



## Penerapan Clustering Terhadap Segmentasi Zonasi Gangguan Layanan Pelanggan Dengan Menggunakan Analisis RFM *Clustering Implementation of Zones User Segmentation Using RFM Analysis*

Nurhadi<sup>1</sup>, Yan Puspitarani<sup>2</sup><sup>1,2</sup> Teknik Informatika, Universitas Widyatama<sup>1</sup>Nurhadi@widyatama.ac.id, <sup>2</sup>Yan.puspitarani@widyatama.ac.id**Abstract**

*The service area and the wide distribution of customers often require quite a long time in terms of handling disturbances, it can reduce the level of customer trust and loyalty in using the network services provided. Customer segmentation by region can help make it easier to analyze disruptions that occur to customer service in customer area zoning so that customer trust and loyalty can be maintained. This research was conducted in order to understand the characteristics of disturbances that occur in several customer zoning areas. In this study, RFM (Recency, Frequency, Monetary) analysis was performed. Clustering using the K-Means modelling technique, and also performance checking using Cluster Distance Performance to determine the performance of each cluster performed. The results of this study indicate that K-Means Clustering with a value of K=2 is the with the smallest Davies Bouldin value. It meaning that K=2 has the best performance when compared to other clusters formed..*

*Keywords: Clustering, Data Mining, Customer Segmentation, K-Means, RFM.*

**Abstrak**

Area pelayanan dan persebaran pelanggan yang luas seringkali membutuhkan waktu yang cukup lama dalam hal penanganan gangguan, hal ini dapat memicu menurunnya tingkat kepercayaan dan kesetiaan pelanggan dalam menggunakan jasa layanan jaringan yang diberikan. Segmentasi pelanggan berdasarkan wilayah dapat membantu mempermudah dalam melakukan analisis gangguan yang terjadi terhadap layanan pelanggan pada zonasi wilayah pelanggan agar kepercayaan dan kesetiaan pelanggan dapat tetap terjaga. Penelitian ini dilakukan agar dapat memahami karakteristik dari gangguan yang terjadi di beberapa wilayah zonasi pelanggan. Pada penelitian ini dilakukan analisis RFM (Recency, Frequency, Monetary). Clustering dengan menggunakan Teknik pemodelan K-Means dan melakukan pengecekan performa menggunakan *Cluster Distance Performance* untuk mengetahui performa dari masing – masing cluster yang terbentuk. Hasil dari penelitian ini menunjukkan bahwa K-Means Clustering dengan nilai K=2 adalah Cluster dengan nilai Davies Bouldin terkecil. Hal ini berarti bahwa K=2 memiliki performansi terbaik bila dibandingkan dengan cluster lain yang terbentuk.

Kata kunci : Data Mining, Clustering, Segmentasi Pelanggan, RFM, K-Means.

**1. Pendahuluan**

Data merupakan asset penting bagi perusahaan, karena dari data – data yang ada kita dapat memproyeksikan apa yang perlu dilakukan untuk pengembangan bisnis dengan cara menganalisa data yang ada. Maka data perlu digali untuk kemudian diolah menjadi informasi dan pengetahuan yang bermanfaat bagi perusahaan. PT. XYZ, merupakan perusahaan yang bergerak di bidang jasa penyedia layanan jaringan / internet. Perusahaan ini menyediakan layanan jaringan telekomunikasi berbasis

serat optic sebaga medianya. Area pelayanan dan persebaran pelanggan yang luas seringkali membutuhkan waktu yang cukup lama dalam hal penanganan gangguan, hal ini dapat menurunkan tingkat kepercayaan dan kesetiaan pelanggan dalam menggunakan jasa layanan jaringan yang disediakan oleh PT. XYZ. Tentunya perusahaan memiliki tujuan untuk mempererat hubungan antara pelanggan dengan perusahaan untuk mendapatkan keuntungan yang signifikan dalam persaingan pasar. Maka dari itu

perusahaan perlu terus mengembangkan keterampilan dan mengidentifikasi pelanggan untuk kemudian menyediakan kebutuhan pelanggan[1].

Oleh karena itu melakukan segmentasi pelanggan berdasarkan wilayah dapat membantu dalam hal pengelompokan pelanggan (clustering) untuk kemudian dianalisa berdasarkan wilayah dari pelanggan yang sering terjadi gangguan dan juga penyebab dari gangguan yang sering terjadi di wilayah tersebut.

Penelitian ini bertujuan untuk menganalisa hasil pengelompokan zonasi pelanggan yang sering terjadi gangguan dan juga penyebab dari penanganan gangguan yang membutuhkan waktu recovery yang lama dan juga penyebab dari gangguan yang sering muncul pada zonasi pelanggan.

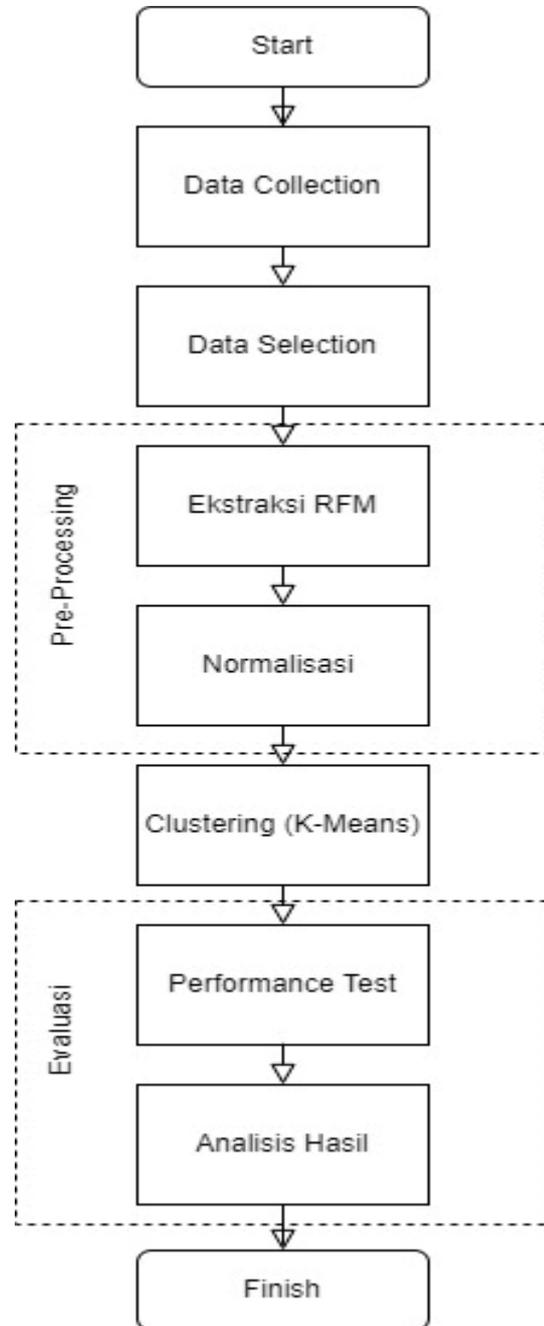
Cara untuk melakukan pengelompokan pelanggan adalah dengan menggunakan teknik penggalian data (data mining). Salah satu metode dari data mining yang dapat dilakukan adalah metode Clustering. Metode Clustering dikenal juga dengan unsupervised learning yang membagi data menjadi kelompok – kelompok atau cluster berdasarkan kemiripan atribut – atribut yang ada dalam data tersebut dan tidak memiliki label[2].

Dalam penelitian ini akan dilakukan analisa terhadap interval gangguan layanan terakhir, frekuensi gangguan, dan durasi gangguan, ketiga variabel ini akan dimodelkan dalam bentuk RFM (Recency, Frequency, Monetary). Analisis terhadap segmentasi pelanggan dengan menggunakan pemodelan RFM ini sebelumnya pernah dilakukan oleh B.E. Adiana, S. Indah, dan P.A Erna (2018) yang melakukan penelitian untuk mengidentifikasi pelanggan yang berpotensi memberikan keuntungan yang tinggi[2].

Penelitian lain dengan menggunakan analisis RFM dilakukan oleh Muhammad Fadillah (2019) [3], dengan mengukur nilai profitabilitas pelanggan terhadap perusahaan berdasarkan transaksi yang dilakukan. Hasil dari penelitian sebelumnya memiliki nilai validasi cluster yang cukup baik dalam pengelompokan pelanggan. Selain Penelitian yang dilakukan oleh Adiana dan M. Fadilah, segmentasi pelanggan dengan menggunakan analisis RFM juga dilakukan oleh Jamal dan Yanto (2019) [4] dalam mengelompokkan Loyalitas Customer. Analisis RFM berdasarkan segmentasi pelanggan dapat menghasilkan kemampuan segmentasi antara 75% sampai dengan 85% [5]. Berdasarkan permasalahan yang terjadi maka akan dilakukan “Penerapan Clustering Terhadap Segmentasi Zonasi Gangguan Layanan Pelanggan dengan menggunakan Analisis RFM”. Penelitian ini diharapkan dapat membantu dalam melakukan Analisa terhadap layanan pelanggan untuk memudahkan dalam hal melakukan penanganan gangguan maupun melakukan peningkatan kualitas kehandalan layanan pelanggan.

## 2. Metodologi Penelitian

Metodologi yang dilakukan dalam penelitian ini terdiri dari 4 tahapan, yaitu Pengumpulan Data, Pemilihan Data, Pre-Processing, Simulasi (Clustering), dan Evaluasi Hasil. Tahapan dalam metodologi tersebut digambarkan dalam flowchart berikut :



Gambar 1. Flowchart Metodologi Penelitian

### 2.1. Pengumpulan Data dan Pemilihan Data

Pengumpulan data dilakukan dengan cara mengunduh RAWDATA gangguan dari aplikasi ticketing gangguan PT. XYZ. Data yang di unduh merupakan data dengan format file \*.xls, dengan waktu periode gangguan 1

Januari 2021 – 30 Juni 2021 dengan jumlah data sebanyak 3384 gangguan.

Data Selection merupakan proses seleksi atau pemilihan data yang relevan terhadap analisis untuk diterima dari koleksi data yang ada [6]. Dari sekian banyak atribut yang ada pada data yang sebelumnya telah di unduh, maka diambil 3 atribut untuk dijadikan sebagai atribut RFM (Recency, Frequency, dan Monetary). Atribut yang akan digunakan sebagai nilai R adalah tanggal terakhir terjadinya gangguan layanan, nilai F adalah frekuensi terjadinya gangguan, sedangkan nilai M (Monetary) diambil dari durasi gangguan (satuan menit) pada suatu wilayah atau zonasi.

## 2.2. Pre-processing Data

Tahapan ini dilakukan agar data mentah dapat diubah menjadi data yang terstruktur. Adapun langkah pertama dalam pre-processing data adalah dengan melakukan penerapan dan ekstraksi RFM.

Pada tahapan ini, akan dipilih tiga variable untuk dijadikan sebagai variable RFM (*Recency*, *Frequency*, dan *Monetary*). Data yang telah dikumpulkan dalam data selection kemudian akan diubah kedalam variable RFM. Atribut *Recency* diambil dari data waktu / tanggal terakhir terjadi gangguan, sedangkan atribut *Frequency* diambil dari seberapa sering terjadi gangguan, dan atribut *Monetary* diambil dari data durasi gangguan.

## 2.3. Ekstraksi Model RFM (*Recency*, *Frequency*, *Monetary*)

### Ekstraksi Variabel Recency

*Recency* merupakan selisih antara tanggal yang ditentukan dengan tanggal transaksi pelanggan terakhir, dalam hal ini adalah tanggal terakhir pelanggan mengalami gangguan pada layanannya. Variabel yang digunakan untuk menghasilkan variabel *Recency* adalah *ARDate*. Setelah mendapatkan nilai dari hasil ekstraksi variable *Recency*, tahapan selanjutnya adalah dengan melakukan ekstraksi variabel *frequency*.

Ekstraksi variabel *Frequency* dilakukan dengan cara menghitung jumlah frekuensi gangguan yang terjadi di suatu wilayah dalam periode waktu tertentu. Kemudian lakukan ekstraksi variable *monetary*. Pada penelitian ini yang dijadikan sebagai variabel *monetary* adalah durasi gangguan dengan satuan menit.

## 2.4. Normalisasi (Min-Max)

Tahapan selanjutnya setelah melakukan ekstraksi nilai RFM adalah melakukan penerapan normalisasi (Min-Max). Normalisasi data digunakan untuk menskalakan nilai data dalam rentang yang ditentukan untuk memudahkan tahapan – tahapan perhitungan seperti perhitungan kesamaan atau operasi *clustering*[7][8].

Salah satu metode untuk melakukan normalisasi adalah metode *Min-Max* yang merupakan Teknik sederhana dalam melakukan penskalaan nilai berdasarkan batas – batas yang telah ditentukan[9].

Normalisasi data juga dilakukan agar jarak dari ketiga variable RFM tidak terlalu jauh. Persamaan untuk melakukan metode normalisasi (min-max) adalah sebagai berikut [10]:

$$Rci = \frac{o_i^R - \min^R}{\max^R - \min^R} (\text{newmax}^R - \text{newmin}^R) + \text{newmin}^R \quad (1)$$

$$Fci = \frac{o_i^F - \min^F}{\max^F - \min^F} (\text{newmax}^F - \text{newmin}^F) + \text{newmin}^F \quad (2)$$

$$Mci = \frac{o_i^M - \min^M}{\max^M - \min^M} (\text{newmax}^M - \text{newmin}^M) + \text{newmin}^M \quad (3)$$

Oi R , Oi F , dan Oi M adalah nilai mentah yang akan dinormalisasi untuk masing – masing *Recency*, *Frequency*, dan *Monetary*. Sedangkan untuk  $\min^R$  ,  $\min^F$  , dan  $\min^M$  adalah nilai minimum dari masing – masing R,F,M. Begitu juga dengan  $\max^R$  ,  $\max^F$  , dan  $\max^M$  yang menunjukkan nilai maksimal dari R,F,M.

$\text{Newmax}^R$  ,  $\text{newmax}^F$  , dan  $\text{newmax}^M$  adalah rentang nilai maksimal R, F, dan M baru yang mempunyai nilai 1, sedangkan  $\text{newmin}^R$  ,  $\text{newmin}^F$  ,  $\text{newmin}^M$  adalah rentang baru nilai minimal masing – masing R, F, dan M yang bernilai 0. Hasil dari persamaan 1, 2, dan 3 akan memiliki nilai yang rentangnya antara 0 sampai dengan 1.

## 2.5. Clustering

*Clustering* adalah salah satu sub kategori data mining dan merupakan proses dimana sampel yang sama dibagi menjadi kelompok – kelompok yang disebut *cluster*. Setiap *cluster* termasuk sampel dimana anggota yang mirip satu sama lain dan berbeda dengan sampel yang tersedia dari kelompok lain [11], proses *clustering* akan mengelompokkan data kedalam suatu kelompok yang memiliki kemiripan maupun karakteristik yang sama [12].

Setelah selesai melakukan ekstraksi RFM dan normalisasi (Min-Max), langkah selanjutnya adalah melakukan pengelompokkan (*clustering*). Pada penelitian ini teknik *clustering* yang digunakan adalah teknik pemodelan K-Means menggunakan aplikasi Rapid miner.

*K-Means* adalah salah satu algoritma *clustering* yang banyak digunakan pada system rekomendasi dan data mining[13][14]. Huruf K pada K-Means *Clustering* dimaksudkan sebagai konstanta jumlah *cluster* yang diinginkan, Bagian penting dari algoritma K-Means yaitu *centroid* yang merupakan titik tengah sebagai pusat *cluster*[15]. Metode K-Means berusaha mengelompokkan data yang ada kedalam beberapa kelompok[16].

Setelah selesai melakukan pengelompokkan dengan menggunakan Teknik pemodelan K-Means Clustering, maka Langkah selanjutnya adalah melakukan evaluasi dan pengujian performa cluster dari pengelompokkan tersebut.

#### 2.4. Davies Bouldin Index

Davies Bouldin Index (DBI) adalah sebuah metode yang diperkenalkan oleh David L. Davies dan Donald W. Bouldin yang digunakan untuk mengevaluasi cluster [17]. Landasan dari metode ini yaitu nilai kohesi dan separasi [18]. Pada proses pengelompokkan, kohesi adalah jumlah dari kedekatan data terhadap *centroid* dari *cluster* yang diikuti. Sedangkan separasi didasarkan pada jarak antar *centroid* dari *cluster* tersebut. Evaluasi dengan menggunakan DBI ini memiliki skema evaluasi internal cluster, dimana baik atau tidaknya hasil cluster dilihat dari kuantitas dan kedekatan antar data hasil cluster [19].

### 3. Hasil dan Pembahasan

Sebelum melakukan simulasi clustering, perlu dilakukan Ekstraksi nilai RFM dari RAWDATA gangguan yang sebelumnya telah di unduh dari aplikasi ticketing gangguan PT. XYZ. Data mentah ini memiliki atribut sebagai berikut :

Tabel 1. Tabel Atribut RAWDATA Gangguan

Atribut	Tipe Data	Keterangan
Description	Text	Deskripsi dari nama dan alamat pelanggan yang mengalami gangguan layanan.
ID Zonasi	Number (integer)	Kode dari wilayah zonasi, terdapat 17 wilayah zonasi
Zonasi	Text	Nama wilayah zonasi, terdapat 17 wilayah zonasi
AR Date	Date	Tanggal dan waktu tiket laporan gangguan dibuat
Complete	Date	Tanggal dan waktu selesainya perbaikan gangguan
Durasi	Real (decimal)	Durasi Gangguan (Complete – AR Date) dengan satuan menit
Root Cause	Text	Jenis dari penyebab gangguan, terdapat 5 Root Cause yaitu : Cable, peripheral, PS (Power Supply), Software, dan Bukan gangguan
Basecamp	Text	Lokasi basecamp tim penanganan gangguan. Terdapat 9 lokasi basecamp yaitu : Bandung, Sumedang, Cirebon, Garut, Tasik, Cianjur, Karawang, Purwakarta, dan Subang.

#### 3.1. Ekstraksi RFM

Setelah mendapatkan data mentah pada proses sebelumnya, Langkah selanjutnya adalah menyiapkan data tersebut agar siap diolah dan dilakukan segmentasi dengan menghilangkan atribut yang tidak diperlukan dalam perhitungan nilai RFM. Hasil pencarian nilai RFM dapat dilihat pada tabel berikut :

Tabel 2. Tabel Hasil Pencarian Nilai RFM

ID	Zonasi	R	F	M
1	Bandung	270	630	1952873,48
2	Cimahi	270	108	1948820,07
3	Sumedang	272	139	163505
4	Majalengka	270	105	1890608,27
5	Cirebon	270	153	1952578,18
6	Kuningan	272	40	1920210,90
7	Indramayu	270	193	1965924,73
8	Garut	270	184	1935906,98
9	Tasik	270	96	1945990,67
10	Ciamis	273	103	1923522,80
11	Pangandaran	278	38	1916526,47
12	Cianjur	270	145	1961393,45
13	Sukabumi	270	205	141018,00
14	Pelabuhan Ratu	280	45	1900457,93
15	Subang	270	313	1933382,40
16	Purwakarta	272	504	1935701,25
17	Karawang	270	383	1962768,33

#### 3.2. Normalisasi

Langkah selanjutnya setelah mendapatkan nilai RFM adalah melakukan normalisasi (Min-Max). Normalisasi ini dilakukan agar rentang perbedaan nilai di antara ketiga variable RFM tidak terlalu jauh karena hasil dari normalisasi ini akan membuat nilai dari masing – masing RFM dalam rentang 0 – 1. Nilai terendah dari normalisasi adalah 0, sedangkan nilai tertinggiya adalah 1.

Tabel 3. Tabel Hasil Normalisasi (Min-Max)

ID	Zonasi	Rci	Fci	Mci
1	Bandung	0,964286	1	0,993361
2	Cimahi	0,964286	0,171429	0,991299
3	Sumedang	0,971429	0,220635	0,08317
4	Majalengka	0,964286	0,166667	0,961689
5	Cirebon	0,964286	0,242857	0,993211
6	Kuningan	0,971429	0,063492	0,976747
7	Indramayu	0,964286	0,306349	1
8	Garut	0,964286	0,292063	0,984731
9	Tasik	0,964286	0,152381	0,98986
10	Ciamis	0,975	0,163492	0,978432
11	Pangandaran	0,992857	0,060317	0,974873
12	Cianjur	0,964286	0,230159	0,997695
13	Sukabumi	0,964286	0,325397	0,071731
14	Pelabuhan Ratu	1	0,071429	0,966699
15	Subang	0,964286	0,496825	0,983447
16	Purwakarta	0,971429	0,8	0,984626
17	Karawang	0,964286	0,607937	0,998394

Tabel 4. Jumlah Penyebab Gangguan Pada Setiap Zonasi

ID Zona	Cable	Peripherals	PS	Bukan Gangguan	Software	Total
1	376	125	71	54	4	630

ID Zona	Cable	Peripherals	PS	Bukan Gangguan	Software	Total
2	60	19	19	9	1	108
3	111	18	6	3	1	139
4	87	6	6	5	1	105
5	81	35	24	11	2	153
6	144	21	13	14	1	193
7	144	21	13	14	1	193
8	90	25	54	11	4	184
9	38	29	16	12	1	96
10	73	16	11	3	0	103
11	21	7	6	4	0	38
12	78	18	44	3	2	145
13	145	27	19	14	0	205
14	33	6	5	1	0	45
15	232	46	14	20	1	313
16	330	72	78	24	0	504
17	271	57	32	23	0	383

### 3.3. K-Means Clustering

Tahapan selanjutnya adalah membuat pemodelan dari data yang telah di olah. Pada penelitian ini dilakukan pemodelan clustering dengan menggunakan Teknik K-Means . Ketika diberikan nilai K=2 maka terbagi menjadi 2 cluster yaitu cluster 0 dan cluster 1. Dari 17 zonasi, 2 zonasi wilayah masuk ke cluster 0, sedangkan 15 zonasi wilayah masuk ke cluster 1.

Tabel 5. Tabel Centroid K=2

Atribut	Cluster_0	Cluster_1
Rci	0.968	0.970
Fci	0.273	0.322
Mci	0.077	0.985

Dari Tabel 5, nampak bahwa nilai atribut Rci dari Cluster\_0 dan Cluster\_1 tidak terpaut jauh, hal ini dikarenakan nilai Recency yang berdekatan, sedangkan nilai dari atribut Mci antara Cluster\_0 dan Cluster\_1 cukup jauh, hal ini disebabkan karena durasi gangguan yang terjadi pada zonasi wilayah yang tergabung ke Cluster\_1 jauh lebih besar dibandingkan dengan zonasi wilayah yang tergabung ke Cluster\_0.

Tabel 6. Hasil Clustering K=2

ID Zonasi	Rci	Fci	Mci	Cluster
1	0,964286	1	0,993361	1
2	0,964286	0,171429	0,991299	1
3	0,971429	0,220635	0,08317	0
4	0,964286	0,166667	0,961689	1
5	0,964286	0,242857	0,993211	1
6	0,971429	0,063492	0,976747	1
7	0,964286	0,306349	1	1
8	0,964286	0,292063	0,984731	1
9	0,964286	0,152381	0,98986	1
10	0,975	0,163492	0,978432	1
11	0,992857	0,060317	0,974873	1
12	0,964286	0,230159	0,997695	1
13	0,964286	0,325397	0,071731	0
14	1	0,071429	0,966699	1
15	0,964286	0,496825	0,983447	1
16	0,971429	0,8	0,984626	1
17	0,964286	0,607937	0,998394	1

Setelah sebelumnya memberikan nilai K=2, untuk mencari jumlah cluster yang paling baik, maka diberikan kembali pembobotan K=3, K=4, dan K=5.

Ketika K=3, maka terbagi menjadi 3 cluster yaitu, cluster 0 yang terdiri dari 4 zonasi, cluster 1 yang terdiri dari 2 wilayah zonasi, dan cluster 2 yang terdiri 11 wilayah zonasi.

Tabel 7. Tabel Centroid K=3

Atribut	Cluster_0	Cluster_1	Cluster_2
Rci	0.966	0.968	0.972
Fci	0.726	0.273	0.175
Mci	0.990	0.077	0.983

Ketika diberikan nilai K=4, maka terbagi menjadi 4 cluster yaitu, cluster 0 yang terdiri dari 2 zonasi, cluster 1 yang terdiri dari 2 wilayah zonasi, cluster 2 yang terdiri dari 11 wilayah zonasi, dan cluster 3 yang terdiri 2 wilayah zonasi.

Tabel 8. Tabel Centroid K=4

Atribut	Cluster_0	Cluster_1	Cluster_2	Cluster_3
Rci	0.968	0.964	0.972	0.968
Fci	0.273	0.552	0.175	0.900
Mci	0.077	0.991	0.983	0.989

Ketika K=5, maka terbagi menjadi 5 cluster yaitu, cluster 0 yang terdiri dari 2 zonasi, cluster 1 yang terdiri dari 7 wilayah zonasi, cluster 2 yang terdiri dari 2 wilayah zonasi, cluster 3 yang terdiri 2 wilayah zonasi, dan cluster 4 yang terdiri dari 4 wilayah zonasi.

Tabel 9. Tabel Centroid K=5

Atribut	Clstr_0	Clstr_1	Clstr_2	Clstr_3	Clstr_4
Rci	0.968	0.976	0.968	0.964	0.964
Fci	0.273	0.121	0.900	0.552	0.268
Mci	0.077	0.977	0.989	0.991	0.994

Setelah dilakukan modelling dengan menggunakan Teknik K-Means Clustering, langkah selanjutnya adalah melakukan evaluasi hasil clustering dengan menggunakan operator Performance Vector (Cluster Distance Performance) pada tools Rapid miner. Hasil evaluasi menggunakan Performance Vector adalah sebagai berikut :

Tabel 10. Tabel Performance Vector K=2

Description	Value
Average within centroid distance	-0.066
Average within centroid distance_cluster_0	-0.003
Average within centroid distance_cluster_1	-0.075
Davies Bouldin Index	-0.297

Tabel 11. Tabel Performance Vector K=3

Description	Value
Average within centroid distance	-0.014
Average within centroid distance_cluster_0	-0.007
Average within centroid distance_cluster_1	-0.037
Average within centroid distance_cluster_2	-0.003
Davies Bouldin Index	-0.371

Tabel 12. Tabel Performance Vector K=4

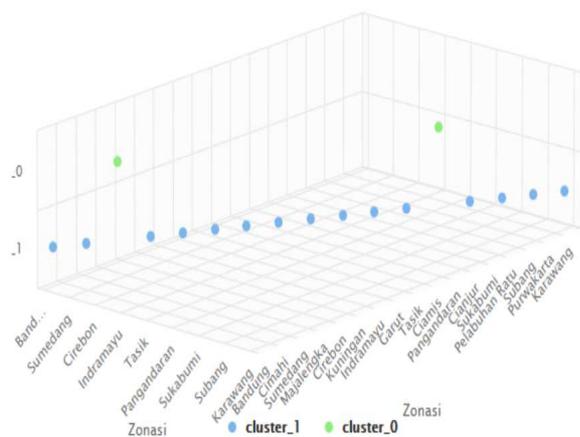
Description	Value
Average within centroid distance	-0.007
Average within centroid distance_cluster_0	-0.003
Average within centroid distance_cluster_1	-0.010
Average within centroid distance_cluster_2	-0.007
Average within centroid distance_cluster_3	-0.003
Davies Bouldin Index	-0.344

Tabel 13. Tabel Performance Vector K=5

Description	Value
Average within centroid distance	-0.003
Average within centroid distance_cluster_0	-0.003
Average within centroid distance_cluster_1	-0.003
Average within centroid distance_cluster_2	-0.010
Average within centroid distance_cluster_3	-0.003
Average within centroid distance_cluster_4	-0.001
Davies Bouldin Index	-0.432

3.4. Analisis Hasil

Setelah melakukan clustering, dan pengecekan performa dengan menggunakan bantuan operator Cluster Distance Performance, maka dapat dilakukan analisis dengan mengecek hasil dari Performance Vector yang memiliki nilai Davies Bouldin Index terkecil. Berdasarkan hasil pengamatan table Performance Vector diatas, Performance dari Vector K=2 memiliki nilai Davies Bouldin terkecil. Hal ini menunjukkan bahwa pengelompokkan zonasi menjadi 2 cluster (K=2) adalah yang terbaik.



Gambar 2. Plott Scatter 3D K=2

Gambar 2 menunjukkan pengelompokkan cluster menjadi dua cluster yang memiliki nilai performa terbaik berdasarkan hasil dari *Performance Vector*, Cluster 0 yang berisikan Zonasi wilayah Sumedang dan Sukabumi adalah cluster dengan performa terbaik.

Tabel 14. Hasil Analisis Cluster untuk Atribut Recency

Cluster	Hasil	Analisis
0	Cluster 0 memiliki nilai rentang Recency dari 0.96429 – 0.97143.	Baik Cluster 0 maupun cluster 1 memiliki nilai Recency yang sama pada batas bawah nilai Recency, hal ini disebabkan karena beberapa zonasi mengalami gangguan terakhir pada tanggal yang sama. Sedangkan zonasi Pelabuhan Ratu pada Cluster 1 memiliki nilai 1 yang berarti bahwa zonasi Pelabuhan Ratu memiliki nilai Recency paling besar karena tanggal terjadinya gangguan terakhir paling jauh waktunya dibandingkan dengan zonasi yang lain.
1	Cluster 1 memiliki nilai rentang Recency dari 0.96429 – 1.	

Tabel 15. Hasil Analisis Cluster untuk Atribut Frequency

Cluster	Hasil	Analisis
0	Cluster 0 memiliki nilai rentang Frequency dari 0.22063 – 0.3254.	Jika dilihat dari rentang nilai Frequency antara Cluster 0 dan Cluster 1 tidak terlihat perbedaan yang mencolok yang memisahkan kedua cluster tersebut.
1	Cluster 1 memiliki nilai rentang Frequency dari 0.06032 – 1.	

Tabel 16. Hasil Analisis Cluster untuk Atribut Monetary

Cluster	Hasil	Analisis
0	Cluster 0 memiliki nilai rentang Monetary dari 0.07173 – 0.08317.	Zonasi yang termasuk ke dalam cluster 0 memiliki dua nilai Monetary terendah jika dibandingkan dengan keseluruhan nilai monetary. Hal ini menunjukkan bahwa cluster 0 terbentuk dari kelompok zonasi yang memiliki durasi gangguan paling kecil. Sedangkan cluster 1 memiliki kemiripan dalam nilai monetary dari setiap anggota clusternya dengan nilai yang lebih besar dibandingkan dengan cluster 0 yang menunjukkan bahwa zonasi yang tergabung dalam cluster 1 adalah zonasi dengan durasi gangguan yang lama.
1	Cluster 1 memiliki nilai rentang Monetary dari 0.96169 – 1.	

Dilihat dari durasi gangguan yang terjadi pada Cluster 0 adalah zonasi dengan durasi gangguan paling kecil, namun jumlah gangguannya lebih banyak dari beberapa zonasi yang tergabung ke dalam Cluster 1. Hal ini menunjukkan bahwa penanganan gangguan yang

terjadi di Cluster 0 lebih baik dari penanganan gangguan yang terjadi di Cluster 1. Sedangkan beberapa zonasi yang tergabung kedalam Cluster 1 seperti zonasi Pangandaran, Kuningan, dan Pelabuhan Ratu memiliki total jumlah gangguan yang lebih sedikit bila dibandingkan dengan zonasi pada Cluster 0, namun durasi gangguannya dua kali lipat lebih lama dibandingkan zonasi pada Cluster 0. Hal ini disebabkan karena lokasi Basecamp yang terlalu jauh dengan lokasi gangguan.

Zonasi Subang, Karawang, Purwakarta, dan Bandung yang tergabung ke dalam Cluster 1 merupakan 4 wilayah zonasi dengan durasi gangguan yang lama, dan jumlah gangguan terbanyak walaupun lokasi basecamp dengan zonasi gangguan berada di satu wilayah. Hal ini mengindikasikan bahwa 4 wilayah zonasi tersebut perlu penambahan tim dikarenakan 4 wilayah zonasi tersebut juga merupakan top four dalam hal jumlah gangguan

#### 4. Kesimpulan

Berdasarkan hasil analisis di atas, zonasi Sumedang, dan Sukabumi memiliki nilai yang lebih baik dibandingkan dengan 15 wilayah zonasi yang lainnya. Oleh karena itu 15 wilayah zonasi yang lain memerlukan perhatian lebih dalam hal penanganan gangguan agar dapat tetap memberikan kualitas pelayanan yang andal dan kepercayaan pelanggan tetap terjaga. Perbaikan kualitas dalam penanganan gangguan dapat dilakukan dengan cara penambahan jumlah tim gangguan dan menambah lokasi basecamp tim gangguan yang terlalu jauh dengan lokasi pelanggan agar dapat mengurangi durasi perjalanan dan recovery gangguan.

Perlu dilakukan perbaikan terhadap infrastruktur jaringan terhadap zonasi yang gangguannya banyak disebabkan karena karena kabel. Perbaikan dapat dilakukan dengan cara melakukan pemeliharaan infrastruktur jaringan maupun melakukan pemeliharaan rutin terhadap infrastruktur jaringan.

#### Daftar Rujukan

- [1] R. W. S. Brahmana, F. A. Mohammed, and K. Chairuang, "Customer Segmentation Based on RFM Model Using K-Means, K-Medoids, and DBSCAN Methods," *Lontar Komput. J. Ilm. Teknol. Inf.*, vol. 11, no. 1, pp. 32–43, Apr. 2020, doi: 10.24843/LKJITI.2020.V11.I01.P04.
- [2] B. E. Adiana, I. Soesanti, and A. E. Permasari, "ANALISIS SEGMENTASI PELANGGAN MENGGUNAKAN KOMBINASI RFM MODEL DAN TEKNIK CLUSTERING," *J. Terap. Teknol. Inf.*, vol. 2, no. 1, pp. 23–32, Apr. 2018, doi: 10.21460/JUTEL.2018.21.76.
- [3] M. Fadillah, "IMPLEMENTASI CLUSTERING TERHADAP SEGMENTASI PELANGGAN DENGAN MENGGUNAKAN ANALISIS RFM," 2019, Accessed: Dec. 08, 2021. [Online]. Available: <http://repository.widyatama.ac.id/xmlui/handle/123456789/12941>.
- [4] J. Jamal and D. Yanto, "Analisis RFM dan Algoritma K-Means untuk Clustering Loyalitas Customer," *Energy*, vol. 9, no. 1, pp. 1–8, 2019.
- [5] A. T. Widiyanto and A. Witanti, "Segmentasi Pelanggan Berdasarkan Analisis RFM Menggunakan Algoritma K-Means Sebagai Dasar Strategi Pemasaran (Studi Kasus PT Coversuper Indonesia Global)," *KONSTELASI Konvergensi Teknol. dan Sist. Inf.*, vol. 1, no. 1, pp. 204–215, Apr. 2021, doi: 10.24002/KONSTELASI.V1I1.4293.
- [6] "Data Cleansing: Apa Itu, Manfaat, dan Cara Melakukannya - Glints Blog." <https://glints.com/id/lowongan/data-cleansing-cleaning/#.YbMXGVkxXIU> (accessed Dec. 10, 2021).
- [7] R. Sistem, "Perbandingan Metode Clustering dalam Pengelompokan Data Puskesmas," vol. 1, no. 10, pp. 5–12, 2021.
- [8] J. Sangeetha and V. Sinthu Janita Prakash, "An Efficient Inclusive Similarity Based Clustering (ISC) Algorithm for Big Data," *Proc. - 2nd World Congr. Comput. Commun. Technol. WCCCT 2017*, pp. 84–88, Oct. 2017, doi: 10.1109/WCCCT.2016.29.
- [9] A. Chusyairi and P. Ramadar Noor Saputra, "Pengelompokan Data Puskesmas Banyuwangi Dalam Pemberian Imunisasi Menggunakan Metode K-Means Clustering," *Telematika*, vol. 12, no. 2, pp. 139–148, 2019, doi: 10.35671/telematika.v12i2.848.
- [10] F. T. Informasi, "Segmentasi Pelanggan Menggunakan Analisis RFM Dan Algoritma Fuzzy C-Means Untuk Membantu Pengelolaan Hubungan Pelanggan Customer Segmentation Using Rfm Analysis And Fuzzy-C-Means Algorithm To Help Customer Relationship Management At PT . XYZ," 2015.
- [11] A. Fauzi Sistem Informasi, F. H. Universitas Buana Perjuangan Karawang Jl Ronggowaluyo, T. Timur, and K. priati, "Data Mining dengan Teknik Clustering Menggunakan Algoritma K-Means pada Data Transaksi Superstore," 2017, Accessed: Apr. 16, 2022. [Online]. Available: <http://community.tableau.com>.
- [12] D. sarjon Juliansa hengki and Sumijan, "Uji Validasi Algoritma Self Organizing Map (SOM) dan K-Mens untuk Pengelompokan Pegawai," *Resti*, vol. 1, no. 1, pp. 19–25, 2017.
- [13] H. Zarzour, Z. Al-Sharif, M. Al-Ayyoub, and Y. Jararweh, "A new collaborative filtering recommendation algorithm based on dimensionality reduction and clustering techniques," *2018 9th Int. Conf. Inf. Commun. Syst. ICICS 2018*, vol. 2018-January, pp. 102–106, May 2018, doi: 10.1109/IACS.2018.8355449.
- [14] D. Andra and A. B. Baizal, "E-commerce Recommender System Using PCA and K-Means Clustering," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 6, no. 1, pp. 57–63, 2022, doi: 10.29207/resti.v6i1.3782.
- [15] B. Basri, W. Gata, and R. Risnandar, "Analisis Loyalitas Pelanggan Berbasis Model Recency, Frequency, dan Monetary (RFM) dan Decision Tree pada PT. Solo," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 5, p. 943, 2020, doi: 10.25126/jtiik.2020752284.
- [16] W. Lestari, S. Bina, and B. Kendari, "Clustering Data Mahasiswa Menggunakan Algoritma K-Means Untuk Menunjang Strategi Promosi (Studi Kasus: STMIK Bina Bangsa Kendari)," *J. Sist. Inf. dan Sist. Komput.*, vol. 4, no. 2, pp. 35–48, Jul. 2019, doi: 10.51717/SIMKOM.V4I2.37.
- [17] S. Nawrin, M. Rahatur, and S. Akhter, "Exploreing K-Means with Internal Validity Indexes for Data Clustering in Traffic Management System," *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 3, 2017, doi: 10.14569/IJACSA.2017.080337.

- [18] R. K. Dinata, H. Novriando, N. Hasdyna, and S. Retno, "Reduksi Atribut Menggunakan Information Gain untuk Optimasi Cluster Algoritma K-Means," *J. Edukasi dan Penelit. Inform.*, vol. 6, no. 1, p. 48, 2020, doi: 10.26418/jp.v6i1.37606.
- [19] "Peningkatan Hasil Evaluasi Clustering Davies-Bouldin Index dengan Penentuan Titik Pusat Cluster Awal Algoritma K-Means." <https://repositori.usu.ac.id/handle/123456789/3827> (accessed Jun. 07, 2022).