

Metode K-Means *Clustering* dengan Atribut RFM untuk Mempertahankan Pelanggan

<http://dx.doi.org/10.28932/jutisi.v4i3.878>

Stephen Aprius Sutresno^{#1}, Ade Iriani^{#2}, Eko Sedyono^{#3}

[#]Magister Sistem Informasi, Universitas Kristen Satya Wacana
Jl. Diponegoro 52-60 Salatiga

¹stephen.aprius@gmail.com

²ade.iriiani@staff.uksw.edu

³eko@uksw.edu (corresponding)

Abstract — The effort to keep customers is one of the important CRM strategies in each business that can increase profits for the company. The XYZ workshop which became the case study in this study focused more on attracting customers than implementing customer retention strategies. The aim of this research was to analyze customer transaction data in the XYZ workshop using the K-Means clustering method with RFM attributes to classify customers and determine appropriate strategies to retain customers. This research was conducted using a descriptive research method with a quantitative approach, whereas data analysis was carried out through the stages of data selection, preprocessing, transformation, processing and continued with RFM strategy analysis. The results of this study obtained 5 clusters with different strategies according to the RFM score obtained. This strategy can be used by XYZ workshop as a strategy to retain customers to provide more benefits.

Keywords — CRM, Customer retention, K-Means clustering, RFM.

I. PENDAHULUAN

Pada segala bidang bisnis, upaya meningkatkan pertumbuhan usaha sangat penting. Pelanggan menjadi salah satu faktor utama dalam upaya meningkatkan pertumbuhan ekonomi. Relasi terhadap pelanggan diperlukan dalam meningkatkan kualitas pelanggan sehingga memberikan keuntungan yang lebih dalam perusahaan [1].

Customer Relationship Management (CRM) adalah sebuah strategi dalam membangun relasi antara penjual dengan pelanggan [2], [3]. Tingkat kesuksesan CRM ini bergantung dari tingkat keseimbangan antara 3 sumber daya dalam perusahaan, yaitu pelanggan, teknologi dan proses. Suksesnya CRM dapat memberikan dampak antara lain meningkatkan kepuasan pelanggan, kesuksesan mempertahankan pelanggan dan loyalitas pelanggan [4]. Strategi CRM ini baiknya diterapkan sebagai inti bisnis pada tiap tempat usaha, karena pentingnya untuk membangun interaksi terhadap pelanggan dan memenuhi

kebutuhan pelanggan [2].

Salah satu tempat usaha bengkel XYZ yang berdiri sejak tahun 2013, terhitung lebih dari 22.000 pelanggan yang telah melakukan transaksi. Bengkel XYZ dalam menjalankan bisnisnya lebih berfokus untuk menarik pelanggan baru dan belum menerapkan strategi untuk mempertahankan pelanggan. Sedangkan pada dasarnya usaha mempertahankan pelanggan membutuhkan biaya lebih murah dibandingkan dengan menarik pelanggan baru [5]. Selain itu, upaya mempertahankan pelanggan dapat memberikan dampak yang signifikan terhadap peningkatan keuntungan perusahaan. Hal ini dikarenakan: 1) mendapatkan pelanggan baru memerlukan biaya yang cukup signifikan dan membutuhkan beberapa tahun untuk mengubah pelanggan baru menjadi pelanggan yang menguntungkan; 2) meningkatnya kepuasan dan kepercayaan pelanggan dalam hubungan kepada pemasok, mereka akan memberikan proporsi yang lebih besar pula dalam bisnisnya yang dikenal istilah “*share of wallet*”; 3) hubungan yang terus berkembang akan memberikan kolaborasi yang lebih besar dan menurunkan biaya operasi; 4) pelanggan yang memiliki tingkat kepuasan tinggi akan semakin besar kemungkinan pelanggan dapat mempromosikan kepada orang lain, sehingga menurunkan biaya untuk menarik pelanggan baru; 5) pelanggan setia memiliki sensitivitas yang lebih rendah ketika terjadi kenaikan harga [6].

Terdapat beberapa strategi mempertahankan pelanggan yang dapat dilakukan oleh pebisnis, salah satunya menggunakan *analytical CRM*. *Analytical CRM* adalah salah satu tipe CRM yang paling kritis karena dapat memberikan pengetahuan mengenai pelanggan dengan menganalisis *data* histori transaksi dan *data* mengenai latar belakang pelanggan [7]. *Analytical CRM* pada penelitian ini diterapkan melalui analisis *data* transaksi pelanggan di bengkel XYZ selama 3 tahun (2015 sampai 2017) menggunakan metode K-Means *clustering*.

K-Means *clustering* adalah sebuah proses untuk membagi objek-objek yang memiliki kemiripan menjadi satu kelompok [8]. Pada penelitian ini *clustering* pelanggan didasarkan pada nilai dari 3 atribut: *recency*, *frequency* dan *monetary* (RFM). Ketiga atribut tersebut dapat digunakan untuk melihat nilai pelanggan dalam menentukan tingkat loyalitas pelanggan [9].

Berdasarkan deskripsi di atas, penelitian ini menerapkan metode K-Means *clustering* dengan atribut RFM untuk menganalisis *data* transaksi pelanggan di bengkel XYZ, sehingga menghasilkan pengelompokan pelanggan yang memiliki kemiripan. Hasil dilanjutkan dengan pengkategorian pelanggan berdasarkan skor RFM yang didapatkan dan penentuan strategi yang cocok dalam mempertahankan pelanggan, mengingat pentingnya mempertahankan pelanggan di suatu perusahaan untuk peningkatan keuntungan perusahaan.

II. LITERATUR REVIEW

A. Penelitian Terdahulu

Penelitian yang dilakukan oleh Syakur dkk [10], menerapkan K-Means *clustering* menggunakan beberapa atribut profil dari tiap pelanggan yaitu umur, profesi, pendapatan, edukasi, kualitas dan jenis kelamin. Objek penelitian yang digunakan adalah *data* pelanggan di salah satu toko batik selama 1 bulan yang didapatkan melalui kuesioner sebanyak 500 *data* pelanggan. Penelitian membandingkan dan menghitung performa penggunaan metode *Elbow* dalam menentukan jumlah *cluster* (K) pada 100 *data* dan 300 *data*. Hasil menunjukkan K yang sama yaitu 3 *cluster* pada 100 *data* dan 300 *data*, selain itu juga menghasilkan performa yang baik dari penggunaan metode *Elbow*.

Penelitian lain yang dilakukan oleh Rajeh dkk [11], membahas pengelompokan pelanggan berdasar nilai atribut RFM. Pengelompokan dilakukan menggunakan metode K-Means *clustering* yang dilanjutkan dengan menentukan tingkat loyalitas pelanggan berdasarkan skor RFM. Loyalitas pelanggan dibagi menjadi 4 tingkatan yaitu loyal, *disloyal*, *spurious* loyal dan *latent* loyal. Bank kemudian mengimplementasikan layanan *card payment* untuk melihat perpindahan loyalitas dan kenaikan nilai *monetary* selama waktu 1 bulan. Hasil menunjukkan kombinasi loyalitas sikap dan loyalitas perilaku pelanggan memberikan tingkat loyalitas lebih tinggi, maka didapatkan kenaikan nilai *monetary* selama waktu 1 bulan.

Penelitian ini menerapkan metode K-Means *Clustering* dengan atribut RFM yang mengacu pada hasil penelitian oleh Rajeh, skor RFM ini digunakan untuk menentukan tingkatan loyalitas pelanggan sebagai penentuan strategi dalam mempertahankan pelanggan. Mengacu pada hasil penelitian oleh Syakur, penelitian ini juga menggunakan metode *Elbow* untuk menentukan K *optimum*, guna mendapatkan hasil *clustering* yang baik. Perbedaan dengan

penelitian sebelumnya adalah objek *data*, metode perhitungan jarak di dalam K-Means *clustering* menggunakan rumus jarak Manhattan *Distance* dan teknik penentuan skor RFM menggunakan penghitungan kuintil.

B. Customer Relationship Management (CRM)

CRM mulai populer digunakan sebagai pengatur strategi pemasaran sekitar akhir tahun 1990. CRM juga sering dikenal dengan manajemen pelanggan, yaitu pendekatan bisnis untuk membuat, mempertahankan dan meningkatkan relasi terhadap pelanggan dengan tujuan meningkatkan nilai pelanggan dan memaksimalkan keuntungan perusahaan. Pemanfaatan teknologi informasi menjadi salah satu ikatan terkuat CRM dalam mengimplementasikan strategi pemasaran [6].

CRM dibagi menjadi 3 tipe:

- *operational* CRM – tipe CRM yang memiliki interaksi terhadap pelanggan secara langsung. Otomatisasi adalah kunci utama dalam tipe CRM ini, dimana semua bisnis penjualan, bisnis pemasaran dan layanan pelanggan dibuat secara otomatis oleh sistem,
- *analytical* CRM – melibatkan pengumpulan, penyimpanan, pengaturan, analisis, interpretasi dan penggunaan *data* yang didapatkan dari sisi operasional bisnis,
- *collaborative* CRM – melibatkan penggunaan layanan dan infrastruktur untuk membuat interaksi antara perusahaan dan beberapa sumber lainnya seperti pemasok. Hal ini memungkinkan interaksi antara perusahaan, karyawan dan pelanggan.

Pada penelitian ini digunakan *analytical* CRM untuk menganalisis *data* histori transaksi pelanggan. Penggunaan *data mining* sebagai salah satu *tools analytical* CRM untuk membantu dalam analisis *data*.

C. Loyalitas Pelanggan

Strategi pemasaran dalam tingkat bisnis ke pelanggan dilakukan pembagian atau pengelompokan berdasarkan beberapa kategori yaitu *geographic*, *demographic and socioeconomic*, *psychographic*, *benefit*, *usage*, *loyalty* dan *occasion* [6]. Loyalitas pelanggan menunjukkan seberapa kuat hubungan antara pelanggan dengan pemasok. Loyalitas pelanggan tidak muncul secara tiba-tiba, namun dibangun oleh pemasok melalui pendekatan yang berpusat pada produk dan layanan yang menjadi keinginan pelanggan. Pelanggan yang memiliki loyalitas tinggi ini akan berpikir dua kali untuk membeli dari pemasok lainnya [12].

Kualitas layanan, kualitas produk, strategi harga dan atribut penjualan adalah 4 variabel utama yang mempengaruhi loyalitas pelanggan. Pencapaian keempat faktor tersebut dibutuhkan strategi yang kompleks. Strategi yang dapat dilakukan adalah dengan melihat histori transaksi pelanggan [13].

D. K-Means Clustering

Clustering adalah salah satu metode *data mining* yang digunakan untuk mengelompokkan *data* menjadi bagian-bagian yang memiliki kemiripan. Metode K-Means merupakan satu tipe *clustering* dengan konsep membagi n objek atau *data* menjadi K cluster [14]. Output dari metode K-Means ini bergantung pada jumlah K dan posisi titik *centroid* yang ditentukan di awal [10].

Proses untuk metode K-Means *clustering* seperti berikut:

- *input*:
 - K: jumlah *cluster*
 - Ci: nilai *centroid* awal
 - D: *data set* mengandung n objek
- *output*:
 - Sejumlah *set* dari K cluster
- langkah:
 - a) Masukkan K cluster
 - b) Masukkan nilai *centroid* C untuk setiap cluster
 - c) Hitung jarak setiap *data* objek ke setiap *centroid* cluster
 - d) Posisikan *data* objek ke cluster terdekat
 - e) Ubah nilai *centroid* dengan menghitung nilai rata-rata dari semua objek untuk setiap cluster
 - f) Ulangi langkah c ke e, sampai tidak ada *data* objek yang berpindah di setiap K cluster

E. Deteksi Outlier

Outlier adalah *data* anomali yang mungkin muncul akibat kesalahan mekanis, perubahan perilaku sistem, perilaku curang dan *human errors*. Deteksi outlier penting dalam analisis *data mining*, salah satunya untuk K-Means *clustering*. Melihat teknik K-Means *clustering* yang mengelompokkan *data* berdasarkan jarak terdekat, maka outlier menjadi permasalahan dalam K-Means *clustering* yang dapat mempengaruhi hasil akhir [15]. Salah satu metode deteksi outlier yang berdasarkan K-Means cocok dalam *data* ukuran yang besar [16], sehingga digunakan metode deteksi outlier berdasarkan jarak Manhattan *Distance* seperti berikut:

1. mencari jarak setiap *data* menggunakan rumus Manhattan *Distance*,
2. menghitung nilai *threshold* yaitu yang didapatkan melalui rumus: nilai *threshold* = (jarak maksimum + jarak minimum) / 2,
3. jika jarak > nilai *threshold* maka *data* dianggap sebagai outlier dan dihapus, sedangkan jika jarak \leq nilai *threshold* maka *data* dianggap bukan outlier dan tetap dapat digunakan.

Melakukan tahapan deteksi dan penghapusan *data outlier* dapat memberikan hasil akurasi dan *silhouette index* yang lebih tinggi dalam K-Means *clustering*, dibandingkan melakukan analisis tanpa menghapus outlier [15].

F. Recency, Frequency and Monetary (RFM)

RFM adalah sebuah teknik analisis yang mengidentifikasi perilaku pelanggan dan merepresentasikan karakteristik

perilaku pelanggan dengan tiga atribut: 1) *Recency*: mengacu pada waktu terakhir pembelian dengan waktu yang sedang berjalan sekarang; 2) *Frequency*: mengacu pada jumlah transaksi yang telah dilakukan oleh pelanggan; 3) *Monetary*: mengacu pada jumlah uang yang telah dikeluarkan oleh pelanggan dalam semua transaksi [11].

RFM sering digunakan sebagai teknik pengelompokkan pelanggan dilihat dari kombinasi nilai tiga *digit* dari tiap atribut, dimana tiap atribut memiliki nilai berdasar kuintil dengan *interval* 1 sampai 5 yang pembagiannya sama rata. Nilai yang diberikan pada tiap atribut menunjukkan baik atau buruknya skor pelanggan dengan nilai 5 yang terbaik dan 1 yang terburuk, maka akan terbentuk kombinasi angka seperti 555, 554, 553, ..., 111, dengan total kombinasi 125 ($5 \times 5 \times 5$) skor RFM [17].

III. METODE PENELITIAN

Penelitian ini menggunakan metode penelitian deskriptif dengan pendekatan kuantitatif, yang dilakukan dalam beberapa tahapan: 1) identifikasi masalah; 2) pengumpulan *data*; 3) *literatur review*; 4) analisis *data*; 5) analisis strategi RFM. Tahapan identifikasi masalah dan pengumpulan *data* dilakukan melalui observasi dan wawancara kepada salah satu pengurus bengkel XYZ. *Data* yang didapatkan adalah *data* transaksi dan pelanggan dari tahun 2015 sampai 2017 dalam bentuk *database* Ms.Access. Tahapan *literatur review* dilakukan dengan mempelajari penelitian sebelumnya dan mencari teori-teori pendukung dilakukan. Teori-teori yang dipelajari mengenai CRM, mempertahankan pelanggan, loyalitas pelanggan, K-Means *clustering* dan RFM.

Tahapan analisis *data* dilakukan dengan beberapa langkah dari proses *data mining*: 1) *data selection*: memilih *data* yang akan dianalisis dari *database* Ms.Access dan didapatkan *data* sejumlah 15.913 *data* pelanggan, 40.361 *data* transaksi dan 76.214 *data* transaksi detail; 2) *preprocessing*: dilakukan *query* untuk mendapatkan baris *data* yang dibutuhkan untuk langkah *processing*, dilanjutkan *data cleaning* untuk *data outlier* menggunakan algoritma berdasarkan jarak, sehingga didapatkan *data* akhir sejumlah 14.979 *data* pelanggan, 30.774 *data* transaksi dan 56.208 *data* transaksi detail; 3) *transformation*: *data* yang dihasilkan ditampung ke dalam format excel; 4) *processing*: pemrosesan *data* menggunakan tools IBM SPSS versi 16.0 dengan metode K-Means *clustering*. Pada tahapan analisis strategi RFM, hasil analisis *data* dilanjutkan dengan penghitungan skor RFM untuk tiap cluster dan penentuan strategi yang cocok dari skor RFM yang telah didapatkan.

IV. HASIL DAN PEMBAHASAN

Setelah dilakukan proses *data selection*, *preprocessing* dan transformasi *data mining*, didapatkan *data* pelanggan yang terlihat pada Tabel I.

TABEL I
DATA PELANGGAN BENGKEL XYZ

ID Pelanggan	Recency	Frequency	Monetary
85	1	1	51.000
97	833	3	239.000
105	702	3	276.000
74	16	2	150.000
120	493	2	47.000
.....

Tabel I adalah *data set* pelanggan yang digunakan untuk pemrosesan K-Means *clustering*. *Data set* berjumlah 14.979 baris dan diambil dari *data* tahun 2015, 2016 dan 2017. *Recency* menunjukkan waktu pembelian terakhir pelanggan yang dibandingkan tanggal 1 Januari 2018 dengan satuan hari. *Frequency* menunjukkan jumlah transaksi yang telah dilakukan pelanggan dihitung dari tanggal 1 Januari 2015 sampai dengan 31 Desember 2017. *Monetary* merupakan total uang yang telah dikeluarkan pelanggan dari semua transaksi dihitung dari tanggal 1 Januari 2015 sampai dengan 31 Desember 2017.

Proses berikutnya adalah melakukan normalisasi *data*. Normalisasi dilakukan karena skala perbandingan nilai dari ketiga atribut *recency*, *frequency* dan *monetary* berbeda jauh. Perbedaan skala menjadi salah satu masalah untuk penghitungan jarak *data* ke titik *centroid* di K-Means *clustering*, dimana skala terbesar akan mendominasi lainnya [18]. Normalisasi dilakukan menggunakan *Z-Score normalization*, yang berdasarkan nilai rata-rata dan standar deviasi [19]. Rumus untuk penghitungan *Z-Score normalization* dapat dilihat pada Rumus 1.

$$d' = \frac{d - \text{mean}(p)}{\text{std}(p)} \quad \dots\dots(1)$$

dimana:

d = *data*

p = atribut *data*

mean(p) = nilai rata-rata di atribut p

std(p) = nilai standar deviasi di atribut p

Perhitungan normalisasi menggunakan Rumus 1 untuk semua atribut dari *data set* pelanggan menghasilkan *data* yang terlihat pada Tabel II.

TABEL III
HASIL NORMALISASI

ID Pelanggan	ZRecency	ZFrequency	ZMonetary
85	-0,66568	-0,49490	-0,48483
97	2,27653	0,14935	-0,00149
105	1,81327	0,14935	0,09363

74	-0,61264	-0,17277	-0,23030
120	1,07418	-0,17277	-0,49511
.....

Tabel II adalah hasil dari normalisasi *data* untuk setiap atribut *recency*, *frequency* dan *monetary* menggunakan *Z-Score normalization*. Terlihat jelas perbedaan skala nilai dari Tabel I dan Tabel II, selain itu juga nilai yang semula bilangan bulat berubah menjadi bilangan pecahan.

Langkah selanjutnya yang dilakukan setelah normalisasi adalah *clustering data*. Penentuan metode penghitungan jarak ke titik *centroid* diperlukan untuk proses *clustering*. Terdapat banyak metode penghitungan jarak seperti Minkowski, Manhattan *distance*, Average *distance*, Euclidean *distance*, Cosine *deasure* dan masih banyak lagi. Metode penghitungan jarak yang cocok untuk diaplikasikan dalam K-Means *clustering* dengan *data* numerik adalah Euclidean *distance* dan Manhattan *distance* [18], [20]. Pada penelitian ini digunakan Manhattan *distance* untuk metode penghitungan jarak karena memiliki akurasi dan performa lebih unggul jika dibandingkan dengan Euclidean *distance* [18]. Rumus untuk metode Manhattan *distance* dapat dilihat pada Rumus 2.

$$d_{man} = \sum_{i=1}^n (x_i - y_i) \quad \dots\dots(1)$$

dimana:

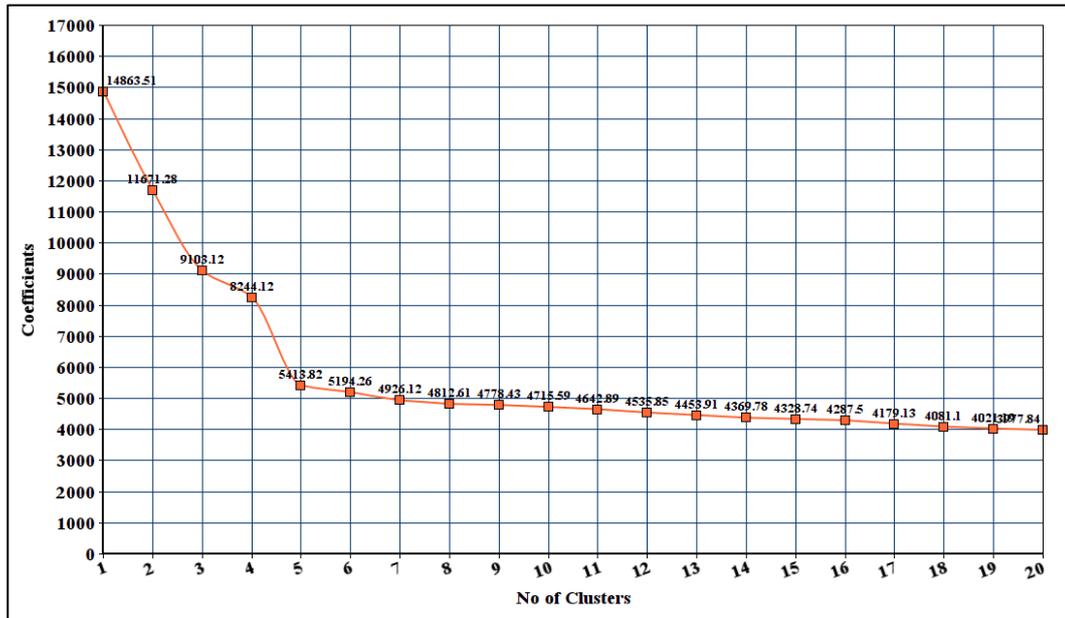
n = jumlah atribut *data*

i = atribut *data*

x_i = nilai *data* di atribut i

y_i = nilai *centroid* di atribut i

Pada metode K-Means *clustering*, nilai K dan nilai *centroid* awal menentukan hasil akhir dari *clustering*. Diperlukan sebuah pendekatan untuk menentukan kedua nilai tersebut sehingga memberikan hasil *clustering* yang baik [10]. Nilai K merupakan penentu jumlah *cluster* atau jumlah pengelompokan *data*. Pendekatan untuk penentuan nilai K *optimum* dapat dilakukan dengan metode *Elbow*. Metode *Elbow* adalah metode visualisasi yang melihat perubahan nilai koefisien dari variansi K. Iterasi dimulai dari nilai K=2 dan menambah sebanyak 1 setiap iterasinya, dimana setiap iterasi akan dilakukan perhitungan K-Means *clustering*. Hasil nilai koefisien setiap iterasi dibandingkan dan dicari selisih nilai yang memiliki perubahan drastis dari tinggi ke rendah, sehingga jika digambarkan dengan grafik akan terlihat seperti sebuah siku [21]. Hasil dari pendekatan menentukan nilai K dengan metode *Elbow* dapat dilihat pada Gambar 1.



Gambar 1. Grafik Hasil Metode Elbow

Terlihat pada Gambar 1, *cluster* nomor 5 memiliki selisih koefisien yang memiliki perubahan drastis dari tinggi ke rendah, juga secara visual terlihat sebagai titik pusat siku. Disimpulkan bahwa nilai *K optimum* untuk *data* yang digunakan dalam penelitian ini adalah 5. Selanjutnya ditentukan nilai *centroid* awal yang dilakukan melalui metode berikut:

1. hitung nilai rata-rata dari semua objek *data*, maka nilai tersebut sebagai nilai titik awal *centroid* ($K=1$),
2. hitung jarak antara nilai *data* dengan semua nilai *centroid* yang telah terpilih, kemudian pilih *data* dengan jarak terjauh untuk dijadikan sebagai *centroid* berikutnya ($K=2, 3, 4$, dst),
3. ulangi langkah 2 untuk mendapatkan nilai *centroid* sejumlah K .

Pendekatan menentukan nilai *centroid* awal dengan metode tersebut memiliki performa lebih baik jika dibandingkan dengan nilai *centroid* awal yang ditentukan secara acak, karena metode tersebut memiliki tingkat akurasi yang lebih tinggi dan jumlah iterasi lebih sedikit [22]. Hasil dari pendekatan menentukan nilai *centroid* awal dengan metode tersebut dapat dilihat pada Tabel III.

TABEL III
NILAI *CENTROID* AWAL

Cluster	ZRecency	ZFrequency	ZMonetary
1	-0,14052	-0,15522	-0,15211
2	2,27299	2,40427	2,49745
3	-0,66568	-0,49490	-0,61594
4	2,43566	2,40427	2,21208
5	-0,66568	-0,49490	-0,61337

Setelah mendapatkan semua variabel *input* yang dibutuhkan untuk *K-Means clustering* yaitu *data set* pada Tabel II, jumlah *cluster* $K = 5$ dan nilai *centroid* awal pada Tabel III, dilanjutkan dengan tahap akhir dari *processing data mining*. Semua variabel *input* yang telah disiapkan dimasukkan dan diproses menggunakan *tools* IBM SPSS versi 16.0, yang menghasilkan nilai *centroid* akhir yang dapat dilihat pada Tabel IV.

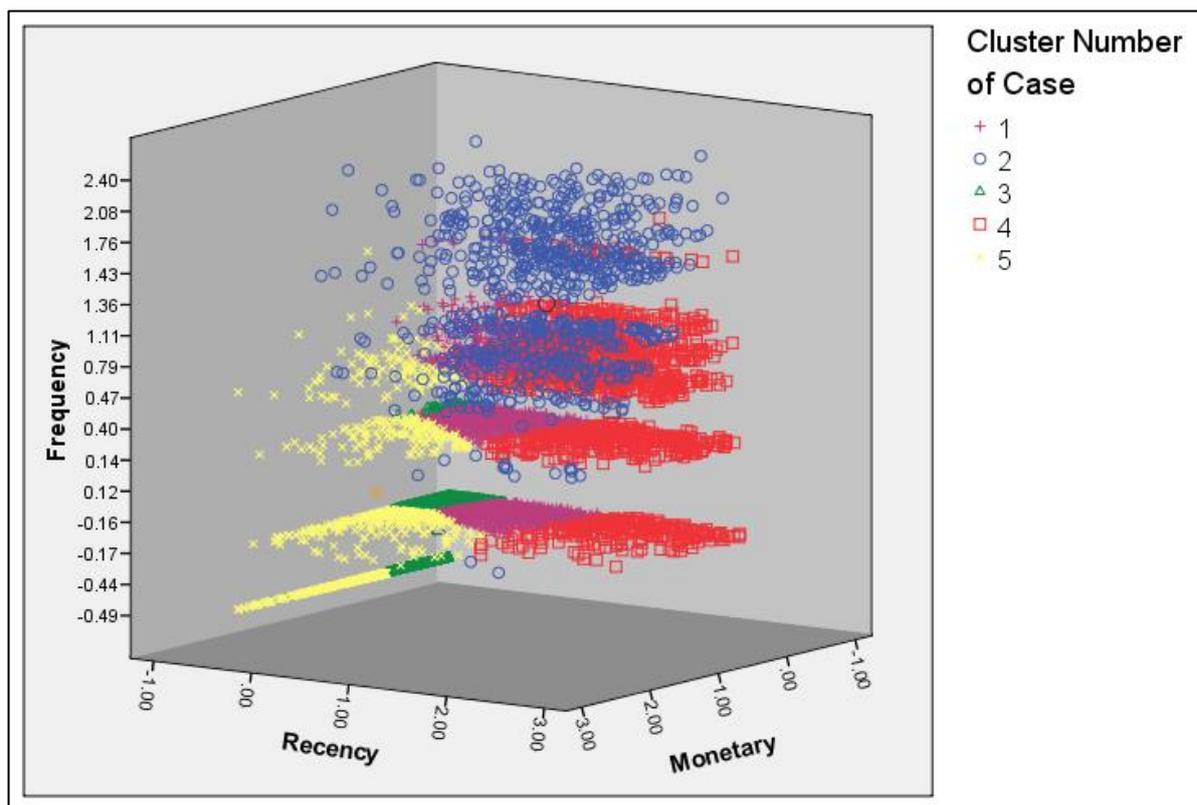
TABEL IV
NILAI *CENTROID* AKHIR

Cluster	ZRecency	ZFrequency	ZMonetary
1	0,27357	0,12196	-0,14784
2	1,59190	1,36840	1,17737
3	-0,63118	-0,44696	-0,41611
4	1,52449	0,40366	0,13440
5	-0,37712	-0,16832	0,85532

Nilai *centroid* akhir yang terlihat pada Tabel IV merupakan nilai *centroid* yang didapatkan pada iterasi ke-39. Hasil juga menunjukkan pembagian 14.979 *data* pelanggan dalam kelompok tiap *cluster* seperti terlihat pada Tabel V.

TABEL V
JUMLAH PELANGGAN TIAP *CLUSTER*

Cluster	Jumlah
1	2.470
2	830
3	9.333
4	1.406
5	940



Gambar 2. Diagram Tebar 3D Clustering Data Pelanggan

Data pelanggan yang telah dikelompokkan sesuai *clusternya* pada Tabel V selanjutnya dipetakan dalam bentuk diagram tebar 3D pada Gambar 2.

Gambar 2 menunjukkan hasil dari pengelompokkan *data* pelanggan menggunakan K-Means *clustering*, dimana masing-masing *clusternya* dipetakan dengan simbol. Masing-masing *cluster* selanjutnya dilakukan penentuan skor RFM yang sesuai dengan batasan pada Tabel VI.

TABEL VI
BATASAN SKOR RFM

Skor	ZRecency	ZFrequency	ZMonetary
1	$1,86632 \leq x \leq 2,49932$	$-0,49490 \leq x \leq 0,08492$	$-0,61594 \leq x \leq 0,00673$
2	$1,23332 \leq x \leq 1,86631$	$0,08493 \leq x \leq 0,66476$	$0,00674 \leq x \leq 0,62941$
3	$0,60032 \leq x \leq 1,23331$	$0,66477 \leq x \leq 1,24460$	$0,62942 \leq x \leq 1,25209$
4	$-0,03267 \leq x \leq 0,60031$	$1,24461 \leq x \leq 1,82443$	$1,25210 \leq x \leq 1,87477$
5	$-0,66568 \leq x \leq -0,03268$	$1,82444 \leq x \leq 2,40427$	$1,87478 \leq x \leq 2,49745$

Batasan skor RFM pada Tabel VI didapatkan melalui penghitungan kuintil yaitu pembagian sama rata sebanyak 5 bagian (20%) dari nilai minimum dan nilai maksimum tiap atribut. Masing-masing *cluster* akan diberikan skor sesuai

dengan batasannya, dengan nilai 5 adalah terbaik dan 1 yang terburuk [17]. Kombinasi skor RFM yang didapatkan digunakan sebagai penentu tingkatan loyalitas pelanggan, yang juga menentukan strategi yang dapat dilakukan oleh tempat usaha bengkel XYZ.

Berikut pembagian tingkatan loyalitas pelanggan dalam kategori berdasar skor RFM dapat dilihat pada Tabel VII. Untuk atribut frequency dan monetary dapat dilakukan kombinasi karena kedua atribut tersebut memiliki keterkaitan dengan jumlah pembelian yang dilakukan oleh pelanggan. Skor kombinasi frequency dan monetary ini didapatkan melalui penghitungan rata-rata dari kedua atribut tersebut.

TABEL VII
KATEGORI PELANGGAN BERDASARKAN SKOR RFM [23]

Kategori	Skor Recency	Skor Kombinasi Frequency & Monetary	Strategi
<i>Champions</i>	4 – 5	4 – 5	Beri imbalan atau hadiah, bisa dijadikan sebagai pengadopsi awal untuk produk baru dan sebagai pelaku promosi produk
<i>Loyal Customers</i>	2 – 5	3 – 5	Lakukan <i>up-selling</i> terhadap produk dengan nilai tinggi, meminta ulasan dan libatkan dalam berbagai kegiatan
<i>Potential Loyalist</i>	3 – 5	1 – 3	Menawarkan program keanggotaan dan merekomendasikan produk lain atau <i>cross-selling</i>
<i>Recent Customers</i>	4 – 5	0 – 1	Berikan dukungan <i>on-boarding</i> , memberikan kesuksesan awal dan mulai membangun hubungan dengan pelanggan
<i>Promising</i>	3 – 4	0 – 1	Ciptakan <i>brand awareness</i> dan tawarkan uji coba gratis
<i>Customer Needing Attention</i>	2 – 3	2 – 3	Buat penawaran dengan waktu terbatas dan rekomendasikan berdasarkan histori pembelian pelanggan sebelumnya dan aktifkan kembali pelanggan
<i>About to Sleep</i>	2 – 3	0 – 2	Membagikan produk yang berharga, merekomendasikan produk populer atau pembaruan diskon dan membangun relasi kembali kepada pelanggan
<i>At Risk</i>	0 – 2	2 – 5	Kirim <i>email</i> yang bersifat personal untuk membangun relasi kembali kepada pelanggan, menawarkan produk baru dan memberikan produk yang bermanfaat
<i>Can't Lose Them</i>	0 – 1	4 – 5	Menarik pelanggan kembali melalui pembaruan produk, jangan sampai kehilangan persaingan dari tempat usaha lainnya dan berbicara kepada pelanggan
<i>Hibernating</i>	1 – 2	1 – 2	Tawarkan produk lain yang relevan dan diskon khusus dan menciptakan ulang nilai dari merek produk

Melihat batasan skor RFM pada Tabel VI dan pembagian kategori pelanggan pada Tabel VII, maka skor RFM yang didapatkan untuk tiap *cluster* dapat dilihat pada Tabel VIII.

Hasil analisis skor RFM menunjukkan *cluster* 1 sebanyak 2.470 pelanggan, *cluster* 3 sebanyak 9.333 pelanggan dan *cluster* 5 sebanyak 940 pelanggan termasuk dalam kategori *Potential Loyalist*. Kategori *Potential Loyalist* merupakan kategori pelanggan yang masih aktif dalam waktu dekat namun frekuensi pembelian dan uang yang dikeluarkan pelanggan masih di bawah rata-rata. Pelanggan yang termasuk dalam kategori ini memiliki loyalitas tingkatan sedang dan dapat dilakukan strategi untuk meningkatkan loyalitas pelanggan. Strategi yang dapat dilakukan adalah menawarkan program keanggotaan. Strategi lainnya dengan melakukan *cross-selling*, yaitu menjual tambahan produk kepada pelanggan, dimana tambahan produk didapatkan

dengan melihat seringnya produk tambahan dibeli ketika produk utama dibeli pada satu transaksi dalam histori transaksi pelanggan [24].

Hasil pada *cluster* 2 sebanyak 830 pelanggan termasuk dalam kategori *Loyal Customers*. Kategori *Loyal Customers* merupakan kategori pelanggan dengan nilai *recency*, *frequency* dan *monetary* yang di atas rata-rata. Pelanggan yang termasuk dalam kategori ini memiliki loyalitas yang tinggi. Strategi yang dapat dilakukan untuk kategori ini adalah melakukan *up-selling* terhadap produk yang memiliki nilai tinggi, juga melibatkan pelanggan dalam berbagai kegiatan seperti pengisian formulir kepuasan pelanggan, uji coba produk baru dan sebagainya.

TABEL VIII
SKOR RFM PER CLUSTER

Cluster	ZRecency	ZFrequency	ZMonetary	Jumlah	Skor RFM	Kategori
1	0,27357	0,12196	-0,14784	2.470	421	<i>Potential Loyalist</i>
2	1,59190	1,36840	1,17737	830	243	<i>Loyal Customers</i>
3	-0,63118	-0,44696	-0,41611	9.333	511	<i>Potential Loyalist</i>
4	1,52449	0,40366	0,13440	1.406	222	<i>Customer Needing Attention</i>
5	-0,37712	-0,16832	0,85532	940	513	<i>Potential Loyalist</i>

Hasil lainnya untuk *cluster* 4 sebanyak 1.406 pelanggan termasuk dalam kategori *Customer Needing Attention*. Kategori *Customer Needing Attention* merupakan kategori pelanggan yang memiliki skor RFM dan tingkat loyalitas rata-rata ke bawah. Kategori ini tidak memberikan keuntungan yang cukup berarti bagi bengkel XYZ, namun dapat ditingkatkan dengan menerapkan beberapa strategi. Strategi yang dapat dilakukan untuk kategori ini adalah membuat penawaran dengan waktu terbatas seperti promosi atau diskon sebuah produk. Strategi lainnya yaitu merekomendasikan produk kepada pelanggan sesuai dengan kebutuhannya, yang dapat dilihat dari histori pembelian pelanggan sebelumnya.

V. KESIMPULAN

Berdasarkan hasil penelitian yang didapatkan, penerapan K-Means *clustering* dengan atribut RFM pada *data* transaksi pelanggan di bengkel XYZ mengelompokkan 14.979 *data* pelanggan dalam 5 *cluster*. Hasil menunjukkan *cluster* 1, 3 dan 5 merupakan dominan pelanggan di bengkel XYZ dengan total 12.743 (85%) pelanggan. Pelanggan di *cluster* 1, 3 dan 5 yang termasuk dalam kategori *Potential Loyalist* ini dapat dijadikan prioritas utama bagi pemilik bengkel XYZ dalam menjalankan strategi mempertahankan pelanggan. Bengkel XYZ dapat melakukan beberapa strategi yaitu penawaran program keanggotaan dengan berbagai keuntungan untuk pelanggan seperti pemberian harga khusus atau diskon, sistem poin dan lain sebagainya. Strategi lainnya dengan melakukan *cross-selling* untuk produk tertentu dalam bentuk promo ataupun paket, dimana produk yang ditawarkan masih menjadi kebutuhan pelanggannya dengan cara melihat histori dari pelanggan pada transaksi sebelumnya. Beberapa strategi tersebut dapat diterapkan di bengkel XYZ guna mempertahankan pelanggan yang memberikan manfaat untuk meningkatkan keuntungan bagi bengkel XYZ.

DAFTAR PUSTAKA

- [1] K. O. Joseph, A. Omotayo, A. Mosunmola, & B. Taiye, "Marketing Concept and The Satisfaction of Consumer Needs The Nigerian Consumers Experience," *Research Journal of Marketing*, vol. 4, pp. 1–15, Feb. 2016.
- [2] C. H. Adikaram, A. Khatibi, & M. S. Yajib, "The Relationship between Customer Relationship Management and Customer Satisfaction," *International Journal of Arts and Commerce*, vol. 5, pp. 69–95, Mar. 2016.
- [3] O. O. Oluseye & B. T. Tairat, "Customer Relationship Management Approach and Student Satisfaction in Higher Education Marketing," *Journal of Competitiveness*, vol. 6, pp. 49–62, Sep. 2014.
- [4] N. M. Manzuma-Ndaaba, Y. Harada, A. R. Romle, & A. S. Shamsudin, "Cognitive, Affective and Conative Loyalty in Higher Education Marketing: Proposed Model for Emerging Destinations," *International Review of Management and Marketing*, vol. 6, pp. 168–175, 2016.
- [5] L. J. Kiplagat & L. Oyugi, "Influence of Customer Retention on Market Share in Chemelil Sugar Company Limited, Kenya," *The International Journal Of Business & Management*, vol. 4, pp. 304–311, Sep. 2016.
- [6] A. Payne, *Handbook of CRM: Achieving Excellence in Customer Management*, Oxford: Elsevier, 2005.
- [7] J. Reynolds, *A Practical Guide to CRM Building More Profitable Customer Relationships*, Boca Raton: CRC Press, 2002.
- [8] J. Han, M. Kamber, & J. Pei, *Data Mining Concepts and Techniques*, 3rd ed., Waltham: Elsevier, 2012.
- [9] C. Naik, A. Kharwar, & N. Desai, "A Review: RFM Approach on Different Data Mining Techniques," *International Journal of Emerging Technology and Advanced Engineering*, vol. 3, pp. 725–728, Okt. 2013.
- [10] M. A. Syakur, B. K. Khotimah, E. M. S. Rochman, & B. D. Satoto, "Integration K-Means Clustering Method and Elbow Method For Identification of The Best Customer Profile Cluster", *IOP Conf. Ser.: Mater. Sci. Eng.*, 2018, paper 012017, p. 336.
- [11] S. M. Rajeh, F. A. Koudehi, S. M. Seyedhosseini, & R. Farazmand. "A Model for Customer Segmentation Based On Loyalty Using Data Mining Approach and Fuzzy Concept in Iranian Bank," *International Journal of Business and Behavioral Sciences*, vol. 4, pp. 118–136, Sep. 2014.
- [12] B. Thomas & J. Tobe, *Anticipate: Knowing What Customer Need Before They Do*, New Jersey: John Wiley & Sons, Inc, 2012.
- [13] K. Khadka & S. Maharjan, "Customer Satisfaction and Customer Loyalty," Degree Program in Business Management. thesis, Centria University of Applied Sciences Pietarsaari, Nov. 2017.
- [14] K. D. Joshi & P. S. Nalwade, "Modified K-Means for Better Initial Cluster Centres," *International Journal of Computer Science and Mobile Computing*, vol. 2, pp. 219–223, Jul. 2013.
- [15] A. Barai & L. Dey, "Outlier Detection and Removal Algorithm in K-Means and Hierarchical Clustering," *World Journal of Computer Application and Technology*, vol. 5, pp. 24–29, 2017.
- [16] K. Modi & B. Oza, "Outlier Analysis Approaches in Data Mining," *International Journal of Innovative Research in Technology*, vol. 3, pp. 6–12, Des. 2016.
- [17] M. Mohammadian & I. Makhani. "RFM-Based customer segmentation as an elaborative analytical tool for enriching the creation of sales and trade marketing strategies," *International Academic Journal of Accounting and Financial Management*, vol. 3, pp. 21–35, 2016.
- [18] A. S. Shirshorshidi, S. Aghabozorgi, & T. Y. Wah, "A Comparison Study on Similarity and Dissimilarity Measures in Clustering Continuous Data," *PLoS ONE*, vol. 10, Des. 2015.
- [19] C. Saranya & G. Manikandan, "A Study on Normalization Techniques for Privacy Preserving Data Mining," *International Journal of Engineering and Technology*, vol. 5, pp. 2701–2704, Jul. 2013.
- [20] D. Kaur, "A Comparative Study of Various Distance Measures for Software Fault Prediction," *International Journal of Computer Trends and Technology*, vol. 17, Nov. 2014.
- [21] T. M. Kodinariya & P. R. Makwana, "Review on determining number of Cluster in K-Means Clustering," *International Journal of Advance Research in Computer Science and Management Studies*, vol. 1, pp. 90–95, Nov. 2013.
- [22] A. M. Baswade & P. S. Nalwade, "Selection of Initial Centroids for k-Means Algorithm," *International Journal of Computer Science and Mobile Computing*, vol. 2, pp.161–164, Jul. 2013.
- [23] A. Nair. (2017) RFM Analysis For Successful Customer Segmentation. [Online]. Tersedia: <https://www.putler.com/rfm-analysis/>
- [24] T. S. Walia & G. K. Gulati, "Conceptual Framework Of Cross Selling," *International Journal of in Multidisciplinary and Academic Research*, vol. 3, pp. 145–150, Okt. 2014.