

# Segmentasi dan Pembentukan Model Regresi Nasabah Berbasis Analisis *Recency*, *Frequency* dan *Monetary*

<http://dx.doi.org/10.28932/jutisi.v8i2.5075>

Riwayat Artikel

Received: 04 Juli 2022 | Final Revision: 02 Agustus 2022 | Accepted: 02 Agustus 2022

Creative Commons License 4.0 (CC BY – NC)



Ronaldo Cristover Octavianus <sup>#1</sup>, Hapnes Toba <sup>#2</sup>, Bernard Renaldy Suteja <sup>#3</sup>

<sup>#</sup> Program studi Magister Ilmu Komputer, Universitas Kristen Maranatha  
Jl. Prof. Drg. Suria Sumantri No. 65, Bandung 40164, Indonesia

<sup>1</sup>ronaldochristover@gmail.com

<sup>2</sup>hapnes.toba@it.maranatha.edu

<sup>3</sup>bernard.rs@it.maranatha.edu

✉ Corresponding author: hapnes.toba@it.maranatha.edu

**Abstrak** — Di masa pandemi ini, peningkatan jumlah nasabah suatu perusahaan sekuritas meningkat cukup tinggi. Hal tersebut mengharuskan perusahaan sekuritas untuk melakukan analisis terkait data nasabah sekuritas terhadap data transaksi agar perusahaan tersebut dapat mengetahui segmentasi dari nasabah yang sudah terdaftar dan juga agar perusahaan dapat memprediksi terkait pola transaksi dari nasabah yang ada di perusahaan tersebut. Dalam melakukan pengolahan data transaksi dapat menggunakan model RFM (*Recency*, *Frequency*, *Monetary*) sebagai cara untuk pengelompokan nasabah sesuai dengan nilai bisnis yang dimiliki. Setelah melakukan pemodelan dengan menggunakan RFM, data yang tersebut dilakukan *Clustering* dengan menggunakan algoritma *K-Means* untuk mengetahui segmentasi yang ada dalam model RFM dalam setiap kelompoknya. Model RFM yang sudah dilakukan *clustering* akan menghasilkan suatu segmen-segmen berdasarkan kelompok RFM. Pada data tersebut dilakukan proses analisis regresi linier dimana pada setiap kelompok dan segmen dianalisis dan diprediksi terkait variabel-variabel seperti *recency*, *monetary* dan *frequency*. Hasil dari pengelompokan data, segmentasi pelanggan dan juga prediksi dengan regresi linier ini dapat menjadi salah satu acuan perusahaan untuk membuat suatu keputusan bisnis. Dari proses regresi linier yang dilakukan atas atribut-atribut RFM, dihasilkan suatu prediksi nilai moneter dari nilai *recency* yang sudah ada dan juga dapat diketahui nilai moneter dari *frequency* dengan tingkat error yang cukup baik.

**Kata kunci** — *Frequency*; *Monetary*; *Recency*; Regresi Linier; Segmentasi Pelanggan.

## Segmentation and Formation of Customer Regression Model Based on *Recency*, *Frequency* and *Monetary* Analysis

**Abstract** — During this pandemic, the number of customers of a securities company has increased quite high. This requires securities companies to conduct analysis related to secure customer data against transaction data so that the company can find out the segmentation of registered customers and also so that companies can predict the transaction patterns of customers in the company. In processing transaction data, the RFM (*Recency*, *Frequency*, *Monetary*) model can be used as a way to group customers according to

their business values. After doing the modeling using RFM, the data is clustered using the K-Means algorithm to find out the segmentation in the RFM model in each group. The RFM model that has been clustered will produce segments based on the RFM group. In this data, a linear regression analysis process is carried out where each group and segment is analyzed and predicted related to variables such as recency, monetary and frequency. The results of data grouping, customer segmentation and also predictions with linear regression can be one of the company's references to make a business decision. From the linear regression process carried out on the RFM attributes, a prediction of the monetary value of the existing recency value is generated and the monetary value of the frequency can also be known with a fairly good error rate.

Keywords — Customer Segmentation; Frequency; Linear Regression; Monetary; Recency.

## I. PENDAHULUAN

Pada masa pandemi terjadi peningkatan jumlah nasabah yang mendaftar di perusahaan sekuritas dan melakukan transaksi di pasar saham. Peningkatan yang terjadi juga sejalan dengan peningkatan jumlah transaksi serta volume pembelian dan penjualan saham yang dilakukan oleh nasabah. Dengan meningkatnya volume dan kuantitas transaksi, maka dirasa menjadi penting bagi perusahaan. Berdasarkan hal tersebut, diperlukan suatu analisis terkait segmentasi dari nasabah yang telah terdaftar dan analisis terkait transaksi yang sudah dilakukan oleh nasabah tersebut. Data transaksi nasabah perlu dilakukan suatu pemilihan atribut yang cocok untuk menghasilkan suatu output yang baik sehingga dalam model RFM (*Recency*, *Frequency* dan *Monetary*) dapat digunakan sebagai acuan untuk pemilihan atribut dalam proses penelitian. Penelitian dengan model RFM digunakan mendapatkan informasi awal terkait behaviour seorang nasabah. Proses pengelompokan nasabah yang dilakukan dengan model RFM, sering digunakan dalam proses marketing dan juga segmentasi nasabah untuk kebutuhan bisnis [1], [2], [3]. Menurut John R. Miglautsch, nasabah yang dikelompokkan dengan menggunakan RFM, dapat diberikan skor dengan membagi nasabah menjadi beberapa bagian berdasarkan nilai yang diurutkan dari nilai *recency*, *frequency* dan juga *monetary* [2]. Proses *clustering* dapat dilakukan dengan mengelompokkan data yang sudah dilakukan pemrosesan model RFM dimana proses *clustering* digunakan untuk melihat kemiripan antara objek satu dengan yang lain berdasarkan atribut RFM [1], [4], [5].

Pada penelitian ini, data transaksi yang digunakan adalah tahun 2021 dimana pada data digunakan merupakan data secara umum dan tidak memiliki informasi terkait nasabah, emiten dan lain-lainnya. Dengan menggunakan data transaksi yang ada, perusahaan memiliki tujuan untuk mendapatkan informasi terkait kelompok dan segmentasi dari nasabah yang sudah melakukan transaksi dan mendapatkan masukan terkait hal-hal yang bisa dijadikan sebagai bahan untuk melakukan follow-up kepada nasabah lama dan nasabah baru. Untuk itu diperlukan suatu perhitungan dengan menggunakan model RFM untuk mengelompokkan nasabah berdasarkan nilai dari transaksi yang sudah ada sebelumnya. Untuk melakukan analisis nasabah dilakukan proses *clustering* dengan algoritma *K-Means* dimana metode ini merupakan suatu proses untuk membagi objek-objek yang memiliki kemiripan menjadi suatu kelompok [1].

Pada penelitian ini, kluster nasabah yang telah dihitung akan didasarkan pada tiga (3) atribut yaitu *recency*, *frequency* dan juga *monetary* atau disebut juga RFM. Berdasarkan deskripsi diatas, penelitian ini menerapkan metode *K-Means Clustering* dengan menggunakan model RFM pada data transaksi nasabah. Dari hasil segmentasi yang ada, dilakukan suatu prediksi dengan menggunakan regresi linier apabila data yang dimiliki memiliki hubungan linier antara satu dengan yang lain. Dari hasil yang didapatkan dengan regresi, dilakukan prediksi untuk mencari hubungan antara atribut yang nantinya dapat digunakan sebagai bahan *follow-up* perusahaan terhadap nasabah dengan menggunakan hasil prediksi nilai *monetary*, *recency* ataupun *frequency*.

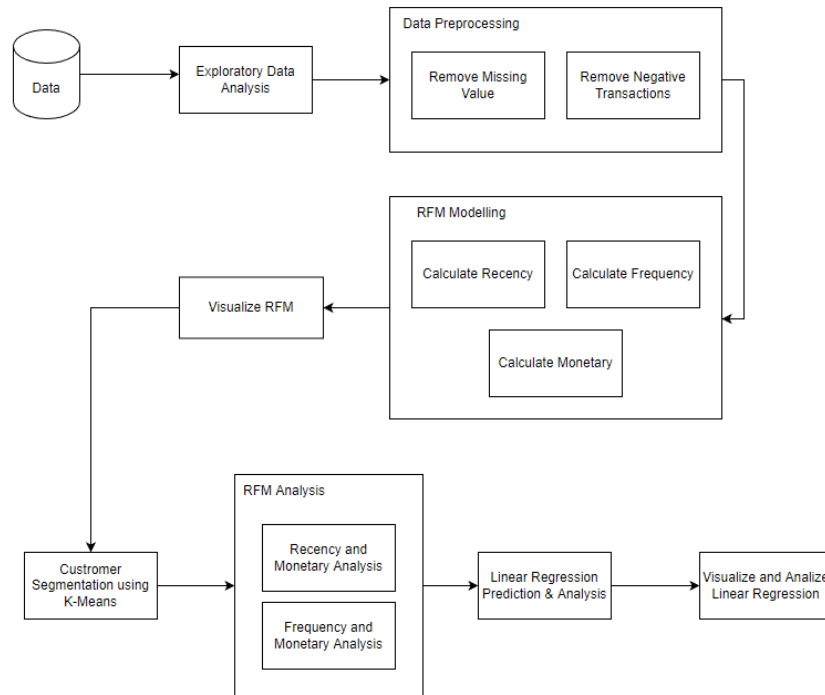
## II. METODE PENELITIAN

### A. Metodologi

Dalam melakukan penelitian, dibuat desain penelitian yang terdiri dari beberapa tahap. Tahap-tahap tersebut digambarkan pada Gambar 1. Pada Gambar 1, Data dilakukan proses *exploratory data analysis (EDA)*, proses ini dilakukan untuk melakukan investigasi terhadap data untuk menemukan pola, anomali dan juga melakukan pemeriksaan terhadap data. Data yang sudah dianalisis akan dilakukan *preprocessing* dimana data tersebut akan dilakukan pembersihan agar dapat digunakan untuk proses *modelling* dengan model RFM. Pada model RFM akan dilakukan perhitungan *recency*, *frequency* dan juga *monetary*. Setelah dilakukan perhitungan, nasabah akan dilakukan pengelompokan berdasarkan skor dan bobot yang nantinya akan dibagi berdasarkan perhitungan kuantil sehingga terbentuk suatu kelompok nasabah berdasarkan model RFM.

Hasil pengolahan data tersebut akan digambarkan dan dilanjutkan dengan proses segmentasi dengan menggunakan algoritma *K-Means*. Segmentasi yang dilakukan dengan *K-Means* akan membentuk suatu *cluster* terhadap kelompok-kelompok nasabah yang sebelumnya diolah dengan model RFM. Dari proses analisa data RFM yang sudah dilakukan segmentasi, dilakukan proses prediksi dengan menggunakan regresi linier. Dengan menggunakan regresi linier, data akan dilakukan perhitungan dan prediksi dengan menggunakan variabel *recency*, *frequency* dan *monetary*. Hasil dari perhitungan

tersebut, perusahaan dapat melakukan prediksi nilai *recency* dengan menggunakan nilai *monetary*, memprediksi nilai *monetary* dengan menggunakan *frequency*, sehingga perusahaan dapat melakukan *follow-up* kepada nasabah berdasarkan variabel tersebut.



Gambar 1. Alur metodologi penelitian

### B. Dataset

Data yang digunakan merupakan data transaksi nasabah secara umum yang terjadi selama tahun 2021. Koleksi data memiliki kolom *customer\_id*, *date*, *buyamount*, *buycomm*, *sellamount*, *sellcomm*, *trxvalue*, *sales\_id*, *created\_at*, *created\_by*, *updated\_at*, *updated\_by*. Jumlah data yang ada digunakan 70.488 baris data dan jumlah nasabah yang melakukan transaksi adalah 1597. Cuplikan koleksi data dapat dilihat pada Gambar 2.

	customer_id	date	buyamount	buycomm	sellamount	sellcomm	trxvalue	sales_id	created_at	created_by	updated_at	updated_by
id												
1	18	05/01/2021	35000000.0	0.0025	0.0	0.0025	35000000.0	8	21/12/2021 11:59	1	16/03/2022 14:39	1
2	19	05/01/2021	35400000.0	0.0025	0.0	0.0025	35400000.0	8	21/12/2021 11:59	1	16/03/2022 14:39	1
3	20	05/01/2021	12200000.0	0.0025	15680000.0	0.0025	27880000.0	8	21/12/2021 11:59	1	16/03/2022 14:39	1
4	21	05/01/2021	108346500.0	0.0025	101960000.0	0.0025	210306500.0	8	21/12/2021 11:59	1	16/03/2022 14:39	1
5	22	05/01/2021	7905000.0	0.0025	12807000.0	0.0025	20712000.0	8	21/12/2021 11:59	1	16/03/2022 14:39	1

Gambar 2. Cuplikan himpunan data transaksi nasabah 2021

### C. Data Preprocessing

Pada kegiatan ini, dilakukan proses pemilahan terkait apa saja yang akan digunakan dalam penelitian. *Preprocessing* ini penting untuk dilakukan karena dapat memberikan fungsi dan manfaat pada saat melakukan pengolahan data. Proses ini penting dilakukan untuk memastikan kualitas data baik sebelum digunakan untuk analisis data. Dalam proses ini perlu dipastikan terkait akurasi data, kelengkapan data, konsistensi, ketepatan waktu, terpercaya, dan juga interpretasi yang baik. Berikut adalah proses yang dilakukan pada saat melakukan *preprocessing* :

#### 1. Data Cleaning

Pada proses *data cleaning*, data yang ada dilakukan penghapusan terhadap yang dianggap salah, tidak lengkap ataupun tidak akurat. Pada data transaksi yang ada, dilakukan penghapusan baris data yang tidak memiliki nilai, merubah nilai kosong menjadi numerik nol (0), merubah data *minus* menjadi nol (0).

## 2. Data Reduction

Pada proses *data reduction*, data yang ada dilakukan proses pengurangan. Proses pengurangan ini dilakukan untuk memastikan data yang sudah ada, merupakan data yang seragam seperti memiliki tahun yang sama (tahun 2021), selain itu dilakukan proses penghapusan kolom yang tidak digunakan, yaitu: *sales\_id*, *created\_at*, *created\_by*, *updated\_at*, *updated\_by*. Kolom-kolom tersebut dihapus karena tidak digunakan pada saat analisis.

## 3. Data Transformation

Pada proses *data transformation*, dilakukan proses perubahan data seperti dilakukan perhitungan total komisi dari setiap transaksi dengan melakukan perhitungan dengan persamaan 1.

$$\text{Total komisi} = (\text{buycomm} + \text{sellcomm}) * \text{trxvalue} \quad (1)$$

Keterangan:

*buycomm* = Nilai desimal dari persentase komisi pembelian

*sellcomm* = Nilai desimal dari persentase komisi penjualan

*trxvalue* = Total jumlah total pembelian dan penjualan

Setelah merubah data tersebut dilakukan perubahan tipe data pada kolom *date* dimana tipe data sebelumnya merupakan tipe data *String* dan diubah menjadi tipe data *Datetime*. Perubahan dilakukan agar dapat dilakukan analisis terkait waktu yang ada di dalam data transaksi tersebut [3].

## D. Pemodelan dengan Recency, Frequency dan Monetary (RFM)

Setelah dilakukan *preprocessing data*, dilakukan proses pengolahan data dengan menggunakan model *Recency, Frequency, Monetary* (RFM) dimana data yang ada diperiksa transaksi terakhir yang dilakukan oleh nasabah, jumlah frekuensi melakukan transaksi berdasarkan nasabah dan jumlah transaksi yang dilakukan oleh nasabah. Pemodelan RFM adalah suatu pemodelan segmentasi nasabah yang sangat populer dan teknik ini sering digunakan untuk identifikasi dalam pemasaran. RFM menilai suatu data menjadi tiga faktor, yaitu:

- *Recency*: ini mengacu pada jumlah hari sebelum tanggal referensi ketika nasabah melakukan transaksi terakhir.
- *Frequency*: ini mengacu pada jumlah transaksi yang dilakukan oleh nasabah sebelum tanggal referensi.
- *Monetary*: ini mengacu pada jumlah nominal transaksi jual dan beli yang dilakukan nasabah sebelum tanggal referensi [2] [3] [4].

## E. Visualisasi Data dengan model RFM

Data yang sudah diolah dengan menggunakan model RFM, dilakukan penggambaran secara visual dengan menggunakan grafik dan juga chart. Pada visual yang ditampilkan tersebut, digambarkan terkait sebaran yang dimiliki oleh setiap variabel, dan juga hubungan nilai *recency*, *frequency* dan juga *monetary*.

## F. Segmentasi pelanggan dengan menggunakan Cluster K-Means

Pada tahap ini, dilakukan proses segmentasi nasabah berdasarkan data yang sudah diolah menjadi data RFM dimana dilakukan untuk mengetahui segmentasi terkait nasabah dan juga transaksi terakhir (*Recency*) dan mengetahui pengelompokan nasabah berdasarkan jumlah yang dihasilkan dengan transaksi yang sering dilakukan [2] [3].

## G. Analisis RFM terkait data segmentasi

Pada tahap ini dilakukan proses analisis terhadap data RFM, dimana dari data yang sudah diolah tersebut dilakukan proses *cluster* dengan menggunakan *K-Means*. Dari data hasil *cluster* tersebut lalu dilihat jumlah *cluster* yang dimiliki secara umum dan juga *cluster* yang dimiliki dari tiap-tiap kelas yang ada pada data RFM tersebut, setiap *cluster* dan kelompok dilihat pola dimiliki.

## H. Analisis dan Prediksi dengan Regresi Linier

Data yang sudah dilakukan proses segmentasi serta *clustering*, dilihat apakah data tersebut bersifat linier atau tidak. Data yang bersifat linier akan dilakukan proses analisis dan perhitungan untuk mengetahui terkait hubungan antara variabel *recency*, *frequency* dan juga *monetary*. Dari proses analisis ini juga dilakukan proses prediksi terhadap data yang sudah memiliki *cluster* dan kelompok. Pengolahan data dengan menggunakan regresi linier akan digunakan untuk melakukan prediksi antara variabel *recency*, *frequency* dan juga *monetary* yang akan terjadi di masa yang akan datang. Dari data yang sudah diprediksi dengan menggunakan regresi linier, dilakukan suatu pengujian untuk melihat akurasi dengan menggunakan MSE untuk mengetahui besaran kesalahan yang terjadi dalam memprediksi data [6].

### III. HASIL DAN PEMBAHASAN

#### A. Preprocessing Data

Pada penelitian ini, data yang dipergunakan adalah data di tahun 2021 dengan total data 70488 baris data. Pada tahap *preprocessing* dilakukan pembersihan data dengan menghapus baris dengan nilai *customer\_id* yang kosong, dan merubah semua nilai *null* dan *NaN* dan juga *minus* satu (-1). Setelah melakukan penghapusan data kosong, dilakukan proses transformasi dengan membuang kolom-kolom yang selain kolom *customer\_id*, *trxvalue*, dan juga kolom tanggal.

Kolom pada Tabel 1 dipertahankan karena merupakan kolom yang diperlukan untuk menerapkan model RFM. Setelah dilakukan proses transformasi data, data tersebut akan dikonversi menjadi file dengan format *csv* agar bisa dikelola di proses selanjutnya

TABEL 1  
HASIL PREPROCESSING

Kolom	Keterangan
<i>customer_id</i>	ID dari nasabah
<i>date</i>	Tanggal nasabah melakukan transaksi
<i>trxvalue</i>	Jumlah nilai gabungan antara nilai transaksi penjualan dan juga pembelian

#### B. Pengelompokan Data Dalam Tabel RFM

Dari himpunan data transaksi yang sudah tersedia kemudian dilakukan analisis dengan menggunakan RFM. Analisis dilakukan dengan mengelompokkan data pada kolom *Recency*, *Frequency* dan *Monetary*. Kolom *Recency* diisi dengan jumlah hari sekarang dengan transaksi terakhir. Kolom *Frequency* diisi dengan berapa kali nasabah tersebut melakukan transaksi. Kolom *Monetary* diisi dengan total nominal transaksi yang dilakukan oleh nasabah. Pada Tabel 2, dapat dilihat data hasil *preprocessing* yang akan diolah dengan menggunakan model RFM [2] [7] [8].

TABEL 2  
DATA SEBELUM PEMROSESAN DENGAN MODEL RFM

<i>date</i>	<i>customer_id</i>	<i>trxvalue</i>
2021-05-01	18	35000000
2021-05-01	19	35400000
2021-05-01	20	27880000
2021-05-01	21	210306500
2021-05-01	22	20712000

Tabel 2 dilakukan pemrosesan dan kalkulasi dengan menggunakan model *RFM* dimana nilai *recency* dihitung dengan:\ menggunakan persamaan (2).

$$R = \text{current\_date} - \text{last\_row\_date} \quad (2)$$

Keterangan:

*R* = Nilai *recency*, nilai *recency* yang semakin kecil akan semakin baik dan menandakan bahwa nasabah baru saja melakukan transaksi dalam waktu dekat.

*current\_date* = Tanggal transaksi saat ini

*last\_row\_date* = Tanggal transaksi terakhir nasabah

Data pada kolom *frequency* didapat dari total transaksi yang dilakukan oleh *customer\_id*, dan *monetary* merupakan penjumlahan dari *trxvalue*, sehingga dihasilkan tabel dengan model *RFM* seperti pada Tabel 3.

TABEL 3  
HASIL PENGOLAHAN MODEL RFM

<i>customer_id</i>	<i>recency</i>	<i>frequency</i>	<i>monetary</i>
1	100	2	1066000
2	0	3	456500
5	103	4	609900

<i>customer_id</i>	<i>recency</i>	<i>frequency</i>	<i>monetary</i>
7	18	22	244879500
9	1	2	4942500

Setelah dilakukan pemrosesan dengan model RFM, dilakukan perhitungan skor dan juga bobot dari setiap nilai *recency*, *frequency* dan juga *monetary*. Proses skoring *recency*, *frequency* dan *monetary* dilakukan pelabelan bahwa nilai tersebut masuk ke dalam kuantil berapa. Nilai kuantil pada *recency* berbeda dengan nilai *frequency* dan *monetary*, nilai *recency* yang besar menandakan bahwa nasabah sudah lama tidak melakukan transaksi sehingga skor satu akan diberikan pada *recency* terbesar dan kuantil blok pertama. Nilai *frequency* dan *monetary* akan mendapatkan skor 1 apabila nilai yang dimiliki terkecil dan pada kuantil blok pertama. Data hasil perhitungan bobot dapat dilihat pada Tabel 4 [3].

TABEL 4  
HASIL PERHITUNGAN SKOR RFM BERDASARKAN KUANTIL

<i>customer_id</i>	<i>recency</i>	<i>frequency</i>	<i>monetary</i>
1	1	1	1
2	3	2	1
5	1	1	1
7	3	3	3
9	3	2	2

Setelah mendapatkan skor pada setiap kolom *recency*, *frequency* dan *monetary*, dilakukan proses penilaian dari setiap skor yang sudah dibuat. John R. Miglautsch menulis bahwa RFM dapat membantu bisnis dengan menghitung bobot dengan menggunakan persamaan 3 [6] [9] [10]:

$$Bobot = R * 3 + F * 2 + M * 1 \quad (3)$$

Keterangan:

- Bobot* = Nilai pembobotan dari perhitungan dari tiap nilai RFM
- R* = *Recency*
- F* = *Frequency*
- M* = *Monetary*

Dari hasil kalkulasi pada persamaan (3), nilai maksimum yang didapat dari bobot adalah 30. Dengan skor tersebut menandakan bahwa nasabah tersebut merupakan nasabah sering melakukan transaksi, dan memiliki nilai *monetary* yang tinggi. Sedangkan nilai terkecil adalah 6, yang menandakan nasabah tersebut jarang melakukan transaksi dan juga memiliki nilai *monetary* rendah. Hasil dari perhitungan dapat dilihat pada Tabel 5.

TABEL 5  
HASIL PERHITUNGAN BOBOT

<i>customer_id</i>	<i>recency</i>	<i>frequency</i>	<i>monetary</i>	Bobot RFM
1	1	1	1	6
2	3	2	1	18
5	1	1	1	6
7	3	3	3	22
9	3	2	2	19

Pada bobot yang sudah dihitung pada Tabel 6 maka bisa terlihat nilai dari seorang nasabah dalam melakukan transaksi. Dari nilai tersebut dapat dilakukan pembagian kembali dengan menggunakan kuantil, sehingga menghasilkan lima kelompok dengan nilai-nilai sesuai blok. Dengan nilai maksimum 30, maka setiap blok memiliki *range* 6. Dengan cara seperti ini, dapat diberikan label pada setiap kelompok seperti pada Tabel 6.

TABEL 6  
PENGELOMPOKAN NASABAH BERDASARKAN KUANTIL BOBOT RFM

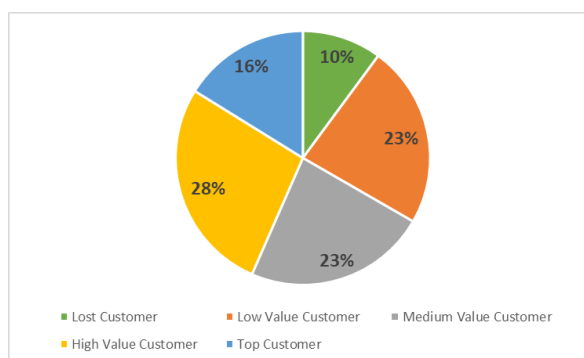
Bobot	Kelompok Nasabah
1-6	Lost Customer
7-11	Low Value Customer
12-16	Medium Value Customer
17-22	High Customer
> 22	Top Customer

Pada Tabel 7, dapat dilihat label pembobotan dari setiap kelompok yang sudah ada. Lalu dilakukan *labeling* terhadap data yang sudah ada sehingga dapat dibentuk pengelompokan nasabah pada Tabel 7.

TABEL 7  
KELOMPOK NASABAH BERDASARKAN BOBOT

customer_id	bobot	Kelompok nasabah
1	6	Lost Customer
2	18	High Value Customer
5	6	Lost Customer
7	22	High Customer
9	19	High Customer

Dari Tabel 7, diketahui kelompok dari setiap nasabah. Dan dari data yang sudah diolah tersebut, lalu dapat digambarkan dengan menggunakan *pie chart* untuk mengetahui persentase penyebaran nasabah yang sudah melakukan transaksi selama tahun 2021 seperti pada Gambar 3.



Gambar 3. Kelompok nasabah berdasarkan bobot RFM

### C. Cluster Data dengan K-means

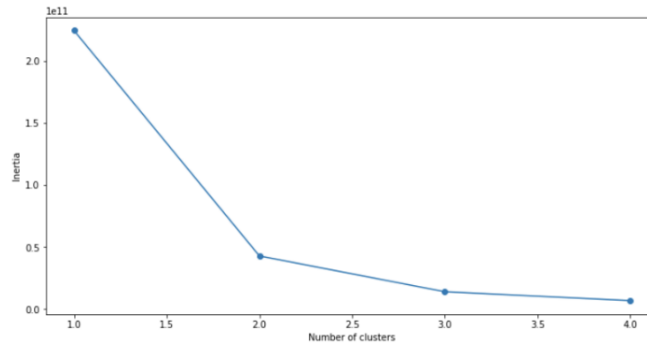
Data RFM yang sudah dimiliki, dilakukan pengecekan jumlah *cluster* yang dimiliki dengan menggunakan algoritma *K-Means*. Untuk menghitung jumlah *cluster* yang optimal, perhitungan optimum *cluster* dilakukan dengan menggunakan *elbow method*. Tujuan dari metode *elbow method* ini adalah untuk memilih nilai *k* yang kecil. Penentuan nilai optimal jumlah *cluster* dilakukan dengan menggunakan *elbow method*. Hal ini dilakukan dengan memperhatikan nilai perbandingan antara jumlah *cluster* yang pembentukan suatu 'siku' (sudut 90 derajat) dalam suatu titik. Metode siku ini memperlihatkan bahwa semakin besar jumlah *cluster k* maka nilai *error* (berupa jarak) antar clusternya akan semakin kecil. Metode statistik yang digunakan untuk mengukur nilai selisih total dari nilai sebenarnya adalah *Sum Squared Error (SSE)*. Persamaan *SSE* dapat dilihat pada persamaan 4.

$$SSE = \sum_{i=1}^n d^2 \quad (4)$$

Keterangan:

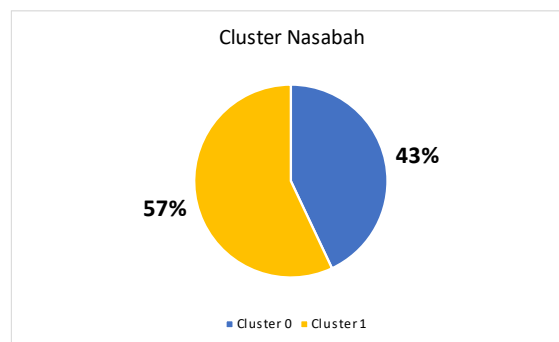
*d* = jarak antara *cluster* dengan pusat *cluster*

Perhitungan dengan *SSE* digunakan untuk mengukur data yang diperoleh dengan model prediksi sebelumnya dan dijadikan referensi penelitian untuk menentukan jumlah *cluster* optimal. Gambar 4 menunjukkan variasi *inertia* terhadap banyaknya *cluster* *k*. Pada Gambar 4, dapat dilihat proses tekuk terjadi pada nilai *cluster* nol (0) sehingga nilai jumlah *optimum cluster* yang didapat adalah dua *cluster*.



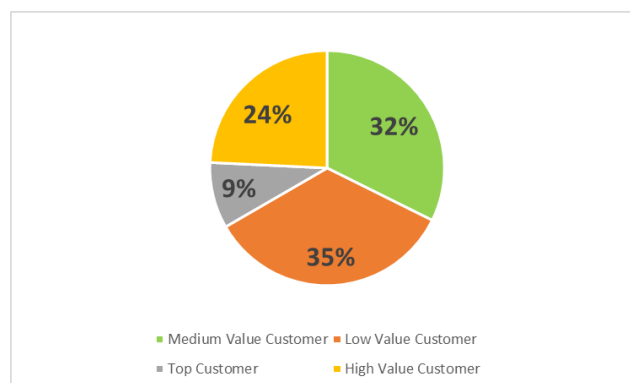
Gambar 4. Hasil perhitungan jumlah optimum *cluster* dengan *Elbow Method*

Dari hasil proses segmentasi dengan menggunakan *K-means* pada data RFM yang sudah dilakukan kalkulasi, didapati bahwa 57% dari data RFM merupakan segmen 0 dan 43% dari data tersebut merupakan segmen 1 dengan visualisasi seperti Gambar 5.



Gambar 5. Persentase *cluster* nasabah

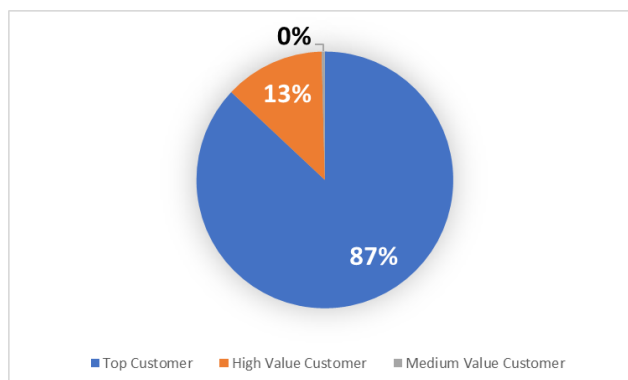
Dari data segmentasi yang dimiliki, maka dilakukan proses visualisasi yang membantu untuk mengetahui kelompok apa saja yang ada di dalam *cluster* satu (1) dan nol (0). Sebaran kelompok RFM berdasarkan *cluster* nol (0) dapat dilihat pada Gambar 6 dan *cluster* satu (1) pada Gambar 7.



Gambar 6. Persentase sebaran kelompok RFM pada *Cluster* Nol (0)



Pada Gambar 6 diketahui sebaran dari kelompok RFM yang ada di dalam *cluster* satu (1), dimana pada *cluster* satu (1) hanya terdapat 9% *Top Customer*, dan didominasi oleh *Low Value* dan *Medium Value Customer*. Untuk sebaran kelompok RFM pada *cluster* nol (0), dapat dilihat pada Gambar 7. Pada Gambar 7, dapat diketahui didominasi oleh *Top Customer* sebesar 87% , dimana pada *cluster* ini diisi dengan nasabah yang memiliki nilai besar [11].



Gambar 7. Persentase sebaran kelompok RFM pada Cluster Satu (1)

Dari peta sebaran *cluster* terhadap data dengan model RFM, dapat ditarik kesimpulan sementara bahwa nasabah dapat dibagi menjadi 2 *cluster* yang memiliki karakteristik unik. *Cluster* nol (0) diisi dengan nasabah yang memiliki nilai tinggi secara RFM karena didominasi oleh *Top Customer* dan *High Value Customer*. Sedangkan *cluster* satu (1) diisi dengan nasabah yang lebih beragam namun didominasi oleh *Medium Value Customer* dan *Low Value Customer*.

#### D. Analisis Data dengan Regresi Linier

Analisis data dengan menggunakan regresi linier digunakan untuk mencari hubungan sebab akibat antara variabel  $x$  dengan variabel  $y$ . Dilakukan proses pencocokan pada data RFM yang sebelumnya dianalisis dalam bentuk *cluster*. Faktor penyebab pada regresi linier pada umumnya dilambangkan dengan  $x$  atau disebut juga prediktor sedangkan variabel akibat dilambangkan dengan  $y$  disebut juga dengan respon. Persamaan regresi linier dapat dilihat pada persamaan (5), (6), dan (7).

$$y = a + bx \quad (5)$$

$$a = \frac{(\sum y)(\sum x^2) - (\sum x)(\sum xy)}{n(\sum x^2) - (\sum x)^2} \quad (6)$$

$$b = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2} \quad (7)$$

Keterangan:

$y$  = variabel *response* atau variabel akibat (*dependent*)

$x$  = variabel prediktor atau variabel faktor penyebab (*independent*)

$a$  = konstanta

$b$  = koefisien regresi (kemiringan); besaran *response* yang ditimbulkan oleh faktor

$n$  = jumlah data

Pada proses ini dilakukan perhitungan antara *recency* dan *monetary* dimana dengan memanfaatkan salah satu variabel ini diharapkan dapat menjadi salah satu masukan bagi perusahaan untuk melakukan *follow-up* nasabah berdasarkan *frequency* transaksi yang dilakukan oleh nasabah. Dengan cara demikian dapat diketahui dalam rentang berapa waktu (satuan: hari) nasabah dalam suatu kelompok tertentu perlu untuk dilakukan *follow-up*.

Dari himpunan data pada setiap *cluster* dilakukan proses kalkulasi regresi linear untuk memprediksi rentang waktu *follow-up*. Dari data yang ada, dihitung *intercept* dan koefisien. Nilai koefisien yang dihasilkan adalah  $-2.78611013e-18$   $1.00000000e+00$   $-1.11022302e-16$  dan nilai *intercept* 0 terhadap *regressor* [9], [10]. Dari hasil prediksi terhadap *monetary*, didapatkan data sebenarnya dan data prediksi dimana dapat dilihat pada Gambar 8:

	Data Sebenarnya	Data Prediksi
0	0	-7.199563e-15
1	155	1.550000e+02
2	3	3.000000e+00
3	32	3.200000e+01
4	155	1.550000e+02
...	...	...
334	31	3.100000e+01
335	0	-2.711232e-14
336	62	6.200000e+01
337	354	3.540000e+02
338	153	1.530000e+02

Gambar 8. Data prediksi Monetary dari nilai Recency

MSE atau disebut dengan Mean Squared Error merupakan metode yang digunakan untuk pengecekan error. Nilai MSE yang rendah mendekati nol menandakan bahwa hasil peramalan sesuai dengan hasil yang aktual. MSE biasanya digunakan untuk mengevaluasi metode pengukuran dengan model regresi atau model peramalan. Rumus MSE dijelaskan pada persamaan 8 [12] [13].

$$MSE = \frac{\sum_{t=1}^n (At - Ft)^2}{n} \quad (8)$$

Keterangan:

- At = Nilai aktual permintaan
- Ft = Nilai hasil peramalan
- N = banyaknya data

MSE biasanya digunakan untuk mengevaluasi metode pengukuran dengan model regresi atau model peramalan. Dari data recency dan monetary dilakukan proses perhitungan MSE dimana hasil dari MSE data training adalah 0.358 dan MSE data uji adalah 0.353. Setelah dilakukan proses pengecekan error dengan MSE, dilakukan pengujian dengan menggunakan determinasi dengan maksud mengukur kemampuan model dalam menerangkan seberapa pengaruh variabel independen secara bersama-sama mempengaruhi variabel dependen yang diindikasikan dalam R-Squared.

Nilai R-Squared yang baik, adalah nilai yang mendekati satu. Menurut Warren [14], R-Squared lebih dari 0,67 dikategorikan kuat, dikategorikan moderat apabila lebih besar dari 0,33 hingga 0,67. Pada Gambar 9 dapat dilihat R-Squared yang didapatkan dari perhitungan hasil uji regresi linier adalah 0,720 sehingga dapat dikategorikan kuat [14].

```

MSE train RF : 0.358; MSE test: 0.353
                                OLS Regression Results
=====
Dep. Variable:                  y      R-squared (uncentered):      0.720
Model:                        OLS      Adj. R-squared (uncentered):  0.720
Method:                        Least Squares      F-statistic:                1634.
Date:                            Mon, 04 Jul 2022      Prob (F-statistic):         0.00
Time:                             09:37:38      Log-Likelihood:             -650.87
No. Observations:                1271      AIC:                        1306.
Df Residuals:                     1269      BIC:                        1316.
Df Model:                          2
Covariance Type:                  nonrobust
=====
                                coef      std err      t      P>|t|      [0.025      0.975]
-----+-----
recency                1.7051      0.030      57.163      0.000      1.647      1.764
monetary              -0.0053      0.374      -0.014      0.989      -0.739      0.729
=====
Omnibus:                    100.671      Durbin-Watson:              1.795
Prob(Omnibus):              0.000      Jarque-Bera (JB):           123.363
Skew:                       0.736      Prob(JB):                   1.63e-27
Kurtosis:                   3.405      Cond. No.                   12.5
=====

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

```

Gambar 9. Hasil Perhitungan R-Squared Recency dan Monetary

#### IV. KESIMPULAN

Data transaksi yang dimiliki perusahaan sekuritas secara *general*, dapat digunakan untuk mengetahui kelompok nasabah dengan menggunakan model RFM dimana dari data yang ada dapat diketahui bahwa terdapat dua *cluster* dengan lima kelompok yang didominasi oleh nasabah dengan *Medium*, *Low* dan *Lost Customer*. Berdasarkan data ini, perusahaan dapat memikirkan untuk melakukan pendekatan terhadap nasabah-nasabah tersebut. Dengan menggunakan data tersebut juga, dapat dilakukan proses prediksi dengan menggunakan regresi linier, perusahaan dapat melakukan prediksi dengan menggunakan data es dan juga data prediksi. Berdasarkan hasil pengujian dengan menggunakan regresi linier, hasil dari perhitungan tersebut cukup baik karena memiliki nilai MSE 0,358 dan mendekati 0. Untuk nilai *R-Squared* yang didapatkan adalah 0.720, nilai tersebut didapatkan terhadap variabel *Recency* dan juga *Monetary*.

#### UCAPAN TERIMA KASIH

Ucapan terima kasih kami haturkan kepada perusahaan Mandiri Sekuritas Cabang Puri Indah karena telah bersedia untuk membantu menyediakan data dalam penelitian ini. Kami juga berterima kasih kepada segenap anggota tim terkait karena sudah membantu dalam banyak hal sehingga penelitian dan penulisan jurnal ini dapat berjalan dengan baik.

#### DAFTAR PUSTAKA

- [1] P. Anitha and M. M. Patil, "RFM model for customer purchase behavior using K-Means algorithm," *Journal of King Saud University*, vol. 34, no. 5, pp. 1785-1792, 2019.
- [2] J. R. Miglautsch, "Thoughts on RFM Scoring," *Journal of Database Marketing & Customer Strategy Management*, vol. 8, no. 1, pp. 67-72, 2000.
- [3] M. Tsoi and V. Shekoldin, "RFM-analysis as a Tool for Segmentation of High-Tech Products Consumers," *13th International Scientific-Technical Conference on Actual Problems of Electronics Instrument Engineering (APEIE)*, vol. 3, pp. 290-293, 2016.
- [4] 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 Teknologi dan Sistem Informasi*, vol. 1, no. 1, pp. 204-215, 2021.
- [5] R. Parangin-angin, R. Nainggolan and E. Simarmata, "Improved the performance of the K-means cluster using the sum of squared error (SSE) optimized by using the Elbow method.," *Journal of Physics: Conference Series*, vol. 1361, pp. 1-6, 2019.
- [6] G. N. Ayuni and D. Fitriah, "Penerapan Metode Regresi Linear Untuk Prediksi Penjualan Properti pada PT XYZ," *Jurnal telematika*, vol. 14, pp. 79-86, 2019.
- [7] Basri, W. Gata and Risnandar, "Analisis Loyalitas Pelanggan Berbasis Model Recency, Frequency, dan Monetary (RFM) dan Decision Tree pada PT. SOLO," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, vol. 7, pp. 943-950, 2020.
- [8] C. H. Cheng and Y. S. Chen, "Classifying the Segmentation of Customer Value via RFM model and RS theory," *Expert Systems with Applications*, vol. 36, no. 3, pp. 4176-4184, 2009.
- [9] P. Katemba and R. K. Djoh, "Prediksi Tingkat Produksi Kopi Menggunakan Regresi Linear," *Jurnal Ilmiah Flash*, vol. 3, no. 1, pp. 42-41, 2017.
- [10] A. A. Suryanto, "Penerapan Metode Mean Absolute Error (MEA) Dalam Algoritma Regresi Linear Untuk Prediksi Produksi Padi," *SAINTEKBU: Jurnal Sains dan Teknologi*, vol. 11, no. 1, pp. 78-83, 2019.
- [11] W. Jing and L. Zheng, "Research on Customer Segmentation Model by Clustering," *Proceedings of the 7th international conference on Electronic commerce*, pp. 316-318, 2005.
- [12] N. T. Hartanti, "Metode Elbow dan K-Means Guna Mengukur Kesiapan Siswa SMK Dalam Ujian Nasional," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 6, no. 2, pp. 82-89, 2020.
- [13] A. Febriani and S. A. Putri, "Segmentasi Konsumen Berdasarkan Model Recency, Frequency, Monetary," *JIEMS (Journal of Industrial Engineering and Management Systems)*, vol. 13, no. 2, pp. 52-57, 2020.
- [14] D. Chicco, M. Warrens and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," *PeerJ Computer Science*, vol. 7, pp. 1-24, 2021.