

KERANGKA PENGAMBILAN KEPUTUSAN UNTUK PEMASARAN PRESISI MENGGUNAKAN METODE RFM, ALGORITMA K-MEANS DAN DECISION TREE

Jupriyanto¹; Siti Nurlela²

Program Pascasarjana Magister Ilmu Komputer
STMIK Nusa Mandiri Jakarta
www.nusamandiri.ac.id

¹jupriyanto.kahar@gmail.com; ²siti.sie@nusamandiri.ac.id

Abstract— Precision marketing provides the ability for companies to offer products made specifically to customers and provide the ability for companies to attract customers with specially made marketing messages. This paper presents a new decision-making framework using data mining techniques. First, this study presents a trend model for accurately predicting monthly supply quantities; second, using the RFM (Recency, Frequency, Monetary) model to choose attributes to group customers into different groups; third, using the K-Means Algorithm to create customer clustering based on RFM data of each customer, fourth, using Decision Tree to identify attribute values that are important for distinguishing different customer groups; and finally, from the process of data mining that researchers do create various bidding strategies that target each customer cluster. Sales data from Syifamart in Subang, West Java, were collected and used in a case study to illustrate how to implement the proposed framework. The research shows that the process of data mining from sales transaction history is 351,158 rows, with customer-based aggregation using the RFM method and extracted using the k-means clustering algorithm to form 4 (four) optimal clusters. The four (four) clusters are classified using a decision tree algorithm so Syifamart can find out which potential customers and potential customers. For the availability of stock supply, management predicts the need for product inventory using a trend method where the stock in the following month is predicted by using the sales history in the previous month.

Intisari— Pemasaran presisi memberikan kemampuan pada perusahaan untuk menawarkan produk-produk yang dibuat secara khusus kepada pelanggan dan memberikan kemampuan kepada perusahaan untuk menarik minat pelanggan dengan pesan-pesan pemasaran yang dibuat secara khusus. Penelitian ini menyajikan kerangka kerja pengambilan keputusan baru menggunakan teknik data mining. Pertama, penelitian ini menyajikan model tren untuk memprediksi secara akurat

kuantitas pasokan bulanan; kedua, menggunakan model RFM (Recency, Frequency, Monetary) untuk memilih atribut untuk mengelompokkan pelanggan ke dalam kelompok sesuai history transaksi belanjanya; ketiga, menggunakan Algoritma K-Means untuk membuat clustering pelanggan berdasarkan data RFM masing-masing pelanggan, keempat, menggunakan Decision Tree untuk mengidentifikasi nilai atribut penting untuk membedakan kelompok pelanggan yang berbeda; dan akhirnya, dari proses data mining yang peneliti lakukan menciptakan berbagai strategi penawaran yang menargetkan setiap cluster pelanggan. Data penjualan dari Syifamart di Subang Jawa Barat, dikumpulkan dan digunakan dalam studi kasus untuk menggambarkan bagaimana mengimplementasikan kerangka yang diusulkan. Dari penelitian yang telah dilakukan menunjukkan bahwa proses data mining dari history transaksi penjualan 351,158 rows, dengan agregasi berdasarkan pelanggan menggunakan metode RFM dan diekstraksi dengan menggunakan algoritma clustering k-means membentuk 4 (empat) cluster optimal. Keempat (empat) cluster tersebut diklasifikasikan dengan menggunakan algoritma decision tree sehingga Syifamart dapat mengetahui mana pelanggan potensial dan mana pelanggan yang tidak potensial. Untuk ketersediaan pasokan stok, manajemen memprediksi kebutuhan persediaan produk dengan menggunakan metode tren dimana stok di bulan selanjutnya di prediksi dengan menggunakan history penjualan di bulan sebelumnya.

Kata Kunci: Pemasaran Presisi, Metode Tren, RFM, K-Means, Decision Tree.

PENDAHULUAN

Dalam menjalankan roda bisnis, perusahaan membutuhkan strategi pemasaran yang cocok dan efektif dengan bisnis yang sedang mereka jalankan. Hal itu bertujuan untuk menarik pelanggan lebih banyak lagi serta meningkatkan penjualan

sehingga dapat mendatangkan keuntungan dan juga kesuksesan. Laju percepatan globalisasi ekonomi dan meningkatnya persaingan pasar, tekanan ekonomi dan persaingan usaha telah menyebabkan perusahaan menghadapi masalah memilih kebijakan pengambilan keputusan strategis yang tepat untuk menjual produk yang tepat kepada pelanggan yang tepat pada waktu yang tepat, sehingga perusahaan dapat meningkatkan laba mereka. Baru-baru ini di kenal bahwa pemasaran presisi telah menjadi sarana untuk menghasilkan keuntungan lebih dan menjadi semakin penting karena pelanggan menjadi lebih tahu tentang produk dan hak mereka sebagai pelanggan. Ketersediaan data pelanggan dan catatan transaksi memberikan pemahaman yang lebih baik tentang perilaku dan preferensi konsumsi pelanggan.

Usaha retail, sejumlah biaya dihabiskan untuk menarik pelanggan baru, akan tetapi seiring waktu pelanggan tersebut dapat menghentikan pembelian mereka karena berbagai alasan (Wardani et al., 2018).

Syifamart merupakan sebuah usaha yang selalu meningkatkan kualitasnya baik secara internal maupun eksternal, namun tinggat pemasaran dan penjualan pada syifamart masih sangat rendah.

Model RFM merupakan model yang membedakan pelanggan penting dari data yang besar oleh tiga variabel yaitu recency, frequency dan monetary (Hardiani, Hartanto, & Mada, 2017). *A decision tree is an efficient data mining algorithm with a strong explanatory capability* (You et al., 2015)

Penelitian ini bertujuan untuk membuat kerangka kerja pemasaran presisi dengan cara melakukan pengelompokan dan clustering pelanggan menggunakan metode Recency Frequency and Monetary (RFM) dan Algoritma K-means serta menggunakan metode prediksi tren sehingga dapat memprediksi kebutuhan persediaan produk yang akan dipasarkan sesuai dengan kebutuhan pelanggan.

BAHAN DAN METODE

A. Datasets

Syifamart adalah salah satu Unit Bisnis yang mendukung program dan kegiatan di Yayasan As-Syifa Al-Khoeriyah. Saat ini Syifamart memiliki sistem yang terintegrasi dengan Smartcard untuk metode pembayarannya. Lokasi Syifamart berada di sekitar kompleks lembaga pendidikan As-Syifa, dengan harapan semua kebutuhan terpenuhi di Syifamart mulai dari makanan ringan sehari-hari, kebutuhan sekolah, kebutuhan kantor hingga kebutuhan rumah tangga bulanan, tidak perlu

keluar dari kompleks untuk memenuhi kebutuhan internal lembaga.

Penelitian ini menggunakan data yang diambil dari sistem penjualan Syifamart selama 6 bulan, periode Juli 2018 hingga Desember 2018. Dataset untuk 6 bulan adalah 351.158 berdasarkan transaksi penjualan yang ada. Dataset sampel penjualan dapat dilihat pada Tabel 1.

Tabel 1. Dataset Penjualan

NO TRANS	NAMA PRODUK	QTY	TGL BUKTI	HARGA	PLG
J-2018-220429	slay olay blubbery	20	7/1/2018	1,000	2017300006
J-2018-221862	keripik pusaka 3000	1	7/6/2018	3,000	2017300010
J-2018-221862	Twistko jagung bakar 80gr	1	7/6/2018	6,000	2017300010
J-2018-220400	rinso molto 900g	1	7/1/2018	19,000	2017300022
J-2018-220400	molto pink all in one pouch 300ml	1	7/1/2018	9,500	2017300022
J-2018-220402	sedaap minyak goreng 1liter	1	7/1/2018	13,500	2017300046
J-2018-220402	Kecap Bango 60 ml	1	7/1/2018	3,000	2017300046
J-2018-220402	Goodday Mocacino	2	7/1/2018	1,250	2017300046
J-2018-220404	Frestea Melati cup	1	7/1/2018	2,000	2017300061
J-2018-220406	teh javana 350ml	7	7/1/2018	3,000	2017300068
J-2018-220410	Batre Abc Kecil	4	7/1/2018	2,500	UMUM
J-2018-220412	axo 600ml	1	7/1/2018	2,500	UMUM

Sumber: (Jupriyanto. & Nurlala, 2019)

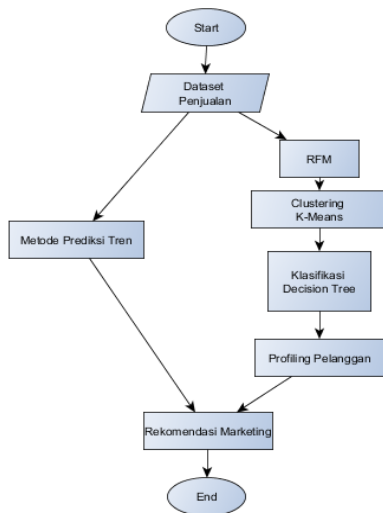
B. Metode

Menurut (widiarina, 2013) Data mining mengusulkan proses standar data mining yang disebut dengan Cross-Industry Standard Process for Data Mining (CRISP-DM).

(Amborwati Armadyah, 2014) menyebutkan bahwa Algoritma data mining yang sering digunakan untuk segmentasi adalah K-Means Clustering dan atributatribut yang digunakan untuk proses mining pada segmentasi konsumen adalah data konsumen, produk, demografi, perilaku konsumen, transaksi, RFMDC, menurut (Tsipstsis K & A., 2009) K-means menggunakan

ukuran Euclidean distance dan secara iteratif menentukan tiap record dari cluster asal. RFM (Recency, Frequency Monetary), dan LTV (Life Time Value). Dan penting untuk menggabungkan algoritma clustering dengan beberapa algoritma data mining lainnya seperti Classification, Association, dan CPV matriks untuk mendapatkan nilai potensial dari tiap cluster.

Pada penelitian ini menyajikan model tren untuk memprediksi kuantitas pasokan bulanan, dan menggunakan gabungan metode RFM, Clustering K-Means dan Klasifikasi Decision Tree untuk profiling pelanggan. Metode penelitian sesuai Gambar 1 di bawah ini :



Sumber: (Jupriyanto. & Nurlela, 2019)
 Gambar 1. Metode Penelitian

HASIL DAN PEMBAHASAN

Penelitian ini menyajikan kerangka kerja pengambilan keputusan baru menggunakan teknik data mining. Pertama, menyajikan model tren untuk memprediksi secara akurat kuantitas pasokan bulanan; kedua, menggunakan model RFM (Recency, Frequency, Monetary), model ini untuk memilih atribut yang akan mengelompokkan pelanggan ke dalam kelompoknya sesuai history transaksi belanjanya; ketiga, menggunakan Algoritma K-Means untuk membuat clustering pelanggan berdasarkan data RFM masing-masing pelanggan, keempat, menggunakan Decision Tree untuk mengidentifikasi nilai atribut penting untuk membedakan kelompok pelanggan yang berbeda; dan akhirnya, dari proses data mining yang peneliti lakukan menciptakan berbagai strategi penawaran yang menargetkan setiap cluster pelanggan.

A. Model Tren

Model tren ditunjukkan oleh perubahan nilai rata-rata penjualan produk “keripik pusaka 3000”

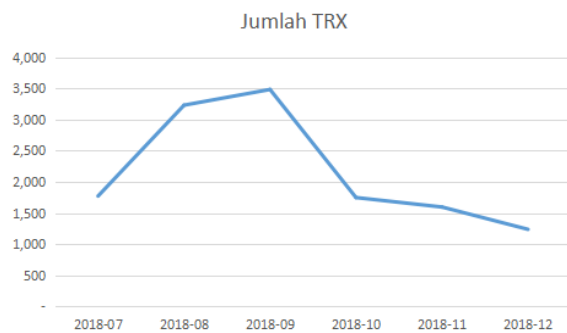
selama periode 1 July 2018 sampai dengan 31 Desember 2018. Berikut tren penjualan penulannya selama 6 bulan :

Tabel 2. Tren 6 bulan menjual produk "keripik pusaka 3000"

Period	Jumlah trx
2018-07	1,773
2018-08	3,231
2018-09	3,483
2018-10	1,749
2018-11	1,597
2018-12	1,239

Sumber: (Jupriyanto. & Nurlela, 2019)

Dari tabel di atas akan dibuat deskripsi data ke dalam bentuk poligon agar dapat memudahkan menganalisis data. Berikut ini adalah poligon data dari data hasil penjualan produk "keripik pusaka 3000" :



Sumber: (Jupriyanto. & Nurlela, 2019)
 Gambar 2. Grafik poligon Tren 6 bulan menjual produk " keripik pusaka 3000"

Dari grafik poligon di atas, dapat dilihat bahwa setiap bulan penjualan produk "keripik pusaka 3000" fluktuasi dalam hal jumlah transaksi, puncak dari transaksi terbanyak adalah pada bulan September 2018.

B. Model RFM (Recency, Frequency, Monetary)

Untuk mengklasifikasikan pelanggan, penelitian ini berfokus pada pelanggan yang sebelumnya telah belanja produk. Pra-proses data sebelum membangun model pengambilan keputusan adalah menggunakan metode RFM:

1. Membuang Data Transaksi yang bukan dari member, penelitian ini fokus ke transaksi yang dilakukan oleh member yang memiliki Smartcard RFID.
2. Menentukan produk yang akan dilakukan penelitian, untuk mengakomodir aspek pemasaran presisi dari sisi ketersediaan produk bisa diteliti juga, penelitian ini akan fokus meneliti produk “keripik pusaka 3000” dengan

total transaksi pembelian oleh member sejumlah 6.744 transaksi dan total nominal transaksi sejumlah Rp 22,424,400.

- Selanjutnya, kami memilih atribut yang akan digunakan dalam prosedur data mining dari penelitian ini, yaitu atribut RFM (recency, frequency, monetary), sehingga diperoleh data sesuai table 3 berikut ini :

Tabel 3. Data Atribut RFM

No	Pelanggan	RFM		
		R (Hari)	F (Kali)	M (Rp)
1	2017300006	42	14	42,000
2	2017300008	31	18	60,000
3	2017300010	13	20	60,000
4	2017300018	13	46	138,000
5	2017300022	53	10	33,000
6	2017300027	13	10	30,000
7	2017300029	17	33	99,000
8	2017300036	11	32	96,000
9	2017300043	28	26	93,000
10	2017300045	11	34	105,000
11	2017300046	125	3	12,000
12	2017300061	11	30	93,000
13	2017300068	49	26	81,000
14	2017300074	27	1	3,000
15	2017300075	103	20	60,000

Sumber: (Jupriyanto. & Nurlela, 2019)

- Menentukan kriteria pembobotan berdasarkan variabel RFM. Pembobotan dibagi menjadi 5 skala / skor seperti yang tercantum pada Tabel 4.

Tabel 4. Skala RFM

Score	R (hari)		F (kali)		M (hari)	
	Min	Max	Min	Max	Min	Max
	5	0	52	48	1000	146,568
4	53	105	36	47	109,926	146,567
3	106	158	24	35	73,284	109,925
2	159	211	12	23	36,642	73,283
1	212	1000	0	11	0	36,641

Sumber: (Jupriyanto. & Nurlela, 2019)

- Setelah menentukan skala RFM, selanjutnya melakukan Skoring Pelanggan berdasar RFM dengan cara melakukan transformasi data RFM menjadi seperti table 5 di bawah ini :

Tabel 5. Skoring Pelanggan

No	Pelanggan	RFM		
		R	F	M
1	2017300006	5	2	2
2	2017300008	5	2	2
3	2017300010	5	2	2
4	2017300018	5	4	4

No	Pelanggan	RFM		
		R	F	M
5	2017300022	4	1	1
6	2017300027	5	1	1
7	2017300029	5	3	3
8	2017300036	5	3	3
9	2017300043	5	3	3
10	2017300045	5	3	3
11	2017300046	3	1	1
12	2017300061	5	3	3
13	2017300068	5	3	3
14	2017300074	5	1	1
15	2017300075	4	2	2

Sumber: (Jupriyanto. & Nurlela, 2019)

C. Algoritma K-Means

Setelah semua data transaksi diubah menjadi angka, maka data tersebut dapat dikelompokkan dengan menggunakan algoritma K-means. Untuk dapat mengelompokkan data ke dalam beberapa kluster diperlukan beberapa langkah sebagai berikut:

- Menentukan jumlah kluster yang diinginkan. Dalam penelitian ini data yang ada akan dikelompokkan menjadi empat kelompok.
- Menentukan titik pusat awal dari setiap cluster. Dalam penelitian ini titik pusat awal ditentukan secara acak dan diperoleh titik pusat setiap cluster yang dapat dilihat pada tabel 6

Tabel 6. Titik Pusat Cluster

Pusat Cluster	Pelanggan	R	F	M
CLUSTER 1 (C1)	2017300029	5	3	3
CLUSTER 2 (C2)	2017300312	5	5	5
CLUSTER 3 (C3)	2017300375	5	2	2
CLUSTER 4 (C4)	2017300668	4	2	2

Sumber: (Jupriyanto. & Nurlela, 2019)

- Tempatkan setiap data dalam kluster. Dalam penelitian ini, metode k-means digunakan untuk mengalokasikan setiap data ke dalam sebuah cluster yang paling dekat dengan titik pusat setiap cluster. Untuk mengetahui cluster mana yang paling dekat dengan data, diperlukan untuk menghitung jarak setiap data dengan titik pusat masing-masing cluster. Misalnya, menghitung jarak dari data transaksi pelanggan = 2017300027 ke pusat klaster pertama:

$$D(i, j) = \sqrt{(x1i - x1j)^2 + (x2i - x2j)^2 + \dots + (xki - xkj)^2} \dots (1)$$

diperoleh jarak ke masing-masing cluster adalah :

- D(C1) = 2.828427125
- D(C2) = 5.656854249

- 3. $D(C3) = 1.414213562$
- 4. $D(C4) = 1.732050808$

Dari perhitungan di atas, jarak terdekat pelanggan 2017300027 adalah ke cluster 3 (1.414213562), sehingga member 2017300027 di masukan ke cluster 3. Hasil perhitungan lengkap dengan pusat cluster inisial bisa di lihat di table 7.

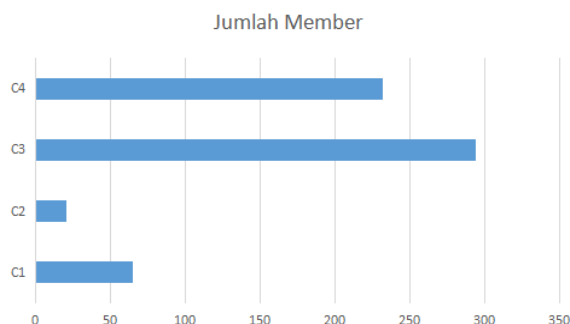
Tabel 7. Jarak ke masing-masing cluster

PELANGGAN	JARAK KE				DEKAT KE
	C1	C2	C3	C4	
2017300006	1.41	4.24	-	1.00	C3
2017300008	1.41	4.24	-	1.00	C3
2017300010	1.41	4.24	-	1.00	C3
2017300018	1.41	1.41	2.83	3.00	C1
2017300022	3.00	5.74	1.73	1.41	C4
2017300027	2.83	5.66	1.41	1.73	C3
2017300029	-	2.83	1.41	1.73	C1
2017300036	-	2.83	1.41	1.73	C1
2017300043	-	2.83	1.41	1.73	C1
2017300045	-	2.83	1.41	1.73	C1
2017300046	3.46	6.00	2.45	1.73	C4
2017300061	-	2.83	1.41	1.73	C1
2017300068	-	2.83	1.41	1.73	C1

Sumber: (Jupriyanto. & Nurlela, 2019)

- d. Setelah semua data ditempatkan ke dalam cluster terdekat, kemudian kalkulasi ulang pusat klaster baru berdasarkan rata-rata anggota dalam klaster.
- e. Setelah mendapatkan titik pusat baru dari masing-masing cluster, kembali dari langkah c hingga titik pusat setiap cluster tidak berubah lagi dan tidak ada data yang bergerak dari satu cluster ke cluster lainnya.

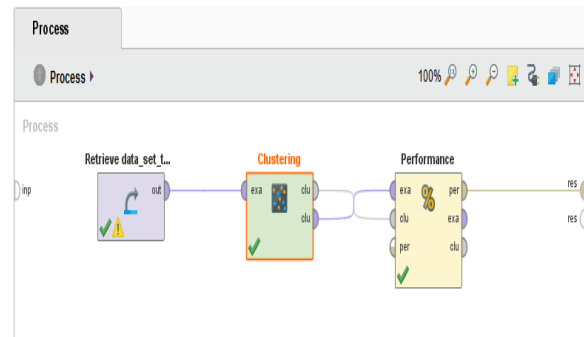
Dari hasil pengolahan data yang dilakukan berdasarkan dataset transaksi pelanggan menggunakan K-Means melalui 4 iterasi, maka terbentuklah cluster seperti yang ditunjukkan pada Gambar 3 yang menunjukkan bahwa hasil clustering diperoleh 65 anggota cluster 1, 21 anggota cluster 2, 294 anggota cluster 3, dan 232 anggota cluster 4.



Sumber: (Jupriyanto. & Nurlela, 2019)

Gambar 3. Grafik Sebaran Cluster

Dari hasil pengelompokan, perlu untuk menguji validitas cluster dengan menghitung Indeks Davies Bouldin untuk menentukan apakah jumlah cluster yang terbentuk optimal atau tidak.



Sumber: (Jupriyanto. & Nurlela, 2019)

Gambar 4. Design pengecekan performace cluster

Cluster adalah konvergen jika tidak ada perubahan atau pergerakan anggota dari satu cluster ke cluster lainnya. Selain itu, gugus konvergen juga ditandai dengan tidak ada perubahan dalam nilai-nilai indeks DB. Dari lima skenario pengujian yang telah dilakukan, peringkat akan dibuat berdasarkan nilai indeks-DB terbaik. Tingkat optimal dari suatu cluster dapat diukur dengan Nilai Indeks DB. Davies dan Bouldien menyebutkan bahwa jumlah cluster terbaik adalah yang memiliki indeks terkecil di antara kelompok lain.

Tabel 8. Nilai Indeks DB

Jumlah Cluster	DB-index
2	-0.626
3	-0.673
4	-0.518
5	-0.552

Sumber: (Jupriyanto. & Nurlela, 2019)

Hasil penelitian menunjukkan bahwa jumlah cluster 4 memiliki Indeks Davies Bouldien (DI) terkecil di antara cluster lain dari 0,518 yang tercantum pada Tabel 2

D. Decision Tree

Setelah menemukan segmentasi pelanggan, berikutnya akan dianalisis pada setiap segmen untuk menentukan karakteristik pelanggan menggunakan Algoritma Decision Tree. Tujuan klasifikasi adalah untuk mengetahui informasi penting yang terdapat dalam data transaksi pelanggan yang didasarkan pada Keterkinian, Frekuensi, Moneter (RFM) dan Cluster. Tabel 9 menunjukkan jumlah lengkap data transaksi Pelanggan. Dalam proses ini, di teliti sebanyak 612 dataset yang terdiri dari 4 kluster.

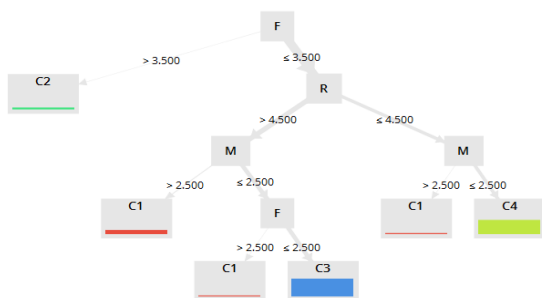
Tabel 9. Member Cluster

No	Cluster	Jumlah Member
1	Cluster 1	65
2	Cluster 2	21
3	Cluster 3	294
4	Cluster 4	232

Sumber: (Jupriyanto. & Nurlela, 2019)

Pembentukan model klasifikasi dilakukan dengan mengimplementasikan data ke Rapidminer Studio 9.1 dengan dataset adalah hasil Clsutering menggunakan K-Means yang berjumlah 612 data menghasilkan pohon model dan aturan penentuan karakteristik pelanggan.

Hasil dari model klasifikasi dapat dilihat sebagai pohon pada Gambar 4 dan model klasifikasi ini melibatkan 4 atribut yaitu Keterkinian, Frekuensi, Moneter (RFM) dan Cluster.



Sumber: (Jupriyanto. & Nurlela, 2019)

Gambar 5. Model Decision Tree untuk Menentukan Karakteristik Pelanggan

Hasil prediksi dari penetapan parameter karakteristik pelanggan pada 612 data menghasilkan prediksi sebesar 98,86%.

accuracy: 98.86% +/- 1.05% (micro average: 98.86%)

	true C3	true C2	true C4	true C1	class precision
pred C3	294	0	0	3	98.99%
pred C2	0	21	0	0	100.00%
pred C4	0	0	232	4	98.31%
pred C1	0	0	0	58	100.00%
class recall	100.00%	100.00%	100.00%	89.23%	

Sumber: (Jupriyanto. & Nurlela, 2019)

Gambar 6. Deskripsi dari Model Decision Tree untuk Menentukan Karakteristik Pelanggan

Profil pelanggan dilakukan dengan menggunakan aturan yang dihasilkan dari proses sebelumnya. Karakteristik pelanggan yang digunakan dalam profil pelanggan adalah karakteristik menurut (Maryani & Riana, 2017) Tabel 10 adalah karakteristik pelanggan yang dihasilkan dalam penelitian ini.

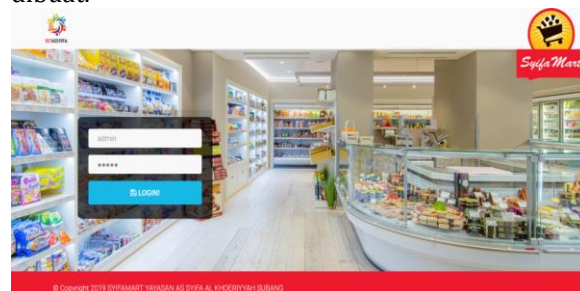
Tabel 10. Karakteristik Pelanggan

Type Pelanggan	Cluster	Karakteristik Pelanggan
Pelanggan Berulang	1	Pelanggan dalam Cluster 1 disebut sebagai pelanggan tetap, yaitu pelanggan yang telah melakukan pembelian suatu produk 2 kali atau lebih.
Klien	2	Pelanggan dalam Cluster 2 disebut sebagai tipe pelanggan Klien. Mereka membeli secara teratur, dan hubungan dengan pelanggan jenis ini telah kuat dan akan bertahan lama yang membuat mereka tidak terpengaruh oleh tarikan pesaing dari produk lain.
Penganjur	3	Pelanggan dalam Cluster 3 melakukan pembelian rutin, dengan nominal yg terbatas. Ini sangat menguntungkan bagi perusahaan. Pelanggan jenis ini juga dapat digunakan sebagai pasar perusahaan karena jenis pelanggan ini melakukan pemasaran untuk perusahaan dengan mendorong teman-teman mereka untuk membeli barang / jasa di perusahaan.
Pelanggan Pertama Kali	4	Pelanggan dalam Cluster 4 memiliki karakteristik tingkat intensitas rendah. Mereka adalah pelanggan baru yang tingkat kedatangannya sangat rendah. Pelanggan jenis ini adalah pelanggan yang membeli untuk pertama kalinya, mereka masih pelanggan baru.

Sumber: (Maryani & Riana, 2017)

E. Implementasi System

Untuk kebutuhan manajemen dan tim marketing Syifamart, dibawah ini implementasi dari kerangka kerja Pemasaran Presisi yang telah dibuat.



Sumber: (Jupriyanto. & Nurlela, 2019)

Gambar 6. Halaman login Implementasi System

Tim marketing atau manajemen Syfamart di berikan akses username dan password, sehingga bisa masuk ke system yang peneliti buat.

ID	NAMA	ALAMAT	NO TELP	REVISI
1	NILA FATMA ALFA	SESUA BARU (SMA PUTRI GULU)	82323297911	
2	NURHANA ASHIM HUSNIA	SESUA BARU (SMA PUTRI GULU)	82386499287	
3	NILA NURHANA	SESUA BARU (SMA PUTRI PRESTASI)	82313302895	
4	AZULAN SAZIDAH PRANONO	SESUA BARU (SMA PUTRI PRESTASI)	82311883811	
5	INDITHY NUR AZULAN	SESUA BARU (SMA PUTRI GULU)	82389420989	
6	ZAKWEDAH SHAMSERAH HUSNIAK AL THOULU	SESUA BARU (SMA PUTRI PRESTASI)	82323297911	
7	SARDYAN BHINDA ARJUNINA	SESUA BARU (SMA PUTRI PRESTASI)	82323249869	
8	FAZKA SHIFA MEDINA	SESUA BARU (SMA PUTRI PRESTASI)	82323291811	
9	HAFIDIN ALFA REBEKHA SILANDITA	SESUA BARU (SMA PUTRI PRESTASI)	82386497359	
10	UMADHAN ABDURRAHMAN PRANSARI	SESUA BARU (SMA PUTRI SMPK)	82323299385	
11	IMAN NURDIAN PRANSI	SESUA BARU (SMA PUTRI PRESTASI)	82323291811	
12	M. HENRI FIDELIUSARAHAN	SESUA BARU (SMA PUTRI PRESTASI)	82323299385	
13	FALDAN ALFIDY FATMA	SESUA BARU (SMA PUTRI GULU)	82323291811	
14	RAHMAT SHAFIA AZZUMAR YUSUF	SESUA BARU (SMA PUTRI SMPK)	82323299481	
15	M. ALIHAQI RAHMAT TANAKA	SESUA BARU (SMA PUTRI SMPK)	82323291811	

Sumber: Jupriyanto dan Nurlela (2019)

Gambar 7. Data pelanggan untuk preparation penelitian

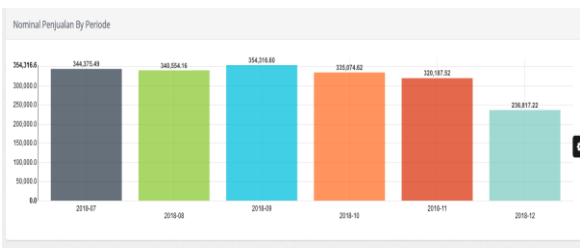
Dalam penelitian ini data pelanggan / member akan di gunakan sebagai subyek. Berdasarkan data pembelian yang ada di Syfamart, pelanggan / member akan ketahuan di cluster mana posisinya berada.

ID	TARIKH	WAKTU	NAMA BARANG	QTY	TIME STAMP	STATUS	STATUS BAYAR	PELUNGAN	KETERANGAN
1	2018-12-20	09:00:00	Min. paku 100ml	1,00	2018-12-20 09:00:00	2018-12-20	3000		
2	2018-12-20	09:00:00	Larutan Peningkat Kaki 3 200ml	1,00	2018-12-20 09:00:00	2018-12-20	3000		
3	2018-12-20	09:00:00	Yoghurt	1,00	2018-12-20 09:00:00	2018-12-20	3000	MAACHERA ALFITA CANTONO	
4	2018-12-20	09:00:00	Min. Biji di Krim Susu	1,00	2018-12-20 09:00:00	2018-12-20	3000	MAACHERA ALFITA CANTONO	
5	2018-12-20	09:00:00	Min. Biji di Krim Susu	1,00	2018-12-20 09:00:00	2018-12-20	4000	MAACHERA ALFITA CANTONO	
6	2018-12-20	09:00:00	Min. Biji di Krim Susu	1,00	2018-12-20 09:00:00	2018-12-20	3000	MAACHERA ALFITA CANTONO	
7	2018-12-20	09:00:00	Puding Biji Vanila 180g	1,00	2018-12-20 09:00:00	2018-12-20	3000	MAACHERA ALFITA CANTONO	
8	2018-12-20	09:00:00	Min. paku 100ml	1,00	2018-12-20 09:00:00	2018-12-20	3000	MAACHERA ALFITA CANTONO	
9	2018-12-20	09:00:00	Min. paku 100ml	1,00	2018-12-20 09:00:00	2018-12-20	3000		
10	2018-12-20	09:00:00	Min. paku 100ml	1,00	2018-12-20 09:00:00	2018-12-20	3000		
11	2018-12-20	09:00:00	Min. paku 100ml	1,00	2018-12-20 09:00:00	2018-12-20	3000		
12	2018-12-20	09:00:00	Min. paku 100ml	1,00	2018-12-20 09:00:00	2018-12-20	3000	NILA ALFA AZZUMARA PUTRI	
13	2018-12-20	09:00:00	Min. paku 100ml	1,00	2018-12-20 09:00:00	2018-12-20	3000	MESHA RANAZ AZ ZAHRA	
14	2018-12-20	09:00:00	Min. paku 100ml	1,00	2018-12-20 09:00:00	2018-12-20	3000	MESHA RANAZ AZ ZAHRA	
15	2018-12-20	09:00:00	Min. paku 100ml	1,00	2018-12-20 09:00:00	2018-12-20	7000	MESHA RANAZ AZ ZAHRA	

Sumber: (Jupriyanto. & Nurlela, 2019)

Gambar 8. Dataset penjualan Syfamart

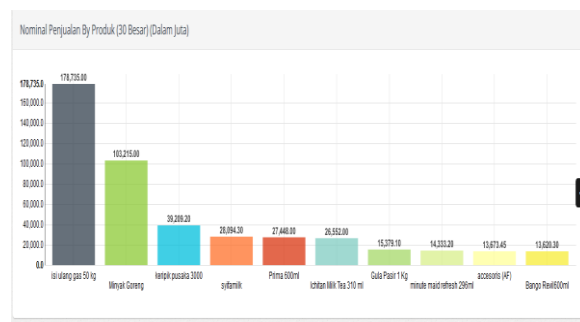
Dataset penjualan Syfamart yang di gunakan dalam penelitian ini adalah data penjualan periode July 2018 sampai dengan Desember 2018, dengan fokus penelitian adalah penjualan yang menggunakan smart card. Penjualan pelanggan umum di keluarkan dari dataset untuk kevalidan dalam membuat clustering pelanggan.



Sumber: (Jupriyanto. & Nurlela, 2019)

Gambar 9. Penjualan by periode bulan

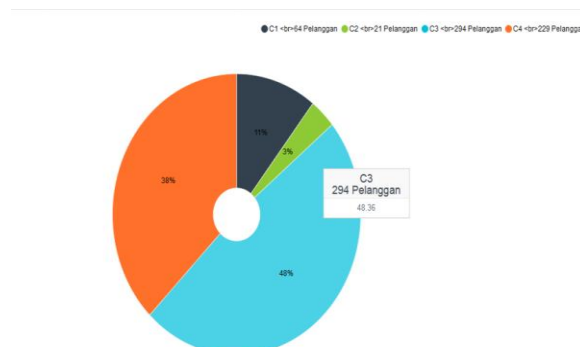
System / tool yang di buat, menampilkan historikal penjualan setiap bulannya untuk memudahkan tim Marketing atau manajemen dalam menentukan ketersediaan stok produk yang ada di Syfamart.



Sumber: (Jupriyanto. & Nurlela, 2019)

Gambar 10. Penjualan by produk

System / tool yang di buat juga menampilkan historikal penjualan setiap produk yang ada, untuk memudahkan tim Marketing atau manajemen dalam menentukan stok produk yang mana yang harus di prioritaskan ada di Syfamart.



Sumber: (Jupriyanto. & Nurlela, 2019)

Gambar 11. Jumlah Member Clustering Pelanggan

Jumlah member dalam masing-masing cluster di tampilkan dalam system / tool, sehingga tim marketing atau manajemen syfamart bisa dengan mudah membuat terobosan-terobosan baru dengan existing cluster yang ada.

PROFILLE PELANGGAN	
1. PELANGGAN	MOHAMMAD LANTIPRENGINS
2. RELAS	IS ALI EIN ARI THALIB (MAMARU)
3. NO TELP	82323232470
4. CLUSTER	C3
5. TYPE CLUSTER	Pengguna
6. KARAKTERISTIK CLUSTER	Pelanggan dalam Cluster 3 melakukan pembelian rutin, dengan nominal yg terbatas. Ini sangat menguntungkan bagi perusahaan. Pelanggan jenis ini juga dapat digunakan sebagai marketing perusahaan karena jenis pelanggan ini melakukan pemesanan dengan membunyikan tanda-tanda mereka untuk membeli barang / jasa di perusahaan.

Sumber: (Jupriyanto. & Nurlela, 2019)

Gambar 12. Profile Pelanggan / Member

Profile pelanggan di buat untuk memudahkan tim marketing atau manajemen Syifamart dalam memberikan service yang tepat kepada masing-masing pelanggan Syifamart.

KESIMPULAN

Proses data mining dari history transaksi penjualan 351,158 rows, kemudian dilakukan agregasi berdasarkan pelanggan menggunakan metode RFM. Setelah mendapatkan data RFM diekstraksi dengan menggunakan algoritma clustering k-means sehingga membentuk 4 (empat) cluster optimal. Keempat (empat) cluster tersebut kemudian diklasifikasikan dengan menggunakan algoritma decision tree untuk mengetahui setiap karakteristik pelanggan (customer profiling) sehingga perusahaan dapat mengetahui mana pelanggan potensial dan mana pelanggan yang tidak potensial.

Manajemen syifamart bisa meprediksi kebutuhan persediaan akan produk, penelitian ini menggunakan metode tren dimana penjualan / stok di bulan selanjutnya bisa di prediksi dengan menggunakan history penjualan di bulan sebelumnya.

Untuk memfasilitasi perusahaan / manajemen dalam mengambil keputusan hasil penelitian ini di implementasikan dengan cara membuat satu tool aplikasi sehingga baik tim marketing dan pihak manajemen bisa dengan mudah melihat hasil clustering dan bisa dengan mudah membuat rekomendasi model pemasaran yang terbaik berdasarkan dari hasil clustering dan profile pelanggan.

REFERENSI

Amborwati Armadyah, edi winarko. (2014). REVIEW PEMANFAATAN TEKNIK DATA MINING DALAM SEGMENTASI KONSUMEN. *Prosiding Seminar Ilmiah Nasional Komputer Dan Sistem Intelijen (KOMMIT 2014)*, 8(Kommit), 66-73.

Hardiani, T., Hartanto, R., & Mada, U. G. (2017). Segmentasi Nasabah Tabungan Menggunakan Model RFM (Recency , Frequency , Monetary) dan K-Means Pada Lembaga Keuangan Mikro, (May), 463-468.

Jupriyanto., & Nurlela, S. (2019). *Laporan Akhir Penelitian: Kerangka Pengambilan Keputusan Untuk Pemasaran Presisi Menggunakan Metode RFM, Algoritma K-Means Dan Decision Tree*. Jakarta.

Maryani, I., & Riana, D. (2017). Clustering and

profiling of customers using RFM for customer relationship management recommendations. *2017 5th International Conference on Cyber and IT Service Management*, 2-7. <https://doi.org/10.1109/CITSM.2017.8089258>

Tsitsis K, C., & A. (2009). *Data Mining Techniques in CRM*.

Wardani, N. W., Dantes, G. R., Indrawan, G., Studi, P., Informasi, T., Studi, P., ... Pelanggan, S. (2018). Prediksi Customer Churn Dengan Algoritma Decision Tree C4 . 5. *JURNAL RESISTOR*, 1(1), 16-24.

widiarina. (2013). Algoritma Cluster Dinamik Untuk Optimasi Cluster Pada Algoritma K-Means Dalam Pemetaan Nasabah Potensial Algoritma Cluster Dinamik Untuk Optimasi Cluster Pada Algoritma K-Means Dalam. *Prosiding SNATIF*, 1(1), 33-36.

You, Z., Si, Y. W., Zhang, D., Zeng, X., Leung, S. C. H., & Li, T. (2015). A decision-making framework for precision marketing. *Expert Systems with Applications*, 42(7), 3357-3367. <https://doi.org/10.1016/j.eswa.2014.12.022>