



PEMANFAATAN DATA TRANSAKSI UNTUK DASAR MEMBANGUN STRATEGI BERDASARKAN KARAKTERISTIK PELANGGAN DENGAN ALGORITMA *K-MEANS CLUSTERING* DAN MODEL RFM

Carudin¹

¹ Manajemen Informatika, Sekolah Tinggi Manajemen Informatika dan Komputer Bani Saleh
Bekasi Timur, Jawa Barat, Indonesia 17113
carudin@stmik-banisaleh.ac.id

Abstract

Every time there is a transaction process carried out by a customer, that process adds to the data collection in a database. This study uses transaction data to determine customer segmentation and build a strategy based on customer characteristics with the RFM and K-Means model approach. K-Means Clustering is an algorithm that can produce a visual cluster model with the Rapidminer application version 9.9, using the RFM attribute to represent the number of customers from each cluster. The transaction data for the last three years, 2017, 2018, and 2019 with 4,332 transactions, were then managed based on the RFM model resulting in 1898 customers. Furthermore, a cluster analysis carries out using the K-Means algorithm with 319 customers in cluster 1, 314 customers in cluster 2, 316 customers in cluster 3, 317 customers in cluster 4, 315 customers in cluster 5, and 317 customers in cluster 6. The company can use the results of this study to determine customer characteristics and as a consideration for making a new strategy.

Keywords: *Clustering, Recency, Frequency, Monetary, K-Means, Customer*

Abstrak

Setiap waktu terdapat proses transaksi yang dilakukan oleh pelanggan, proses tersebut menambah koleksi data pada sebuah *database*. Pada penelitian ini dengan melakukan pemanfaatan data transaksi untuk mengetahui segmentasi pelanggan dan membangun strategi berdasarkan karakteristik pelanggan dengan pendekatan model RFM dan *K-Means*. *K-Means Clustering* adalah sebuah algoritma yang dapat menghasilkan suatu model *cluster visual* dengan aplikasi Rapidminer versi 9.9, dengan menggunakan atribut RFM berfungsi untuk mewakili jumlah pelanggan dari setiap *cluster*. Dari data transaksi 3 tahun terakhir 2017, 2018, dan 2019 dengan jumlah 4.332 transaksi yang kemudian diolah berdasarkan model RFM menghasilkan 1898 pelanggan. Selanjutnya dilakukan analisis *cluster* dengan menggunakan algoritma *K-Means* dengan hasil *cluster 1* memiliki 319 pelanggan, *cluster 2* memiliki 314 pelanggan, *cluster 3* memiliki 316 pelanggan, *cluster 4* memiliki 317 pelanggan, *cluster 5* memiliki 315 pelanggan, dan *cluster 6* memiliki 317 pelanggan. Dari hasil penelitian ini dapat dimanfaatkan oleh perusahaan untuk mengetahui karakteristik pelanggan dan sebagai bahan pertimbangan membuat suatu strategi baru.

Kata kunci: *Clustering, Recency, Frequency, Monetary, K-Means, Pelanggan*

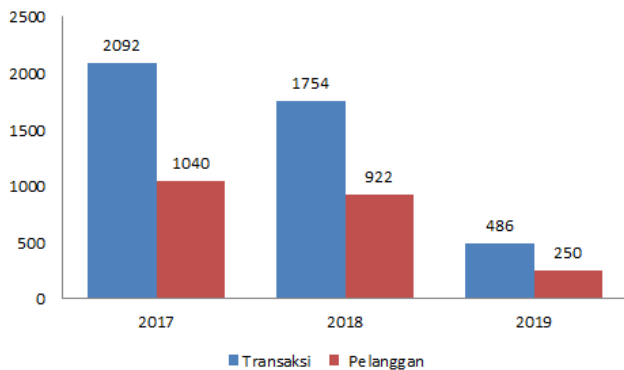
1. PENDAHULUAN

Perusahaan yang bergerak di bidang retail data transaksi penjualan merupakan data yang sangat berharga di dalam suatu perusahaan. Karena dari data tersebut dapat diolah menjadi beragam informasi yang bisa dimanfaatkan oleh perusahaan, seperti mengetahui prediksi perkembangan perusahaan di masa depan, melihat omset perusahaan, mengetahui karakteristik pelanggan dan lainnya.

Dewasa ini sebagian perusahaan besar mulai memanfaatkan data transaksi untuk diolah dengan menggunakan suatu metode tertentu, sehingga menghasilkan informasi penting yang bisa memprediksi atau membangun suatu strategi bisnis yang bisa memberikan kemajuan pada perusahaan. Salah satunya adalah PT. XYZ yang bergerak di jasa penjualan mobil Jepang, yang mulai memanfaatkan data transaksi penjualan untuk mengelompokan pelanggan berdasarkan karakteristik-karakteristiknya, sehingga dari data yang

diolah dapat dijadikan acuan untuk membangun suatu strategi pemasaran yang dapat meningkatkan jumlah transaksi penjualan di masa yang akan datang.

Permasalahan yang muncul saat ini adalah dari 3 tahun terakhir ini terjadi penurunan jumlah transaksi, tahun 2017 sebanyak 2092 transaksi dari 1040 pelanggan, tahun 2018 1754 transaksi dari 922 pelanggan, sedangkan jumlah transaksi di tahun 2019 hanya 486 transaksi dari 250 pelanggan, bentuk penurunan jumlah transaksi terlihat di gambar 1 di bawah ini:



Gambar 1. Grafik Transaksi Penjualan

Dari data transaksi yang disajikan, terlihat jelas semakin banyak pelanggan maka semakin banyak menambah jumlah transaksi. Berdasarkan kondisi dan masalah yang telah dijelaskan di atas, diidentifikasi terkait sebab-sebab yang mempengaruhi penurunan jumlah transaksi. Sehingga perlu dilakukan pengendalian pendekatan terhadap pelanggan dan strategi baru. Pada penelitian ini digunakan data transaksi penjualan 3 tahun terakhir yaitu 2017, 2018 dan 2019 dengan total jumlah transaksi 4332.

Bagaimana memanfaatkan data transaksi dengan model RFM dan mengelompokkan berdasarkan karakteristik pelanggan dengan pendekatan algoritma *K-Means*, sehingga dari hasil clustering dapat dijadikan suatu strategi baru yang tepat dan akurat.

Untuk meningkatkan jumlah transaksi penjualan maka perlu adanya proses pengendalian pendekatan antara perusahaan dan pelanggan, karena hubungan antara perusahaan dan pelanggan merupakan elemen bisnis utama yang menunjang peningkatan transaksi penjualan [1]. Proses pengendalian hubungan antara perusahaan dan pelanggan disebut dengan istilah *Customer Relationship Management (CRM)*. Cakupan konsep CRM meliputi seperangkat metode dan strategi untuk mengembangkan hubungan jangka panjang yang dapat menguntungkan pelanggan dan meningkatkan transaksi penjualan. Menurut [2], CRM dibagi menjadi empat bagian yang berbeda: (1) Identifikasi Pelanggan, (2) Daya Tarik Pelanggan, (3) Retensi Pelanggan, dan (4) Pengembangan Pelanggan.

Segmentasi pelanggan adalah salah satu bagian paling penting dari identifikasi pelanggan dan mencakup mengelompokkan seluruh pelanggan sesuai dengan

karakteristik pelanggan berdasarkan perilaku pembelian, informasi demografis dan geografis, serta atribut psikografis. Model RFM adalah yang banyak digunakan untuk menentukan segmentasi pelanggan.

Model RFM

Model RFM dikenal sebagai teknik data mining berdasarkan perilaku yang menggambarkan profil pelanggan dengan memanfaatkan data transaksi dengan mengambil poin *recency*, *frequency*, dan *monetary* [3].

Nilai R, F, dan M dibedakan menjadi lima dengan nilai 1 sampai 5. Nilai R dihasilkan dari waktu transaksi terakhir dengan waktu interval penelitian. Nilai 5 untuk pelanggan yang memiliki waktu transaksi paling dekat dengan waktu penelitian, nilai 1 pelanggan yang memiliki waktu transaksi jauh dengan waktu penelitian. Kemudian nilai F, didapat dari pelanggan yang paling banyak bertransaksi dengan nilai 5, pelanggan yang sedikit bertransaksi diberikan nilai 1. Nilai M didapat dari pelanggan yang memiliki transaksi dengan jumlah uang terbanyak dengan nilai 5. Sedangkan, pelanggan yang melakukan transaksi dan jumlah uang sedikit mempunyai nilai-nilai 1 [4]. Berikut rumus untuk mendapatkan nilai RFM:

$$RFM = ((\text{Nilai Recency} \times 100) + (\text{Nilai Frequency} \times 0) + \text{Nilai Monetary})$$

Discretization

Proses menentukan bobot di setiap fitur data sehingga membentuk data diskrit kategorik. Sehingga dihasilkan *dataset* indek rasio data dibagi ke dalam 6 bagian dalam pembobotan 6,5,4,3,2,1 dengan konversi 6= "Superstar", 5= "Golden Customer", 4= "Typical Customer", 3= "Occasional Customer", 2= "everyday shopper", 1= "Dormant Customer". Pada bagian dengan nilai rasio terbesar diberi bobot 6 karena menandakan *customer superstar* [3]. Demikian selanjutnya sampai customer dengan nilai rasio terkecil diberi bobot 1. Hasil yang didapat dari proses pembobotan ini menjadi *dataset* dasar untuk diolah pada pengolahan berikutnya yakni mengklasifikasikan *customer* dengan menerapkan metode K-means. Interval data pembobotan ditunjukkan pada tabel 1, 2 dan 3 berikut:

Tabel 1. Interval Pembobotan R

Recency	Dari (Jumlah Hari)	Sampai (Jumlah Hari)
6	0	60
5	61	120
4	121	180
3	181	240
2	241	300
1	301	>360

Tabel 2. Interval Pembobotan F

Frequency	Dari (Jumlah Hari)	Sampai (Jumlah Hari)
1	1	10
2	11	20
3	21	30
4	31	40
5	41	50
6	51	>60

Tabel 3. Interval Pembobotan M

Monetary	Dari (Jumlah Hari)	Sampai (Jumlah Hari)
1	140.000.000	200.000.000
2	200.000.001	260.000.000
3	260.000.0001	310.000.000
4	310.000.001	500.000.000
5	500.000.001	700.000.000
6	700.000.001	>52.780.050.000

Data Mining

Untuk mendapatkan model yang berguna pada data yang memiliki ukuran yang besar maka perlu menggunakan suatu teknik penambangan data atau dengan istilah *data mining* [5]. Data yang berasal dari suatu basis data, seperti data transaksi, *data warehouse* atau data lainnya. *Data mining* adalah suatu tahapan penting dalam proses KDD (*Knowledge Discovery in Database*) terpenting berhubungan dengan ekstraksi dan perhitungan model yang diteliti.

Clustering

Clustering atau disebut juga dengan klasterisasi yaitu suatu metode yang berfungsi mengelompokkan objek data menjadi beberapa kelompok, sehingga dari setiap kelompok mempunyai tingkat kesamaan maksimum dan data antar *cluster* mempunyai tingkat kesamaan minimum [6].

Nilai kemiripan bersumber pada nilai atribut yang melukiskan suatu objek. *Clustering* juga membuat partisi suatu set objek kedalam kelompok yang dinamakan dengan *cluster* [7]. Objek yang berada di dalam *cluster* mempunyai kesamaan karakteristik dengan lainnya dan memiliki perbedaan dengan *cluster* lainnya. Proses mempartisi data menggunakan suatu algoritma *clustering* [8].

Algoritma K-Means

K-Means yaitu algoritma yang memiliki tujuan mengelompokkan data dan membentuk beberapa kelompok. Pemrosesan *clustering K-Means*, dikerjakan oleh sistem komputer yang berfungsi mengelompokkan secara otomatis data-data yang menjadi inputan tanpa menentukan target kelompoknya [9]. Pada masing-masing *cluster* memiliki titik pusat (*centroid*) yang merepresentasikan *cluster* tersebut. Proses perhitungan *K-Means* sebagai berikut [10]:

- Menetapkan nilai k sebagai jumlah *cluster* yang dibuat.
- Menetapkan nilai *centroid* pertama atau titik pusat *cluster*. Proses ini menentukan nilai *centroid* dengan cara random, akan tetapi untuk proses selanjutnya menggunakan rumus berikut:

$$v_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} X_{kj}$$

Dimana :

V_{ij} = *Centroid* ke-I untuk *variable* ke-j

N_i = Jumlah cluster ke-i

i, k = indeks dari *cluster*

j = indeks dari *variabel*

X_{kj} = nilai data ke-k di *cluster* untuk *variable* ke-j

- Jarak *centroid* dengan titik tiap-tiap objek, bisa dihitung dengan menerapkan rumus *Euclidean Distance* (nilai kedekatan) berikut:

$$D(x_i, y_j, z_k) = \sqrt{(x_{i1} - m_{j1})^2 + (x_{i2} - m_{j2})^2 + (x_{i3} - m_{j3})^2}$$

- Mencari jarak minimum antar *cluster*
 $a_k = d_k = \min \{ D(X_k, C_i), i = 1, 2, 3, 4, 5, 6 \}$ $k = 1, 2, \dots, n$
 Keterangan:
 $a_k = d_k$ = Jarak minimum tiap *cluster*
 X_k = Keanggotaan data ke-k
 C_i = Nilai *centroid cluster* ke-i

- Mencari nilai *centroid* baru

$$v_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} X_{kj}$$

Dimana :

v_{ij} = *centroid* rata-rata *cluster* ke-i untuk variabel ke-j

N_i = jumlah anggota *cluster* ke-I i, k = indeks dari *cluster*

j = indeks dari variabel

X_{kj} = nilai data ke-k variabel ke-j dalam *cluster* tersebut.

- Jika hasil perhitungan pada iterasi berikutnya tidak ada perubahan pada anggota *cluster*, maka perhitungan dinyatakan selesai.

Beberapa penelitian sebelumnya menggunakan model RFM untuk menganalisis data penjualan seperti yang dilakukan oleh Tavakoli, Molavi dkk. di perusahaan Digikala, perusahaan *E-Commerce* terbesar di Timur Tengah, dengan membandingkan model RFM dan model

Customer Quantile dari hasil perbandingan diterapkan ke dalam bentuk promosi melalui media SMS sesuai dengan strategi menunjukkan bahwa model segmentasi RFM mampu meningkatkan jumlah pembelian [1].

Penelitian oleh Ananthi Sheshasaayee dan Logeshwari dari data transaksi pelanggan dilakukan konversi kedalam model RFM kemudian dilakukan diskritisasi dan dikelompokkan menjadi 2, satu cluster 0 dikategorikan sebagai pelanggan yang memiliki frekuensi pembelian tinggi sehingga kelompok ini hanya membutuhkan sedikit konsentrasi dan loyalitas, sedang cluster 1 memiliki frekuensi pembelian rendah sehingga mereka membutuhkan lebih banyak penawaran dan pengingat melalui iklan [11].

Penelitian oleh Ina Maryani, Dwiza Riana dkk, melakukan analisis segmentasi pelanggan dari transaksi pembelian isi ulang pulsa dengan model RFM dan teknik clustering K-Means, dari 82.648 transaksi diolah berdasarkan Model RFM menghasilkan 102 Nasabah, kemudian dilakukan

Pada pengumpulan data ini dengan melakukan studi pustaka, membaca dan mempelajari jurnal-jurnal yang berhubungan dengan topik penelitian, melakukan observasi ke objek penelitian, melakukan wawancara dengan bagian penjualan, melihat dan mempelajari data transaksi penjualan. Pada proses ini bertujuan untuk menghasilkan data dan informasi yang selengkap-lengkapya terkait sistem yang berjalan saat ini.

b. Instrumen Penelitian

Pada instrumen penelitian ini mengadopsi dan memodifikasi pada penelitian sebelumnya yang dilakukan oleh [1] dan [4]. Data transaksi penjualan 3 tahun 2017, 2018 dan 2019 dengan jumlah transaksi 4332 penjualan, terlihat data transaksi penjualan pada gambar 1 berikut:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	No	Filling No.	No Faktur	Tgl Faktur	SLS	NPK Sales	Nama Salesman	KD.	J	Customer	No	Customer	Cus	Leasing	Tipe	Kenda	Chasis			
913	T151-2017000916	T151V1017000916	##	00041477	ADY KURNIAMAN	CSH	7008970703	LIE SUGIARTO	09	RKB BCA	AL30GA/T	JNFGF3DH3H8010495								
914	T151-2017000917	T151V1017000917	##	00053942	LIE YANCHE NUARI	WCSH	7008968718	RENALDY SOETEDJO	09	PT. TOYOTA	JG0N65V2A/T	MHFPG8G52H0846645								
915	T151-2017000918	T151V1017000918	##	00088529	FIK MULYADI	CR5	7008946033	GERALDUS FARMAN	09	GM42GM/TL	MHFJF8EH11019072									
916	T151-2017000919	T151V1017000919	##	00049393	SYARIEP HIDAYATULL	CSH	7008949992	DIAN SUTARNA	09	ACC KWIANG	NSP707M/T00	MHFZ28H3H80025209								
917	T151-2017000920	T151V1017000920	##	00044568	NOVA RIMA CHRISTIN	CSH	7008969878	FT.ASMA JANEHA SAKTI SAKTI	10	PT BANK GMS	F703A/741TK	MHFZFA3CK3JH044611								
918	T151-2017000921	T151V1017000921	##	00057969	IMAM ARYA KUSUMA	CSH	7008989892	KAMILIA EDU, SE	09	ACC KWIANG	F35SM/T	MHM5EA3JH0069317								
919	T151-2017000922	T151V1017000922	##	00017577	INES GANESHA	CR4	7008981453	PT VENTURA INVESTASI UTAMA	10	GM65V2A/TEHMFPG8G5H0846461										
920	T151-2017000923	T151V1017000923	##	00052513	RENDY FERNANDO	CR5	7008986146	MAMTAR ROMANA SIAGIAN, SE	09	PT.MITSUBI	LHSP70QA/T00	MHFZ28H3H80034746								
921	T151-2017000924	T151V1017000924	##	00049393	SYARIEP HIDAYATULL	CR3	7008737016	CV AMBULANCE PINTAR INDONESIA	10	HD22M/T03C0R	JTFSS22PH0165364									
922	T151-2017000925	T151V1017000925	##	00053943	ROVERNANDO SIAHAN	CSH	7008974837	ROVERNANDO SIAHAN	09	ACC KWIANG	F35SM/T	MHM5EA3JH0232172								
923	T151-2017000926	T151V1017000926	##	00041477	ADY KURNIAMAN	CR3	7007110834	ADY KURNIAMAN	09	F54VM/T	MHM5EA3JH0230226									
924	T151-2017000927	T151V1017000927	##	00041176	LASEFRANOTO	CR5	7008943535	KAMIRAN	09	PT. TOYOTA	JF54VM/T	MHM5EA3JH0230394								
925	T151-2017000928	T151V1017000928	##	00054919	TITO PRABASTO	CR5	7008965581	PT NASUTA GRAHA TRIKENCANA	10	PT.BCA FINA	GM42GA/T02	MHFABSEH00064259								
926	T151-2017000929	T151V1017000929	##	00053332	RANI PERMATA SARI	CR3	7008983846	FERIANDI	09	PT.BCA FINA	GM65V2A/T	MHFPG8G5H08464518								
927	T151-2017000930	T151V1017000930	##	00038157	FELIX	CR2	7008989674	ANDREAS TJIA	09	PT.MITSUBI	LEA550A/TL1M	MRO53AK50H4509747								
928	T151-2017000931	T151V1017000931	##	00009793	SILVIA	CR1	7008990034	PT SRIREJEKI PERDANA STEEL	10	PT.BCA FINA	AL30GA/T	JNFGF3DH7H8010404								
929	T151-2017000932	T151V1017000932	##	00049393	SYARIEP HIDAYATULL	CR1	7008984688	RAHMIAH	09	PT. TOYOTA	JF54VA/T	MHM5FB4JH012414								
930	T151-2017000933	T151V1017000933	##	00009793	SILVIA	CSH	7008984666	PT CHINA HARBOUR INDONESIA	10	ASSVOA/TL1M	MRO53AK50H4509700									
931	T151-2017000934	T151V1017000934	##	00009793	SILVIA	CR3	7008987853	SILVIA	09	PT. TOYOTA	JVE30GA/T	JNFGF3DH6H8010619								
932	T151-2017000935	T151V1017000935	##	00017577	INES GANESHA	CR3	7008974160	FT EQUITY FINANCE INDONESIA	10	GM65V2A/TEHMFPG8G5H0846405										
933	T151-2017000936	T151V1017000936	##	00054963	ERIS SUSTIANANDI	CR2	7041007210	BERRY VONSIN BAY	09	PT.BCA FINA	TGH40GA/TL	MHFV8EM2H2334029								
934	T151-2017000937	T151V1017000937	##	00017577	INES GANESHA	CR5	7041004957	BOWO PRIYATMOKO	09	PT Toyota	AzGM65V2A/T	MHFPG8G5H0847301								
935	T151-2017000938	T151V1017000938	##	00017577	INES GANESHA	CR3	7008963986	PT CSM CORPORATAMA	10	F35SM/T	MHM5EA3JH0230298									
936	T151-2017000939	T151V1017000939	##	00053942	LIE YANCHE NUARI	WCSH	7008977431	AMBRIATY SIHOMBING	09	PT. TOYOTA	F35SM/T	MHM5EA3JH0232290								
937	T151-2017000940	T151V1017000940	##	00044568	NOVA RIMA CHRISTIN	CSH	7041009222	MARJI WASHI	09	PT Toyota	AF54VM/T	MHM5FA3JH0282972								
938	T151-2017000941	T151V1017000941	##	00049393	SYARIEP HIDAYATULL	CSH	7041015662	RISKI RISMANDI	09	F54VM/T	MHM5FA3JH013006									
939	T151-2017000942	T151V1017000942	##	00049145	VIVIELLA	CSH	7041002213	FT GRITA MUTIARA ABADI	10	PT. TOYOTA	JGM42VA/TL	MHFPG8G5H041355								
940	T151-2017000943	T151V1017000943	##	00017577	INES GANESHA	CSH	7008986246	PT CSM CORPORATAMA	10	F35SM/T	MHM5EA3JH0069862									
941	T151-2017000944	T151V1017000944	##	00017577	INES GANESHA	COP	7008985873	PT CSM CORPORATAMA	10	F35SM/T	MHM5EA3JH0069128									
942	T151-2017000945	T151V1017000945	##	00017577	INES GANESHA	CR5	7041014010	PT CSM CORPORATAMA	10	F35SM/T	MHM5EA3JH0069869									
943	T151-2017000946	T151V1017000946	##	00017577	INES GANESHA	CR5	7041013786	PT CSM CORPORATAMA	10	F35SM/T	MHM5EA3JH0069583									
944	T151-2017000947	T151V1017000947	##	00017577	INES GANESHA	CR5	7008980205	PT CSM CORPORATAMA	10	F35SM/T	MHM5EA3JH0069808									
945	T151-2017000948	T151V1017000948	##	00017577	INES GANESHA	CR5	7008954408	PT CSM CORPORATAMA	10	F35SM/T	MHM5EA3JH0069897									
946	T151-2017000949	T151V1017000949	##	00017577	INES GANESHA	CR1	7008947729	PT CSM CORPORATAMA	10	F35SM/T	MHM5EA3JH0069978									
947	T151-2017000950	T151V1017000950	##	00017577	INES GANESHA	CR5	7041009413	PT CSM CORPORATAMA	10	F35SM/T	MHM5EA3JH0070154									

Gambar 2. Dataset Transaksi Penjualan

klasterisasi kedalam 2 cluster dengan hasil 63 pelanggan di cluster 1 dan 39 pelanggan di cluster 2 [4].

Pada penelitian ini terdapat perbedaan dari penelitian-penelitian sebelumnya, pada penelitian oleh Ina Maryani dkk [4] dan penelitian oleh Ananthi Sheshasaayee dkk [11], pengelompokan segmentasi hanya menggunakan 2 cluster saja, sedangkan pada penelitian saat ini menggunakan 6 cluster mengacu pada teori karakteristik pelanggan, penentuan nilai centroid awal dilakukan secara acak sehingga menghasilkan performa clustering yang baik.

2. METODE PENELITIAN

Tahapan ini berisi penjelasan tentang jenis penelitian/ desain penelitian.

2.1. Metode Pengumpulan Data, Instrumen Penelitian dan Metode Pengujian

a. Metode Pengumpulan Data

Variabel-variabel instrumen yang akan digunakan dalam model RFM, terlihat pada tabel 1 berikut:

Tabel 4. Daftar Instrumen Penelitian

Variabel	Keterangan
Recency	Perbandingan waktu transaksi terakhir dengan waktu penelitian. Apabila jarak semakin dengan dengan waktu penelitian maka nilai R semakin besar.
Frequency	Jumlah banyaknya transaksi oleh pelanggan pada suatu waktu tertentu. Jika jumlah transaksi semakin banyak, nilai F semakin besar.
Monetary	Banyaknya uang yang dikeluarkan pelanggan dalam bertransaksi pada suatu waktu tertentu, jika uang yang dikeluarkan banyak maka nilai M

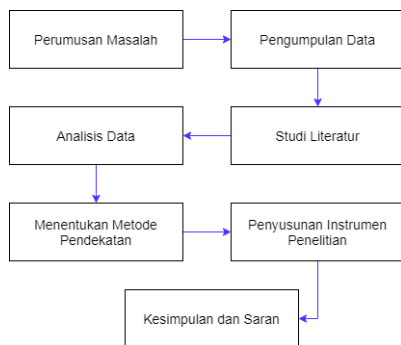
Variabel	Keterangan
	semakin besar.

c. Metode Pengujian

Pada tahap ini dilakukan pengujian *Cluster Distance Performance* dengan menggunakan Rapidminer [12]. Dimana pengujian ini akan menghitung kinerja jarak antar *cluster*, hasil pengujian menunjukkan nilai lebih kecil jarak antar *cluster* maka dinyatakan baik [13].

2.2. Tahapan penelitian

Pada tahapan penelitian dituangkan dalam bentuk bagan alur, terlihat pada gambar 3:



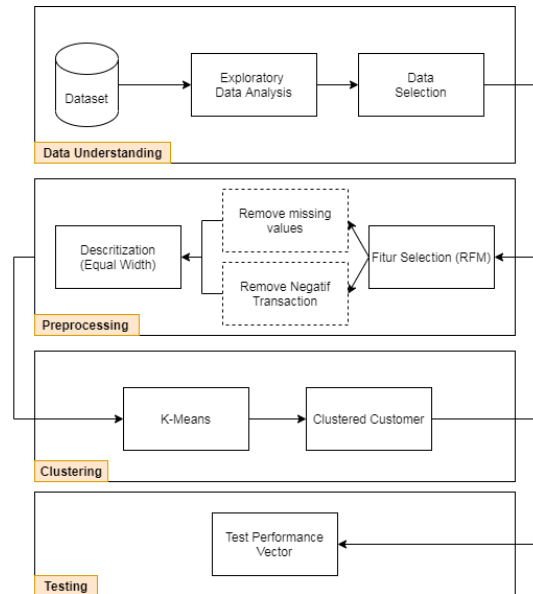
Gambar 3. Tahapan Penelitian

Pada tahapan penelitian dapat dijelaskan sebagai berikut:

- Perumusan Masalah**
Peneliti melakukan pencarian pokok permasalahan yang akan diteliti.
- Pengumpulan Data**
Pada pengumpulan data peneliti menggali data dan informasi dari berbagai sumber, dari teknik wawancara sampai dokumen pendukung lainnya seperti data transaksi penjualan.
- Studi Literatur**
Peneliti melakukan pencarian teori dan model dari berbagai sumber, seperti jurnal, buku, hasil karya ilmiah dan artikel lainnya yang memiliki kesamaan dalam pembahasan.
- Analisis Data**
Pada tahapan analisis data dapat dihasilkan dari penggalian data dan informasi. Dari hasil penggalian data dan informasi kemudian akan dijadikan bahan untuk menganalisis sesuai permasalahan yang ada.
- Penyusunan Instrumen Penelitian**
Instrumen penelitian diadopsi dari hasil penelitian sebelumnya yang telah dilakukan oleh [4] dan [1]
- Menentukan Metode Pendekatan**
Peneliti menentukan metode pendekatan yang cocok untuk bisa memecahkan permasalahan yang ada, yaitu menggunakan Model RFM dan *K-Means Clustering*
- Kesimpulan dan Saran**
Tahapan terakhir penelitian memberikan kesimpulan dan saran dari hasil penelitian yang dilakukan.

3. HASIL DAN PEMBAHASAN

Berikut adalah hasil dan pembahasan dari penelitian, dituangkan dalam bentuk diagram terlihat pada gambar 4 berikut:



Gambar 4. Alur Penelitian

Berikut adalah rincian penjelasan setiap alur prosesnya.

Data Understanding (Pemahaman Data)

Dataset yang digunakan dalam penelitian ini adalah *dataset* transaksi penjualan tahun 2017, 2018, dan 2019 pada PT. XYZ. Dengan jumlah transaksi penjualan 4332.

Preprocessing

Dari hasil penggalian data, terdapat data transaksi dengan bentuk struktur seperti pada tabel 2 berikut:

Tabel 5. Struktur Tabel Transaksi

No	Nama Filed	Keterangan
1	TglTransaksi Customer	Tanggal pada saat customer melakukan pembelian
2	NoFaktur	Nomor urut transaksi
3	KdCustomer	Kode Customer yang melakukan transaksi
4	NamaCustomer	Nama Customer yang melakukan transaksi pembelian
5	TypeKendaraan	Kode setiap kendaraan yang dibeli customer
6	KaroseriSpec	Nama kendaraan dan jenisnya
7	Qty	Jumlah beli
8	HargaUnit	Harga satuan kendaraan
9	TotalHarga	Jumlah harga hasil kali Qty dan Harga Unit

Selanjutnya mereduksi filed atau menghapus filed yang tidak digunakan dalam model RFM, sehingga terbentuk hasil yang terlihat pada tabel 3 berikut:

Tabel 6. Hasil Reduksi Filed

No	Nama Filed	Keterangan
1	TglTransaksi	Tanggal pada saat customer melakukan pembelian
2	KdCustomer	Kode customer yang melakukan transaksi
3	NamaCustomer	Nama customer yang melakukan transaksi
4	TotalHarga	Jumlah harga kendaraan yang dibeli

Berikut adalah bentuk data transaksi yang telah dilakukan reduksi filed atau penghapusan filed yang tidak dibutuhkan, terlihat pada tabel 7 berikut:

Tabel 7. Hasil Reduksi Data Transaksi

Tgl Transaksi	Kd Customer	Nama Customer	Total Harga
2019-01-02	2100004679	PT. SERASI AUTORAYA	227450000
2019-01-02	5300024425	ANDRI JAUHARI	176450000
2019-01-03	5300127855	PT CSM CORPORATA MA	162000000
2019-01-04	5301143940	PT.TIMAS SUPLINDO	191100000
2019-01-07	5302006658	AJB BUMI PUTERA 1912	329000000
...
2019-04-30	7042449321	BUDI HARTONO	452250000

Selanjutnya dilakukan transformasi data kedalam model RFM, yang dituangkan pada tabel 5 berikut:

Tabel 8. Hasil Transformasi Data ke dalam Model RFM

Kd Customer	Nama Customer	R	F	M
2100004679	PT. SERASI AUTORAYA	557	10	6493400000
5300024425	ANDRI JAUHARI	818	1	297800000
5300127855	PT CSM CORPORATA MA	245	327	519826950000
5301143940	PT.TIMAS	802	4	4047000000

Kd Customer	Nama Customer	R	F	M
SUPLINDO				
5302006658	AJB BUMI PUTERA 1912	102	1	1112600000
...
7042449321	BUDI HARTONO	1	245	191.100.000

Berikutnya dari hasil proses reduksi dilakukan proses *Descritization (Equal Width)*, terlihat pada tabel 9:

Tabel 9. Hasil *Descrititation*

No	Kd Customer	Nama Customer	R	F	M
1	2100004679	PT. SERASI AUTORAYA	1	2	5
2	5300024425	ANDRI JAUHARI	1	1	4
3	5300127855	PT CSM CORPORATA MA	2	5	5
4	5301143940	PT.TIMAS SUPLINDO	1	2	5
5	5302006658	AJB BUMI PUTERA 1912	1	1	5
..
4332	7042449321	BUDI HARTONO	2	1	2

Clustering

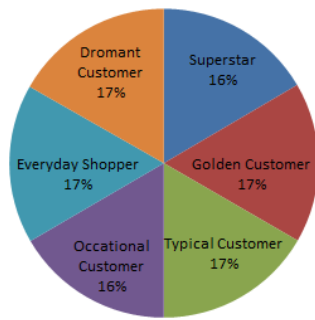
Tahapan ini menggunakan algoritma *K-means Clustering* dengan aplikasi Rapidminer, berikut hasil *clustering* terlihat pada gambar 5:



Gambar 5. Hasil Klustering *K-Means*

Hasil *cluster model* kemudian dimasukkan kedalam tabel kelas pelanggan, untuk mengelompokan pelanggan ke dalam 6 *Cluster*, yaitu C5 = *Dormant Customer*, C4 = *Everyday shopper*, C3 = *Occasional Customer*, C2 = *Typical Customer*, C1 = *Golden Customer*, C0 =

Superstar, sehingga menghasilkan presentasi kelompok seperti pada gambar 10 berikut ini:



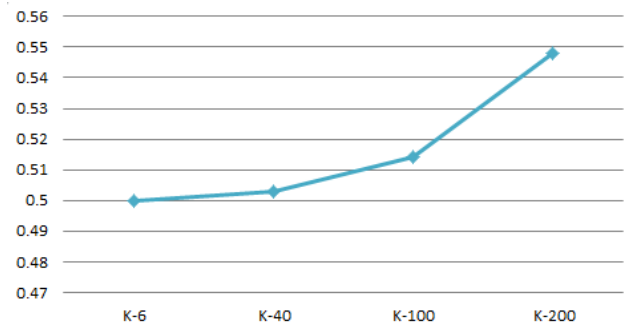
Gambar 6. Persentase Berdasarkan Kelas Pelanggan

Tabel 10. Kelas Pelanggan

Kelas Pelanggan	Jumlah Pelanggan	Keterangan
<i>Superstar</i>	314	<ul style="list-style-type: none"> • Pelanggan memiliki loyalitas yang tinggi. • Pelanggan yang mengeluarkan banyak uang dalam melakukan transaksi. • Pelanggan yang memiliki kedekatan rentang waktu dengan penelitian • Memiliki jumlah transaksi paling banyak.
<i>Golden Customer</i>	319	<ul style="list-style-type: none"> • Pelanggan yang mengeluarkan banyak uang tertinggi kedua • Memiliki kedekatan dengan rentang waktu dengan penelitian • Memiliki nilai rata-rata transaksi
<i>Typical Customer</i>	316	Memiliki nilai rata-rata uang dan rata-rata transaksi
<i>Occasional Customer</i>	315	<ul style="list-style-type: none"> • Memiliki jumlah uang rendah • Memiliki nilai rentang waktu paling rendah • Memiliki transaksi tertinggi
<i>Everyday shopper</i>	316	<ul style="list-style-type: none"> • Mempunyai peningkatan pada jumlah transaksi • Memiliki nilai transaksi rendah • Memiliki jumlah uang sedang sampai dengan rendah
<i>Dormant Customer</i>	318	<ul style="list-style-type: none"> • Memiliki jumlah transaksi dan jumlah uang yang paling rendah • Nilai rentang waktu yang paling rendah

Testing (Test Performance Cluster)

Berikut gambar 7. grafik hasil uji performa *vector* :



Gambar 7. Hasil Uji Performa

Hasil uji kinerja *cluster* menunjukkan bahwa pengelompokan dengan menggunakan 6 *cluster* menghasilkan nilai *Davies Bouldin*: 0.500, sedangkan pengelompokan dengan menggunakan lebih dari 40 *cluster* menghasilkan nilai *Davies Bouldin*: > 0.500. sehingga dapat disimpulkan bahwa dengan menggunakan 6 *cluster* memiliki kinerja *cluster* terbaik dengan menggunakan jumlah data 1898.

4. KESIMPULAN

Berdasarkan uraian sebelumnya dan hasil dari simulasi data, maka pada penelitian ini dapat ditarik kesimpulan diantaranya:

Dari hasil penggalian data penjualan tahun 2017, 2018, dan 2019 dihasilkan segmentasi pelanggan berdasarkan kelas pelanggan, *Dormant Customer* sebanyak 318, *Everyday Shopper* sebanyak 316, *Occasional Customer* sebanyak 315, *Typical Customer* sebanyak 316, *Golden Customer* sebanyak 319 dan *Super Start* sebanyak 314.

DAFTAR PUSTAKA

[1] M. Tavakoli, M. Molavi, V. Masoumi, M. Mobini, S. Etemad, and R. Rahmani, "Customer Segmentation and Strategy Development Based on User Behavior Analysis, RFM Model and Data Mining Techniques: A Case Study," *Proc. - 2018 IEEE 15th Int. Conf. E-bus. Eng. ICEBE 2018*, pp. 119–126, 2018, doi: 10.1109/ICEBE.2018.00027.

[2] D. S. AH Kracklauer, DQ Mills, "Customer Management as The Origin of Collaborative Customer Relationship Management," 2004.

[3] B. E. Adiana, I. Soesanti, and A. E. Permanasari, "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.

[4] I. Maryani, D. Riana, R. D. Astuti, A. Ishaq, Sutrisno, and E. A. Pratama, "Customer Segmentation Based on RFM Model and Clustering Techniques with K-Means Algorithm," *Proc. 3rd Int. Conf. Informatics Comput. ICIC 2018*, pp. 1–6, 2018, doi: 10.1109/IAC.2018.8780570.

[5] A. A. Fajrin, A. Maulana, T. Informatika, U. P.

- Batam, and J. R. Soeprapto, "Penerapan Data Mining untuk Analisis Pola Pembelian Konsumen dengan Algoritma *Fp-Growth* pada Data Transaksi Penjualan *Spare Part* Motor," *J. Ilmu Komput. Vol. 05, No.01*, vol. 05, no. 01, pp. 27–36, 2018.
- [6] C. D. Rumiarti, I. Budi, P. Studi, M. Teknologi, F. I. Komputer, and U. Indonesia, "Segmentasi Pelanggan pada *Customer Relationship Management* di Perusahaan Ritel: Studi Kasus Pt Gramedia Asri Media," *Jurnal Sistem Informasi*, Vol. 13, No. 1, pp. 1-10, 2017. DOI: <http://dx.doi.org/10.21609/jsi.v13i1.525>
- [7] Sudriyanto, "Clustering Loyalitas Pelanggan dengan Metode RFM (*Recency, Frequency, Monetary*) dan Fuzzy C-Means," *Pros. SNATIF Ke-4*, pp. 815–822, 2017.
- [8] M. L. Sibuea and A. Safta, "Pemetaan Siswa Berprestasi menggunakan Metode *K-Means Clustering*," *Jurteks*, vol. 4, no. 1, pp. 85–92, 2017, doi: 10.33330/jurteks.v4i1.28.
- [9] F. Hadi, M. Mustakim, D. O. Rahmadia, F. H. Nugraha, N. P. Bulan, and S. Monalisa, "Penerapan *K-Means Clustering* berdasarkan RFM Mofek sebagai Pemetaan dan Pendukung Strategi Pengelolaan Pelanggan (Studi Kasus: PT. Herbal Penawar Alwahidah Indonesia Pekanbaru)," *J. Sains dan Teknol. Ind.*, vol. 15, no. 1, pp. 69–76, 2017.
- [10] M. Reza and T. Akter, "Segmentation of Mobile Customers using Data Mining Techniques," *IJERT Vol.7 Issue 10*, vol. 7, no. 10, pp. 251–255, 2018.
- [11] A. Sheshasaayee and L. Logeshwari, "Implementation of Clustering Technique Based RFM Analysis for Customer Behaviour in Online Transactions," *Proc. 2nd Int. Conf. Trends Electron. Informatics, ICOEI 2018*, no. Icoei, pp. 1166–1170, 2018, doi: 10.1109/ICOEI.2018.8553873.
- [12] R. Gustriansyah, N. Suhandi, and F. Antony, "Clustering Optimization in RFM Analysis Based On *K-Means*," *Indones. J. Electr. Eng. Comput. Sci.*, Vol. 18, No. 1, pp. 470–477, 2019. doi: 10.11591/ijeecs.v18.i1.pp470-477.
- [13] T. A. A. Sandi, M. Raharjo, J. L. Putra, and R. Ridwan, "Clustering Kesetiaan Pelanggan dengan Model RFM (*Recency, Frequency, Monetary*) dan *K-Means*," *J. Pilar Nusa Mandiri*, vol. 14, no. 2, p. 239, 2018. doi: 10.33480/pilar.v14i2.950.