

SEGMENTASI PELANGGAN SALON NUII BEAUTY GLOW MENGGUNAKAN K-MEANS CLUSTERING

Sajidah, Ruli Herdiana, Dodi Solihudin

Program Studi D3 Komputerisasi Akuntansi, STMIK IKMI Cirebon
Jl. Perjuangan No 10B, Majasem, Kec. Kesambi, Cirebon Indonesia
Sajidahzidah2002@gmail.com

ABSTRAK

Persaingan dalam industri kecantikan menuntut setiap pelaku bisnis di bidang ini untuk menentukan strategi pemasaran agar manajemen tidak melakukan kesalahan dalam memberikan pelayanan kepada pelanggan. Salon Nuii Beauty Glow belum dapat mengklasifikasikan pelanggan mana yang menguntungkan bagi perusahaan. Oleh karena itu, perusahaan mengalami kesulitan untuk menentukan strategi pemasaran yang tepat bagi pelanggannya. Peneliti menggunakan model clustering data mining yang sangat efisien dalam menentukan karakteristik kelompok pelanggan yang terbentuk, dengan menggunakan algoritma K-Means yang memiliki keunggulan dalam pengelompokan. Algoritma k-means clustering merupakan salah satu teknik data mining yang dapat membagi data dalam suatu himpunan menjadi beberapa kelompok, dimana kesamaan data pada satu kelompok lebih besar dibandingkan dengan kesamaan data tersebut dengan data pada kelompok lain. Dari penelitian ini diperoleh 3 cluster pelanggan. Cluster 0 menempati jumlah pelanggan terbanyak yaitu 48 pelanggan, cluster 1 memiliki 10 pelanggan, dan cluster 3 memiliki 12 pelanggan. Ketiga cluster tersebut dikategorikan sesuai dengan karakteristik masing-masing data dengan melakukan uji indeks Davies Bouldin untuk menentukan cluster terbaik. Dan model RFM (Kebaruan, frekuensi, moneter) digunakan untuk memfasilitasi pengelompokan..

Kata kunci: : segmentasi, clustering, pelanggan, Model RFM, salon, k-means

1. PENDAHULUAN

Pertumbuhan industri kecantikan dapat menyebabkan persaingan pelanggan yang ketat di salon kecantikan. Persaingan yang ketat memaksa perusahaan menyusun strategi untuk memenangkan persaingan atau setidaknya mampu mempertahankannya.[1] Dalam dunia bisnis, persaingan memungkinkan perusahaan untuk tumbuh dan berkembang, banyak perubahan positif dapat terjadi jika mereka bereaksi dengan cara yang benar. Setiap bisnis harus memperhatikan strategi pemasaran, salah satunya adalah segmen pelanggan yang dapat mendukung posisi bisnis dalam persaingan. Agar tidak kalah bersaing, perusahaan atau organisasi bisnis menciptakan pasar dan beradaptasi dengan perkembangan saat ini dengan menerapkan strategi bisnis yang dihasilkan dari analisis kebutuhan pelanggan. Customer Insights didasarkan pada manajemen hubungan pelanggan yang menentukan strategi dan proses untuk memperoleh, mempertahankan, dan bekerja dengan pelanggan unik untuk menciptakan nilai bagi bisnis dan pelanggan.

Salon kecantikan Nuii Beauty Glow adalah salon kecantikan baru yang terletak di pusat kota Indramayu, dimana terdapat 277 transaksi dan 71 pelanggan yang belum diketahui jasa apa yang paling laris serta pelanggan yang profitable dan potensial. Dengan mengumpulkan pelanggan, akan diketahui jenis layanan pelanggan yang paling sering dibeli pelanggan, serta membuat strategi pemasaran selanjutnya dan menentukan promosi, pendapatan, dan harga.

Metode clustering dari berbagai literatur yang

telah dilakukan sebelumnya dan telah direview, metode clustering menggunakan algoritma K-Means memberikan hasil yang baik dan memuaskan. Dengan menganalisis hasil clustering, dapat menentukan pelanggan mana yang potensial atau menguntungkan bagi perusahaan dan dapat mengembangkan strategi pemasaran yang tepat untuk setiap kelompok pelanggan dengan menyesuaikan promosi, pendapatan, dan harga sesuai dengan karakteristik pembelian pelanggan.

2. LANDASAN TEORI

2.1. Data Mining

Data mining adalah proses ekstraksi informasi yang berharga dan tersembunyi dari dalam dataset yang besar, yang dilakukan dengan menggunakan teknik dan algoritma analisis data untuk mengidentifikasi pola dan tren yang dapat membantu dalam pengambilan keputusan. Proses data mining ini dapat mencakup beberapa tahap, mulai dari pre-processing data, eksplorasi data, pemodelan, evaluasi, hingga interpretasi hasil.[2]

Proses pencarian dan pemurnian informasi dari tumpukan data yang besar merupakan proses utama dari data mining, tujuan utama dari pengolahan data adalah untuk menghasilkan informasi baru. Data mining disebut juga Knowledge Discovery in Database (KDD) yaitu suatu proses yang secara otomatis mencari data dalam ruang memori yang sangat besar dari data untuk menemukan pola dengan menggunakan teknik seperti klasifikasi asosiasi atau clustering. Oleh karena itu data mining dapat digunakan untuk mengklasifikasikan data pelanggan

dengan menggunakan algoritma yang ada pada data mining, dengan mencoba menggali pengetahuan dalam menentukan strategi pemasaran Nui Beauty Glow.

2.2. Clustering

Clustering adalah teknik pengumpulan data yang digunakan dengan menganalisis data untuk memecahkan masalah dalam pengelompokan data atau lebih partisi yang efisien dari dataset ke subset. Dalam proses clustering, targetnya adalah kasus distribusi (benda, orang, peristiwa dan lain-lain) dalam kelompok, hingga tingkat hubungan antara anggota cluster yang sama yaitu antara kuat dan lemah antara anggota cluster yang berbeda. Teknik cluster memiliki dua jenis metode dalam pengelompokannya yaitu hierarchical clustering dan non-hierarchical clustering. Hierarchical clustering merupakan metode pengelompokan data yang bekerja dengan menggabungkan dua atau lebih set data memiliki kemiripan atau persamaan, maka proses dilanjutkan dengan hal yang lain kedekatan kedua, proses ini berlanjut hingga cluster membentuk semacam pohon di suatu tempat ada tingkat atau tingkat yang jelas antara benda-benda yang mirip dengan yang terkecil serupa. Namun secara logika, semuanya hanya akan membentuk satu cluster pada akhirnya.[3]

2.3. Algoritma K-Means

Algoritma K-Means adalah salah satu metode clustering yang paling umum digunakan dalam analisis data tanpa pengawasan atau supervisi. Tujuannya adalah untuk mempartisi data ke dalam kelompok atau cluster sehingga data dengan karakteristik yang sama akan dikelompokkan ke dalam satu kelompok yang sama, sementara data dengan karakteristik yang berbeda akan dikelompokkan ke dalam kelompok yang berbeda pula[4]. Menurut [5] Metode K-Means termasuk dalam kategori pengelompokan nonhierarchical, yang artinya data akan langsung dipartisi ke dalam kelompok-kelompok yang diinginkan tanpa harus melewati tahapan pengelompokan yang lebih tinggi atau lebih rendah. Dalam prosesnya, metode K-Means mengidentifikasi centroid atau pusat cluster dari data yang ada, kemudian mengelompokkan data ke dalam cluster berdasarkan jarak Euclidean antara setiap data dan centroid

metode clustering yang berbasis jarak dan hanya bekerja pada data numerik atau angka saja. Metode ini dapat bekerja dengan sangat efisien pada data yang memiliki jumlah yang banyak, dan juga dapat mengatasi keberadaan data outlier.[6]

Untuk melakukan algoritma K-means secara umum dilakukan dengan algoritma dasar sebagai berikut :

1. Tentukan jumlah cluster atau nilai k.
2. Alokasikan data ke dalam cluster secara random atau menentukan titik centroid awal secara

sembarang.

3. Hitung jarak antara setiap data dengan centroid masing-masing cluster menggunakan rumus euclidean distance: $d = \sqrt{\sum (x_i - a_i)^2}$, di mana x_i adalah nilai dari fitur ke- i pada data, a_i adalah nilai dari fitur ke- i pada centroid.
4. Alokasikan setiap data ke dalam cluster dengan centroid terdekat (jarak terkecil).
5. Hitung kembali posisi centroid dari masing-masing cluster berdasarkan rata-rata posisi data di dalamnya.

Ulangi langkah 3-5 hingga tidak ada perubahan dalam posisi centroid dan/atau alokasi data ke dalam cluster. Pada langkah 3 dan 5, rumus Euclidean Distance digunakan untuk mengukur jarak antara setiap data dengan centroid masing-masing cluster. Setelah itu, data akan dialokasikan ke dalam cluster dengan centroid terdekat (jarak terkecil). Pada langkah 5, posisi centroid dari masing-masing cluster dihitung kembali berdasarkan rata-rata posisi data di dalamnya. Proses ini diulang hingga tidak ada perubahan dalam posisi centroid dan/atau alokasi data ke dalam cluster, yang menunjukkan bahwa model sudah konvergen.

2.4. Model RFM

segmentasi menjadi konsep penting dalam pemasaran, terutama dalam konteks relationship marketing. Dalam relationship marketing, perusahaan fokus pada membangun hubungan yang kuat dan berkelanjutan dengan pelanggan. Melalui segmentasi, perusahaan dapat memahami kebutuhan dan preferensi pelanggan secara lebih baik, sehingga dapat memberikan layanan yang lebih personal dan relevan untuk meningkatkan hubungan dan kepuasan pelanggan.

Dengan segmentasi, perusahaan dapat memecah pasar menjadi kelompok yang lebih kecil dengan karakteristik yang serupa. Hal ini memungkinkan perusahaan untuk menyesuaikan produk, harga, promosi, dan saluran distribusi mereka untuk setiap kelompok dengan cara yang lebih efektif dan efisien. Dalam hal ini, segmentasi membantu perusahaan memfokuskan sumber daya mereka pada segmen pelanggan yang paling menguntungkan dan potensial untuk meningkatkan retensi dan loyalitas pelanggan.

[7] berpendapat bahwa Segmentasi pelanggan yang akurat. Oleh karena itu, proses pemilihan atribut atau fitur yang tepat sangat penting dalam segmentasi pelanggan. Beberapa faktor yang biasanya dipertimbangkan dalam pemilihan fitur termasuk keunikan, relevansi, ukuran, keterjangkauan, dan stabilitas fitur. Selain itu, teknik clustering yang tepat juga perlu diterapkan dalam proses segmentasi untuk menghasilkan hasil yang lebih akurat dan bermakna. Clustering adalah proses pengelompokan data yang serupa dalam kelompok-kelompok yang berbeda untuk memudahkan analisis dan pemahaman. Dengan teknik clustering yang tepat, segmentasi pelanggan

dapat menjadi lebih efektif dan efisien dalam menghasilkan strategi pemasaran yang tepat dan akurat.

Untuk mengatasi masalah pemilihan kualitas terbaik, disarankan untuk menggunakan model RFM. RFM (*Recency, Frequency, Monetary*) merupakan metode yang biasa digunakan untuk mengumpulkan data loyalitas pelanggan dengan membagi aktivitas pelanggan menjadi 3 karakteristik yaitu *Recency, Frequency* dan *Monetary*. pelanggan yang baru saja melakukan transaksi (*Recency*), pelanggan yang sering melakukan transaksi (*Frequency*), dan pelanggan yang menghabiskan uang banyak dalam transaksi (*Monetary*). Dengan memahami karakteristik setiap segmen, bisnis dapat menyesuaikan strategi pemasaran mereka untuk meningkatkan retensi pelanggan, meningkatkan frekuensi transaksi, atau meningkatkan nilai transaksi. Dalam praktiknya, model RFM dapat dikombinasikan dengan algoritma clustering seperti K-means untuk menghasilkan segmentasi pelanggan yang lebih akurat dan efektif.[8]

3. METODE

Penerapan Knowledge Discovery and Database (KDD) dalam penelitian ini meliputi pemahaman bisnis, pemilihan data, preprocessing data, transformasi data, mining data, evaluasi hasil, dan interpretasi. Dalam konteks segmentasi pelanggan, proses KDD dapat digunakan untuk mengidentifikasi pola atau kategori pelanggan yang berbeda berdasarkan atribut tertentu, seperti RFM. [9]

Langkah-langkah Metodologi Knowledge Discovery in Database (KDD) dijelaskan sebagai berikut

3.1. Selection

Pada tahap ini, data yang relevan dan sesuai dengan kebutuhan penelitian dipilih dari sumber data yang tersedia.

3.2. Pre-processing

Tahap pre-processing dilakukan untuk membersihkan data dari nilai-nilai yang hilang, duplikat, atau tidak valid. Hal ini bertujuan untuk meningkatkan kualitas data dan memastikan bahwa data yang digunakan dalam proses KDD adalah data yang berkualitas.

3.3. Transformation

Tahap ini melibatkan transformasi data ke dalam bentuk yang sesuai dengan algoritma dan teknik data mining yang akan digunakan dalam penelitian.

Transformasi ini dapat meliputi normalisasi, pengkodean variabel, dan reduksi dimensi.

3.4. Data Mining

Pada tahap ini, teknik data mining digunakan untuk mengeksplorasi data dan menemukan pola atau hubungan yang mungkin terdapat di dalam data. Metode yang digunakan dalam proses ini dapat berupa clustering, klasifikasi, regresi, asosiasi, dan lain-lain.

3.5. Interpretation/Evaluation

Setelah proses data mining selesai, hasil yang ditemukan akan dievaluasi dan diinterpretasikan. Hasil interpretasi ini akan digunakan untuk membuat keputusan atau tindakan yang relevan.

4. HASIL DAN PEMBAHASAN

proses segmentasi pelanggan umumnya dilakukan dengan metode clustering yang dimulai dengan menyiapkan data pelanggan, melakukan pemilihan fitur (*feature selection*) atau reduksi dimensi (*dimensionality reduction*) jika diperlukan, menentukan jumlah cluster yang diinginkan, menentukan pusat (*centroid*) secara acak, menghitung jarak antara setiap data pelanggan dengan setiap *centroid*, mengelompokkan data pelanggan ke dalam cluster terdekat, dan kemudian menghitung kembali pusat dari setiap cluster.

Proses ini diulang-ulang hingga tidak terjadi perpindahan data pelanggan dari satu cluster ke cluster lainnya. Tujuan utama dari proses segmentasi pelanggan adalah untuk memisahkan pelanggan menjadi beberapa kelompok berdasarkan kesamaan karakteristik, sehingga dapat digunakan untuk pengambilan keputusan dalam strategi pemasaran dan layanan pelanggan yang lebih efektif. [10]

4.1. Data Mining

Data Discovery (KDD) adalah proses komprehensif untuk menganalisis data yang akurat, terkini, dan berwawasan yang pada akhirnya akan berguna bagi manajemen[9]. Pada proses KDD ini terdapat proses pemilihan data atau atribut yang akan digunakan dalam analisis. Pemilihan ini dilakukan berdasarkan persyaratan KDD yang diperlukan dalam penelitian, serta relevansi dan keterkaitan atribut dengan tujuan analisis yang ingin dicapai. Dalam pemilihan atribut, juga perlu dipertimbangkan aspek keamanan dan privasi data yang terkait dengan pengumpulan dan penggunaan data tersebut.

Tabel 1. Data Transaksi Salon Nui Beauty Glow

Tanggal	No Pelanggan	Nama	Jenis Perawatan	Nominal	Alamat
2/2/2022	1	ADE	FACIAL DETOX	125,000.00	pasekan
2/2/2022	1	ADE	FACIAL WHITENING	85,000.00	pasekan
2/2/2022	11	DANIRIH	BB GLOW	225,000.00	paoman
2/2/2022	20	ESIH	FACIAL DERMABRASI	100,000.00	pabean udik

Tanggal	No Pelanggan	Nama	Jenis Perawatan	Nominal	Alamat
2/2/2022	23	FUJI	FACIAL ACNE	150,000.00	paoman
2/2/2022	24	HENI	FACIAL DETOX	125,000.00	pabean udik
2/2/2022	24	HENI	FACIAL WHITENING	85,000.00	pabean udik
2/3/2022	10	ATIN	FACIAL PENGENCANGAN	100,000.00	margadadi
2/3/2022	23	FUJI	FACIAL DETOX	125,000.00	paoman
2/3/2022	23	FUJI	LULUR	100,000.00	paoman
2/3/2022	30	IPAH	FACIAL WHITENING	85,000.00	paoman
2/3/2022	33	IYOL	MESO PRP	355,000.00	paoman
2/4/2022	27	IPAH	FACIAL CROMOSOM	180,000.00	margadadi

Tabel 1 terdapat 7 atribut yang terdapat dalam data transaksi Salon Nui Beauty Glow. Dalam proses pemilihan data, maka perlu ditinjau kembali atribut yang diperlukan dan atribut yang kurang diperlukan serta menambahkan atribut baru

4.2. Pre-Processing

Pada tahap ini dilakukan berbagai macam tindakan seperti membersihkan data dari duplikasi, noise, dan kesalahan data. Selain itu, juga dilakukan reduksi dimensi atau pemilihan atribut yang relevan untuk mempercepat proses clustering pada tahap selanjutnya. Pemilihan atribut yang dibutuhkan dalam proses clustering pelanggan menggunakan model RFM (Recency, Frequency, dan Monetary).

Tabel 2. Detail Penggunaan Atribut

Atribut	Detail Penggunaan
Tanggal	✓
NoPelanggan	✗
Nama	✓
JenisPerawatan	✗
Nominal	✓
Alamat	✓

Tabel 3. Perubahan Atribut

Perubahan Atribut	Atribut Awal
Facial Detox	Jenis Perawatan
Facial Whitening	Jenis Perawatan
Meso PRP	Jenis Perawatan
BB glow	Jenis Perawatan
facial pengencangan	Jenis Perawatan
Dna salmon	Jenis Perawatan
meso lips	Jenis Perawatan
lulur	Jenis Perawatan
spfso	Jenis Perawatan
facial cromsom	Jenis Perawatan
facial vit c	Jenis Perawatan
baby botox	Jenis Perawatan
dermabrasi	Jenis Perawatan
facial acne	Jenis Perawatan
tanam benang	Jenis Perawatan

Atribut NoPelanggan dan JenisPerawatan Pada Tabel 2 merupakan tahap awal pembersihan data, menghilangkan atribut yang kurang penting dan menambahkan atribut yang diperlukan seperti terlihat

pada Tabel 3, ada perubahan atribut JenisPerawatan menjadi beberapa atribut yang sebelumnya ada di anggota atribut JenisPerawatan. ata yang berjenis nominal seperti alamat dan tanggal harus diinisialisasi terlebih dahulu ke dalam bentuk angka atau numerikal sebelum dilakukan proses analisis data. Hal ini dilakukan untuk memudahkan proses clustering.

Proses inisialisasi data dapat dilakukan dengan beberapa teknik seperti label encoding atau one-hot encoding. Teknik label encoding mengganti nilai nominal pada atribut dengan nilai numerik terurut dari 1 sampai n, sedangkan teknik one-hot encoding mengubah setiap nilai nominal menjadi kolom baru dan memasukkan nilai 1 atau 0 untuk setiap baris tergantung pada nilai nominal yang dimiliki. Untuk melakukan inisialisasi alamat terlihat pada tabel dibawah ini:

Tabel 4. Inisialisasi Atribut Alamat

Alamat	Nilai
Pabean udik	8
Pasekan	7
Karangturi	6
Pagirikan	5
Karangsong	4
Paoman	3
Margadadi	2
Terusan	1

Tabel 5. Perubahan atribut Hari

Tanggal	Tanggal Penelitian	Hari
2/2/2022	1/1/2023	333
3/23/2022	1/1/2023	284
5/11/2022	1/1/2023	235

Hasil perubahan atribut Tanggal menjadi Hari terlihat pada tabel 4, dengan menghitung jarak dari tanggal transaksi hingga tanggal penelitian yang peneliti lakukan yaitu tanggal 01 Januari 2023.

Pada tahap ini merupakan tahap untuk memastikan data transaksi penjualan yang dipilih telah layak untuk dilakukan proses pengolahan. Setelah melakukan selection data, terdapat 19 atribut yang dapat digunakan dalam proses data mining. Terlihat pada tabel 6 dibawah ini:

Tabel 6. Hasil Data Cleaning

Hari	NAMA	Alamat	Nominal	facial detox	facial whitening			tanam benang
333	ADE	7	210,000.00	1	1			0
333	DANIRIH	3	350,000.00	1	0			0
333	ESIH	8	100,000.00	0	0			0
333	FUJI	3	150,000.00	0	0			0
333	HENI	8	210,000.00	1	1			0
333	HENI	8	100,000.00	0	0			0
332	ATIN	2	225,000.00	1	0			0
332	FUJI	3	210,000.00	1	1			0
332	FUJI	3	100,000.00	0	0			0
332	IPAH	3	85,000.00	0	1			0
332	IYOL	3	480,000.00	1	0			0
331	IPAH	3	355,000.00	0	0			0
331	ISTI	2	305,000.00	1	0			0
331	LELI	2	225,000.00	1	0			0
331	MAMAH KIKI	4	225,000.00	1	0			0
331	NURLAELA	2	100,000.00	0	0			0
330	ANGGUN	8	225,000.00	1	0			0
330	NISA	1	125,000.00	1	0			0

Tabel 7. keterangan atribut

Atribut	Keterangan
Hari	menandakan Recency, jarak transaksi terakhir dengan waktu penelitian
Nama	Nama Pelanggan Salon Nui Beauty Glow
Alamat	Alamat Pelanggan Salon Nui Beauty Glow
Nominal	biaya transaksi yang menandakan Monetary
Facial Detox Facial whitening Tanam Benang	Facial Detox sampai tanam benang menandakan Frequency, jumlah transaksi selama perawatan

Menurut [11] Untuk memodelkan kriteria RFM pada atribut di dalam dataset, diperlukan atribut-atribut yang dapat merepresentasikan setiap kriteria tersebut. Berikut adalah penjelasan untuk masing-masing kriteria:

a. Kriteria Recency (R):

Kriteria Recency mengukur seberapa baru terakhir kali pelanggan melakukan transaksi dengan periode penelitian yang diberikan. Untuk merepresentasikan kriteria ini, dibutuhkan atribut yang menunjukkan adanya rentang waktu transaksi terakhir pelanggan dengan periode penelitian. Salah satu cara untuk menghitung atribut ini adalah

dengan menghitung selisih waktu antara tanggal transaksi terakhir pelanggan dengan tanggal akhir periode penelitian. Dalam hal ini, atribut yang dibutuhkan adalah atribut Tanggal yang dirubah menjadi Hari menggunakan rumus pengurangan excel.

b. Kriteria Frequency (F):

Kriteria Frequency mengukur seberapa sering pelanggan melakukan transaksi dengan periode penelitian yang diberikan. Untuk merepresentasikan kriteria ini, dapat dilakukan dengan menghitung jumlah transaksi yang dilakukan oleh pelanggan selama periode penelitian. Dalam hal ini, atribut yang dibutuhkan adalah jumlah transaksi dari setiap jenis perawatan yang dilakukan oleh pelanggan. Atribut frequency dapat dihitung dengan menjumlahkan anggota dari atribut Facial Detox, Facial Whitening, sampai Tanam Benang, menggunakan rumus sum excel.

c. Kriteria Monetary (M):

Kriteria Monetary mengukur seberapa besar jumlah biaya yang dikeluarkan pelanggan selama periode penelitian yang diberikan. Untuk merepresentasikan kriteria ini, dibutuhkan atribut yang berhubungan dengan biaya yang telah dihabiskan pelanggan selama melakukan transaksi. Salah satu atribut yang dapat digunakan adalah atribut Nominal yang berhubungan dengan biaya yang telah dihabiskan pelanggan selama melakukan transaksi.

Tabel 8. Hasil Atribut sesuai model RFM

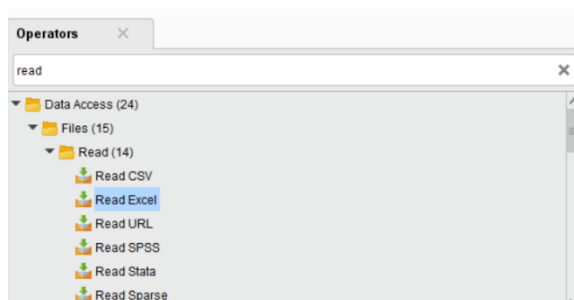
Nama	Alamat	Recency (Hari)	Frequency (Sum F.Detox Dll)	Monetary (Nominal)
ADE	7	333	2	210,000.00
DANIRIH	3	333	2	350,000.00
ESIH	8	333	1	100,000.00
FUJI	3	333	1	150,000.00
HENI	8	333	2	210,000.00

Nama	Alamat	Recency (Hari)	Frequency (Sum F.Detox Dll)	Monetary (Nominal)
HENI	8	333	1	85,000.00
ATIN	2	332	2	225,000.00
FUJI	3	332	2	210,000.00
FUJI	3	332	1	100,000.00
IPAH	3	332	1	85,000.00
IYOL	3	332	2	480,000.00
IPAH	3	331	1	355,000.00
ISTI	2	331	2	305,000.00

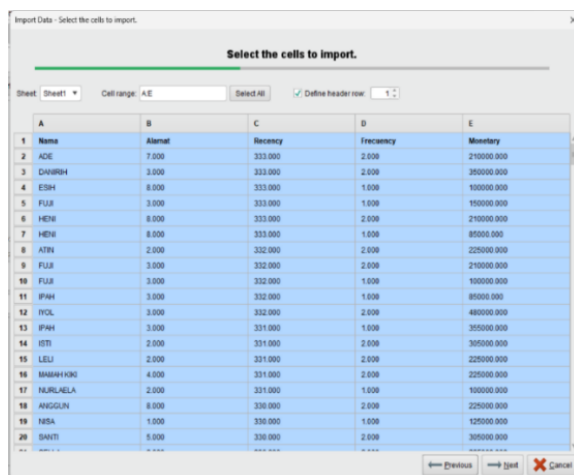
4.3. Transformations

Transformasi data digunakan untuk mengubah data menjadi bentuk yang sesuai dalam proses pengumpulan data. Teknik transformasi data yang paling umum adalah normalisasi, pemilihan fitur, dan discretization. Adapun data transformation pada penelitian ini dapat dilihat pada gambar berikut ini:

Sebelum melakukan transformasi data, operator read excel diperlukan dalam proses ini, untuk menampilkan data yang akan digunakan.



Gambar 1. Opeartor Read Excel



Gambar 2. pemilihan sel dataset

Gambar 4. 2 pemilihan sel dataset, Sel data yang digunakan adalah sel dari kolom A ke kolom E, selanjutnya dapat dilakukan proses transformasi.

Perubahan role atribut Nama menjadi ID agar

atribut Nama memiliki peranan yang unik pada saat proses pengumpulan data berlangsung karena Nama bertipe polynomial sedangkan dalam clustering metode k-means hanya bisa mengolah data yang berbentuk angka atau numerik.

Gambar 3. Hasil Transformasi Data

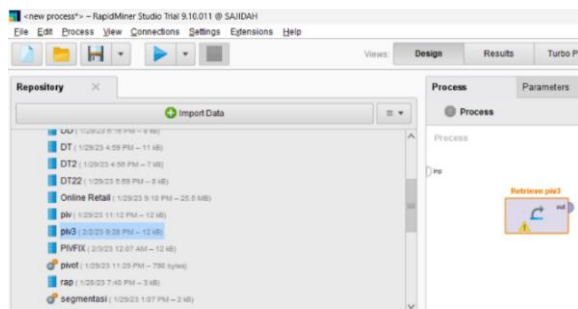
Gambar 3 menunjukkan desain proses transformations data dengan hasil proses diketahui setiap atribut memiliki value numeric, date dan polynomial. Data yang telah di transformations, siap dilakukan clustering menggunakan Algoritma K-means.

4.4. Proses Data Mining

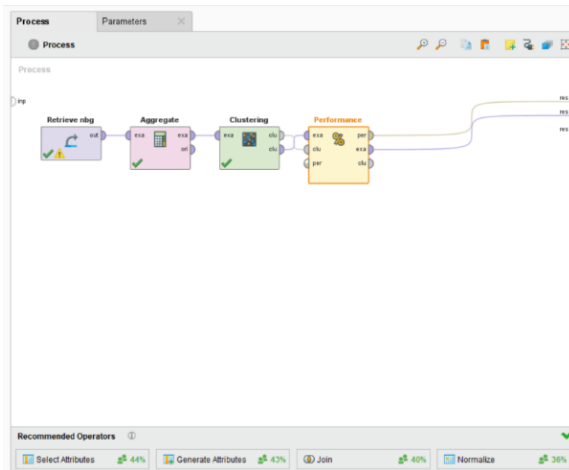
Teknik Data Mining Menyimpan data untuk menghasilkan informasi yang berguna. Pada penelitian ini, data mining dilakukan dengan metode K-Means Clustering menggunakan software Rapidminer. Sebelum memproses data, hal pertama yang perlu Anda lakukan adalah mengimpor database Rapidminer dan memasukkannya ke halaman konfigurasi process seperti yang ditunjukkan pada Gambar 4.5 di bawah ini.

Untuk dapat melakukan proses data yang sesuai dengan teknik data mining maka digunakan operator-operator yang telah tersedia pada tools RapidMiner.

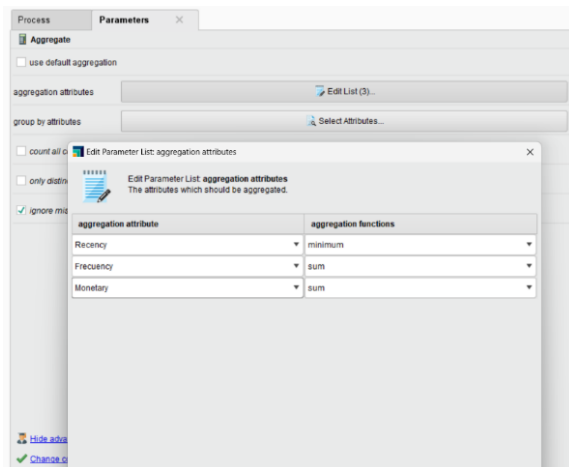
Dibawah ini adalah model proses penggunaan tools RapidMiner, Masing-masing operator tersebut memiliki fungsi yang berbeda-beda.



Gambar 4. proses Impor Data

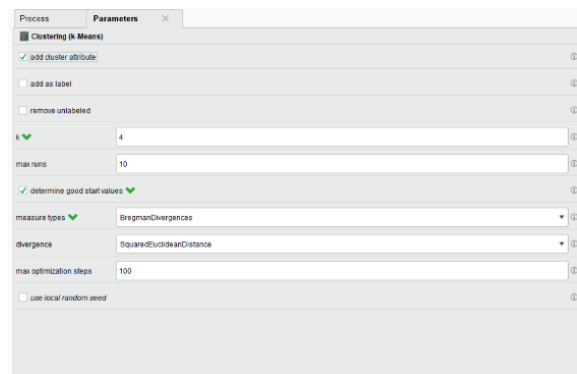


Gambar 5. Operator Data Mining



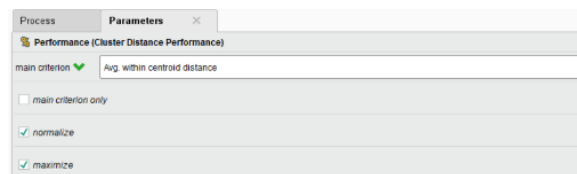
Gambar 6. Operator Agregate

Pada gambar 6 parameter operator agregate digunakan untuk menghitung anggota atribut yang baru ditambahkan sesuai *function* yang akan dibuat untuk atribut baru. Atribut *Recency* menggunakan *function* minimum, *Frequency* menggunakan *function* sum, *Monetary* menggunakan *function* sum dan *group by* menggunakan atribut Nama dan Alamat sebagai acuan dalam melakukan agregasi tersebut.



Gambar 7. Operator clustering K-means

Operator clustering pada gambar 7. digunakan untuk proses pen-golahan data dengan nilai teknik K-Means, dengan menentukan cluster secara acak dari $k=2$ hingga $k=6$ agar mendapatkan pusat cluster yang terbaik.

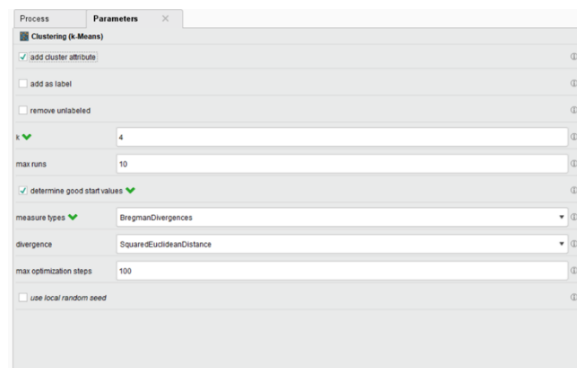


Gambar 8. Operator Performance

Operator performance pada gambar 8 ini digunakan untuk evaluasi kinerja metode Clustering berbasis centroid. Operator ini mem-berikan daftar nilai kriteria kinerja berdasarkan centroid cluster.

4.5. Evaluation

Untuk menentukan cluster terbaik dalam penelitian ini dilakukan analisis cluster dengan memilih k terbaik menggunakan uji *Davies Bouldin Index*. Dari memasukkan $K=2$ sampai $K=6$ dan melihat cluster yang baik untuk digunakan dalam pengolahan data. Seperti yang ditunjukkan pada Gambar 9 di bawah ini:



Gambar 9. Operator K-means

Setelah melakukan 5 percobaan untuk menentukan nilai k , Pada Percobaan kedua dengan menggunakan nilai $k=3$ menghasilkan nilai

Performance Vektor (Davies Bouldin Index) yang lebih optimal dibandingkan dengan percobaan pertama dan ketiga hingga seterusnya. Nilai yang dihasilkan yaitu nilai yang mendekati 0 dengan nilai -0.379.

Nilai K	Nilai DBI
2	-0.433
3	-0.379
4	-0.472
5	-0.519
6	-0.526

Gambar 10. Hasil Nilai DBI

Cluster Model

```
Cluster 0: 49 items
Cluster 1: 10 items
Cluster 2: 12 items
Total number of items: 71
```

Gambar 11. hasil cluster model

Cluster	Alamat	minimum(Recency)	count(Frequency)	sum(Monetary)	Rank
Cluster_0	4,714	173,898	1,939	Rp 379,489.80	3
Cluster_1	4,100	20,300	10,700	Rp 2,185,000.00	1
Cluster_2	3,750	107,917	6,250	Rp 1,195,000.00	2

Gambar 12. Nilai Centroid Setiap Atribut

Berdasarkan Gambar 12 terlihat bahwa terdapat tiga cluster yang diperoleh. cluster 0 beranggotakan 49 pelanggan, cluster 1 beranggotakan 10 pelanggan dan cluster 2 beranggotakan 12 pelanggan. Dan total pelanggan sebanyak 71 dari hasil clustering tersebut.

Row No.	id	cluster	Nama	Alamat	minimum(R...	sum(frecue...	sum(Moneta...
1	1	cluster_2	NISA	1	6	9	1290000
2	2	cluster_2	SISKA	1	57	7	950000
3	3	cluster_2	ATRI	2	97	5	890000
4	4	cluster_0	AYANI	2	141	3	435000
5	5	cluster_0	AYANG	2	18	4	580000
6	6	cluster_0	EDAH	2	73	3	435000
7	7	cluster_2	FAWAZ	2	284	3	325000
8	8	cluster_2	ISTI	2	18	8	1325000
9	9	cluster_1	LELI	2	12	12	2150000
10	10	cluster_0	NITA	2	284	1	85000
11	11	cluster_0	NUNG	2	204	2	400000
12	12	cluster_1	NURLAELA	2	11	13	2435000
13	13	cluster_0	NURUL	2	3	4	535000
14	14	cluster_0	OLIN	2	97	4	545000
15	15	cluster_0	PIYAH	2	97	3	505000
16	16	cluster_0	RINI	2	236	2	195000
17	17	cluster_0	RIRI	2	328	1	100000
18	18	cluster_0	RISWATI	2	291	1	225000

Gambar 13. Hasil Clustering

Berdasarkan hasil dari Gambar 13 maka dapat disimpulkan bahwasannya cluster 1 merupakan pelanggan yang menempati peringkat 1 dengan rata-rata Monetary 2.185.000,00 dan Frequency 10,7. cluster 2 merupakan pelanggan yang menempati peringkat ke 2 yang memiliki nilai rata-rata Monetary 1.195.000,00, dan Frequency 6,250 sedangkan cluster 0 merupakan pelanggan yang menempati peringkat ke 3 dengan rata-rata Monetary 379.489,80 dan Frequency 1,939.

Data transaksi pelayanan Salon Nui Beauty

Glow sebanyak 277 setelah melalui proses clustering dan model RFM maka dihasilkan segmentasi pelanggan pada masing-masing cluster sesuai tipe pelanggan. Berdasarkan kategori model RFM Pelanggan yang masuk kategori loyal terdapat pada cluster 1 dengan jumlah 10 pelanggan karena memiliki jarak pelayanan atau kunjungan terdekat dengan hari penelitian (Recency), memiliki jumlah perawatan yang tinggi (frequency) dan Total Biaya perawatan paling besar (Monetary). Berbeda dengan cluster 2 yang memiliki jarak kunjungan yang lumayan jauh dari waktu penelitian, namun Frequency dan Monetary pada cluster 2 cukup besar atau berada pada rata-rata. maka pada cluster 2 dikategorikan sebagai pelanggan potensial yang berpotensi meningkat menjadi pelanggan yang loyal. Sedangkan pada cluster 0 berjumlah 48 pelanggan dapat dikategorikan sebagai pelanggan yang kurang potensial dengan rata-rata Recency cukup jauh, serta memiliki Frequency dan Monetary yang rendah.

Cluster	R	F	M	Kategori
Cluster_0	174 hari bertransaksi dengan jarak waktu yang jauh	1-2 kali kunjungan atau perawatan di Salon	379,489.80 Total Pembelian Bernilai Kecil	pelanggan lama yang yang jarang melakukan kunjungan dan perawatan di salon.
Cluster_1	20 hari bertransaksi dengan jarak waktu yang dekat	10 kali kunjungan atau perawatan di salon	2,185,000.00 Total Pembelian Paling Tinggi	Pelanggan yang memiliki loyalitas tinggi, dan pelanggan yang memberikan nilai kepada salon karena sering melakukan transaksi dengan nominal yang besar dengan jarak waktu yang dekat.
Cluster_2	107 hari bertransaksi dengan jarak waktu yang cukup jauh	6 kali kunjungan atau perawatan di salon	1,195,000.00 Total Pembelian bernilai Rendah	Pelanggan baru yang berpotensi menjadi pelanggan yang loyal, dengan frekuensi dan nominal yang cukup tinggi

Gambar 14. Kategori RFM dari hasil clustering

5. KESIMPULAN

Berdasarkan penerapan algoritma K-means untuk menentukan segmentasi pelanggan, dapat disimpulkan bahwa Metode tersebut telah diuji dalam sistem dan terdapat 49 pelanggan yang termasuk dalam cluster pertama atau cluster 0 dengan kategori pelanggan tidak loyal dengan nilai RFM rendah. Oleh karena itu, diperlukan strategi promosi dengan memberikan potongan harga atau kupon untuk perawatan salon Nui Beauty Glow. 10 pelanggan termasuk dalam klaster kedua atau klaster 1 dengan kategori pelanggan setia, dengan nilai RFM tinggi. 12 pelanggan termasuk dalam klaster ketiga atau klaster 2 dengan kategori pelanggan rata-rata karena merupakan pelanggan baru yang baru saja melakukan transaksi dengan nilai RFM rata-rata.

Jumlah klaster optimal ditentukan dengan menggunakan Davies Bouldin Index (DBI) yang menghasilkan 3 klaster dengan karakteristiknya masing-masing, sehingga membagi data pelanggan menjadi 3 segmen. Model RFM diterapkan untuk mengkategorikan data berdasarkan Recency, Frequency, dan Monetary agar mudah diimplementasikan menggunakan algoritma K-means

clustering. Model RFM juga memiliki kemampuan untuk menganalisis keanggotaan setiap cluster untuk menemukan karakteristik dari setiap cluster.

DAFTAR PUSTAKA

- [1] M. Tamia and W. Soediono, "SERVICE QUALITY, CUSTOMER SATISFACTION, DAN CORPORATE IMAGE TERHADAP CUSTOMER LOYALTY PADA KLINIK KECANTIKAN DI JAKARTA," 2019. [Online]. Available: <http://jurnaltsm.id/index.php/MB>
- [2] A. Nugraha, O. Nurdiawan, and G. Dwilestari, "PENERAPAN DATA MINING METODE K-MEANS CLUSTERING UNTUK ANALISA PENJUALAN PADA TOKO YANA SPORT," 2022.
- [3] S. Gizi Balita Di Kelurahan Jumpat Rejo Sukodono Sidoarjo Amir Ali, R. Medis dan Informasi Kesehatan Stikes Yayasan Rumah Sakit DrSoetomo DrMoestopo, and A. Surabaya, "Clustering Data Antropometri Balita Untuk Menentukan," 2020. [Online]. Available: <http://jurnal.mdp.ac.id>
- [4] A. Sulistiyawati and E. Supriyanto, "Implementasi Algoritma K-means Clustering dalam Penentuan Siswa Kelas Unggulan," vol. 15, no. 2.
- [5] I. Nyoman and M. Adiputra, "CLUSTERING PENYAKIT DBD PADA RUMAH SAKIT DHARMA KERTI MENGGUNAKAN ALGORITMA K-MEANS," *INSERT: Information System and Emerging Technology Journal*, vol. 2, no. 2, p. 99, 2021.
- [6] A. Febrian, Nana Suarna, and Gifthera Dwilestari, "PENERAPAN ALGORITMA K-MEANS UNTUK MENGELOMPOKKAN DATA PENGIRIMAN PAKET DI KANTOR POS CIREBON," *JURNAL TEKNOLOGI TECHNOSCIENTIA*, pp. 23–27, Sep. 2022, doi: 10.34151/technoscience.v15i1.3858.
- [7] M. Tamia and W. Soediono, "SERVICE QUALITY, CUSTOMER SATISFACTION, DAN CORPORATE IMAGE TERHADAP CUSTOMER LOYALTY PADA KLINIK KECANTIKAN DI JAKARTA," 2019. [Online]. Available: <http://jurnaltsm.id/index.php/MB>
- [8] S. E. Jamal, "Terbit online pada laman web jurnal: <https://ejournal.upm.ac.id/index.php/energy/> Jurnal Energy (Jurnal Ilmiah Ilmu-ilmu Teknik) Analisis RFM dan Algoritma K-Means untuk Clustering Loyalitas Customer RFM Analysis and K-Means Algorithm for Customer Loyalty Clustering," *Kom Jurnal ENERGY (Jurnal Ilmiah Ilmu-ilmu Teknik)*, vol. 9, no. 1, p. 18, 2019, doi: 10.51747/energy.
- [9] A. Madani, A. Rahmah, F. Nurunnisa, and A. Elia, "SENTIMAS: Seminar Nasional Penelitian dan Pengabdian Masyarakat Customer Segmentation at BC HNI 2 Pekanbaru by Applying the K-Medoids Algorithm and Recency, Frequency, Monetary (RFM) Model Segmentasi Pelanggan pada BC HNI 2 Pekanbaru dengan Menerapkan Algoritma K-Medoids dan Model Recency, Frequency, Monetary (RFM)." [Online]. Available: <https://journal.irpi.or.id/index.php/sentimas>
- [10] R. Y. Firmansah, J. Dedy Irawan, and N. Vendyansyah, "ANALISIS RFM (RECENCY, FREQUENCY AND MONETARY) PRODUK MENGGUNAKAN METODE K-MEANS," 2021.
- [11] B. E. Adiana, I. Soesanti, and A. E. Permanasari, "Analisis Segmentasi Pelanggan Menggunakan Kombinasi RFM Model dan Teknik Clustering," no. 2, 2018, doi: 10.21460/jutei.2017.21.76.