengan nilai 0.38.[1]. Penelitian lain terkait penerapan data mining dengan metode *clustering* juga pernah dilakukan Supriyadi, analisis dilakukan dengan permasalahan terkait evaluasi armada truk dalam produktivitas kinerjanya. Penggunaan algortima *K-Means* dan *K-Medoids* dilakukan sebagai perbandingan hasil dalam analisis data. Dari kedua algoritma yang dipakai dilakukan pengujian dengan *cluster* k=3 sampai *cluster* k=10. Hasil analisis diperoleh dengan evaluasi menggunakan *Davies Bouldin Index* (DBI) yaitu *cluster* terbaik terdapat pada *cluster* k=3 dengan algoritma *K-Means* diperoleh nilai DBI sebesar 0,67 sedangkan pada algoritma *K-Medoids* diperoleh nilai DBI sebesar 1,78. Dengan hasil perbandingan kedua algoritma tersebut, maka algoritma *K-Means* dipilih sebagai metode pengelompokkan terbaik dengan nilai DBI lebih kecil.[2]. Adapun penelitian lain yang dilakukan oleh Triyandana, penelitian ini melakukan analisis terkait permasalahan penumpukan bahan stok makanan. Oleh sebab itu, dilakukan pengelompokkan data menu pada Dpom Coffe. Analisis dilakukan dengan menggunakan algoritma *K-Means* dengan evaluasi *Davies Bouldin Index* (DBI). Hasil analisis yang dihasilkan dari penelitian ini yaitu *cluster* terbaik terdapat pada k=3 dengan nilai DBI sebesar 0,457.[3].

Penelitian ini bertujuan untuk melakukan pengelompokkan data terhadap kejadian kecelakaan lalu lintas di Kota Cirebon. Pengelompokkan data ini dilakukan dengan menerapkan metode *data mining* guna mendapatkan informasi yang dapat memberikan wawasan baru dalam penanganan kasus kecelakaan lalu lintas. Analisis yang dilakukan dalam penelitian ini menerapkan *data mining* dengan menggunakan algoritma *K-Means* serta menggunakan teknik *Knowledge Discovery in Database* (KDD). Algoritma *K-Means* memiliki kelebihan seperti kecepatan perhitungan yang lebih tinggi, kemudahan implementasi, dan hasil yang lebih akurat dibandingkan algoritma lainnya [4]. Penelitian ini diharapkan dapat memberikan pengetahuan baru terkait karakteristik yang ada dalam kejadian kecelakaan lalu lintas. Hasil dari penelitian ini juga dapat memberikan wawasan dan pengetahuan mengenai upaya pencegahan terhadap kecelakaan lalu lintas. Dengan analisis yang dilakukan menggunakan teknik data mining menggunakan algoritma *K-Means* clustering dapat memberikan literatur referensi bagi pembaca penelitian ini untuk mengembangkkan analisis pada penelitian berikutnya.

2. TINJAUAN PUSTAKA

2.1. Kecelakaan Lalu Lintas

Kecelakaan lalu lintas merupakan peristiwa yang terjadi di jalan raya baik yang tidak diduga maupun tidak disengaja yang dapat mengakibatkan korban juga kerugian harta benda, menurut Undang-Undang Republik Indonesia No. 22 Tahun 2009 tentang Lalu Lintas dan Angkutan Jalan. Kecelakaan lalu lintas disebabkan oleh kesalahan manusia itu sendiri *(human error)*[5].

2.2. *Data Mining*

Data Mining adalah proses analisis data yang bertujuan untuk menemukan hubungan yang signifikan dan mencapai kesimpulan yang sebelumnya tidak diketahui dengan menggunakan metode terkini, memberikan manfaat bagi pemilik data tersebut. *Data Mining* juga diartikan sebagai proses menemukan pola yang menarik dan tersembunyi dari kumpulan data yang besar[6].

2.3. *Algoritma K-Means*

Algoritma *K-Means* adalah metode pengelompokan data non-hierarki yang membagi data menjadi beberapa kelompok berdasarkan karakteristik serupa[7].

2.4. *Davies Bouldin Index (DBI)*

Davies Bouldin Index (DBI) adalah cara validasi cluster yang diperkenalkan oleh D.L. Davies. *Davies Bouldin Index* (DBI) digunakan untuk mengevaluasi hasil *cluster* yang terbentuk. DBI merupakan metode evaluasi *cluster* dalam suatu pengelompokkan data yang menitikberatkan pada penilaian kohesi dan separasi[8].

2.5. *Knowledge Discovery in Database (KDD)*

Knowledge Discovery in Database (KDD) adalah suatu pendekatan penelitian yang bertujuan untuk memperoleh informasi dari dalam suatu basis data[9]. KDD juga merupakan suatu metode analisa data yang biasa digunakan dalam proses *data mining.*

2.6. *RapidMiner*

RapidMiner ialah perangkat lunak yang dapat diakses oleh siapapun dan memiliki sifat terbuka, artinya bersifat *open source*. *RapidMiner* merupakan perangkat lunak yang digunakan untuk melakukan pemrosesan data berdasarkan prinsip-prinsip dan algoritma *data mining*.

3. METODE PENELITIAN

3.1. Studi Pendahuluan

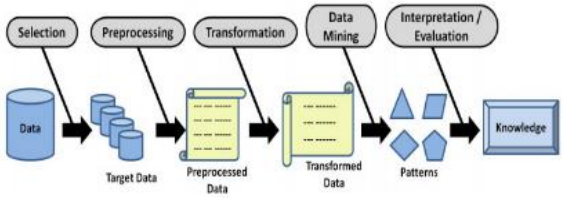
Pada tahap ini dilakukan analisa terhadap topik permasalahan yaitu kecelakaan lalu lintas. Serta pada tahap ini juga data dikumpulkan dan diperoleh untuk dijadikan bahan analisis pada proses penelitian.

3.2. *Literature Review*

Tahapan *literature review* merupakan proses yang dilakukan untuk melakukan tinjauan terhadap teori-teori yang berkaitan dengan topik penelitian.

3.3. Analisis Data

Pada tahap analisa data digunakan teknik *Knowledge Discovery in Database* (KDD). Adapun proses tahapan KDD seperti tampak pada gambar berikut.



Gambar 1. Proses Tahapan KDD

Sebelum masuk dalam proses tahapan menggunakan teknik *Knowledge Discovery in Database* (KDD), proses mempersiapkan dataset diperlukan untuk dilakukan pada proses selanjutnya. Dataset yang digunakan yaitu data kecelakaan lalu lintas di Kota Cirebon tahun 2020-2023 dengan total 406 data *record.* Setelah *dataset* siap, kemudian data dimasukan kedalam proses dengan menggunakan aplikasi *Rapidminer* dengan operator *read excel.* Data yang digunakan berformat *xlsx.* Setelah *data* dimasukan dengan menggunakan operator *read excel,* selanjutnya masuk dalam tahapan KDD.

Teknik KDD ini memiliki tahapan terstruktur yaitu *Selection, Preprocessing, Transformation, Data Mining,* dan *Evaluation*[10]. Adapun tahapan dalam KDD dapat dijelaskan berikut ini.

3.3.1. *Selection*

Langkah pertama dalam proses ini adalah *Selection*. Data awalnya berformat xlsx dan dimasukan menggunakan operator *read excel*. Setelah itu, dilakukan seleksi atribut dari data kecelakaan lalu lintas. Parameter pada operator *read excel* tetap menggunakan parameter *default*. Pada tahap ini, identifikasi *ID* juga dilakukan menggunakan operator *Generate ID* dengan menggunakan parameter *default*.

3.3.2. *Preprocessing*

Setelah melalui tahap pada langkah sebelumnya, data kemudian menjalani proses *preprocessing*. Dalam langkah *preprocessing* ini, dilakukan pembersihan data, seperti menghapus data yang kosong dan mengatasi *missing value* yang dapat menghambat proses pengolahan data berikutnya. Sebelum melakukan pre-processing, dilakukan pemeriksaan statistik data. Jika statistik data menunjukkan tidak adanya *missing value* dan data menunjukkan konsistensi, maka tahap *preprocessing* tidak diperlukan.

3.3.3. *Transformation*

Pada langkah ini, data akan dilakukan transformasi dengan tujuan menyesuaikan tipe data yang cocok untuk pengolahan menggunakan algoritma *K-Means* dalam tahap data mining. Dalam tahap transformasi data, tipe data *Nominal* akan diubah menjadi *Numerical*, dan tipe data Time akan diubah menjadi bentuk *Numerical*. Hal ini bertujuan agar pada tahap berikutnya, proses pengolahan dapat dilakukan dengan menggunakan algoritma *K-Means*. Untuk melakukan transformasi, digunakan operator *Nominal to Numerical* dan *Date to Numerical*. Parameter pada operator *Nominal to Numerical* dan *Date to Numerical* menggunakan parameter *default*. Namun, proses pemilihan parameter dilakukan dengan maksud menganalisis pengaruh atribut waktu terhadap parameter seperti *hour* atau *minute*, yang mempengaruhi penentuan *cluster* dengan nilai optimal. Proses seleksi parameter untuk atribut waktu dilakukan melalui operator *Optimize Parameter Grid*.

3.3.4. *Data Mining*

Langkah ini melibatkan pengolahan data menggunakan algoritma *K-Means* untuk melakukan klasterisasi. Penggunaan algoritma *K-Means* terintegrasi didalam operator *Optimize Parameter Grid*. Pada tahap ini, parameter *Numerical Measure* digunakan dengan semua jenisnya, jumlah klaster ditentukan, nilai evaluasi menggunakan kinerja untuk menentukan hasil *cluster* optimal, dan penerapan parameter pada atribut waktu dilakukan melalui konfigurasi operator *Optimize Parameter Grid*. Dalam operator *Optimize Parameter Grid*, terdapat serangkaian proses termasuk operator *Date to Numerical*, operator *K-Means*, dan operator *Cluster Distance Performance*.

3.3.5. *Evaluation*

Pada langkah ini, dilakukan evaluasi terhadap hasil analisis yang menggunakan algoritma K-Means dengan menganalisis pemodelan klaster yang telah melalui proses pada tahap sebelumnya.

4. HASIL DAN PEMBAHASAN

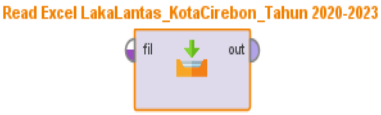
Penelitian mengenai pengelompokkan data kecelakaan lalu lintas di Kota Cirebon pada tahun tahun 2020-2023 menghasilkan *cluster* dengan nilai optimal berdasarkan *Davies Bouldin Index*. Berikut penjelasan dari hasil analisis yang diperoleh dengan proses tahapan teknik *Knowledge Discovery in Database* (KDD).

Pada tahap awal dilakukan memasukan *dataset* yang akan dilakukan pada proses selanjutnya. *Dataset* yang digunakan yaitu data kecelakaan lalu lintas di Kota Cirebon tahun 2020-2023 dengan total 406 data record. *Dataset* ini memiliki 8 atribut diantaranya Tahun, Lokasi Kejadian, Waku Kejadian, Jenis Kendaraan, Kondisi, Jenis Kelamin, Profesi, dan Usia. Adapun atribut dengan tipe data tampak seperti pada tabel berikut.

Tabel 1. Atribut dengan Tipe Data

No.	Atribut	Tipe Data
1.	Tahun	Nominal
2.	Lokasi Kejadian	Nominal
3.	Waktu Kejadian	Time
4.	Jenis Kendaraan	Nominal
5.	Kondisi Korban	Nominal
6.	Jenis Kelamin	Nominal
7.	Profesi	Nominal
8.	Usia	Nominal

Dataset dimasukan kedalam aplikasi *rapidminer* dengan menggunakan operator *Read Excel.* Adapun operator *Read Excel* yang digunakan tampak seperti gambar berikut.

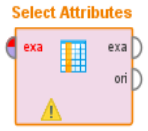


Gambar 2. Operator *Read Excel*

Berdasarkan proses memasukan data menggunakan operator *Read Excel* parameter yang digunakan bersifat *default*. Setelah proses memasukan data, selanjutnya proses analisis dimulai sesuai dengan tahapan proses teknik *Knowledge Discovery in Database* (KDD) sebagai berikut :

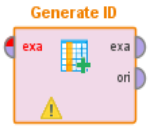
4.1. Selection

Proses awal dalam tahap *selection* dilakukan dengan memilih atribut yang akan digunakan melalui operator *Select Attributes*, seperti tampak pada gambar berikut.



Gambar 3. Operator *Select Attributes*

Berdasarkan proses pemilihan atribut menggunakan operator *Select Attributes* dipilih semua atribut yang ada dalam *dataset* dengan penggunaan parameter *default*. Setelah melakukan pemilihan atribut, selanjutnya dilakukan proses pembentukan *ID* pada *dataset.* Pemebentukan *ID* dilakukan menggunakan operator *Generate ID* seperti pada gambar berikut.



Gambar 4. Operator *Generate ID*

Berdasarkan hasil proses pembentukan *ID* pada operator *Generate ID* dengan parameter *default* didapat hasil seperti pada tabel berikut.

Tabel 2. Hasil Statisik Proses Generate ID

No.	Uraian	Keterangan
1.	Record	406
2.	Special Attribute	1
3.	Reguler Atribute	8
4.	**Attribute :**	
	Tahun	Nominal
	Lokasi Kejadian	Nominal
	Waktu Kejadian	Time
	Jenis Kendaraan	Nominal
	Kondisi Korban	Nominal
	Jenis Kelamin	Nominal
	Profesi	Nominal
	Usia	Nominal

Setelah proses *selection* sudah dilakukan dengan memilih atribut dan pembentukan *ID*, kemudia dilakukan proses pada tahap selanjutnya.

4.2. *Preprocessing*

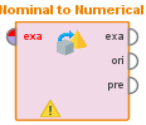
Proses pembersihan data untuk mengatasi nilai yang hilang atau tidak konsisten dilakukan melalui tahap *preprocessing*. Sebelum memulai *preprocessing*, analisis *dataset* pertama-tama dilakukan untuk memverifikasi keberadaan nilai yang hilang atau konsistensi data pada atribut yang terpilih. Informasi tersebut diperoleh dari hasil statistik dataset yang terdapat pada gambar berikut. Setelah meninjau hasil statistik, dapat disimpulkan bahwa tahap *preprocessing* tidak diperlukan karena secara keseluruhan atribut memiliki nilai atau data yang lengkap dan konsisten, tanpa adanya nilai yang hilang.

Setelah mengetahui proses *preprocessing* tidak perlu dilakukan, maka dilakukan pada proses selanjutnya.

Gambar 5. Hasil Statistik Atribut

4.3. *Transformation*

Pada langkah ini dilakukan transformasi data, karena dataset memiliki jenis tipe data yang beragam, termasuk tipe *Nominal* dan *Time.* Sementara itu, algoritma yang digunakan dalam proses data mining adalah *K-Means* yang umumnya membutuhkan tipe data *Numerical*. Proses transformasi diimplementasikan melalui operator *Nominal* to *Numerical* dan *Date to Numerical.* Langkah awal dalam transformasi data adalah Nominal *to Numerical*, seperti terlihat pada gambar berikut.

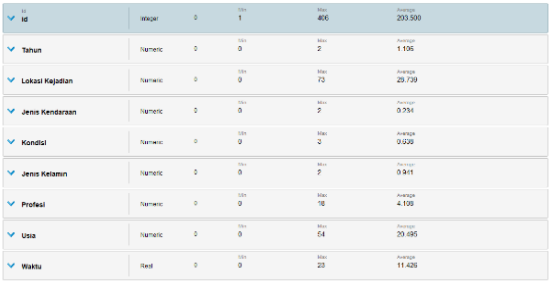


Gambar 6. Operator *Nominal to Numerical*

Setelah melakukan transformasi data dari tipe *Nominal* ke *Numerical*, langkah berikutnya adalah melakukan transformasi pada atribut waktu yang berjenis tipe *Time* menjadi bentuk *Numerical*. Namun, transformasi atribut waktu ini dilakukan secara terperinci dalam operator *Optimize Parameter Grid*. Proses transformasi ini memanfaatkan operator *Date to Numerical*, sebagaimana terlihat pada gambar berikut.

Gambar 7. Operator *Date to Numerical*

Berdasarkan proses transformasi data *Nominal to Numerical* dan *Data to Numerical* diperoleh hasil statistik seperti terlihat pada gambar berikut.



Gambar 8. Hasil Statistik *Transformation*

Setelah proses *transformation* data sudah dilakukan, kemudian dilakukan pada proses selanjutnya.

4.4. *Data Mining*

Setelah melalui proses transformasi data dan memastikan bahwa semua atribut telah memiliki tipe data *Numerical*, langkah berikutnya adalah menjalani proses data mining. Pada tahap data mining, digunakan algoritma *K-Means* untuk mengelola data menjadi bentuk cluster. Proses data mining ini melibatkan operator *Optimize Parameter Grid*, di mana operator *K-Means* terdapat di dalamnya. Operator *K-Means* yang digunakan terlihat pada gambar berikut.

Gambar 9. Operator *K-Means Clustering*

Pada saat proses menggunakan operator *K-Means*, parameter *default* yang digunakan. Namun, seleksi parameter akan dilakukan di dalam konfigurasi operator *Optimize Parameter Grid*.

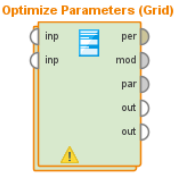
Setelah menerapkan operator *K-Means*, dalam langkah proses yang sama, evaluasi dilakukan untuk menentukan cluster dengan nilai optimal. Evaluasi ini dilakukan dengan memanfaatkan operator *Cluster Distance Performance* di dalam operator *Optimize Parameter Grid*.

Penggunaan operator *Cluster Distance Performance* terlihat pada gambar berikut.

Gambar 10. Operator *Cluster Distance Performance*

Pada proses menggunakan operator *Cluster Distance Performance*, parameter *default* digunakan. Meskipun demikian, seleksi parameter tetap akan dilakukan melalui konfigurasi operator *Optimize Parameter Grid*.

Dengan merujuk pada penerapan operator *Date to Numerical* dalam langkah transformasi data, penggunaan operator *K-Means*, dan penggunaan operator *Cluster Distance Performance*, semua tahapan tersebut dilakukan dalam kerangka operator *Optimize Parameter Grid* sebagaimana telah dijelaskan sebelumnya. Penerapan operator *Optimize Parameter Grid* terlihat pada gambar berikut.



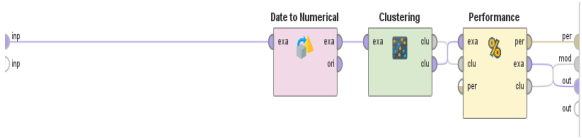
Gambar 11. Operator *Optimize Parameter Grid*

Dalam proses penggunaan operator *Optimize Parameter Grid*, terdapat pemilihan parameter dari setiap operator yang terdapat di dalamnya. Detail mengenai pemilihan parameter dari masing-masing operator dalam proses tersebut dijelaskan dalam tabel berikut.

Tabel 3. Pemilihan Parameter

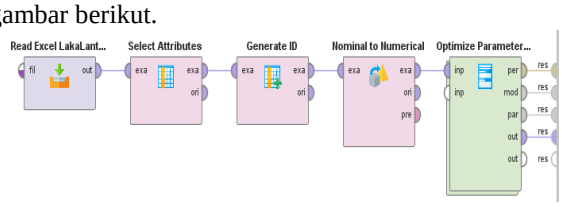
No	Operator	Parameter	Isi Parameter
1	Date to Numerical	time_unit	minute
			hour
2	K-Means	k	2-20
		Measure Types :	
		numerical_measure	EuclideanDistance CamberraDistance ChebychevDistance CorrelationSimilarity CosineSimilarity DiceSimilarity DynamicTimeWarpingDistance InnerProductSimilarity JaccardSimilarity KernelEuclideanDistance ManhattanDistance MaxProductSimilarity
3	Cluster Distance Performance	main_criterion	Davies Bouldin

Sebagaimana telah diuraikan sebelumnya, penerapan algoritma *K-Means*, transformasi data menggunakan operator *Date to Numerical*, dan evaluasi untuk menentukan klaster dengan nilai optimal menggunakan *Davies Bouldin Index* (DBI) melalui operator *Cluster Distance Performance*, semuanya terdapat dalam kerangka proses operator *Optimize Parameter Grid*. Adapun proses tersebut dapat dilihat pada gambar berikut.



Gambar 12. Proses didalam *Optimize Parameter Grid*

Model proses dalam *rapidminer* hingga mencapai tahap *data mining* dengan memanfaatkan operator *Optimize Parameter Grid* terlihat seperti pada gambar berikut.



Gambar 13. Proses Sampai Tahap *Data Mining*

A. Evaluation

Berdasarkan hasil analisis dengan model proses yang dibuat yaitu melakukan uji coba *cluster* k=2 sampai k=20, parameter *Numerical Measure* dengan semua jenisnya, dan penggunaan atribut waktu (*hour* atau *minute*) diperoleh sebagai berikut.

1. Hasil Analisis *cluster* (k=2 sampai k=20) dengan *Euclidean Distance* berdasarkan nilai DBI.

hour : [(0,762), (0,943), (1,037), (0,948), (1,027), (1,096), (1,114), (1,116), (1,159), (1,217), (1,202), (1,234), (1,297),(1,257), (1,217), (1,342), (1,264), (1,220), (1,286)].
minute : [(1,003), (1,050), (1,168), (1,122), (1,139), (1,041), (1,004), (1,018), (0,966), (0,998), (1,009), (1,050), (1,096), (0,999), (1,090), (1,060), (1,061), (1,137), (1,123)].

2. Hasil Analisis *cluster* (k=2 sampai k=20) dengan *ManhattanDistance* berdasarkan nilai DBI.

hour : [(0,762), (0,943), (1,033), (0,948), (1,022), (1,095), (1,113), (1,138), (1,140), (1,166), (1,221), (1,156), (1,230), (1,281), (1,288), (1,207), (1,231), (1,195), (1,244)].
minute : [(1,003), (1,050), (1,168), (1,122), (1,139), (1,048), (1,006), (1,017), (0,966), (1,019), (0,994), (1,073), (1,072), (1,030), (1,081), (1,113), (1,068), (1,068), (1,190)].

3. Hasil Analisis *cluster* (k=2 sampai k=20) dengan *InnerProductSimilarity* berdasarkan nilai DBI.

hour : [(0,762), (0,943), (1,033), (0,948), (1,027), (1,098), (1,129), (1,117), (1,232), (1,257), (1,237), (1,236), (1,187), (1,264), (1,237), (1,218), (1,201), (1,287), (1,261)].
minute : [(1,003), (1,050), (1,177), (1,122), (1,143), (1,042), (1,057), (1,013), (0,967), (0,956), (1,024), (1,003), (1,112), (1,100), (1,129), (1,066), (1,086), (1,142), (1,079)].

4. Hasil Analisis *cluster* (k=2 sampai k=20) dengan *CamberraDistance* berdasarkan nilai DBI.

hour : [(0,762), (0,943), (1,039), (0,948), (1,026), (1,138), (1,153), (1,131), (1,194), (1,164), (1,206), (1,237), (1,161), (1,303), (1,268), (1,240), (1,214), (1,252), (1,230)].
minute : [(1,003), (1,050), (1,174), (1,127), (1,146), (1,043), (1,001), (1,001), (0,966), (0,993), (1,013), (1,044), (1,081), (1,044), (1,112), (1,056), (1,130), (1,089), (1,135)].

5. Hasil Analisis *cluster* (k=2 sampai k=20) dengan *MaxProductSimilarity* berdasarkan nilai DBI.

hour : [(0,762), (0,943), (1,033), (0,948), (1,027), (1,114), (1,129), (1,231), (1,179), (1,186), (1,152), (1,192), (1,260), (1,247), (1,255), (1,226), (1,257), (1,211), (1,265)].
minute : [(1,003), (1,050), (1,168), (1,115), (1,146), (1,043), (1,005), (1,001), (0,961), (0,994), (0,985), (1,090), (1,077), (1,080), (1,125), (1,083), (1,091), (1,133), (1,090)]

6. Hasil Analisis *cluster* (k=2 sampai k=20) dengan *JaccardSimilarity* berdasarkan nilai DBI.

hour : (0,762), (0,943), (1,033), (0,948), (1,036), (1,122), (1,126), (1,132), (1,228), (1,199), (1,231), (1,245), (1,250), (1,214), (1,193), (1,173), (1,293), (1,330), (1,234)].
minute : [(1,003), (1,050), (1,177), (1,121), (1,143), (1,048), (1,049), (1,005), (0,963), (1,006), (1,073), (1,066), (1,086), (1,060), (1,089), (1,079), (1,105), (1,151), (1,133)].

7. Hasil Analisis *cluster* (k=2 sampai k=20) dengan *ChebychevDistance* berdasarkan nilai DBI.

hour : [(0,762), (0,943), (1,033), (0,948), (1,027), (1,098), (1,154), (1,151), (1,117), (1,214), (1,234), (1,229), (1,260), (1,266), (1,237), (1,266), (1,233), (1,202), (1,204)].
minute : [(1,007), (1,050), (1,168), (1,113), (1,139), (1,058), (1,005), (1,019), (0,968), (0,951), (1,035), (1,024), (1,046), (1,107), (1,082), (1,120), (1,090), (1,122), (1,136)].

8. Hasil Analisis *cluster* (k=2 sampai k=20) dengan *OverlapSimilarity* berdasarkan nilai DBI.

hour : [(0,762), (0,943), (1,033), (0,948), (1,027), (1,122), (1,139), (1,133), (1,215), (1,244), (1,197), (1,253), (1,230), (1,216), (1,262), (1,215), (1,241), (1,251), (1,265)].
minute : [(1,003), (1,050), (1,174), (1,122), (1,139), (1,054), (1,003), (1,022), (0,958), (1,035), (0,948), (1,139), (1,036), (1,055), (1,084), (1,148), (1,138), (1,049), (1,080)].

9. Hasil Analisis *cluster* (k=2 sampai k=20) dengan *KernelEuclideanDistance* berdasarkan nilai DBI.

hour : [(0,762), (0,943), (1,036), (0,948), (1,025), (1,091), (1,152), (1,196), (1,146), (1,162), (1,174), (1,246), (1,227), (1,281), (1,239), (1,266), (1,281), (1,225), (1,175)].
minute : [(1,003), (1,050), (1,171), (1,121), (1,150), (1,042), (1,056), (1,015), (0,970), (1,012), (1,066), (1,010), (1,051), (1,084), (1,098), (1,113), (1,103), (1,038), (1,110)].

10. Hasil Analisis *cluster* (k=2 sampai k=20) dengan *CorrelationSimilarity* berdasarkan nilai DBI.

hour : [(0,762), (0,943), (1,033), (0,948), (1,027), (1,118), (1,159), (1,196), (1,149), (1,171), (1,257),

(1,247), (1,320), (1,220), (1,248), (1,316), (1,224), (1,279), (1,238)].
minute : [(1,003), (1,052), (1,171), (1,135), (1,147), (1,042), (1,008), (1,020), (0,962), (0,973), (1,070), (0,989), (1,022), (1,069), (1,138), (1,083), (1,101), (1,098), (1,080)].

11. Hasil Analisis *cluster* (k=2 sampai k=20) dengan *DiceSimilarity* berdasarkan nilai DBI.

hour : [(0,762), (0,943), (1,033), (0,948), (1,027), (1,083), (1,200), (1,108), (1,229), (1,228), (1,228), (1,239), (1,180), (1,266), (1,283), (1,154), (1,253), (1,247), (1,173)].
minute : [(1,007), (1,052), (1,174), (1,121), (1,143), (1,054), (1,080), (1,022), (0,968), (1,029), (1,029), (1,097), (1,046), (1,105), (1,101), (1,022), (1,093), (1,179), (1,063)].

12. Hasil Analisis *cluster* (k=2 sampai k=20) dengan *DynamicTimeWarpingDistance* berdasarkan nilai DBI.

hour : [(0,762), (0,943), (1,037), (0,948), (1,026), (1,096), (1,141), (1,133), (1,095), (1,236), (1,195), (1,183), (1,220), (1,203), (1,275), (1,219), (1,230), (1,243), (1,236)].
minute : [(1,007), (1,050), (1,174), (1,124), (1,142), (1,040), (1,063), (1,014), (0,961), (0,996), (1,027), (1,087), (0,980), (1,087), (1,073), (1,107), (1,077), (1,119), (1,089)].

13. Hasil Analisis *cluster* (k=2 sampai k=20) dengan *CosineSimilarity* berdasarkan nilai DBI.

hour : [(0,762), (0,943), (1,036), (0,948), (1,026), (1,047), (1,022), (1,022), (1,167), (1,152), (1,219), (1,187), (1,262), (1,246), (1,272), (1,088), (1,302), (1,244), (1,228)].
minute : [(1,007), (1,052), (1,174), (1,121), (1,142), (1,090), (1,154), (1,125), (0,961), (0,999), (1,013), (1,006), (1,077), (1,036), (1,123), (1,269), (1,119), (1,161), (1,074)].

Setelah mengetahui hasil eksperimen dapat dilihat bahwa untuk *cluster* dengan nilai optimal terdapat pada *cluster* k=2 dengan nilai DBI sebesar 0,762, serta hasil teresebut berlaku untuk semua jenis *Numerical Measure type* dan terhadap atribut waktu dengan parameter *(hour).*

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian yang diperoleh dapat disimpulkan bahwa *cluster* dengan nilai optimal berdasarkan DBI terdapat pada cluster k=2 dengan nilai DBI sebesar 0,762. Pengaruh parameter yang ada pada *Numerical Measure* dan pengaruh atribut waktu berdasarkan parameter *hour* atau *minute* yaitu terdapat pada semua jenis Numerical Measure *type* dalam menghasilkan cluster dengan nilai optimal diantaranya EuclideanDistance,ManhattanDistance,CamberraDistance,MaxProductSimilarity,JaccardSimilarity,ChebychevDistance,OverlapSimilarity,KernelEuclideanDistance,CorrelationSimilarity,DiceSimilarity,DynamicTimeWarpingDistance, dan CosineSimilarity, serta untuk pengaruh parameter atribut waktu dalam menghasilkan cluster dengan nilai optimal terdapat pada parameter *hour*(jam). Adapun saran yang dapat dilakukan dari hasil penelitian ini dapat berupa pengujian analisis dengan parameter *Measure Type* yang lain seperti *BregmanDivergences*, *MixedMeasures*, dan *NominalMeasures* agar dapat memperoleh hasil yang lebih maksimal dan bervariasi.

DAFTAR PUSTAKA

[1] T. Kurniawan And M. Jajuli, "Clustering Data Kecelakaan Lalu Lintas Di Kecamatan Cileungsi Menggunakan Metode K-Means," 2022.

[2] A. Supriyadi, A. Triayudi, And I. D. Sholihati, "Perbandingan Algoritma K-Means Dengan K-Medoids Pada Pengelompokan Armada Kendaraan Truk Berdasarkan Produktivitas," *Jipi (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)* , Vol. 6, No. 2, Pp. 229–240, 2021.

[3] G. Triyandana, L. A. Putri, And Y. Umaidah, "Penerapan Data Mining Pengelompokan Menu Makanan Dan Minuman Berdasarkan Tingkat Penjualan Menggunakan Metode K-Means," *Journal Of Applied Informatics And Computing (Jaic)*, Vol. 6, No. 1, Pp. 40–46, 2022, [Online]. Available: Http://Jurnal.Polibatam.Ac.Id/Index.Php/Jaic

[4] R. Siagian, Pahala Sirait, And A. Halima, "E-Commerce Customer Segmentation Using K-Means Algorithm And Length, Recency, Frequency, Monetary Model," *Journal Of Informatics And Telecommunication Engineering*, Vol. 5, No. 1, Pp. 21–30, Jul. 2021, Doi: 10.31289/Jite.V5i1.5182.

[5] A. Franseda, "Integrasi Metode Decision Tree Dan Smote Untuk Klasifikasi Data Kecelakaan Lalu Lintas," *Jurnal Sistem Dan Teknologi Informasi (Justin)*, Vol. 8, No. 3, Pp. 282–290, Jul. 2020, Doi: 10.26418/Justin.V8i3.40982.

[6] T. Hardiani, "Analisis Clustering Kasus Covid 19 Di Indonesia Menggunakan Algoritma K-Means," *Jurnal Nasional Pendidikan Teknik Informatika (Janapati)*, Vol. 11, No. 2, Pp. 156–165, Aug. 2022, Doi: 10.23887/Janapati.V11i2.45376.

[7] E. Muningsih, I. Maryani, And V. R. Handayani, "Penerapan Metode K-Means Dan Optimasi Jumlah Cluster Dengan Index Davies Bouldin Untuk Clustering Propinsi Berdasarkan Potensi Desa," *Jurnal Sains Dan Manajemen*, Vol. 9, No. 1, Pp. 95–100, 2021, [Online]. Available: Www.Bps.Go.Id

[8] R. Julianti Hablum, A. Khairan, And Rosihan, "Clustering Hasil Tangkap Ikan Di Pelabuhan Perikanan Nusantara (Ppn) Ternate

Menggunakan Algoritma K-Means," *Jiko (Jurnal Informatika Dan Komputer) Ternate*, Vol. 2, No. 1, Pp. 26–33, 2019.

[9] A. Amrullah, I. Purnamasari, B. Nurina Sari, Garno, And A. Voutama, "Analisis Cluster Faktor Penunjang Pendidikan Menggunakan Algoritma K-Means (Studi Kasus: Kabupaten Karawang)," *Jire (Jurnal Informatika & Rekayasa Elektronika)*, Vol. 5, No. 2, Pp. 244–252, 2022, [Online]. Available: Http://E-Journal.Stmiklombok.Ac.Id/Index.Php/Jirehttp://E-Journal.Stmiklombok.Ac.Id/Index.Php/Jire

[10] N. Nurahman, A. Purwanto, And S. Mulyanto, "Klasterisasi Sekolah Menggunakan Algoritma K-Means Berdasarkan Fasilitas, Pendidik, Dan Tenaga Pendidik," *Matrik : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, Vol. 21, No. 2, Pp. 337–350, Mar. 2022, Doi: 10.30812/Matrik.V21i2.1411.