

SEGMENTASI PELANGGAN TELKOMSEL MENGGUNAKAN METODE CLUSTERING DENGAN RFM MODEL DAN ALGORITMA K-MEANS

TELKOMSEL CUSTOMER SEGMENTATION USING CLUSTERING METHOD WITH RFM MODEL AND K-MEANS ALGORITHM

Mufti Alie Satriawan¹, Rachmadita Andreswari², Oktariani Nurul Pratiwi³

^{1,2,3} S1 Sistem Informasi, Fakultas Rekayasa Industri, Universitas Telkom

¹muftialies@student.telkomuniversity.ac.id, ²andreswari@telkomuniveristy.co.id,

³onurulp@telkomuniversity.ac.id

Abstrak

Pengguna operator seluler di Indonesia setiap tahunnya meningkat. Pada tahun 2018, jumlah pengguna operator seluler di Indonesia telah mencapai 254 juta pelanggan. Dari semua perusahaan operator seluler di Indonesia, PT Telekomunikasi Selular (Telkomsel) yang memiliki jumlah pelanggan terbanyak yaitu 163 juta. Akan tetapi di tahun tersebut, jumlah pelanggan telkomsel menurun sebanyak 17%. Hal tersebut dapat diselesaikan dengan memahami karakteristik dari pelanggannya, sehingga dapat digunakan sebagai langkah untuk strategi pemasaran yang tepat. Salah satu caranya adalah segmentasi pelanggan. Segmentasi pelanggan dapat dilakukan melalui pendekatan data mining dengan teknik clustering. Tujuannya adalah segmentasi pelanggan dan mengetahui karakteristik pelanggan dari setiap segmennya. Algoritma K-Means digunakan untuk pembentukan kluster dan pembentukan kluster didasarkan pada Model RFM (Recency, Frequency, dan Monetary). Metode Elbow dan Davies Bouldin Index (DBI) digunakan untuk menemukan jumlah kluster (k) yang optimal. Dan silhouette digunakan untuk pengujian kualitas kluster. Hasilnya terdapat 4 segmen, segmen 1 yang berjumlah 412773 yang memiliki profile pelanggan baru, segmen 2 yang berjumlah 357175 yang memiliki profile pelanggan potensial, segmen 3 yang berjumlah 125909 yang memiliki profile pelanggan loyal, dan segmen 4 yang berjumlah 16494 yang memiliki profile pelanggan perlu perhatian

Kata Kunci: Segmentasi pelanggan, data mining, clustering, model RFM, algoritma K-Means.

Abstract

Every year, cellular operator users in Indonesia are increasing. In 2018, the number of mobile operator users in Indonesia has reached 254 million subscribers. Of all cellular operator companies in Indonesia, PT Telekomunikasi Selular (Telkomsel) has the highest number of subscribers, namely 163 million. However, in that year, the number of Telkomsel subscribers decreased by 17%. This can be solved by understanding the characteristics of its customers, so that it can be used as a step for the right marketing strategy. One way is customer segmentation. Customer segmentation can be done through a data mining approach with clustering techniques. The goal is customer segmentation and to find out the characteristics of customers from each segment. The K-Means algorithm is used for cluster formation and cluster formation based on the RFM (Recency, Frequency, and Monetary) Model. The Elbow and Davies Bouldin Index (DBI) method was used to find the optimal number of clusters (k). And silhouettes are used for cluster quality testing. The result is that there are 4 segments, segment 1, totaling 412773 with new customer profiles, segment 2 with 357175 potential customer profiles, segment 3 totaling 125909 with loyal customer profiles, and segment 4 with 16494 customer profiles that need attention.

Keywords : *Customer segmentation, data mining, clustering, RFM models, K-Means algorithms.*

1. Pendahuluan

Pengguna operator seluler di Indonesia setiap tahunnya meningkat. Berdasarkan data dari kominfo, jumlah nomor prabayar selular yang telah terdaftar di Indonesia pada tahun 2018 sebesar 254 juta pelanggan [1]. PT Telekomunikasi Selular (Telkomsel) merupakan salah satu operator selular yang memiliki jumlah pelanggan terbanyak yaitu 163 juta [2]. Akan tetapi di tahun 2018, jumlah pelanggan Telkomsel menurun sebanyak 17% [3]. Hal tersebut perlu ditindaklanjuti dengan cara memahami karakteristik dari pelanggannya. Hal tersebut sebagai langkah untuk strategi pemasaran yang tepat. Salah satu caranya adalah segmentasi pelanggan. Segmentasi pelanggan merupakan mengelompokkan pelanggan yang memiliki kesamaan karakteristik [4]. Berdasarkan segmentasi yang diperoleh, telkomsel dapat mengetahui perilaku (*behavior*) konsumen sehingga akan membantu dalam penerapan strategi pemasaran (*marketing*) yang tepat untuk meningkatkan pendapatan perusahaan.

RFM model merupakan teknik segmentasi pelanggan berbasis data yang memungkinkan perusahaan mengambil keputusan secara taktis [5]. RFM menyimpan informasi waktu pembelian paling akhir (*recency*), frekuensi pembelian (*frequency*) dan sejumlah uang yang dihabiskan oleh pelanggan (*monetary*) [5]. Segmentasi tersebut bisa mengidentifikasi dan membagi pelanggan ke dalam kelompok yang homogen dan menargetkan mereka dengan strategi pemasaran berbeda dan disesuaikan, sehingga dapat meningkatkan kepuasan pelanggan dan retensi pelanggan. *Data mining* adalah teknik untuk menemukan informasi yang menarik, tak terduga atau berharga dalam kumpulan data yang berukuran besar [6]. Dalam mengimplementasikan *data mining*, terdapat banyak teknik yang dapat dipilih

salah satunya teknik *unsupervised learning*. Salah satu jenis dari teknik *unsupervised learning* yang paling sesuai dengan segmentasi yaitu *clustering*. *Clustering* merupakan teknik *unsupervised learning* yang membagi data dalam suatu himpunan ke dalam beberapa kelompok yang mana kesamaan data dalam suatu kelompok lebih besar dibandingkan kesamaan data tersebut dengan data dalam kelompok lain [7].

Algoritma *K-Means* merupakan salah satu metode *clustering non hirarki*. Algoritma ini mempartisi data ke dalam bentuk satu atau lebih *cluster*/kelompok, data yang memiliki karakteristik sama dikelompokkan ke dalam satu *cluster*/kelompok yang sama dan data yang memiliki karakteristik yang berbeda dikelompokkan ke dalam *cluster*/kelompok yang lain [8].

Berdasarkan penelitian terdahulu yang dilakukan oleh Adiana dkk [4], yang berjudul “Analisis Segmentasi Pelanggan Menggunakan Kombinasi RFM Model dan Teknik *Clustering*” menyimpulkan bahwa Model RFM dapat digunakan sebagai variabel/atribut data yang digunakan untuk proses *clustering* dapat menemukan jumlah *cluster* yang paling optimal. Selain itu, dalam penelitian ini algoritma *k-means* digunakan karena komputasi yang ringan, kemudian sesuai dengan kebutuhan data yang akan di *cluster*, dan sesuai dengan tujuan untuk menentukan jumlah cluster di awal untuk memperoleh jumlah kelompok pelanggan yang optimal.

Hasil dari *clustering* dengan RFM model ini, menghasilkan hasil pengkasteran dengan berbagai karakteristik pelanggannya. Hal ini dapat dijadikan referensi dalam menentukan pemasaran sesuai dengan karakteristik pelanggannya.

2. Dasar Teori /Material dan Metodologi/perancangan

2.1. Segmentasi Pelanggan

Segmentasi adalah proses membagi pelanggan menjadi beberapa klaster dengan kategori loyalitas pelanggan untuk membangun strategi pemasaran [9]. Dalam konsep pemasaran, segmentasi merupakan hal yang penting dalam *relationship marketing*. Meningkatkan hubungan dengan pelanggan menjadi lebih menarik dan akan menghasilkan pemahaman yang lebih baik tentang kebutuhan pelanggan [4].

Segmentasi pelanggan adalah metode untuk mengetahui berbagai jenis pelanggan untuk membuat keputusan lebih menguntungkan dengan mengenal pelanggan. Hasil dari segmentasi pelanggan ini dapat menjadi pedoman untuk mengembangkan strategi pemasaran dan cross selling produk baru untuk setiap grup, dan pengembangan produk untuk kelompok pelanggan yang paling berharga [10].

2.2 RFM Model

RFM merupakan model yang banyak digunakan dalam mengidentifikasi perilaku pelanggan dan mewakili karakteristik perilaku pelanggan [11]. RFM merupakan singkatan dari: *recency, frequency, monetary*. Menurut Khobzi et al. [5], RFM model memiliki tiga parameter yang harus diperhatikan yaitu:

- a. *Recency*: Interval antara pembelian dan waktu analisis.
- b. *Frequency*: Jumlah pembelian selama periode tertentu.
- c. *Monetary*: Jumlah uang dibelanjakan selama periode tertentu.

2.3 Data Mining

Data mining adalah penemuan struktur yang menarik, tak terduga atau berharga dalam kumpulan data yang berukuran besar. Sumber data mencakup database, data warehouse, web, dan repository lainnya [6]. Secara umum *data mining* memiliki 2 jenis task yaitu prediktif dan deskriptif. Prediktif itu *data mining* yang dilakukan untuk membentuk sebuah model pengetahuan yang digunakan untuk melakukan prediksi. Sedangkan deskriptif itu *data mining* yang dilakukan untuk mencari suatu pola untuk menjelaskan suatu karakteristik data dalam kumpulan data target [6].

2.4 Clustering

Clustering/klasterisasi merupakan teknik *unsupervised learning* yang membagi data dalam suatu himpunan ke dalam beberapa kelompok yang mana kesamaan data dalam suatu kelompok lebih besar dibandingkan kesamaan data tersebut dengan data dalam kelompok lain [7].

2.5 Algoritma K-Means

Algoritma *K-Means* merupakan salah satu metode *clustering non hirarki*. Algoritma ini mempartisi data ke dalam bentuk satu atau lebih *cluster*/kelompok, data yang memiliki karakteristik sama dikelompokkan ke dalam satu *cluster*/kelompok yang sama dan data yang memiliki karakteristik yang berbeda dikelompokkan ke dalam *cluster*/kelompok yang lain [8].

Menurut Larose [12], algoritma *K-Means* menggunakan suatu persamaan untuk mengukur jarak yang paling minimum pada objek yaitu *Distance Space* atau *Euclidean Distance Space*. *Distance space* adalah suatu komponen dari *K-Means Clustering* yang digunakan untuk menghitung jarak antar objek. Dalam hal ini fungsi dari *Euclidean Distance Space* memiliki suatu persamaan yaitu:

$$D_e = \sqrt{\sum_{i=0}^n (x_i - c_i)^2}$$

Dimana: D_e merupakan Euclidean distance, x_i merupakan titik koordinat objek, dan c_i merupakan koordinat centroid.

2.6 Metode Elbow

Menurut merliana dan santoso [13], metode elbow merupakan suatu metode yang digunakan untuk menghasilkan informasi dalam menentukan jumlah cluster terbaik dengan cara melihat persentase hasil perbandingan antara jumlah *cluster* yang akan membentuk siku pada suatu titik. Metode Elbow memberikan ide atau gagasan dengan cara memilih nilai kluster dan kemudian menambah nilai cluster tersebut untuk dijadikan model data dalam menentukan *cluster* terbaik. Dan selain itu persentase perhitungan yang dihasilkan menjadi pembanding antara jumlah kluster yang ditambah. Hasil persentase yang berbeda dari setiap nilai cluster dapat ditunjukkan dengan menggunakan grafik sebagai sumber informasinya. Jika nilai cluster pertama dengan nilai cluster kedua memberikan sudut dalam grafik atau nilainya mengalami penurunan paling besar maka nilai cluster tersebut yang terbaik.

2.7 Davies Bouldin Index

Davies Bouldin Index (BDI) merupakan salah satu metode pendekatan untuk menentukan nilai kluster yang optimal. Pendekatan ini bertujuan untuk memaksimalkan jarak antara kluster yang satu dengan yang lainnya dan pada waktu yang sama mencoba untuk meminimalkan jarak antara objek dalam sebuah kluster [14]. Jumlah kluster yang terbaik memiliki nilai DBI minimum.

2.8 Silhouette

Silhouette merupakan salah satu metode yang digunakan untuk menguji kualitas kluster yang dihasilkan dari proses *clustering*. Metode ini menghitung rata-rata nilai setiap titik pada himpunan data. Perhitungan nilai setiap titik adalah selisih nilai separation dan compactness yang dibagi dengan maksimum antara keduanya [15]. Berikut merupakan persamaan yang digunakan pada metode silhouette [16]:

$$S(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$

Keterangan:

$S(i)$ = nilai silhouette.

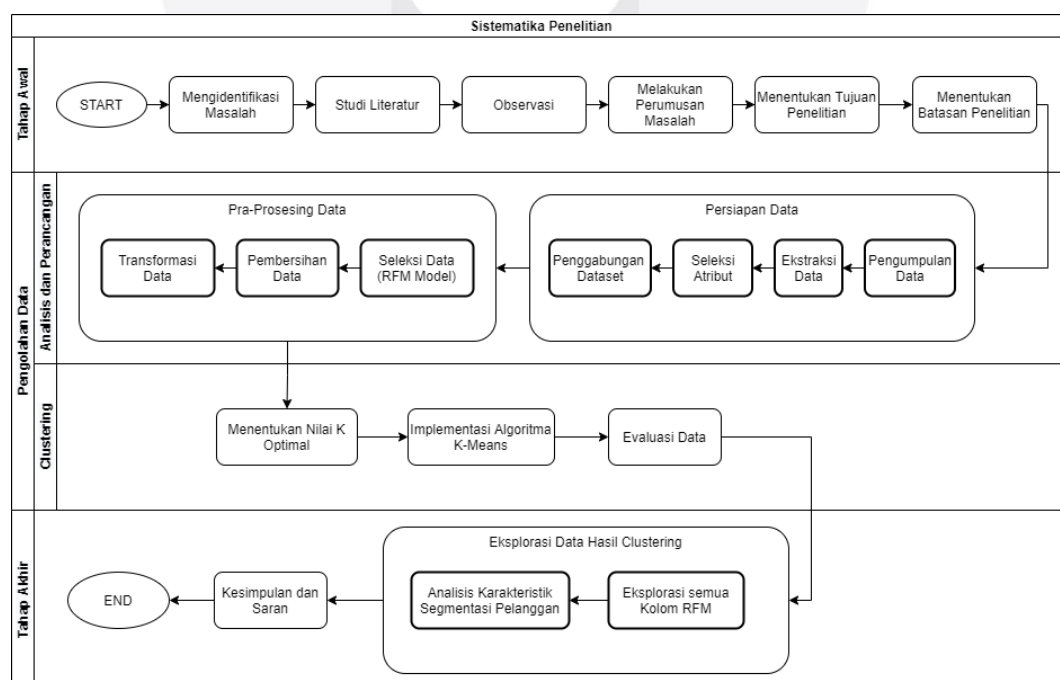
$b(i)$ = rata-rata jarak dari objek i dengan seluruh objek dalam cluster yang sama.

$a(i)$ = rata-rata jarak dari objek i dengan objek berada di cluster yang berbeda.

Dari hasil persamaan tersebut, nilai dari *silhouette* berkisar $-1 \leq S(i) \leq 1$, dimana hasil clustering baik jika nilai silhouette bernilai positif (0-1) [16].

3. Metodologi

Sistematika yang dilakukan pada penelitian ini adalah sesuai dengan rancangan yang telah terbagi menjadi tiga fase, yaitu tahap awal, pengolahan data dan berakhir dengan tahap akhir. Untuk gambaran sistematikanya dapat dilihat pada Gambar 1.



Gambar 1 – Sistematika Penelitian

A. Persiapan Data

Pada tahapan awal ini berfokus pada mempersiapkan data yang masih belum tersambung agar menjadi *dataset*. Data-data ini disambung sehingga membuat suatu *dataset* yang berisi informasi pelanggan dan pola konsumsinya. Data yang akan digunakan merupakan data *basic* yang merupakan data log customer selama 31 hari yang memiliki range dari tanggal 19 juni 2020 sampai 19 juli 2020. Tampilan data basic dapat dilihat pada Gambar 2.

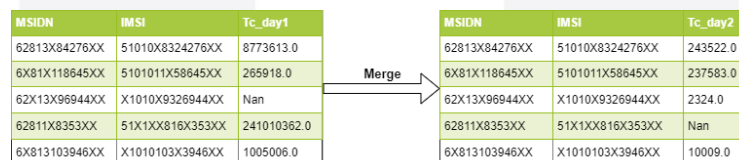
Unnamed: 0	msisdn	imsi	skey	scpaddress	tvendor	ttype	imei	sposname	spotap	lte	card_type	subscriber_type	product_id	frequency	osvendor
0	6281318059XXX	510101832059XXX	110	8.281108e+09	VIVO	S1	8X872504X472XX	ANDROID	YES	YES	USIM	Online Or Prepaid Subscriber	Simpai	GSM 1800 GSM 1900 GSM 850 GSM 900 LTE 1.LTE 3...	GOOGLE
1	X2812X10X80XX	51010X1250X80XX	110	8.281108e+09	XIAOMI	Mi8X	8890870X8590XX	ANDROID	NO	YES	USIM	Online Or Prepaid Subscriber	Simpai	GSM 1800 GSM 1900 GSM 850 GSM 900 LTE 1.LTE 3...	GOOGLE
2	628119229XXX	X1010921321X0XX	310	8.281108e+09	NaN	NaN	NaN	NaN	NaN	NaN	USIM	Hybrid Subscriber	HALO Cek	NaN	NaN
3	628X8XX678XX	5X0X08XX2X878XX	110	8.281108e+09	XIAOMI	REDMIS2	86X85X455XXX	ANDROID	NO	YES	USIM	Online Or Prepaid Subscriber	Simpai	GSM 1800 GSM 1900 GSM 850 GSM 900 LTE 1.LTE 2...	GOOGLE
4	6XXX1X35737XX	51010X38X5737XX	110	8.281108e+09	XIAOMI	REDMI4A	X8903903984XX	ANDROID	YES	YES	USIM	Online Or Prepaid Subscriber	Simpai	GSM 1800 GSM 1900 GSM 850 GSM 900 LTE 1.LTE 3...	GOOGLE

Gambar 2 – Tampilan Data Basic

Data tersebut di seleksi dengan mengambil atribut digunakan yang dalam penelitian yang berisi data pelanggan dan log konsumsi pelanggan. Atribut tersebut antara lain MSIDN, IMSI, dan Total Consumption. Kolom MSIDN dan IMSI digunakan sebagai identifikasi pelanggan (ID pelanggan) dikarenakan isi data yang disensor. Sedangkan kolom *Total Consumption* digunakan untuk melihat pola konsumsi data yang dimiliki oleh customer. Setelah itu, seluruh data dari 31 hari tersebut akan digabungkan untuk menjadi dataset. Ilustrasi penggabungan dapat dilihat pada Gambar 3. Dan hasil penggabungan dapat dilihat pada Tabel 1.

Tabel 1 – Dataset Jadi

MSIDN	IMSI	Tc_day1	Tc_dayN	Tc_day31
62813X84276XX	51010X8324276XX	8773613.0	...	1559700.0
6X81X118645XX	5101011X58645XX	265918.0	...	56093955.0
62X13X96944XX	X1010X9326944XX	Nan	...	114852955.0
62811X8353XX	51X1XX816X353XX	241010362.0	...	Nan
6X813103946XX	X1010103X3946XX	1005006.0	...	35491.0



MSIDN	IMSI	Tc_day1	Tc_day2
62813X84276XX	51010X8324276XX	8773613.0	243522.0
6X81X118645XX	5101011X58645XX	265918.0	237583.0
62X13X96944XX	X1010X9326944XX	Nan	2324.0
62811X8353XX	51X1XX816X353XX	241010362.0	Nan
6X813103946XX	X1010103X3946XX	1005006.0	10009.0

Gambar 3 – Ilustrasi Penggabungan Tabel

B. Pra-Pemrosesan Data

Setelah data menjadi dataset, tahap ini akan berfokus merubah dataset menjadi data yang digunakan untuk melakukan pemodelan. Tahapan ini diawali dengan Data Selection. Data yang diseleksi akan merubah data menjadi atribut Model RFM. Model RFM berisi rentang hari terakhir pelanggan melakukan konsumsi data (*recency*), jumlah frekuensi penggunaan konsumsi data (*frequency*), dan jumlah total konsumsi data (*monetary*). Data tersebut dibuat dengan menggunakan kolom *tc_day1* hingga *tc_day31*. Untuk untuk mendapatkan nilai *Recency*, peneliti mengecek rentang hari terakhir pelanggan melakukan konsumsi data dari 31 hari terakhir. Kemudian untuk mendapatkan nilai *Frequency*, peneliti hanya menjumlahkan total keseluruhan kolom dari *tc-day1* hingga *tc-day31*. Dan untuk mendapatkan nilai kolom *Monetary*, peneliti menggunakan *dataframe* sebelumnya, lalu menjumlahkan total keseluruhan total pemakaian data *customer* selama 31 hari. Lalu ketiga data tersebut akan digabungkan dengan kolom MSIDN dan IMSI. Untuk hasil Model RFM dapat dilihat pada Tabel 2.

Tabel 2 – Penggabungan RFM Model

MSIDN	IMSI	Recency	Frequency	Monetary
62813X84276XX	51010X8324276XX	4	15	2162.10868359
6X81X118645XX	5101011X58645XX	1	10	1023.00244331
62X13X96944XX	X1010X9326944XX	3	10	3312.77339077
62811X8353XX	51X1XX816X353XX	2	11	491.87443066
6X813103946XX	X1010103X3946XX	4	9	1340.62113571

Selanjutnya, tahapan Data Cleansing. Pada dataset tersebut masih memiliki data yang bernilai duplikat pada kolom MSIDN dan IMSI sehingga menyebabkan redundansi data pelanggan. Jadi kolom MSIDN dan IMSI yang sama akan di buang dari dataset.

Dan yang Terakhir, tahapan Data Transformation. Pada tahap ini, peneliti akan melakukan transformasi data dengan atribut yaitu *recency*, *frequency*, dan *monetary*. Ketiga atribut tersebut akan ditransformasi menggunakan metode *zero-mean normalization*. Metode ini akan mentransformasikan atribut dimana nilai dari atribut A tertentu ditransformasikan berdasarkan rata-rata dan standar deviasi dari atribut A [17]. Rumus metode ini dapat dijabarkan pada persamaan dibawah.

$$v' = \frac{v - \bar{A}}{\sigma A}$$

Dimana v' adalah normalisasi dari nilai v , v adalah nilai dari atribut A, \bar{A} adalah nilai rata-rata dari atribut A, dan σA adalah standart deviasi dari atribut A

Data yang telah ditransformasi akan digabungkan dengan *dataset* sebelumnya, seperti pada Tabel 3.

Tabel 3 – Hasil Dataset Preprocessing

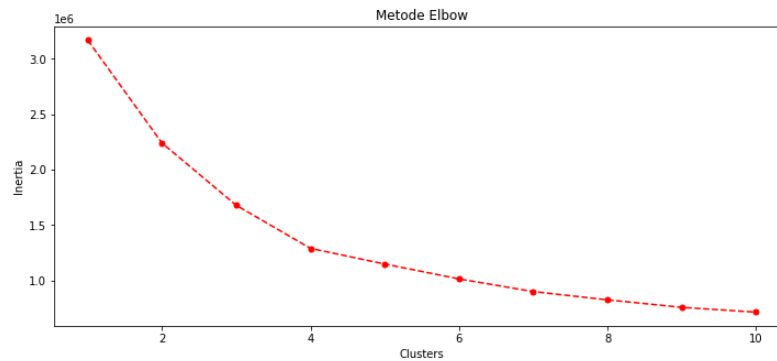
MSIDN	IMSI	Recenc	Frequenc	Monetary	Norm	Norm	Norm
6281318059XXX	510101832059XXX	1	16	3071.193764	-	1.9633	2.8519
X2812X10X80X	51010X1250X80XX	1	16	1212.704412	-	1.9633	0.4024
628119229XX	X1010921321X0XX	2	9	772.705883	-	-	-0.1775
628XX8XX678X	5X0X08XX2X678X	2	12	266.353587	-	0.4545	-0.8449
6XXX1X35737X	51010X36X5737XX	1	12	438.205839	-	0.4545	-0.6184
...
62813X84276XX	51010X8324276XX	4	15	2162.1086835	1.7802	-	-0.5426
6X81X118645XX	5101011X58645XX	1	10	1023.0024433	0.4865	1.5861	1.6537
62X13X96944XX	X1010X9326944XX	3	10	3312.7733907	-	-	0.1524
62811X8353XX	51X1XX816X353X	2	11	491.87443066	-	0.0772	-0.5476
6X813103946XX	X1010103X3946XX	4	9	1340.6211357	0.4865	-	0.5710

C. Penentuan Nilai K

Dalam melakukan *clustering* dengan algoritma k-mean, perlu menentukan terlebih dahulu nilai nilai K. Nilai K dapat diperoleh dengan beberapa metode yang berbeda. Metode yang akan digunakan peneliti terdapat dua metode yaitu metode Elbow dan metode DBI. Hasil kedua metode ini akan menghasilkan nilai K yang optimal. Setelah mendapatkan nilai K, akan dilanjutkan dengan proses implementasi.

1. Metode elbow

Metode elbow merupakan salah satu pengujian dalam *clustering*. Dalam metode elbow, dalam menentukan nilai k optimal diketahui dari penurunan *inertia* sehingga membentuk siku. Hal ini sama seperti yang dilakukan oleh peneliti sebelumnya[11]. Dalam implementasinya, peneliti menggunakan *dataframe* dengan atribut yang telah dinormalisasi sebelumnya yaitu NormR, NormF, dan NormM. Ketiga atribut tersebut akan di klaster menggunakan algoritma k-mean dengan nilai k antara 1 hingga 10. Hasil Klasterisasi dapat dilihat pada Gambar 4.



Gambar 4 – Visualisasi Metode Elbow

Berdasarkan visualisasi dari hasil metode elbow, maka dapat ditarik kesimpulan bahwa nilai k optimal dalam metode elbow adalah 4.

2. Davies Bouldin Index(DBI)

Setelah mendapatkan hasil k optimal dari metode elbow, peneliti akan melakukan pengujian dengan metode *Davies Bouldin Index*(DBI). Pada implementasinya, peneliti menggunakan ketiga atribut yang telah dilakukan pengujian dengan metode elbow dan juga total pengujiannya dilakukan sebanyak 9 kali dengan range 2 hingga 10. Alasan pengujian tersebut di awal dari 2, karena library DBI yang peneliti gunakan tidak menerima pengujian satu kluster, yang mana satu kluster sama dengan tidak ada proses klusterisasi. Hasil perhitungan DBI dapat dilihat pada Tabel 4.

Tabel 4 – Hasil Pengujian DBI

Jumlah Klaster	Nilai DBI
2	1,3473
3	1,1455
4	1,0349
5	1,0958
6	1,0861
7	1,0362
8	1,0411
9	1,0459
10	1,0510

D. Implementasi Algoritma K-Means

Pada tahapan ini, peneliti telah mendapatkan nilai k yang optimal yaitu 4. Selanjutnya akan diimplementasikan algoritma k-mean dengan jumlah pengelompokan sebanyak 4 kluster. Data yang akan diimplementasikan masih menggunakan ketiga atribut yang telah ditransformasikan sebelumnya pada *dataframe*. Dari hasil *clustering*, dapat menemukan nilai centroid dan jumlah baris yang dapat dilihat pada Tabel 5.

Tabel 5 – Nilai Centroid dan jumlah baris hasil Clustering

Centroid	Recency	Frequency	Monetary	Jumlah Baris
0	2.1082	9.0811	566.5946	412773
1	2.0123	13.2248	868.0452	357175
2	2.3634	12.0339	2478.7461	125909
3	7.1473	8.8251	638.6977	160494

E. Evaluasi dan Eksplorasi Data

Pada tahapan ini, data akan dievaluasi dengan metode *Silhouette*, hasil dari perhitungan ini menghasilkan nilai 0.30666957248255655, menunjukkan bahwa hasil kluster sudah optimal. Dibuktikan dengan nilai *silhouette* di atas 0.

Dari pada prose *clustering* sebelumnya, terdapat 4 hasil kluster. Ke empat hasil kluster tersebut merupakan 4 segmen pelanggan yang akan di eksplorasi hasilnya. Hasil eksplorasi segment sebagai berikut:

1. Pada segmen 1, segmen ini dapat diartikan pelanggan baru berdasarkan dari nilai RFM, pasalnya segmen ini memiliki nilai *frequency* dan *monetary* yang rendah tetapi memiliki nilai *recency* yang tinggi. Nilai *recency* yang tinggi dapat diartikan bahwa pelanggan yang baru menggunakan layanan tetapi tidak banyak

menggunakan penggunaan data dalam beberapa hari terakhir. Pelanggan yang baru ini bisa berarti baru menggunakan layanan telkomsel atau pelanggan lama yang baru mengaktifkan layanan telkomsel. Pelanggan pada segmen ini dapat dilakukan layanan khusus seperti promo layanan khusus untuk pelanggan baru agar dapat menambah minat pelanggan untuk menggunakan layanan telkomsel.

2. Pada segmen 2, segmen ini dapat diartikan pelanggan potensial berdasarkan dari nilai RFM, pasalnya segmen ini memiliki nilai *recency* dan *frequency* yang tinggi tetapi memiliki *monetary* yang rendah sehingga pelanggan ini dapat berpotensi menjadi pelanggan yang loyal dengan meningkat nilai *monetary* nya. Salah satu caranya dengan memberikan layanan yang sedikit mahal, tetapi memiliki value yang lebih banyak. Sehingga dengan langkah tersebut, diharapkan pelanggan pada segmen ini dapat meningkatkan *monetary* menjadikan pelanggan yang loyal.
3. Pada segmen 3, segmen ini dapat diartikan pelanggan yang loyal berdasarkan dari nilai RFM, dalam segmen ini sudah dapat terlihat dari ketiga nilai RFM yang tinggi. Pelanggan pada segmen ini sangat perlu dipertahankan oleh telkomsel, pasalnya pelanggan ini dapat menjadi pemasukan tetap yang perlu dipertahankan. Salah satu caranya dengan mempertahankan kualitas layanan pada pelanggan tersebut dan memprioritaskan pelanggan tersebut untuk mendapatkan layanan yang terbaru.
4. Pada segmen 4, segmen ini dapat diartikan pelanggan yang perlu perhatian, perhatian yang dimaksud adalah meningkatkan penggunaan layanan pada telkomsel. Salah satunya dengan memberikan layanan yang memiliki value lebih besar dengan harga yang sama.

Hasil dari rangking segmen dapat dilihat pada Tabel 6.

Tabel 6 – Hasil Profile dan Rangking

Segmen	Jumlah Segmen	Label RFM	Rangking	Profile Segmen
1	412773	R↑F↓M↓	3	Pelanggan Baru
2	357175	R↑F↑M↓	2	Pelanggan Potensial
3	125909	R↑F↑M↑	1	Pelanggan Loyal
4	160494	R↓F↓M↓	4	Pelanggan perlu perhatian

Setelah mengetahui hasil analisis setiap segmen serta *profile* dari segmen tersebut, diharapkan informasi ini dapat membantu pihak telkomsel dalam strategi pemasaran yang sesuai dengan *profile* dan ranking pelanggannya sehingga dapat meningkatkan profit perusahaan.

4. Kesimpulan

Dari hasil penelitian yang dilakukan oleh para peneliti, hasil mendapatkan nilai optimal kluster yaitu 4, dengan nilai pengujian metode *silhouette* mendapatkan score 0.30666957248255655. Hasil *clustering* dari keempat data tersebut memiliki karakteristik pelanggan yang berbeda-beda, antara lain:

1. Pelanggan baru dengan label R↑F↓M↓, yang memiliki jumlah data 412773. Dengan nilai centroid *recency* 2.1082, *frequency* 9.0811, dan *monetary* 566.5946.
2. Pelanggan potensial dengan label R↑F↑M↓ yang memiliki jumlah data 357175. Dengan nilai centroid *recency* 2.0123, *frequency* 13.2248, dan *monetary* 868.0452.
3. Pelanggan loyal dengan label R↑F↑M↑, yang memiliki jumlah data 125909. Dengan nilai centroid *recency* 2.3634, *frequency* 12.0339, dan *monetary* 2478.7461.
4. Pelanggan yang butuh perhatian dengan label R↓F↓M↓, yang memiliki jumlah data 160494. Dengan nilai centroid *recency* 7.1473, *frequency* 8.8251, dan *monetary* 638.6977.

Referensi

- [1] Kominfo, "Inilah rincian jumlah pelanggan prabayar masing-masing operator," *Kominfo Republik Indonesia*, 2018. https://kominfo.go.id/content/detail/13131/inilah-rincian-jumlah-pelanggan-prabayar-masing-masing-operator/0/sorotan_media.
- [2] Katadata, "Telkomsel, Raja Operator Seluler Indonesia," *katadata*, 2019. <https://databoks.katadata.co.id/datapublish/2020/01/23/telkomsel-raja-operator-seluler-indonesia>.
- [3] Telkomsel, "Telkomsel Annual Report 2018: Your Gateway To The Digital World," 2019. [Online]. Available: <https://www.telkomsel.com/about-us/investor-relations>.
- [4] 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, 2018, doi: 10.21460/jutei.2018.21.76.
- [5] H. Khobzi, E. Akhondzadeh-Noughabi, and B. Minaei-Bidgoli, "A New Application of RFM Clustering for Guild Segmentation to Mine the Pattern of Using Banks' e-Payment Services," *J. Glob. Mark.*, vol. 27, no. 3, pp. 178–190, 2014, doi: 10.1080/08911762.2013.878428.
- [6] J. Han, J. Pei, and M. Kamber, *Data Mining: Concepts and Techniques*. USA: Elsevier Inc., 2012.
- [7] D. Zheng, "Application of silence customer segmentation in securities industry based on fuzzy cluster algorithm," *J. Inf. Comput. Sci.*, vol. 10, no. 13, pp. 4337–4347, 2013, doi: 10.12733/jics20102432.
- [8] N. A. Khairani *et al.*, "PENEREAPAN METODE CLUSTERING ALGORITMA K-MEANS SEBAGAI

- PENENTUAN DAERAH RAWAN TITIK API DI PROVINSI KALIMANTAN BARAT APPLICATION OF K-MEANS ALGORITHM CLUSTERING METHOD AS DETERMINATION OF FIRE POINT PRONE AREAS IN WEST KALIMANTAN PROVINCE,” 2019.
- [9] T. Hardiani, S. Sulisty, and R. Hartanto, “Segmentasi Nasabah Tabungan Menggunakan Model RFM (Recency , Frequency , Monetary) dan K-Means Pada Lembaga Keuangan Mikro,” *Semin. Nas. Teknol. Inf. dan Komun. Terap.*, no. May, pp. 463–468, 2015.
- [10] M. Khajvand and M. J. Tarokh, “Estimating customer future value of different customer segments based on adapted RFM model in retail banking context,” *Procedia Comput. Sci.*, vol. 3, pp. 1327–1332, 2011, doi: 10.1016/j.procs.2011.01.011.
- [11] R. H. Zakariyya, “Customer Segmentation by using RFM Model and K-Mean Clustering in PT XYZ,” *Telkom Univ.*, pp. 1–10, 2020.
- [12] D. T. Larose, *Data Mining and Predictive Analytics*. USA: John Wiley & Sons, 2015.
- [13] P. N. E. Merliana and A. J. Santoso, “Analisa Penentuan Jumlah Cluster Terbaik pada Metode K-Means,” pp. 978–979, 2015.
- [14] A. Badruttamam, S. Sudarno, and D. A. I. Maruddani, “PENERAPAN ANALISIS KLASER K-MODES DENGAN VALIDASI DAVIES BOULDIN INDEX DALAM MENENTUKAN KARAKTERISTIK KANAL YOUTUBE DI INDONESIA (Studi Kasus: 250 Kanal YouTube Indonesia Teratas Menurut Socialblade),” *J. Gaussian*, vol. 9, no. 3, pp. 263–272, 2020, doi: 10.14710/j.gauss.v9i3.28907.
- [15] A. Aditya, I. Jovian, and B. N. Sari, “Implementasi K-Means Clustering Ujian Nasional Sekolah Menengah Pertama di Indonesia Tahun 2018/2019,” *J. Media Inform. Budidarma*, vol. 4, no. 1, p. 51, 2020, doi: 10.30865/mib.v4i1.1784.
- [16] P. J. Rousseeuw, “Silhouettes: A graphical aid to the interpretation and validation of cluster analysis,” *J. Comput. Appl. Math.*, vol. 20, no. C, pp. 53–65, 1987, doi: 10.1016/0377-0427(87)90125-7.
- [17] H. Junaedi, H. Budianto, I. Maryati, and Y. Melani, “Data Transformation pada Data Mining,” *Pros. Konf. Nas. Inov. dalam Desain dan Teknol.*, vol. 7, pp. 93–99, 2011.

