

Clustering Panjang Ruas Jalan di BBPJN Sumut Menggunakan Algoritma K-Means

Clustering The Length Of Road Sections In BBPJN Sumut Using The K-Means Algorithm

Mitha Rosadi¹, Dhea Aulia Nurhasanah², Muhammad Siddik Hasibuan³
Program Studi Ilmu Komputer Fakultas Sains Dan Teknologi, Universitas Islam Negeri
Sumatera Utara

E-mail: ¹mitharosadi42@gmail.com, ²dheaulia392@gmail.com,
³muhammadsiddik@uinsu.ac.id

Abstrak

Jalan adalah prasarana perhubungan darat yang menjadi jalur transportasi yang sangat vital. Pengelompokan panjang ruas jalan adalah hal yang belum diketahui pada Data Ruas Jalan di BBPJN Sumut. Oleh karena itu penelitian ini akan membahas *Clustering Panjang Ruas Jalan di Sumatera Utara*. Metode yang digunakan adalah *K-Means Clustering Data Mining*. Dengan metode *K-Means Clustering*, data yang telah didapatkan dapat dikelompokkan menjadi beberapa *Cluster*, dimana proses *K-Means Clustering* diterapkan dengan menggunakan *RapidMiner*. Data yang digunakan adalah data ruas jalan Sumatera Utara, area yang tercatat meliputi Jl. BTS. PROV. ACEH - SIMPANG PANGKALAN SUSU hingga Jl. ONAN RUNGGU - TOMOK. Dapat dibagi menjadi 3 *Cluster*: pendek (C1), sedang (C2) dan panjang (C3). Hasil yang diperoleh adalah terdapat 118 jalan dengan *Cluster* tingkat pendek (C1), 37 jalan dengan *Cluster* tingkat sedang (C2), dan 21 jalan dengan *Cluster* tingkat panjang (C3). Hal ini dapat menjadi masukan bagi BBPJN Sumut untuk mengetahui batas ruas jalan tingkat pendek, tingkat sedang dan tingkat panjang jalan.

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

Abstract

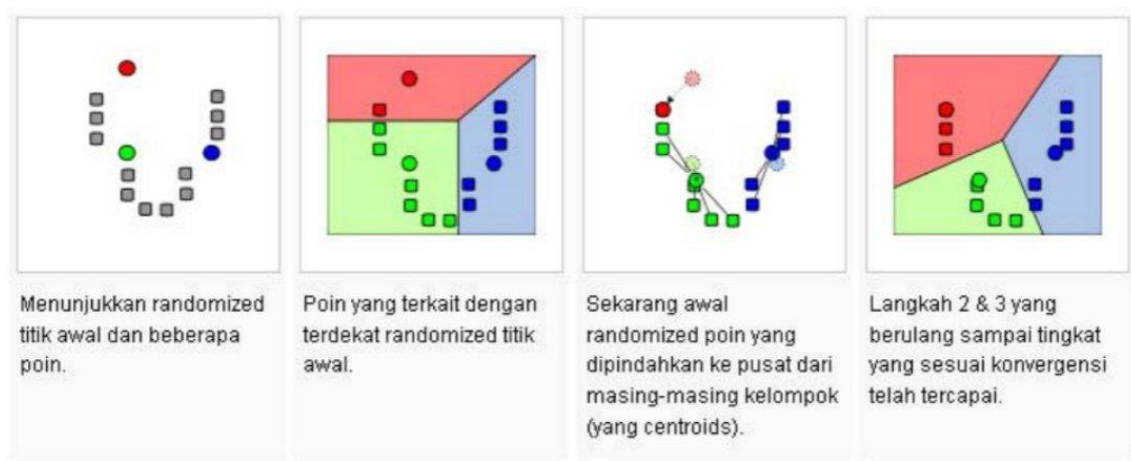
The road is a land transportation infrastructure which is a very vital transportation route. The grouping of road lengths is something that is not yet known in the Road Section Data at North Sumatra BBPJN. Therefore this study will discuss the Long Clustering of Roads in North Sumatra. The method used is *K-Means Clustering Data Mining*. With the *K-Means Clustering* method, the data that has been obtained can be grouped into several clusters, where the *K-Means Clustering* process is implemented using *RapidMiner*. The data used is data for roads in North Sumatra, the area recorded includes Jl. BTS. PROV. ACEH - SIMPANG PANGKALAN SUSU to Jl. ONAN RUNGGU - TOMOK. Can be divided into 3 Clusters: short (C1), medium (C2) and long (C3). The results obtained are that there are 118 roads with short-level clusters (C1), 37 roads with medium-level clusters (C2), and 21 roads with long-level clusters (C3). This can be input for the North Sumatra BBPJN to find out the boundaries of short, medium and long road sections.

Keywords: *Clustering, K-Means, Data Mining, RapidMiner*

1. PENDAHULUAN

Data Mining adalah teknik yang memungkinkan untuk mendapatkan pola atau model dari data yang dikumpulkan. Teknik ini diterapkan di semua jenis lingkungan seperti di bidang biologi, aplikasi pendidikan dan keuangan, industri, polisi, dan proses politik. Dalam *Data Mining* terdapat beberapa teknik, diantaranya adalah induksi aturan dan pohon keputusan yang menurut berbagai penelitian yang dilakukan termasuk yang paling banyak digunakan [1]. *Data Mining* juga disebut proses penggalian informasi dan keahlian dari sejumlah besar data serta kumpulan prosedur untuk menentukan nilai tambah dari suatu kumpulan data berupa informasi yang sebelumnya tidak diketahui dari data tersebut. Dalam dunia *Database*, *Data Mining* dikenal juga dengan *Known as Knowledge Discovery (KKD)*. KKD adalah proses yang melibatkan pengumpulan, analisis, dan interpretasi data, serta penggunaan sejarah untuk mengungkap keteraturan dan pola dalam kumpulan data besar [2]. Metode *Data Mining* terbagi menjadi *Association*, *Classification*, *Prediction*, *Cluster* dan *Data Analysis* [3].

Clustering adalah salah satu teknik dasar *Data Mining* untuk mempartisi *Database* menjadi kelompok-kelompok sehingga visualisasi dan ekstraksi cerdas dari nilai-nilai berguna dari sekumpulan besar kumpulan data terstruktur dan tidak terstruktur dapat dimanifestasikan. Memang, pengelompokan membuat pengelompokan logis volume data tidak terstruktur dalam upaya untuk menganalisisnya. Namun, *Clustering* dianggap sebagai tugas berat dalam domain *Data Mining* dan mendapatkan momentum besar oleh para peneliti di masa lalu [4].



Gambar 1. K-Means Clustering dalam tindakan 2 dimensi

Jalan adalah prasarana perhubungan darat yang menjadi jalur transportasi yang sangat vital [5] serta menjadi pendukung lalu lintas kendaraan, termasuk bangunan pelengkap yang digunakan untuk mendukung pembangunan [6]. Ruas jalan adalah bagian jalan di antara persimpangan jalan. Batas ruas jalan adalah ketentuan yang menjadi patokan antara ruas jalan [7]. Patokan ini yang akan memisahkan antara ruas jalan yang pendek hingga yang panjang. Jika ruas jalan kurang dari sama dengan 15,01 Km maka ruas jalan termasuk ke dalam tingkat pendek (C1), dan jika ruas jalan kurang dari sama dengan 38,44 Km maka ruas jalan termasuk ke dalam tingkat sedang (C2), dan jika tidak maka ruas jalan termasuk ke dalam tingkat panjang (C3). Berdasarkan data ruas jalan di Balai Besar Pelaksanaan Jalan Nasional (BBPJJN) Sumut, ada sekitar 176 ruas jalan yang belum dikelompokkan ke dalam batas ruas jalan. Untuk data ruas jalan di BBPJJN Sumut dapat dilihat pada tabel di bawah ini.

Tabel 1. Data Ruas Jalan di BBPJN Sumut

NO	NAMA RUAS	PANJANG RUAS (KM)
1	BTS. PROV. ACEH - SIMPANG PANGKALAN SUSU	27,15
2	SIMPANG PANGKALAN SUSU - TANJUNG PURA	29,40
3	TANJUNG PURA - BTS. KOTA STABAT	17,13
4	JLN. ZAENAL ARIFIN (STABAT)	1,30
5	BTS. KOTA STABAT - BTS. KOTA BINJAI	10,00
6	JLN. JEND. SUDIRMAN (STABAT)	2,80
7	JLN. AMIR HAMZAH (BINJAI)	7,50
8	JLN. LINGKAR LUAR BINJAI	7,98
9	BTS. KOTA BINJAI - BTS. KOTA MEDAN	8,02
10	JLN. SOEKARNO-HATTA (BINJAI)	4,63
11	JLN. BINJAI RAYA (MEDAN)	2,58
12	BTS. KOTA MEDAN - BTS. KOTA LUBUK PAKAM	14,03
13	JLN. INDUSTRI / JLN. GAGAK HITAM (MEDAN)	5,16
14	JLN. NGUMBAN SURBAKTI (MEDAN)	3,44
...

Berdasarkan data yang disajikan pada tabel 1 terlihat bahwa belum adanya pengelompokan terhadap batas ruas jalan. Oleh karena itu, BBPJN Sumut perlu mengambil langkah untuk mengelompokkan ruas jalan berdasarkan tingkat *Cluster* yang ada.

2. METODE PENELITIAN

1) Pengambilan Data

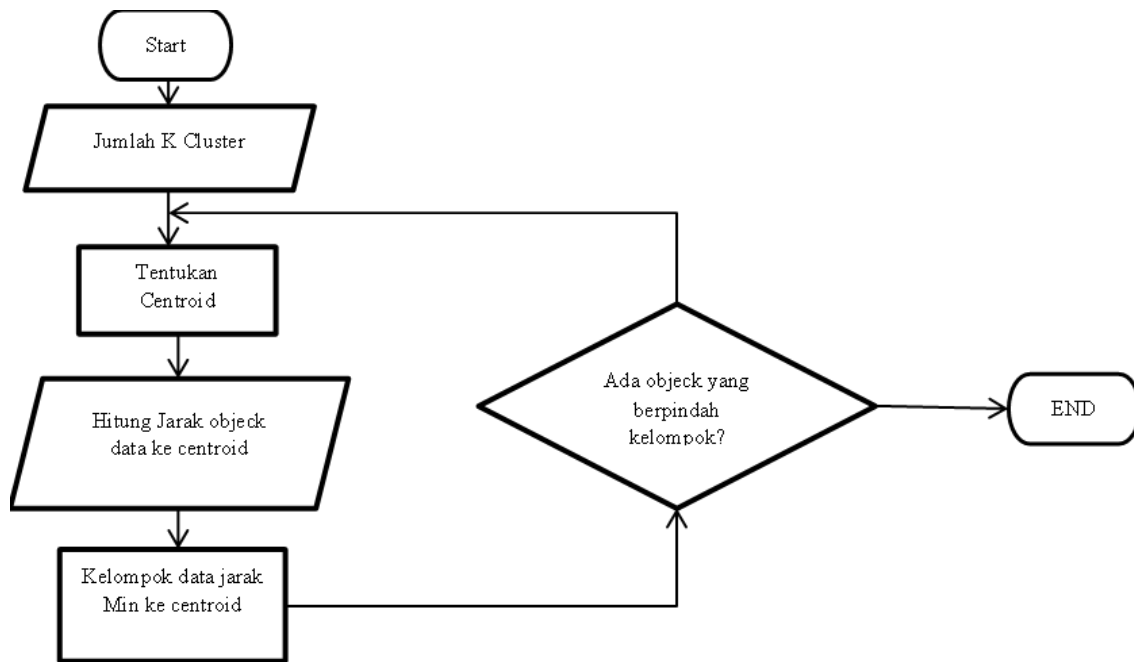
Data yang digunakan dalam penelitian ini adalah data yang diambil dari Balai Besar Pelaksanaan Jalan Nasional (BBPJN) Sumut pada tahun 2022, sebanyak 176 data jalan yang ada di provinsi Sumatera Utara.

2) Metode Penelitian

Penelitian ini menggunakan Algoritma *K-Means Clustering*. *K-Means* merupakan salah satu algoritma yang dapat melakukan *Clustering* dalam *Data Mining* [8], *Clustering* merupakan salah satu metode yang digunakan untuk membagi rangkaian data menjadi beberapa *Group* berdasarkan kesamaan kesamaan (*Similar*) yang telah ditentukan di dalam atribut [9]. Terkadang *Clustering* menjadi masalah yang menantang dari sudut pandang komputasi dan aspek data yang membuatnya menantang [10].

3) Flowchart

Flowchart adalah alur media pembelajaran berdasarkan peta situs untuk memudahkan penataan alur media pembelajaran [11]. *Flowchart* dari algoritma *K-Means* dapat dilihat pada gambar dibawah ini:



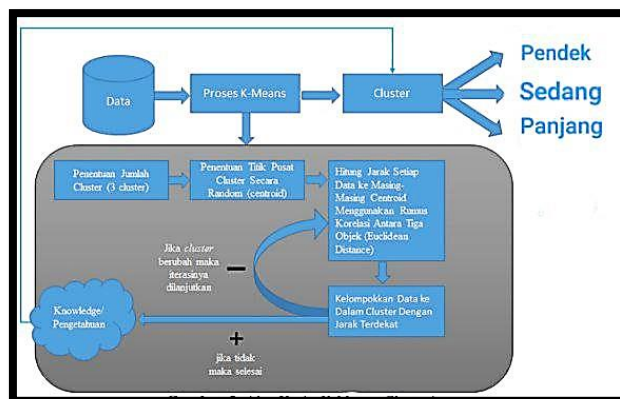
Gambar 2. *Flowchart K-Means* [12]

Adapun cara kerja algoritma K-Means dalam melakukan Clustering adalah sebagai berikut:

- Defenisikan Jumlah K Cluster
- Inisialisasi K pusat cluster (Centroid) yang dapat dipilih dari cluster K atau diperoleh secara acak.
- Untuk setiap record data, hitung dan tandai jarak terdekat ke centroid awal. Untuk melakukan proses pengolahan data pada tiap titik pusat cluster yang dirumuskan sebagai berikut:
$$D(i,j) = \sqrt{(X1i-X1j)^2 + (X2i-X2j)^2 + \dots + (Xki-Xkj)^2}$$

$D(i,j)$ = jarak dari data I kepusat cluster j
 Xki = data ke-I pada atribut data ke-k
 Xkj = titik pusat ke-j pada atribut ke-k
- Hitung dan ubah kembali centorid setiap group cluster,
- Ulangi langkah a ke b, jika anggota cluster tidak berubah (Konvergen), maka berhenti [13].

Adapun langkah-langkah dalam pengolahan data dengan menggunakan metode *K-Means Clustering*:



Gambar 3. Alur kerja *K-Means Clustering* [14].

Gambar 3 merupakan proses algoritma *K-Means Clustering* dimana pada tahapan pertama dilakukan pengumpulan data, *Data Cleaning* dan data transformasi, tahapan selanjutnya adalah proses pengolahan data menggunakan algoritma *K-Means Clustering* dengan menentukan jumlah *Cluster* secara *Random*, titik pusat *Cluster*, menghitung jarak setiap data dan mengelompokkan data pada *Cluster* sehingga menghasilkan *Knowledge* yaitu *Clustering*.

3. HASIL DAN PEMBAHASAN

1) *Centroid*

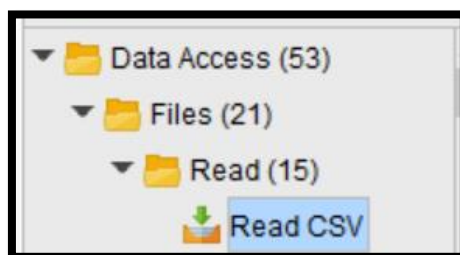
Center of Cluster atau *Centroid* adalah titik awal pengelompokan dalam metode pengelompokan partisi pengelompokan [15]. Nilai *Midpoint*, atau *Data Centroid*, dihasilkan saat mengimplementasikan Algoritma *K-Means*. Metode penentuan nilai titik tengah dicapai dengan dengan mengikuti nilai *Cluster* terendah (minimum) untuk *Cluster* pendek (C1), nilai rata-rata *Cluster* sedang (C2), dan nilai *Cluster* terbesar (C3).

Tabel 2. Data *Centroid* Awal

Data Cluster	Panjang Ruas
(C1) <i>Cluster</i> pendek	0,40
(C2) <i>Cluster</i> sedang	26,20
(C3) <i>Cluster</i> panjang	70,28

2) Implementasi *RapidMiner*

Berdasarkan data pada tabel 1, pengelompokan data ruas jalan di BBPJN Sumut akan dilakukan dengan *RapidMiner*. Pengelompokan metode Gambar berikut mengilustrasikan penggunaan *RapidMiner* untuk *K-Means*.



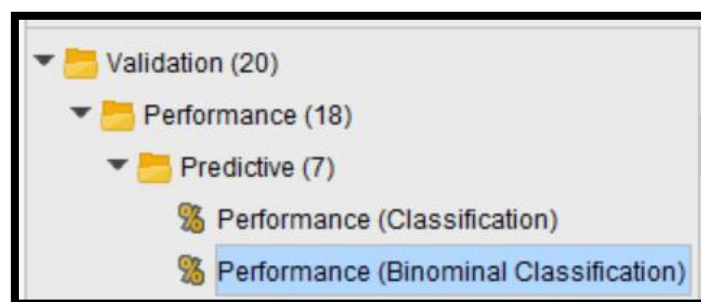
Gambar 4. Operator Pembaca Data CSV

Gambar 4 adalah operator *RapidMiner* yang digunakan untuk menginput data penelitian yang berbasis CSV.



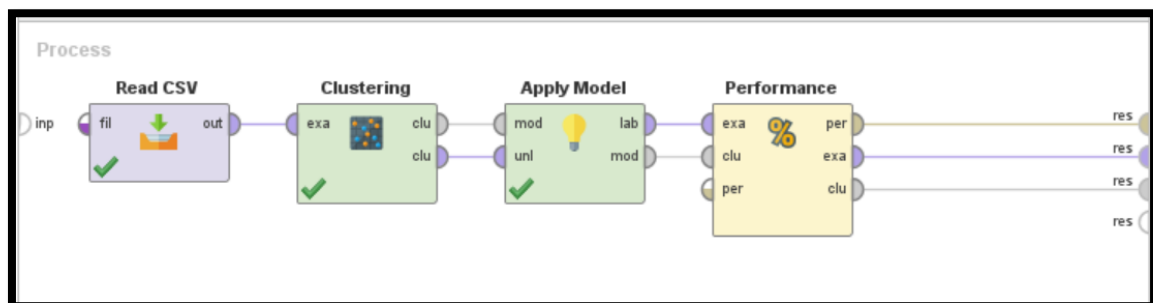
Gambar 5. Operator *Clustering*

Gambar 5 adalah operator *RapidMiner* yang digunakan untuk mengakses *Clustering* dari *K-Means*.



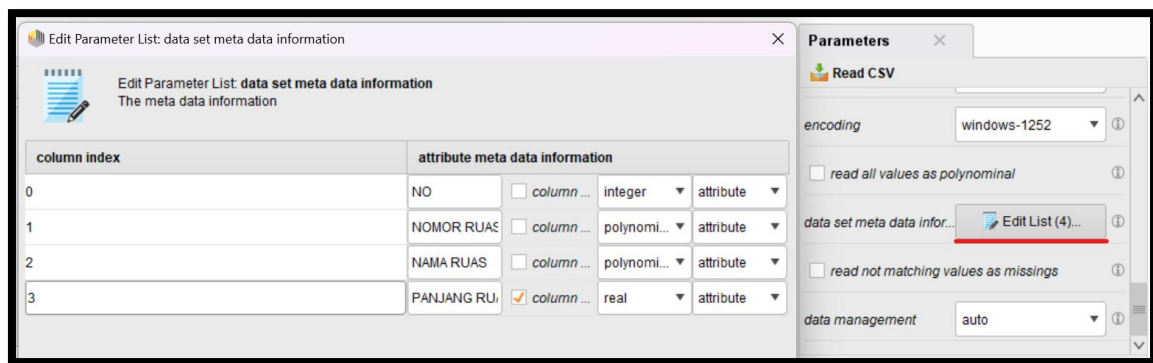
Gambar 6. Operator *Centroid*

Gambar 6 adalah operator *RapidMiner* yang digunakan untuk membuat *Centroid* pada data penelitian yang berbasis CSV.



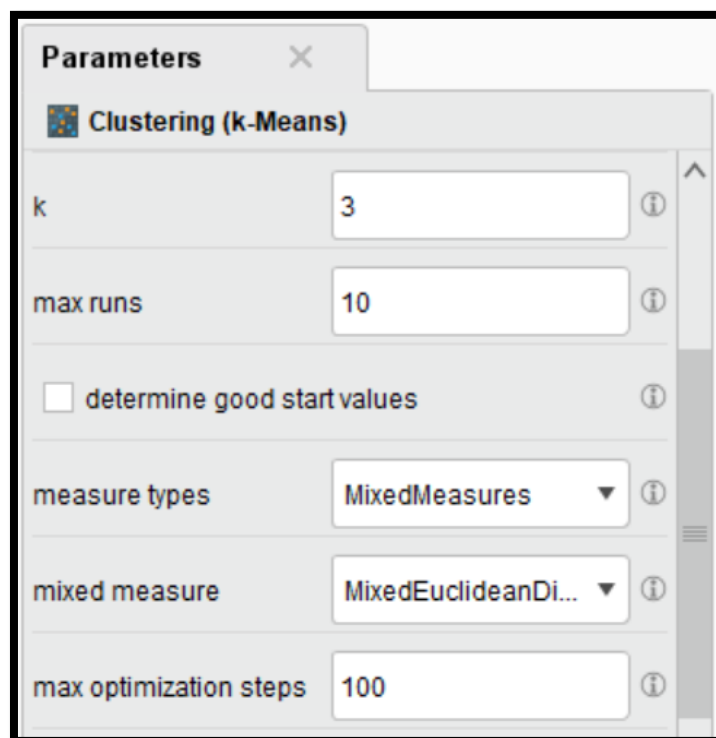
Gambar 7. Urutan Operator

Gambar 7 adalah Urutan Operator *RapidMiner* pada kolom proses.



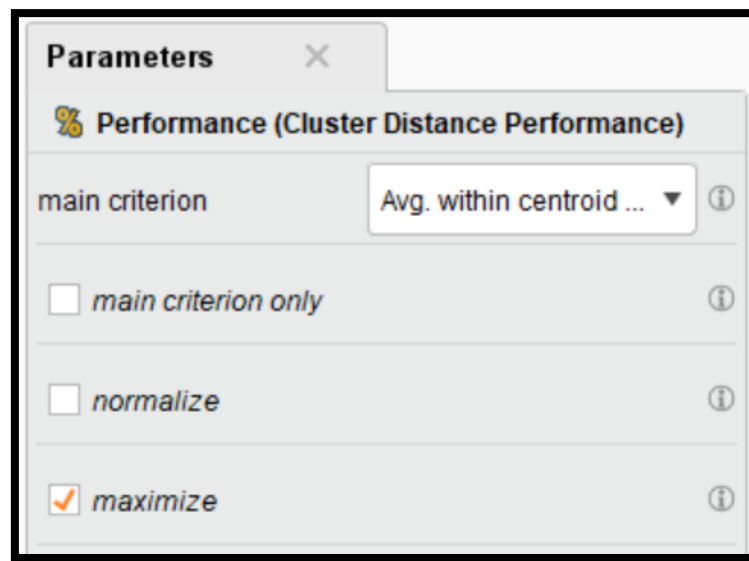
Gambar 8. *Data Set Meta Data Information*

Gambar 8 adalah setingan parameter *List Data Set Meta Data Information*.



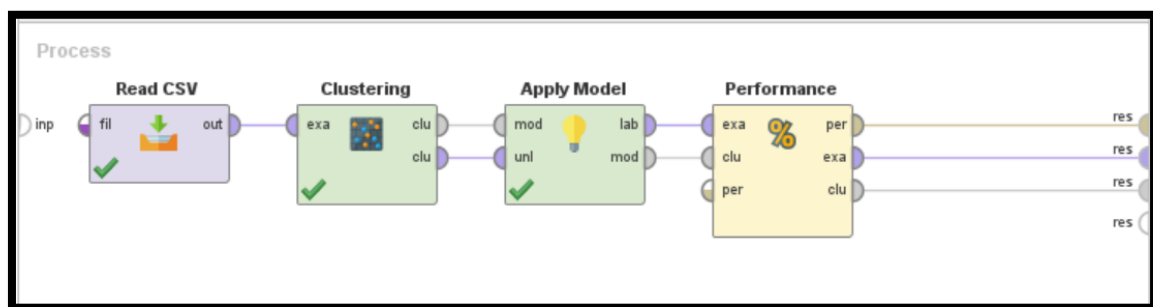
Gambar 9. *Parameter Clustering (K-Means)*

Gambar 9 adalah setingan parameter *Clustering* dimana nilai $k = 3$ dengan *Max Runs* = 10 dan *Measure Types* = *Mixed* dengan *Mixed Measure* = *MixedEuclidean*.



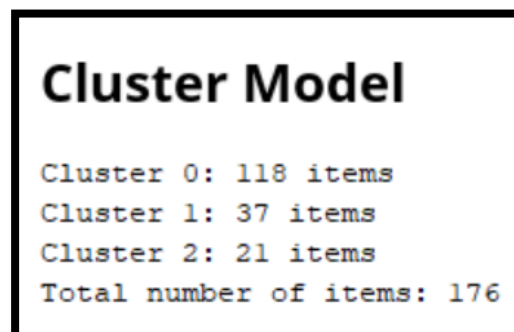
Gambar 10. Parameter *Performance*

Gambar 10 adalah setingan parameter *Centroid* dimana kolom *maximize* di aktifkan.



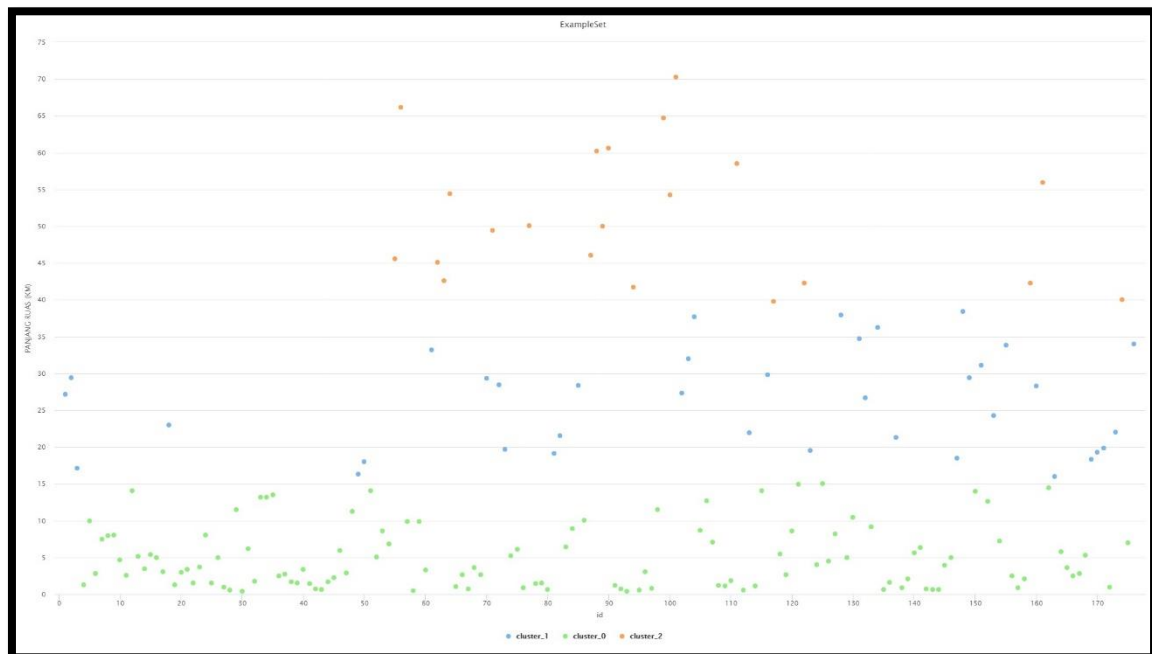
Gambar 11. Gambar 2. Proses Pengelompokan Algoritma K-Means Dengan Nilai K = 3

Gambar 11 adalah proses menghubungkan *Read Excel* dengan *K-Means* dan *Output* yang akan dieksekusi dengan nilai K = 3. Semua parameter yang diperlukan disimpan dalam model objek. Dalam hal ini peneliti menggunakan sampel data ruas jalan di BBPJN Sumut (Berdasarkan Tabel 1) dengan 1 atribut yaitu: Panjang ruas. Pengelompokan data Hasil Akhir dapat dilihat pada gambar di bawah ini.



Gambar 12. Pengelompokan Hasil menggunakan *K-Means*

Dapat diperjelas dari gambar 12 terdapat 3 *Cluster* mulai dari *Cluster* 0, kemudian *Cluster* 1, dan *Cluster* 2. *Cluster* 0 pendek, *Cluster* 1 sedang, dan *Cluster* 2 panjang. Jadi penulis mendapatkan grafik tampilan plot dari pengujian dengan *RapidMiner* sebagai berikut:



Gambar 13. Grafik *Clustering* dalam Tampilan Plot

4. KESIMPULAN DAN SARAN

Pengelompokan panjang ruas jalan di BBPJN Sumut dapat dilakukan dengan *Data Mining*. Metode *Data Mining* yang digunakan adalah *K-means* dengan memanfaatkan Tool *RapidMiner*. Berdasarkan 1 atribut penilaian panjang ruas, pengelompokan menggunakan 3 *Cluster*, yaitu: (C1) *Cluster* pendek, (C2) *Cluster* sedang, dan (C3) *Cluster* panjang. Dari hasil *Clustering*, 118 jalan berada pada (C1) *Cluster* pendek. Jalan-jalan tersebut juga termasuk ke dalam jalan yang sering dilalui oleh transportasi umum maupun pribadi. Sedangkan 21 jalan berada pada (C3) *Cluster* panjang. Jalan-jalan tersebut lebih sering digunakan sebagai jalan lintas antar Kota. Selain itu batas ruas jalan di BBPJN Sumut menjadi semakin jelas karena sudah dikelompokkan menjadi 3 *Cluster*.

DAFTAR PUSTAKA

- [1] A. Vilorio, G. C. Acuña, D. J. Alcázar Franco, H. Hernández-Palma, J. P. Fuentes, and E. P. Rambal, "Integration of Data Mining Techniques to PostgreSQL Database Manager System," *Procedia Comput Sci*, vol. 155, pp. 575–580, 2019, doi: 10.1016/j.procs.2019.08.080.
- [2] P. Edastama, A. S. Bist, and A. Prambudi, "Implementation Of Data Mining On Glasses Sales Using The Apriori Algorithm," *International Journal of Cyber and IT Service Management*, vol. 1, no. 2, pp. 159–172, Oct. 2021, doi: 10.34306/ijcitsm.v1i2.46.

-
- [3] K. Kotecha, V. Piuri, H. N. Shah, and R. Patel, "Data Science and Intelligent Applications: Proceedings of ICDSIA 2020," *Springer Singapore*, 2021.
 - [4] M. Mittal, L. M. Goyal, D. J. Hemanth, and J. K. Sethi, "Clustering approaches for high-dimensional databases: A review," *WIREs Data Mining and Knowledge Discovery*, vol. 9, no. 3, May 2019, doi: 10.1002/widm.1300.
 - [5] M. S. Lauryan and M. Ibrohim, "SISTEM INFORMASI GEOGRAFIS TINGKAT KERUSAKAN RUAS JALAN BERBASIS WEB," *JSiI (Jurnal Sistem Informasi)*, vol. 6, no. 1, p. 20, Mar. 2019, doi: 10.30656/jsii.v6i1.1022.
 - [6] T. Triyanto, S. Syaiful, and R. Rulhendri, "EVALUASI TINGKAT KERUSAKAN JALAN PADA LAPIS PERMUKAAN RUAS JALAN TEGAR BERIMAN KABUPATEN BOGOR," *ASTONJADRO*, vol. 8, no. 2, p. 70, Jan. 2020, doi: 10.32832/astonjadro.v8i2.2628.
 - [7] E. Bhaskara Putra and K. H. Putra, "STUDI KELAS JALAN PADA RUAS JALAN KRIAN-BATAS KABUPATEN MOJOKERTO."
 - [8] F. Nasari and S. Darma, "PENERAPAN K-MEANS CLUSTERING PADA DATA PENERIMAAN MAHASISWA BARU (STUDI KASUS : UNIVERSITAS POTENSI UTAMA)," pp. 6–8, 2015.
 - [9] M. Benri, H. Metisen, and S. Latipa, "ANALISIS CLUSTERING MENGGUNAKAN METODE K-MEANS DALAM PENGELOMPOKKAN PENJUALAN PRODUK PADA SWALAYAN FADHILA," 2015.
 - [10] V. Y. Kiselev, T. S. Andrews, and M. Hemberg, "Challenges in unsupervised clustering of single-cell RNA-seq data," *Nature Reviews Genetics*, vol. 20, no. 5. Nature Publishing Group, pp. 273–282, May 01, 2019. doi: 10.1038/s41576-018-0088-9.
 - [11] M. A. Hamid, L. Yuliawati, and D. Aribowo, "Feasibility of electromechanical basic work e-module as a new learning media for vocational students," *Journal of Education and Learning (EduLearn)*, vol. 14, no. 2, pp. 199–211, May 2020, doi: 10.11591/edulearn.v14i2.15923.
 - [12] M. Ahmed, R. Seraj, and S. M. S. Islam, "The k-means Algorithm: A Comprehensive Survey and Performance Evaluation," *Electronics (Basel)*, vol. 9, no. 8, p. 1295, Aug. 2020, doi: 10.3390/electronics9081295.
 - [13] S. Rudiarto, "IMPLEMENTASI ALGORITMA K-MEANS CLUSTERING PADA APLIKASI PENCARI PELANGGAN POTENSIAL PADA RESTORAN XYZ," *Jurnal Ilmu Teknik dan Komputer*, vol. 2, no. 1, 2018.
 - [14] B. Harahap, "Penerapan Algoritma K-Means Untuk Menentukan Bahan Bangunan Laris (Studi Kasus Pada UD. Toko Bangunan YD Indarung)," 2019.
 - [15] B. Jumadi Dehotman Sitompul, O. Salim Sitompul, and P. Sihombing, "Enhancement Clustering Evaluation Result of Davies-Bouldin Index with Determining Initial Centroid of K-Means Algorithm," in *Journal of Physics: Conference Series*, Jul. 2019, vol. 1235, no. 1. doi: 10.1088/1742-6596/1235/1/012015.