

**PENERAPAN METODE CLUSTERING DENGAN ALGORITMA K-MEANS
UNTUK SEGMENTASI PELANGGAN PADA CV. SINAR SURYA****Maryana¹; Arif Sugianto²; Nurmalasari³; Ester Arisawati⁴**Program Studi Sistem Informasi^{1,2,3,4}Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri^{1,2,3,4}<http://nusamandiri.ac.id>^{1,2,3,4}maryana@nusamandiri.ac.id¹ arsugianto13@gmail.com² numalasari.nmr@nusamandiri.ac.id³ester.err@nusamandiri.ac.id⁴

Abstract— The intense competition in the food distribution sector, requires CV Sinar Surya to manage its customers to the fullest. Knowing loyal customers will help companies determine the right marketing strategy. This study aims to determine potential customers, customers who are loyal to the company. Potential customers are determined by customer segmentation. The model used for segmentation is RFM (Recency, Frequency, and Monetary) and data mining techniques, namely clustering method with the K-Means algorithm. The software used to process the RFM Model is the Microsoft Excel version 2010 while for processing the clustering method is Rapid Miner version 9.0. The results of this segmentation divide the customer into 2 clusters. The best number of clusters is determined based on the davies bouldin index. The first cluster was cluster 0 consisting of 262 customers with RFM Score between 111 - 543. The first cluster included the Everyday Shopper group. The second cluster, cluster 1 consisted of 101 customers with RFM Score 434 - 555. The second cluster included the Golden Customer group.

Keywords: customer segmentation, RFM, clustering, K-Means, davies bouldin index, data mining

Abstrak— Persaingan yang ketat dalam bidang distribusi sembako, mengharuskan CV Sinar Surya untuk mengelola pelanggan secara maksimal. Mengetahui nasabah yang loyal akan membantu perusahaan menentukan strategi pemasaran yang tepat. Penelitian ini bertujuan untuk menentukan pelanggan potensial, pelanggan yang loyal terhadap perusahaan. Pelanggan potensial ditentukan dengan segmentasi pelanggan. Model yang digunakan untuk segmentasi yaitu RFM (Recency, Frequency, and Monetary) dan teknik data mining yaitu metode *clustering* dengan algoritma K-Means. Software yang digunakan untuk mengolah RFM Model yaitu Microsoft Excel versi 2010 sedangkan untuk mengolah metode *clustering* yaitu Rapid Miner versi 9.0. Hasil dari segmentasi ini membagi pelanggan menjadi 2 *cluster*. Jumlah *cluster* terbaik ditentukan berdasarkan davies bouldin index. *Cluster* pertama yaitu *cluster* 0 terdiri dari 261 pelanggan dengan RFM Score antara 111 - 543. *Cluster* pertama termasuk kelompok *Everyday Shopper*. *Cluster* kedua yaitu *cluster* 1 terdiri dari 102 pelanggan dengan RFM Score 443 - 555. *Cluster* kedua termasuk kelompok *Golden Customer*.

Kata kunci: Segmentasi Pelanggan, RFM, Clustering, K-Means, Davies Bouldin Index, Data Mining.

PENDAHULUAN

Setiap lembaga memiliki sistem operasi di mana setiap transaksi kegiatan usahanya selalu tercatat dan tercatat. Pencatatan setiap transaksi sangat berguna untuk semua keperluan organisasi. Data tersebut disimpan dalam database berkapasitas besar. Bagi perusahaan, data yang disimpan di database dapat digunakan untuk membuat laporan penjualan, pengendalian persediaan, dan lain-lain. Selain kebutuhan pada hal tersebut, perusahaan juga perlu mengetahui perkiraan penjualan yang akan didapatkan, pola

pembelian konsumen, analisis kebutuhan pasar dan sebagainya. Semua hal tersebut bisa didapatkan dengan memanfaatkan data data yang sudah ada di dalam basis data.

Beberapa tahun terakhir, data semakin heterogen dan kompleks dengan volume yang meningkat cepat secara eksponensial (Sumiah & Mirantika, 2020). Menurut Turner dalam (Suyanto, 2017) pada tahun 2013 volume data sudah menjadi 44 zettabytes di tahun 2020. Oleh karena itu saat ini dikenal istilah big data, yang menggambarkan volume data sangat besar, terstruktur maupun tidak terstruktur, yang

membanjiri dunia bisnis. Dalam big data, tentu saja akan ditemui kesulitan membaca dan mengetahui pola-pola dan relasi data jika dilakukan secara manual.

Dalam dunia bisnis yang selalu dinamis dan penuh persaingan, para pelaku bisnis harus selalu memikirkan cara - cara untuk terus survive dan jika mungkin mengembangkan bisnis mereka. Untuk mencapai hal itu ada tiga kebutuhan bisnis yang dapat dilakukan, yaitu penambahan jenis maupun peningkatan kapasitas produk, pengurangan biaya operasional perusahaan, dan peningkatan efektivitas pemasaran serta keuntungan (Sutrisno, Afriyudi, & Widiyanto, 2014). Agar bisa memenuhi kebutuhan bisnis diatas banyak cara yang dapat ditempuh salah satunya dengan melakukan analisis data perusahaan (Andriyansyah, Wulanningrum, & Farida, 2019).

CV Sinar Surya merupakan perusahaan yang bergerak dalam bidang distribusi sembako (sembilan bahan pokok). Pangsa pasar CV. Sinar Surya adalah pasar tradisional, pasar kaget, dan agen agen sembako yang berada di wilayah DKI Jakarta dan sekitarnya. Tidak hanya CV. Sinar Surya, cukup banyak perusahaan - perusahaan lain yang bergerak di bidang yang sama. Hal tersebut tentu saja menimbulkan persaingan antar perusahaan. Fokus utama perusahaan untuk bersaing dengan kompetitornya adalah pelanggan (Wulandari, 2014).

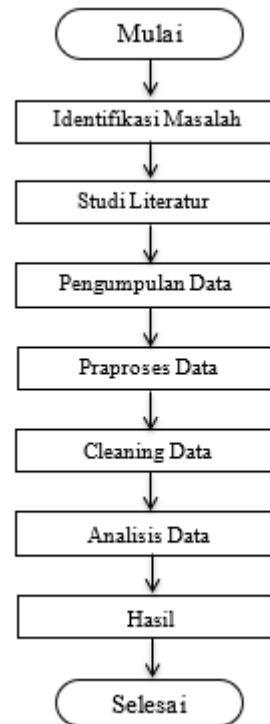
Pelanggan menduduki posisi penting dalam pengembangan strategi bisnis, pelanggan juga merupakan salah satu sumber keuntungan dalam perusahaan (Widiarina & Wahono, 2015). Untuk itu diperlukan suatu pemahaman yang baik mengenai pelanggan. Pemahaman yang baik mengenai pelanggan dapat digunakan perusahaan untuk berinvestasi dengan pelanggan yang potensial. Masalah yang sering dihadapi adalah kesulitan dalam menganalisa nilai pelanggan (Widiarina & Wahono, 2015). Banyak perusahaan mengalami kesulitan untuk mengidentifikasi pelanggan (Zaman & Wahyuni, 2019) (Putra & Wadisman, 2018) atau nasabah yang tepat, hal tersebut dapat mengakibatkan perusahaan kehilangan nasabah potensial (Putra & Wadisman, 2018) dan tentunya akan sangat merugikan perusahaan.

BAHAN DAN METODE

A. Metode Penelitian

Untuk melakukan sebuah penelitian, diperlukan adanya tahapan - tahapan yang tersusun dengan baik dan sistematis agar pelaksanaan penelitian tepat mencapai tujuan yang

diharapkan. Maka dari itu disusunlah tahapan - tahapan metodologi penelitian ini yaitu.



Sumber: Data penelitian (2018)

Gambar 1 Tahapan Penelitian

Penjelasan langkah penelitian Gambar 1.

1. Identifikasi Masalah

Melakukan identifikasi pada suatu masalah merupakan tahap awal pada proses penelitian. Tahap identifikasi masalah dilakukan untuk mengetahui permasalahan yang terjadi pada saat ini. Tahap ini perlu dilakukan agar penelitian yang dilakukan memiliki tujuan yang jelas untuk mengetahui berapa cluster pelanggan yang akurat, yang nantinya berguna bagi perusahaan untuk menentukan kebijakan penjualan yang tepat berdasar hasil clustering.

2. Studi Literatur

Dilakukan dengan mempelajari dan memahami teori - teori yang digunakan yaitu diantaranya data mining, metode k-means clustering dan RFM Model.

3. Pengumpulan Data

Tahap ini merupakan cara mengumpulkan data yang dilakukan dengan cara mengambil populasi data penjualan dari CV. Sinar Surya. Data yang digunakan pada penelitian ini adalah data penjualan selama tahun 2017 pada area pusat.

4. Praproses Data

Praproses data meliputi seleksi data, dan transformasi data. Data yang telah didapat dari tahap pengumpulan data diubah ke dalam bentuk file dataset file excel agar siap diproses ke tahap selanjutnya.

5. Cleaning Data

Untuk cleaning data atau membersihkan data, yaitu melengkapi data, menghapus data duplikat, menghilangkan noise.

6. Analisis Data

Pada tahap ini peneliti melakukan analisis data seperti darimana data didapatkan, menentukan atribut apa saja yang diperlukan dan juga variabel apa yang cocok. Peneliti mendapatkan data ini dari server CV Sinar Surya. Untuk bisa mendapatkan data tersebut perlu adanya hak akses, artinya data server tidak bisa langsung diakses tanpa adanya izin dari administrator server karena data tersebut bersifat rahasia. Pada tahap ini peneliti mendapatkan data penjualan sebanyak 7574 transaksi pada wilayah penjualan area pusat.

7. Hasil

Setelah data dianalisa dan ketersediaannya juga ada di database server CV Sinar Surya, selanjutnya data mentah akan diolah menjadi data yang berbentuk informasi. Data dipilih berdasarkan keperluan penelitian dengan melakukan seleksi data hingga menjadi sebuah dataset. Setelah data dipilih pada tahap penelitian ini selanjutnya yaitu pengolahan data. Pada tahap ini juga dilakukan pengujian data. Jika pengujian belum menghasilkan luaran yang sesuai dengan kebutuhan, maka tahapan penelitian kembali ke pengolahan data. Jika sudah memasuki tahapan ulangan atau kegagalan maka perlu adanya pemecahan masalah terhadap kegagalan tersebut dengan analisa pengolahan data

B. Populasi Penelitian

Populasi dalam penelitian ini adalah pelanggan aktif yang mempunyai transaksi di tahun 2017 pada area penjualan wilayah pusat. Jumlah populasi ada pada penelitian ini sebanyak 363 pelanggan aktif. Sedangkan data transaksi yang tercatat pada populasi penelitian ini sebanyak 7574 transaksi. Dalam penelitian ini peneliti tidak mengambil sampel penelitian dikarenakan sudah menggunakan seluruh populasi yang ada.

C. Metode Analisis Data

Agar dapat menyelesaikan persoalan dan mendapatkan solusi dari permasalahan maka analisis data yang digunakan adalah analisis data kuantitatif. Analisis data kuantitatif merupakan suatu analisa data yang digunakan apabila kesimpulan kesimpulan diperoleh dapat dibuktikan dengan angka angka dan juga perhitungan dalam perhitungan dipergunakan rumus yang ada hubungannya dengan penelitian.

Dalam hal ini peneliti menggunakan metode RFM dan K-Means Clustering.

Dalam penelitian ini peneliti mengambil data transaksi yang terjadi sepanjang tahun 2017 pada area penjualan wilayah pusat. Pola bisnis yang terjadi pada CV Sinar Surya yaitu salesman melakukan kunjungan rutin seminggu sekali ke pasar – pasar sesuai jadwal yang telah ditentukan. Peneliti berhasil mendapatkan data transaksi sebanyak 7574 record dari database server Sinar Surya. Selanjutnya data diolah agar sesuai dengan metode RFM dan menjadi dataset untuk proses K-means clustering. Sebelum dilakukan pengolahan data, terlebih dahulu dilakukan data seleksi data sehingga terbentuk variabel sesuai kebutuhan penelitian. Data yang dibutuhkan adalah customer id, tanggal transaksi terakhir tiap customer, frekuensi pembelian selama satu tahun untuk tiap-tiap customer, dan jumlah saldo penjualan pada tiap customer.

HASIL DAN PEMBAHASAN

A. Tahapan Penelitian

1. Praposes Data

Data yang digunakan dalam penelitian ini yaitu data transaksi 1 tahun selama tahun 2017. Dari hasil tahapan pengumpulan data pada metodologi penelitian, peneliti mendapat data sebanyak 7574 record. Data didapat dari mengakses server database CV sinar surya. Untuk mengakses server database tersebut diperlukan izin dari administrator terkait. Data dari database kemudian di ekspor ke dalam bentuk file excel.

2. Cleaning Data

Dilakukan pembersihan data terhadap data yang diperoleh. Dalam dataset terdapat atribut salesman, tahun, bulan, customer, tg_lj, SumOfNetto, dan invoice. Atribut sales, tahun, bulan, dan invoice dihapus karena data tersebut tidak terpakai pada penelitian ini. Adapun setelah dilakukan pembersihan data maka dilakukan pengolahan data melalui microsoft excel sehingga terbentuk data sesuai kebutuhan penelitian. Dari 7574 record tersebut, diolah dengan menghitung banyaknya transaksi untuk tiap-tiap customer dan menghasilkan 363 record dengan atribut customer, tanggal transaksi terakhir, tanggal akhir periode, total transaksi, dan total saldo penjualan. Selanjutnya dilakukan perubahan atribut agar lebih memudahkan pengolahan RFM Model dengan menghasilkan atribut Recency (R), Frequency (F), dan Monetary (M) bisa dilihat pada Tabel 1.

Tabel 1 Pemetaan data awal RFM

Atribut awal	Atribut akhir
Jarak antara tanggal transaksi terakhir dengan tanggal akhir periode penelitian	<i>Recency (R)</i>
Jumlah transaksi selama satu tahun	<i>Frequency (F)</i>
Jumlah saldo penjualan tiap pelanggan	<i>Monetary (M)</i>

B. Tahap Analisis

1. Proses RFM Model

Setelah data RFM didapat maka proses selanjutnya adalah penentuan RFM score, dimana tiap atribut dihitung score nilainya. Metode penghitungan RFM pada penelitian ini menggunakan simple fixed method. Penentuan scoring dengan metode ini tergantung pada pola bisnis masing masing perusahaan. Perusahaan sudah mempunyai pola bisnis yang tidak berubah selama satu tahun seperti metode pembayaran, pola order customer melalui salesman sehingga dimungkinkan terjadinya pola Recency, Frequency, and Monetary. Selama perusahaan tidak mengubah pola bisnis yang mereka pakai seperti metode pembayaran, pola order customer melalui salesman, kalkulasi RFM melalui metode simple fixed range bisa digunakan. Tahapan penentuan RFM Score adalah sebagai berikut:

a. Penentuan R Score

Pada penentuan Recency Score, Tabel 2, peneliti menggunakan jarak nilai Recency berdasar Recency yang paling lama yaitu 348 hari sampai Recency terbaru yaitu 1hari. Kriteria yang dipakai adalah Recency 348 hari sebagai batas bawah untuk R score 1, rata rata Recency termasuk ke dalam R score 3 atau tengah, Recency 1 hari untuk batas atas R score 5.

Tabel 2 Kriteria penentuan R Score

R Score	Rentang Nilai R
5	1 sampai 4 hari
4	15 sampai 45 hari
3	46 sampai 120 hari
2	121 sampai 179 hari
1	180 hari sampai 348 hari

b. Penentuan F Score

Pada penentuan Frequency Score (Tabel 3), peneliti menggunakan jarak nilai Frequency berdasar Frequency paling rendah yaitu 1 kali sampai Frequency tertinggi yaitu 51 kali. Kriteria yang dipakai adalah Frequency 1 kali untuk batas bawah nilai F score 1, rata rata Frequency menjadi batas atas dalam F score 3, dan Frequency 51 kali untuk batas atas F score.

Tabel 3 Kriteria penentuan F Score

F Score	Rentang Nilai F
5	Diatas 36 sampai 51 kali
4	21 sampai 35 kali
3	9 sampai 20 kali
2	4 sampai 8 kali
1	1 sampai 3 kali

c. Penentuan M Score

Pada penentuan Monetary Score (Tabel 4), peneliti menggunakan jarak nilai Monetary berdasar Monetary paling rendah yaitu 117.600 sampai Monetary tertinggi yaitu 463.664.300. Kriteria yang dipakai adalah Monetary sebanyak 117.600 untuk batas bawah nilai M Score, rata-rata Monetary termasuk ke dalam M Score 3, dan Monetary sebanyak 463.664.300 untuk batas atas M score.

Tabel 4 Kriteria penentuan M Score

Nilai M	Rentang Nilai M
5	150.000.001 sampai 463.664.300
4	50.000.001 sampai 150.000.0000
3	10.000.001 sampai 50.000.000
2	2.000.001 sampai 10.000.000
1	117.600 sampai 2.000.000

d. Hasil RFM Score

Penghitungan RFM Score pada melalui Microsoft Excel, yaitu :

$$\text{RFM Score} = (\text{Nilai Recency} \times 100) + (\text{Nilai Frequency} \times 10) + \text{Nilai Monetary} \dots\dots\dots (1)$$

2. Normalisasi Data

Proses normalisasi dilakukan sebelum dataset diproses melalui aplikasi Rapidminer. Normalisasi dilakukan agar skala dari data tidak terlalu jauh. Ada beberapa metode/teknik yang diterapkan untuk normalisasi data. Pada penelitian ini teknik normalisasi yang digunakan adalah normalisasi Z Score. Disebut juga zero-mean normalization, dimaka nilai dari sebuah atribut A dinormalisasi berdasar nilai rata-rata dan standar deviasi dari atribut A. Sebuah nilai v dari atribut A dinormalisasi menjadi atribut v'.

$$v' = \frac{v - \bar{A}}{\sigma_A} \dots\dots\dots (2)$$

- Dimana
- v' : Nilai normalisasi z score
 - v : Nilai awal
 - \bar{A} : Rata-rata
 - σ_A : Standar Deviasi

Tabel 5 Tabel atribut Z score

	\bar{A} (Mean)	σA (Standar Deviasi)
Recency	35,95	75,46
Frequency	20,87	13,01
Monetary	32.791.863	45.304.807,47

Sumber : Hasil pengolahan data (2018)

Normalisasi z score (Tabel 5) pada data RFM menggunakan software Microsoft Excel. Masing-masing nilai Recency, Frequency, and Monetary ditransformasi menjadi z score normalization.

3. Proses K-Means

Cluster 0 termasuk kelompok Everyday Shopper karena memiliki peningkatan transaksi dan memiliki nilai Monetary sedang sampai rendah dengan memiliki rentang RFM Score antara 111 sampai 543. Cluster 1 termasuk kelompok Golden Customer karena memiliki nilai frekuensi transaksi yang tinggi dan mempunyai nilai Monetary yang tinggi hingga tertinggi. Cluster 1 memiliki rentang RFM Score 443 sampai 555.

Berdasarkan dua cluster pada pengolahan data diatas, maka cluster 1 sebagai Golden Customer yang mempunyai karakteristik frekuensi tinggi dibuktikan pada tabel 6 yaitu 102 customer mempunyai rata-rata frekuensi transaksi 36 kali dalam satu tahun. Rata-rata transaksi pada cluster 1 lebih tinggi dari cluster 0 sebagai Everyday Shopper yang mempunyai rata-rata transaksi 14,95 kali dalam satu tahun. Cluster 1 sebagai Golden Customer dapat membuktikan hipotesa H0 yaitu pelanggan dengan frekuensi order yang tinggi mempunyai pengaruh terhadap pemetaan pelanggan yang potensial.

KESIMPULAN

Segmentasi pelanggan yang dihasilkan pada penelitian melalui proses K-Means Clustering di CV Sinar Surya yaitu sebanyak 2 cluster. Jumlah cluster optimal pada segmentasi pelanggan berdasarkan Davies Bouldin Index. Nilai negatif yang tinggi menunjukkan kinerja yang baik dari Index. Nilai Davies Bouldin Index pada jumlah cluster 2 senilai -0,344 sedangkan untuk jumlah cluster 3 sebanyak -0,244, jumlah cluster 4 sebanyak -0,221, jumlah cluster 5 sebanyak -0,247, dan jumlah cluster 6 sebanyak -0,234. Dua cluster yang dihasilkan yaitu cluster 0 sebanyak 261 pelanggan, dan cluster 1 sebanyak 102 pelanggan. Cluster 0 memiliki RFM Score antara 111 sampai 533 termasuk kelompok Everyday Shopper karena memiliki peningkatan transaksi dan memiliki nilai monetary sedang sampai rendah. Cluster 1 memiliki RFM Score 434 sampai 555 termasuk kelompok Golden Customer karena memiliki nilai frekuensi transaksi yang tinggi mempunyai

mempunyai nilai monetary yang tinggi hingga tertinggi. Hipotesa H0 yaitu pelanggan dengan frekuensi order yang tinggi mempunyai pengaruh terhadap pemetaan pelanggan yang potensial dapat terjawab dengan terbentuknya cluster 1 sebagai Golden Customer yang mempunyai rata-rata transaksi tinggi yaitu sebanyak 36 kali transaksi selama satu tahun. Dengan adanya penelitian pada segmentasi pelanggan ini diharapkan dapat membantu perusahaan dalam mengelompokkan pelanggan-pelanggan CV Sinar Surya sehingga perusahaan dapat menentukan strategi yang tepat untuk setiap kelompok pelanggan.

REFERENSI

- Andriyansyah, A., Wulanningrum, R., & Farida, I. N. (2019). *Penerapan Data Mining Menggunakan Kombinasi Algoritma K-Means Dengan Knn Untuk Menentukan Sasaran Promosi Penjualan Pada CV. Borobudur Prima Sejahtera*. Kediri.
- Putra, R. R., & Wadisman, C. (2018). Implementasi Data Mining Pemilihan Pelanggan Potensial Menggunakan Algoritma K Means. *INTECOMS: Journal of Information Technology and Computer Science*, 1(1), 72-77. Retrieved from <http://journal.ipm2kpe.or.id/index.php/INTECOM/article/view/141>
- Sumiah, A., & Mirantika, N. (2020). Perbandingan Metode K-Nearest Neighbor dan Naive Bayes untuk Rekomendasi Penentuan Mahasiswa Penerima Beasiswa pada Universitas Kuningan. *BUFFER INFORMATIKA*, 6(1), 1-10. Retrieved from <https://journal.uniku.ac.id/index.php/buffer/article/view/2907>
- Sutrisno, S., Afriyudi, A., & Widiyanto, W. (2014). *Penerapan Data Mining Pada Penjualan Menggunakan Metode Clustering Study Kasus Pt. Indomarco Palembang* (No. 78). Palembang. Retrieved from <http://eprints.binadarma.ac.id/78/>
- Suyanto. (2017). *Data Mining untuk Klasifikasi dan Klasterisasi Data*. Bandung, Indonesia: Informatika.
- Widiarina, W., & Wahono, R. S. (2015). Algoritma Cluster Dinamik Untuk Optimasi Cluster Pada Algoritma K-Means Dalam Pemetaan Nasabah Potensial. *Journal of Intelligent Systems*, 1(1), 33-36. Retrieved from

<http://journal.ilmukomputer.org/index.php?journal=jis&page=article&op=view&path%5B%5D=18>

Wulandari, G. F. (2014). Segmentasi Pelanggan Menggunakan Algoritma K-Means Untuk Customer Relationship Management (CRM) Pada Hijab Miulan. Retrieved from eprints.dinus.ac.id website: http://eprints.dinus.ac.id/13155/1/jurnal_13593.pdf

Zaman, B., & Wahyuni, S. (2019). Desain dan

Analisis Implementasi Aplikasi Berbasis Costumer Relationship Management Pada PT. Kareba Intermedia Makassar. *SISITI : Seminar Ilmiah Sistem Informasi Dan Teknologi Informasi*, 6(1), 249–258. Makasar: P4M STMIK Dipanegara Makassar. Retrieved from <https://ejurnal.dipanegara.ac.id/index.php/sisiti/article/view/148>