

SEGMENTASI PELANGGAN BERDASARKAN ANALISIS RFM (*RECENCY, FREQUENCY AND MONETARY INDEXES*) DAN ANALISIS DEMOGRAFI

Kasmari¹⁾, Taryadi²⁾

Universitas STIKUBANK¹⁾, STMIK Widya Pratama²⁾

fkasmari@edu.unisbank.ac.id¹⁾, tari_ball@stmik-wp.ac.id²⁾

Abstrak

Perusahaan mencoba mengenali kebutuhan masyarakat dan pelanggan dengan lebih tepat, salah satu metode yang digunakan adalah CRM. Untuk bisa membedakan kebutuhan banyak pelanggan dan menjalin interaksi antara produsen dan konsumen, harus memahami pelanggan melalui perilaku dan transaksinya. Salah satu metode yang digunakan untuk melakukan segmentasi pelanggan menggunakan model RFM dan metode klustering yaitu K-Means. Sementara demografi digunakan sebagai data pendukung untuk melakukan segmentasi pelanggan. Studi ini menunjukkan bahwa bobot atribut RFM mempengaruhi kinerja asosiasi aturan secara positif. Selain itu, untuk mendapatkan segmen pelanggan yang lebih akurat, disarankan untuk menggunakan kombinasi RFM tertimbang dan atribut demografis. Oleh karena itu, metodologi yang diusulkan menghasilkan hasil dan skor terbaik yaitu sebesar 0.284 dengan jumlah rule yang dikembangkan sebanyak 2491, sehingga pentingnya RFM tertimbang dan data demografi dalam cluster telah terbukti.

Kata kunci: segmentasi, klustering, RFM, demografi.

1. Pendahuluan

Pada tahun 1990-an, dalam bidang bisnis secara bertahap muncul konsep *customer relationship management* (CRM) yang sejak tahun pertama, CRM berlaku, menjadi terkenal sebagai bidang penelitian ilmiah dan meningkatkan minat komunitas bisnis (Soltani & Navimipour, 2016). CRM merupakan model operasional di mana perusahaan memahami dan mempengaruhi perilaku pelanggan melalui interaksi untuk mendapatkan pelanggan baru, mempertahankan pelanggan lama dan meningkatkan loyalitas dan keuntungan pelanggan (Chung & Chen, 2016). Menghadapi kompleksitas dan persaingan yang lebih besar dalam bisnis saat ini, perusahaan perlu mengembangkan aktivitas inovasi untuk memenuhi kebutuhan pelanggan dan meningkatkan kepuasan dan retensi pelanggan (Buraera, Kadir, Alam, 2014). Tujuan utama CRM adalah untuk menciptakan hubungan jangka panjang dan menguntungkan dengan pelanggan. Peningkatan digitalisasi transaksi menghasilkan peningkatan informasi pelanggan yang disimpan dalam database transaksi yang besar (Buraera, Kadir, Alam, 2014). Selain itu, ini adalah pendekatan terkuat dan paling efisien

untuk mempertahankan dan menciptakan hubungan dengan pelanggan (Soltani & Navimipour, 2016).

Untuk memaksimalkan kinerja CRM, perusahaan perlu memiliki kelompok audien dan pelanggan yang tersegmentasi. Segmentasi pertama kali diperkenalkan ke dalam literatur pemasaran oleh Smith (W. Smith 1956). Kemudian, Claycamp dan William (1968), menyebut segmentasi sebagai konsep alternatif dibandingkan strategi diferensiasi produk. Ide utama dari segmentasi atau pengelompokan adalah mengelompokkan pelanggan serupa.

Clustering merupakan proses pengelompokan sekumpulan objek fisik atau abstrak ke dalam kelompok objek serupa (Hosseini & Mohammadzadeh, 2016). Aplikasi pemasaran yang menonjol adalah pengelompokan pelanggan untuk pemasaran langsung, di mana analisis mencoba menemukan kelompok pelanggan yang homogen sehubungan dengan perilaku respons pelanggan melalui data mining (Berry and Linoff, 2018). Dalam hal ini, beberapa penulis mengusulkan metode pengelompokan secara lunak, yang memberikan hasil yang lebih menjanjikan daripada pengelompokan secara keras dan kualitas pengelompokan dalam segmen

yang lebih baik daripada model campuran hingga. Wu dan Cho membandingkan distribusi pelanggan untuk pengelompokan keras dan pengelompokan lunak. Dalam pengelompokan lunak, setiap pelanggan memiliki skor keanggotaan campuran yang terkait dengan setiap kelas laten. Peneliti menekankan pada (1) Segmen mikro untuk perilaku belanja (frekuensi pembelian dan uang yang dikeluarkan), (2) Segmen mikro untuk kepuasan pelanggan online dan (3) Demografi karakteristik karakteristik pelanggan berdasarkan hasil eksperimen (Coussement, Van den Bossche, & De Bock, 2014).

Fallis (2013) menyebutkan bahwa ada banyak algoritma clustering dalam literatur. Sulit untuk memberikan kategorisasi metode pengelompokan yang jelas karena kategori-kategori ini mungkin tumpang tindih sehingga suatu metode yang memiliki fitur dari beberapa kategori dan secara umum, metode pengelompokan dasar utama dapat diklasifikasikan sebagai metode Partisi (seperti algoritma pengelompokan k-Means), Metode hierarki (seperti algoritma Clustering Hierarki Agglomeratif dan Divisif), Metode berbasis kepadatan (Pengelompokan Berbasis Kepadatan Berdasarkan Wilayah Terhubung dengan Kepadatan Tinggi dan Pengelompokan Berdasarkan Fungsi Distribusi Kepadatan adalah metode yang paling populer) dan metode berbasis Grid (Fallis, 2013). Sebagai algoritma pengelompokan lain yang efektif, peta kohonen yang mengatur dirinya sendiri yang terkenal karena kemampuannya memetakan ruang masukan dengan jaringan saraf (Kantardzic, Mehmed, 2011; Monalisa, S. 2017) sangat populer untuk pengelompokan pelanggan saat ini.

Selama beberapa tahun terakhir, teknik pemasaran basis data telah berevolusi dari model RFM (Recency, Frequency and Monetary indexes) sederhana (model yang melibatkan Keterkinian dan Frekuensi pembelian pelanggan, dan pembayaran) ke teknik statistik seperti deteksi interaksi otomatis chisquare (CHAID) dan regresi logistik. Baru-baru ini, model yang menggunakan jaringan saraf tiruan telah digunakan oleh perusahaan untuk melayani CRM melalui analisis informasi pelanggan yang berharga (Cermak, 2015.; Onur, Ejder, and

ZekiAtil, 2018). Mudah digunakan dan implementasi yang cepat adalah alasan pemasar terus menggunakan model RFM meskipun ada teknik penambangan data yang canggih. Selain itu, mudah dimengerti oleh manajer dan pengambil keputusan. Dari sudut pandang lain, ada beberapa titik buta untuk model RFM. Terutama, kurangnya kemampuan eksplorasi dan pandangan ke depan dari metode ini, itulah poin negatif untuk model prediktif. Kedua, hal ini memerlukan perilaku berkelanjutan bagi setiap pelanggan untuk menargetkan pelanggan untuk promosi. Dengan kata lain, analisis ini tidak memperhitungkan dampak transisi tahap kehidupan atau siklus hidup terhadap kemungkinan respons. Terakhir, jika hanya menggunakan segmen RFM yang paling menarik sebagai metode penargetan utama, ada risiko mengabaikan segmen lain yang mungkin menguntungkan.

Penelitian ini mengusulkan algoritma berbeda dalam menggunakan keluaran analisis RFM pelanggan dan data demografis pelanggan ke dalam bentuk skenario yang berbeda. Untuk mengekstraksi aturan berbasis produk dari transaksi pelanggan, segmentasi pelanggan perlu dilakam dalam bentuk kelom sasaran yang tepat. Saat ini segmentasi pelanggan berdasarkan analisis RFM hanya mempertimbangkan indeks kekinian, fekuensi, dan moneter. Dengan kata lain, untuk memodelkan pengelompokan pelanggan menggunakan analisis RFM, faktor pembobotan dan beberapa atribut lain seperti variabel demograsi tidak boleh diabaikan. Tentunya analisis kinerja semua jenis skenario akan mendukung gagasan bahwa kebersamaan atribut transaksi pelanggan untuk menjalan aturan berbasis klaster yang sebenarnya diperlukan.

2. Metode Penelitian

2.1 Karakteristik Data

Pelanggan mempunyai berbagai perbedaan yang ditunjukkan oleh karakteristiknya. Dalam literatur pemasaran konsumen dan industri, dapat ditemukan beberapa variabel segmentasi, seperti geografis, demografi, firmografi, perilaku, variabel terkait proses pengambilan keputusan, perilaku pembelian, faktor situasi, kepribadian, gaya hidup, psikografis, dan sebagainya (He and Li, 2016).

2.2 Data Demografi

Demografi adalah statistik pasti untuk kumpulan data populasi pelanggan. Selain itu, demografi digunakan untuk membedakan studi tentang subset yang dapat diukur. Pemasaran dan jajak pendapat publik merupakan bidang lain yang menggunakan data demografi. Umumnya, etnis, jenis kelamin, usia, status pekerjaan, dan lain-lain dianggap sebagai data demografi. Deskripsi perubahan demografi yang berorientasi waktu dalam suatu populasi disebabkan oleh tren demografi. Misalnya, rata-rata tingkat pembelanjaan suatu populasi dapat meningkat atau menurun seiring berjalannya waktu. (Selvi and Ravi, 2013)

2.3 Data RFM

RFM adalah model yang mendiversifikasi banyak pelanggan dari sejumlah besar data dengan tiga atribut. Interval konsumsi pelanggan, frekuensi dan nilai pembayaran, merupakan ketiga faktor tersebut. Definisi rinci RFM dijelaskan sebagai berikut:

(1) Kekinian pembelian terakhir yang ditunjukkan oleh R adalah jangka waktu antara waktu pembelian terakhir dengan waktu survei tertentu. Disini state yang diinginkan adalah durasi waktu yang lebih pendek yang berarti R lebih besar.

(2) Frekuensi pembelian yang ditunjukkan oleh F menunjukkan jumlah transaksi dalam satu siklus waktu tertentu. keadaan yang diinginkan semakin besar F yang berarti tingginya pengulangan pembelian.

(3) Nilai moneter pembelian yang ditunjukkan oleh M adalah nilai konsumsi uang oleh pelanggan selama jangka waktu tertentu. Keadaan yang diinginkan disini adalah semakin banyak uang maka semakin besar M. Meskipun model RFM adalah metode yang baik yang membedakan pelanggan penting dari data besar dengan tiga variabel, terdapat dua penelitian yang mempunyai pendapat berbeda sehubungan dengan ketiga variabel model RFM. Buttle (2014) menganggap bahwa ketiga variabel tersebut memiliki kepentingan yang sama; oleh karena itu, bobot ketiga variabel tersebut identik. Di sisi lain, Venkatesan dan Kumar (2014) menunjukkan bahwa ketiga variabel tersebut memiliki tingkat kepentingan yang berbeda karena karakteristik industrinya. Dengan demikian, bobot ketiga

variabel tersebut tidak sama (Chung & Chen, 2016).

Moneter dapat diganti dengan durasi, yang dapat digunakan untuk menganalisis perilaku pelanggan, misalnya nilai waktu yang dihabiskan oleh pemirsa radio di saluran tertentu. Sebagian besar bisnis akan menyimpan transaksi pelanggan. Yang diperlukan hanyalah tabel dengan nama pelanggan, tanggal pembelian dan nilai pembelian. Salah satu metodologinya adalah dengan menetapkan skala 1 sampai 10, dimana 10 adalah nilai maksimum dan menetapkan rumus yang dapat digunakan untuk menyesuaikan data dengan skala tersebut.

2.4 Metodologi yang Diusulkan

Dalam beberapa tahun terakhir, data mining tidak hanya mendapatkan popularitas besar di bidang penelitian tetapi juga dalam komersialisasi. Saat ini, dengan memanfaatkan alat data mining untuk membantu CRM, beberapa teknik, yang meliputi *Decision Tree*, *Artificial Neural Network*, *Genetic Algorithm*, *Association Rule Mining* dan lain-lain, biasanya digunakan di beberapa bidang seperti teknik, sains, keuangan dan bisnis, untuk memecahkan masalah terkait dengan pelanggan. Secara umum, tidak ada alat data mining di CRM yang sempurna, karena ada beberapa kekurangan yang tidak pasti di dalamnya. Misalnya, dalam DT, terlalu banyak instance menyebabkan pohon keputusan berukuran besar dan menurunkan tingkat akurasi klasifikasi. Di ANN, khususnya, waktu pelatihannya lama pada kumpulan data yang besar, dan ini merupakan proses coba-coba. Di GA, konvergensinya lambat, waktu komputasinya lama, dan stabilitasnya kurang. Di ARM, hal ini dapat menghasilkan aturan besar yang mungkin merupakan redundansi (Chung & Chen, 2016).

Salah satu tantangan utama bagi setiap pemasar perusahaan adalah mengidentifikasi target bisnis produk atau layanan yang diberikan. Semua pelanggan sebagai target bisnis memiliki profil dan preferensi tertentu. Untuk mengembangkan strategi pemasaran dan rencana pemasaran, profil ini dapat dibuat dan digunakan. Dalam hal ini pertama-tama mengelompokkan komunitas dalam bentuk pelanggan yang dikelompokkan, dan kemudian, mengekstraksi aturan pendukung keputusan yang tepat dengan

menggunakan fitur-fitur cluster yang khas. Algoritma untuk pendekatan segmentasi berbasis ARM dipilih untuk melakukan cluster penambangan aturan. Untuk mendapatkan aturan yang paling efektif dengan redundansi yang lebih sedikit, berbagai kombinasi model skenario telah dirancang, dan performanya akan dievaluasi dengan algoritma pembelajaran mesin.

Dengan menggunakan kemampuan algoritma di atas untuk mengembangkan prosedur CRM berbasis ARM yang logis dan andal. Metodologi utama untuk transaksi pelanggan berbasis ARM berisi lima fase:

1) Fase Data Preprocessing

Tahap preprocessing data yaitu tahap pertama penyiapan data mentah menjadi data yang telah disempurnakan dan dapat digunakan. Pada fase ini dilakukan operasi Pengisian, Penanganan, Transformasi dan Diskritisasi pada data mentah. Langkah penyempurnaan ini mencakup operasi seperti pengaturan nomor atribut, deteksi outlier, normalisasi, diskritisasi, dan pembuatan hierarki konsep. Hal ini mempengaruhi secara langsung keakuratan prediksi dan durasi waktu yang berlalu ke arah yang positif. Untuk itu dilakukan hal-hal sebagai berikut:

- a) Pengurangan Dimensi: Atribut yang tidak perlu harus dihapus, seperti atribut yang hanya memiliki sedikit nilai (yang lainnya null) atau hanya memiliki nilai tunggal.
- b) Pengisian: Nilai yang hilang harus diisi menggunakan pendekatan yang tepat. Dalam percobaan ini, akan mengganti nilai yang bergantung pada Kondisi, Tipe, dan Null dengan ekspresi tertentu menggunakan MATLAB, SPSS.
- c) Penanganan: Nilai outlier dan tidak akurat harus ditangani dan dihapus dari kumpulan data.
- d) Transformasi: Data harus diubah ke dalam format yang sesuai.
- e) Diskritisasi: Sebelum tugas penambangan aturan asosiasi, atribut kontinu harus dikodekan dengan mendiskritisasi nilai asli ke dalam sejumlah kecil rentang nilai. Karena nilai-nilai tersebut hampir berbeda untuk setiap kasus; dengan kardinalitas yang begitu tinggi, tidak memberikan banyak arti pada proses penambangan aturan asosiasi (Asroni, Adrian, 2015).

2) Fase Desain Skenario

Output dari tahap ini adalah data transaksi yang disempurnakan. Selanjutnya menyiapkan database utama untuk analisis. Karena data berisi faktor-faktor tertentu dan gagasan utama dari pekerjaan ini terletak pada alur analisis faktor. Jadi untuk mendapatkan aliran terbaik dan akurat, pada tahap kedua metodologi ini, mencoba mengembangkan seluruh analisis skenario yang mungkin. Setiap skenario individual merupakan kombinasi rangkaian faktor dan spesifikasinya yang berbeda-beda. Pada fase ini, berbagai kombinasi analisis RFM, Segmentasi, dan ARM telah dirancang dan disimpan sebagai kumpulan Skenario Analisis.

Atribut dan tipe data yang berbeda dalam setiap transaksi akan mengarahkan untuk menyediakan tipe skenario komprehensif yang mencakup semua kemungkinan konfigurasi melalui aplikasi data mining dasar. Analisis RFM memiliki tujuh keluaran berbeda; Nilai RFM (Keterkinian, Frekuensi, dan Moneter) dan skor RFM (Skor Kekinian, Skor Frekuensi, dan Skor Moneter) juga merupakan rata-rata skor RFM (rata-rata tertimbang dari skor RFM). Sementara itu, nilai RFM dikelompokkan dalam lima kategori skor dari 1 hingga 5. Selain itu, terdapat kombinasi pembobotan kepentingan yang berbeda untuk analisis Weighted RFM (WRFM) yang digunakan dalam percobaan ini (Coussement et al., 2014).

3) Fase Analisis

Pada tahap analisis, setiap skenario dijalankan dan dianalisis. Hasil analisis tersebut disimpan pada Kumpulan Aturan Berdasarkan Skenario. Fase ini berisi kinerja tiga langkah tetap yaitu Analisis RFM, Segmentasi, dan Analisis ARM. Namun, setiap analisis dijalankan dalam skenario yang berbeda.

a) Analisis RFM

Analisis RFM adalah teknik pemasaran yang digunakan untuk menganalisis perilaku pelanggan seperti kapan pelanggan membeli akhir-akhir ini (kekinian), seberapa sering pelanggan membeli (frekuensi), berapa jumlah pembelian yang dibelanjakan pelanggan (moneter). Metode ini berguna untuk meningkatkan segmentasi pelanggan dengan membagi pelanggan ke dalam berbagai kelompok

untuk layanan personalisasi di masa depan dan untuk mengidentifikasi pelanggan yang kemungkinan besar memiliki kecenderungan untuk merespons promosi. Analisis RFM bergantung pada ukuran Kekinian (R), Frekuensi (F), dan Moneter (M) yang merupakan tiga variabel penting terkait Pembelian yang berpengaruh terhadap kemungkinan pembelian di masa depan. Perhitungan terkait ditunjukkan seperti di bawah ini:

$$R(C_i) = \frac{R_i - R_{Min}}{R_{Max} - R_{Min}} \quad (1)$$

$$F(C_i) = \frac{F_i - F_{Min}}{F_{Max} - F_{Min}} \quad (2)$$

$$M(C_i) = \frac{M_i - M_{Min}}{M_{Max} - M_{Min}} \quad (3)$$

b) Segmentation

Transaksi pembelian dan pelacakan perilaku pembelian pelanggan telah tersegmentasi dalam fase segmentasi pelanggan. Segmentasi pelanggan adalah metodologi penambahan data penting yang digunakan dalam CRM. Ada banyak pendekatan dan algoritma pengelompokan dalam literatur yang akan diperkenalkan dan dibandingkan dua di antaranya untuk memilih algoritma terbaik seperti di bawah ini:

(1) K-Means Clustering Algorithm

Metode pengelompokan K-means menerima D sebagai kumpulan data masukan, sejumlah K segmen yang akan dibentuk, dan fungsi $dist(X_i, X_k)$ yang menunjukkan semacam homogenitas antara setiap pasangan inspeksi.

Misalkan C_h , $h = 1, 2, \dots, K$ adalah cluster yang pusat massanya didefinisikan sebagai titik z_h yang koordinatnya sama dengan mean setiap fitur untuk pengamatan yang termasuk dalam cluster tersebut, yaitu:

$$Z_{ij} = \frac{\sum_{X_i \in C_h} X_{ij}}{card\{C_h\}} \quad (4)$$

3. Hasil dan Pembahasan

Kumpulan data transaksi pelanggan kehidupan nyata yang dievaluasi dengan algoritma yang diusulkan untuk mencapai pola perilaku pembelian pelanggan terbaik dan ekstraksi aturan asosiasi. Untuk ini, rangkaian

aliran yang berbeda dan beragam atribut diperiksa sebagai tipe skenario. Seluruh jenis 42 skenario diterapkan, dijalankan, dan dievaluasi oleh komputer mikroprosesor Intel dual core i7 dan semua kemungkinan hasil diuraikan dalam Tabel 4. Tingkat kluster pelanggan, dukungan, dan kepercayaan telah diatur sebelumnya sama untuk semua jenis skenario, namun jumlah cluster dikurangi menjadi 3 atau 2 pada beberapa skenario. Hal ini disebabkan oleh ketidakmampuan algoritma yang tertanam pada jenis skenario tertentu sehingga tidak mungkin untuk melihat perbedaan dengan algoritma clustering K-Means.

Melalui jenis skenario yang berbeda, jumlah aturan bervariasi antara 329 dan 2491 dan diinginkan jumlah aturan yang lebih tinggi. Waktu berlalu juga berkisar antara 7 hingga 223 menit dan diinginkan waktu berlalu yang lebih singkat. Berdasarkan evaluasi karena skenario 25 memiliki aturan yang paling rinci dengan jumlah cluster yang paling tepat dan waktu berlalu yang lebih singkat, skenario ini terbukti menjadi pilihan terbaik dengan mengacu pada metodologi yang diusulkan (kebersamaan faktor demografi dan RFM saat pengelompokan, menyebabkan aturan asosiasi yang lebih akurat).

Langkah perbandingan telah diselesaikan dengan menggunakan jaringan saraf untuk memeriksa tingkat efektivitas konten cluster untuk memprediksi 5 produk terlaris. Semua tingkat efektivitas untuk lima produk yang sangat direkomendasikan oleh algoritma asosiasi aturan telah dicatat. Semua tingkat efektivitas yang tercatat ini dirata-ratakan dan diilustrasikan pada Tabel 5. Jenis skenario 23, 24 dan 25 memiliki rata-rata tingkat efektivitas yang lebih baik dengan menggunakan NN. Meskipun rata-rata skenario tipe 25 secara eksplisit merupakan yang terbaik. Hasilnya, skenario tipe 25 merupakan skenario terbaik berdasarkan perbandingan Tabel 1 dan 3. Skenario (stream) 25 menekankan pada suatu algoritma, yaitu mengelompokkan pelanggan tidak hanya dengan data RFM tetapi juga komposisi skor RFM dan data Demografi. Selain itu, jenis skenario ini menunjukkan bahwa moneter adalah faktor terpenting di antara hasil analisis RFM untuk pengelompokan.

Tabel 1. Hasil Kinerja dari beberapa skenario

Skenario (1-22)	Jumlah Kluster	Jumlah Rule	Waktu (min)	Skenario (23-42)	Jumlah Kluster	Jumlah Rule	Waktu (min)
1	5	647	127	22	5	517	7
2	5	763	119	23	5	1517	9
3	5	547	113	24	5	1517	8
4	5	547	15	25	5	2491	9
5	5	1647	220	26	5	1175	7
6	5	1647	223	27	5	613	8
7	5	1647	112	28	5	613	10
8	5	1555	35	29	2	522	22
9	5	1555	34	30	5	329	24
10	5	1555	35	31	5	1653	27
11	5	1555	31	32	5	1747	25
12	5	1555	33	33	5	1062	25
13	5	1555	33	34	5	1060	24
14	5	1555	32	35	3	1056	23
15	5	517	10	36	4	517	22
16	5	517	11	37	4	517	24
17	5	517	14	38	4	517	22
18	5	517	13	39	4	517	26
19	5	517	14	40	4	517	26
20	5	517	15	41	4	611	19
21	5	517	15	42	4	423	18

Sebagai hasil dari Association Rules Mining, skenario terbaik (No. 25) telah mengekstraksi 2491 aturan. Aturan-aturan tersebut berbentuk anteseden dan konsekuensi. Antesedennya menunjukkan berbagai faktor dan perilaku pembelian pelanggan. Terdapat 5 cluster berbeda yang terdiri dari keluaran analisis RFM dan data demografi (Isi cluster ditunjukkan pada Tabel 2). Konsekuensinya menunjukkan produk yang telah dibeli oleh pelanggan dengan anteseden tersebut. Pada Tabel 2 menunjukkan sebagian hasil aturan sebagai contoh, khususnya untuk produk yang paling banyak dibeli.

Tabel 2 mendapatkan aturan signifikan untuk mengidentifikasi konsumen produk tertentu dari Skenario No.25 yang merupakan skenario terbaik (di antara 2491 sampel).

Tabel 2. Mendapatkan aturan signifikan untuk mengidentifikasi konsumen produk tertentu dari Skenario No.25 yang merupakan skenario terbaik (di antara 2491 sampel).

Rule	Antecedent	Consequent	Support %	Confidence %
1	Cluster=1, Beverage 2=Yes, Piza 3=Yes	Var0053= Yes	89,8	99,9
2	Cluster=2, Beverage 23=Yes, Piza 11=Yes	Var0065= Yes	83,0	98,7
3	Cluster=5, Beverage 11=Yes, Piza 2=Yes	Var0055= Yes	82,11	97,1

4	Cluster=5, Beverage 7=Yes, Piza 8=Yes	Var0011= Yes	69,21	99,2
5	Cluster=4, Beverage 7=Yes, Piza 2=Yes	Var0039= Yes	70,43	80,42

Cluster 1= (skor R=4, skor F=1, skor M=1, Kota=5, JK:2,umur:2)
Cluster 2= (skor R=5, skor F=2, skor M=2, City:62, JK:1,umur:3)
Cluster 3= (skor R=5, skor F=2, skor M=2, City:12, JK:1,umur:3)
Cluster 5= (skor R=5, skor F=2, skor M=2, City:1, JK:2,umur:1)
Cluster 4= (skor R=5, skor F=2, skor M=2, City:43, JK:2,umur:3)

Tabel 2. Hasil analisis cluster

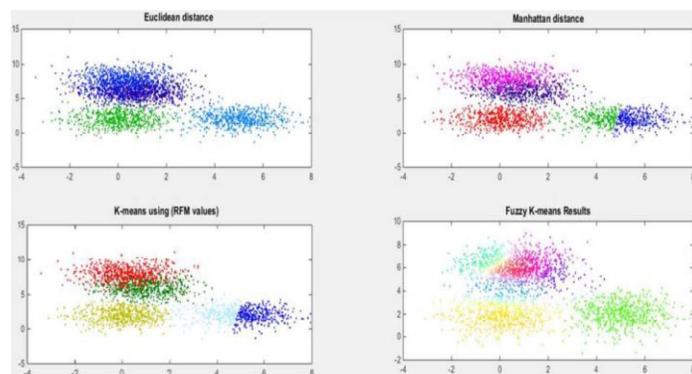
Skenario	Efektifitas produk 0065	Efektifitas produk 0055	Efektifitas produk 0011	Efektifitas produk 0039	Efektifitas produk 0053	Rata2 tingkat efektifitas
1	0.044	0.047	0.060	0.112	0.090	0.070
2	0.060	0.047	0.063	0.049	0.086	0.061
3	0.016	0.132	0.045	0.078	0.035	0.061
4	0.004	0.022	0.032	0.020	0.017	0.017
5	0.013	0.090	0.029	0.001	0.001	0.031
6	0.009	0.019	0.032	0.032	0.018	0.022
7	0.045	0.021	0.028	0.030	0.023	0.029
8	0.006	0.030	0.035	0.062	0.051	0.036
9	0.009	0.047	0.019	0.035	0.044	0.030
10	0.004	0.018	0.037	0.043	0.046	0.019
11	0.008	0.024	0.033	0.038	0.036	0.026
12	0.001	0.016	0.027	0.035	0.050	0.026
13	0.024	0.028	0.026	0.046	0.047	0.034
14	0.003	0.052	0.044	0.030	0.039	0.034
15	0.026	0.014	0.030	0.014	0.024	0.021
16	0.014	0.028	0.021	0.023	0.018	0.021
17	0.018	0.026	0.026	0.012	0.016	0.020
18	0.019	0.019	0.016	0.020	0.035	0.024
19	0.062	0.098	0.013	0.030	0.024	0.045
20	0.091	0.026	0.023	0.033	0.025	0.039
21	0.012	0.010	0.028	0.024	0.021	0.019
22	0.030	0.038	0.016	0.027	0.025	0.027
23*	0.408	0.031	0.012	0.022	0.195	0.133
24*	0.143	0.030	0.014	0.030	0.424	0.128
25**	0.148	0.045	0.177	0.029	0.284	0.136
26	0.003	0.021	0.007	0.012	0.023	0.013
27	0.005	0.029	0.001	0.034	0.024	0.019
28	0.001	0.033	0.019	0.018	0.028	0.020
29	0.021	0.001	0.052	0.006	0.012	0.018
30	0.001	0.023	0.047	0.046	0.058	0.035
31	0.036	0.038	0.037	0.024	0.135	0.054
32	0.041	0.113	0.030	0.019	0.105	0.062
33	0.011	0.022	0.025	0.021	0.061	0.029
34	0.019	0.023	0.017	0.029	0.014	0.022
35	0.068	0.012	0.031	0.021	0.115	0.046
36	0.063	0.023	0.031	0.008	0.021	0.029
37	0.001	0.028	0.014	0.031	0.02	0.019
38	0.049	0.011	0.034	0.035	0.015	0.029
39	0.013	0.026	0.011	0.022	0.014	0.017
40	0.201	0.013	0.031	0.016	0.014	0.019
41	0.024	0.016	0.01	0.024	0.017	0.018
42	0.021	0.017	0.019	0.022	0.018	0.019

3.1 Benchmarking

Untuk meningkatkan keunggulan pekerjaan, skenario ke-25 dari metodologi yang diusulkan yang disimpulkan sebagai jenis skenario terbaik telah dibandingkan dengan pekerjaan serupa yang

telah dilakukan akhir-akhir ini. Algoritma berbeda dari karya serupa telah diterapkan pada kumpulan data. Hasil dari metode yang diusulkan dan hasil tolok ukurnya terdapat di Tabel 4.

Dursun A. dan Caber M., (2016) berfokus pada pembuatan profil pelanggan hotel yang menguntungkan dengan analisis RFM (Dursun & Caber, 2016). Penelitian menggunakan skor RFM ekuiponderansi ($WR=WF=WM$). Pendekatan tersebut digunakan pada fase segmentasi skenario ke-8. Ghodousi et al., (2016), mengumpulkan kumpulan informasi dari kabupaten kota di suatu kota dan menganalisis untuk memprioritaskan kebutuhan perkotaan, juga menilai kepuasan warga (Ghodousi, Alesheikh, & Saeidian, 2016). Empat puluh tiga kebutuhan warga diidentifikasi dan dikategorikan berdasarkan pengelompokan K-means dan Metode Fuzzy Clustering (FCM) berdasarkan nilai RFM yang diberi bobot yang sama ($WR=WF=WM$). Margianti et al., (2016) pada penelitiannya yaitu proses data mining dengan metode yang digunakan adalah clustering dengan model Affinity Propagation dan Recency Frekuensi dan Moneter (RFM) pada 1.000 data pelanggan. Digunakan metode jarak dari algoritma Affinity Propagation dengan dua teknik Euclidean Distance dan Manhattan Distance untuk memudahkan perusahaan menganalisis data transaksi pelanggan (Margianti et al., 2016).



Gambar 1. Grafik klustering dengan teknik benchmarking

3.2 Implikasi

Metode pemasaran modern menggunakan berbagai kriteria untuk mengelompokkan klien, menargetkan kampanye, dan pesan publisitas untuk produk dan layanan (Nettleton, 2014). Memperoleh informasi tentang karakteristik pelanggan, dan menganalisis data yang dicatat sebelumnya untuk menghasilkan aturan asosiasi dan strategi CRM menjadi penting bagi perusahaan karena persaingan di seluruh dunia. Analisis data mining memungkinkan manajer untuk 'menemukan potensi permintaan pelanggan sasaran dengan mengamati perilaku di masa lalu. Studi saat ini menekankan pengelompokan yang merupakan alat penambangan data yang sangat berguna untuk mengurangi kompleksitas kumpulan data transaksional yang sangat besar. Dalam hal ini, keakuratan kluster jalan merupakan faktor penting lainnya yang mempengaruhi secara langsung efektivitas peraturan perkumpulan. Untuk mencapai tingkat akurasi cluster yang dapat diterima, analisis RFM, khususnya, ditambahkan sebagai faktor tambahan pada analisis clustering pelanggan. Karena tingkat efektivitas kluster pelanggan sangat penting, seseorang mungkin perlu memiliki lebih banyak atribut untuk membuat kluster yang diperoleh diperkaya sebanyak mungkin. Data demografi pelanggan adalah faktor lain yang mudah digunakan dan penting yang umumnya mudah diakses oleh perusahaan. Menariknya, data demografi pelanggan meningkatkan tingkat efektivitas dan kualitas kluster dalam aturan yang diekstraksi sementara data ini diintegrasikan dengan faktor skor RFM.

Tabel 4. Kesimpulan hasil benchmarking

	Metode Segmentation	Metode Klustering	Jumlah Klater	Jumlah Rule	Efektivitas
Dursun & Caber, 2016	RFM Skor	K-means	5	1555	0.0519
Ghodousi dkk, 2016	RFM Value	K-means	5	517	0.0465
Margianti, dkk, 2016	RFM Value	Fuzzy K-means	4	1804	0.0097
Metode yang diusulkan	Skor+ RFM	K-means	5	2491	0.284

Jelas bahwa waktu yang digunakan (9 menit), tingkat efektivitas faktor pengelompokan (0,284) dan jumlah aturan (2491) untuk metode yang diusulkan lebih baik daripada pendekatan lainnya. Gambar 1 menyajikan hasil grafis untuk kinerja pengelompokan yang mengadopsi algoritma benchmark untuk kumpulan data menggunakan program MATLAB.

Secara umum, perusahaan menggunakan aturan asosiasi ini untuk membuat penawaran dan kampanye bagi pelanggan saat ini atau calon pelanggan (*Gamification*). Penawaran yang tidak relevan, salah atau pesan yang sering dikirimkan kepada pelanggan akan menurunkan kesan kampanye yang menyebabkan kerugian finansial dan pelanggan. Dengan menggunakan metodologi yang diusulkan untuk pengelompokan pelanggan dan konsekuensinya penambangan aturan asosiasi, pesan yang relevan dan benar akan disampaikan kepada pelanggan sasaran dengan keyakinan lebih tinggi dan lebih tepat.

4. Kesimpulan dan Saran

Perusahaan mencoba mengenali kebutuhan masyarakat dan pelanggan dengan lebih tepat. Jadi untuk membedakan kebutuhan banyak pelanggan dan menjalin interaksi antara produsen dan konsumen, harus memahami pelanggan melalui perilaku dan transaksinya. Dalam hal ini, perlu mengungkap sejumlah besar data pelanggan untuk mendapatkan keunggulan kompetitif. Dalam percobaan ini, mencari dan menguji pendekatan segmentasi pelanggan aliran terbaik berdasarkan RFM dan atribut demografis. Menantang dampak RFM dan atribut demografi untuk memperkaya segmen pelanggan yang dihasilkan dari pengelompokan faktor dan mengekstraksi aturan asosiasi yang lebih andal adalah gagasan utama penelitian ini.

Dalam evaluasi ini, dampak atribut demografis dalam memperkaya faktor segmentasi pelanggan diuji. Berbagai jenis skenario telah dirancang, dilakukan, dan dievaluasi dengan cermat dalam kondisi pengujian yang sama. Metodologi yang efektif dipasang berdasarkan tahap desain skenario yang komprehensif. Hasilnya dievaluasi dan dibandingkan menggunakan teknik pembelajaran mesin yang efisien. Semua hasil tes dicatat dan dirangkum. Studi ini menunjukkan bahwa bobot atribut RFM mempengaruhi kinerja asosiasi aturan secara positif. Selain itu, untuk mendapatkan segmen pelanggan yang lebih akurat, disarankan untuk menggunakan kombinasi RFM tertimbang dan atribut demografis. Oleh karena itu, metodologi yang diusulkan menghasilkan hasil dan skor terbaik,

sehingga pentingnya RFM tertimbang dan data demografi dalam cluster telah terbukti. Meskipun menggunakan data demografi dalam jumlah terbatas sebagai atribut, analisis hasil menunjukkan pentingnya faktor demografi yang digabungkan dengan data RFM.

Untuk percobaan di masa depan, dengan memperluas cakupan atribut akan menggunakan data firmografi dan perilaku pelanggan dalam fase pengayaan input. Selain itu, untuk tahap penyelesaian, mengingat konsep pembelajaran mesin baru akan memiliki peran yang lebih besar sambil menggabungkan teknik pengelompokan dan ekstraksi aturan yang berbeda, sehingga metodologi saat ini akan ditingkatkan dengan menggunakan algoritma pengelompokan metaheuristik dan aturan asosiasi yang dapat meningkatkan kecepatan pelatihan, waktu yang berlalu, dan kemampuan alat penambangan data untuk mengekstraksi aturan yang lebih kuat dan lebih komprehensif.

DAFTAR PUSTAKA

- Astroni, Ronald Adrian. 2015. Penerapan Metode KMeans Untuk Clustering Mahasiswa Berdasarkan Nilai Akademik Dengan Weka Interface Studi Kasus Pada Jurusan Teknik Informatika UMM Magelang. *Jurnal Ilmiah Semesta Teknika*. Vol. 18 (1), 76-82.
- Berry, M. J. A. and Linoff, G. S., 2018. *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management*. Second edition.
- Buttle. 2014. *Customer Relationship Management*. Malang: Bayu Media.
- Cermak P. 2015. *Customer Profitability Analysis and Customer Life Time Value Model: Portfolio Analysis*. University of Economics, Prague.
- Chung, Y.C., & Chen, S.J. (2016). Study on customer relationship management activities in Taiwan tourism factories. *Total Quality Management & Business Excellence*, 27(5-6), 581-594. doi:10.1080/14783363.2015.1019341
- Coussement, K., Van den Bossche, F. A. M., & De Bock, K. W. (2014). Data accuracy's impact on segmentation performance: Benchmarking RFM analysis, logistic regression, and decision trees. *Journal of*

- Business Research*, 67(1), 2751–2758. doi:10.1016/j.jbusres.2012.09.024
- Dursun, A., & Caber, M. (2016). Using data mining techniques for profiling profitable hotel customers: An application of RFM analysis. *Tourism Management Perspectives*, 18, 153–160. doi:10.1016/j.tmp.2016.03.001
- Fallis, A. (2013). *No Title No Title*. *Journal of Chemical Information and Modeling* (Vol. 53). doi:10.1017/CBO9781107415324.004
- Farris P. W., Bendle N. T., Pfeifer P. E., and Reibstein D. J. 2008. *Marketing Metric*. Jakarta: Akademia.
- Ghodousi, M., Alesheikh, A. A., & Saeidian, B. (2016). Analyzing public participant data to evaluate citizen satisfaction and to prioritize their needs via K-means, FCM, and ICA. *Cities*, 55, 70–81.
- He, X. and Li, C., 2016. The Research and Application of Customer Segmentation on Ecommerce Websites.
- Hosseini, Z. Z., & Mohammadzadeh, M. (2016). Knowledge discovery from patients' behavior via clustering-classification algorithms based on weighted eRFM and CLV model: An empirical study in public health care services. *Iranian Journal of Pharmaceutical Research*, 15(1), 355–367
- Kantardzic, Mehmed. 2011. *Data Mining: Concepts, Models, Methods, and Algorithms*. New Jersey: John Wiley & Sons, Inc., Hoboken.
- Kim, E., and Lee, B. 2007. An Economic Analysis of Customer Selection and Leverage Strategies in A Market where Network Externalities Exist. *Decision Support Systems*. 44 (1) 124-134.
- Margianti, E. S., Refianti, R., Mutiara, A. B., Nuzulina, K., Technology, I., & Technology, I. (2016). AFFINITY PROPAGATION AND RFM-MODEL FOR CRM ' S DATA ANALYSIS, 84(2), 272–282
- Monalisa, S. 2017. Klasterisasi Customer Lifetime Value Dengan Model Lrfm menggunakan Algoritma KMeans. *Jurnal Teknologi Informasi dan Komputer*, 5(2).
- Nettleton, D. (2014). Chapter 13 – CRM – Customer Relationship Management and Analysis. In *Commercial Data Mining* (pp. 195–208). doi:10.1016/B978-0-12-416602-8.00013-3
- Onur, D., Ejder, A., and ZekiAtil, B. 2018. Customer Segmentation by Using RFM Model and Clustering Methods: A Case Study in Retail Industry. *International Journal of Contemporary Economics and Administrative Sciences*, 1-20.
- Selvi, K.R and Ravi, R. 2013. The Organizational Achieving Customer Lifetime Value through Customer Relationship Management. *Asia Pacific Journal of Marketing and Management*, 2(6).
- Soltani, Z., & Navimipour, N. J. (2016). Customer relationship management mechanisms: A systematic review of the state of the art literature and recommendations for future research. *Computers in Human Behavior*, 61, 667–688. doi:10.1016/j.chb.2016.03.008
- Venkatesan, R. and Kumar, V. (2014). A Customer Lifetime Value Framework for Customer Selection and Resource Allocation Strategy. *Journal Marketing*, 68, 106-125.