

Segmentasi Pelanggan Menggunakan Algoritma K-Means dan Model RFM (Studi Kasus: Industri Pengolahan Limbah Rumah Tangga)

Agustinus David Susanto

Manajemen Sistem Informasi, Program Pascasarjana, Universitas Gunadarma
Jl. Margonda Raya No. 100 Pondok Cina Depok Jawa Barat
E-mail: daviidsusanto@gmail.com

Abstrak

Pada era bisnis modern saat ini persaingan antar perusahaan dalam menarik *customer* sudah semakin inovatif dan bervariasi. Dengan pemanfaatan teknologi informasi tentu akan menghasilkan data yang akan terus bertambah seiring dengan terus berkembangnya bisnis. Data tersebut seringkali belum dapat dimanfaatkan dengan baik oleh perusahaan untuk dapat membantu pengembangan bisnisnya. ABC (bukan nama sebenarnya) adalah sebuah *startup application* berbasis *mobile* yang memberikan layanan pengumpulan limbah rumah tangga yang dapat didaur ulang seperti plastik kertas dan logam. Agar dapat memberikan pelayanan yang baik kepada seluruh customernya tentu perusahaan pengembang ABC harus memiliki strategi yang dapat diterapkan pada proses bisnisnya. Untuk itu diperlukan suatu analisa terhadap pola atau kebiasaan customer sehingga dapat dijadikan acuan dalam menerapkan strategi bisnis yang maksimal. Segmentasi dari customer dapat dilakukan menggunakan teknik *data mining* dengan algoritma k-means dan model RFM (*Recency, Frequency, Monetary*). Proses yang dilakukan dalam segmentasi di antaranya: pengenalan data, *data cleaning*, *data modeling*, pembentukan *cluster* menggunakan algoritma k-means validasi hasil dengan *elbow method* dan *silhouette coefficient* dan visualisasi hasil. Setelah dilakukan serangkaian proses segmentasi dan validasi terhadap data yang ada didapatkan nilai *silhouette coefficient* yaitu 0,7576967818785827 dan 3 *cluster* optimal dan *Cluster 1* merupakan *cluster* terbaik dengan rata-rata nilai RFM tertinggi dan segmentasi customer kedalam 7 *segment customer*.

Kata kunci : *Segmentasi, K-means, RFM, Data Mining, Industri Limbah Rumah Tangga.*

Pendahuluan

Saat ini teknologi informasi menjadi salah satu faktor yang sangat penting dan tidak dapat dipisahkan dalam segala bidang bisnis. Dengan pemanfaatan teknologi informasi tentu akan menghasilkan data yang akan terus bertambah banyak seiring dengan terus berkembangnya bisnis dan penerapan teknologi informasi. Data yang banyak tersebut seringkali belum dapat dimanfaatkan dengan baik oleh perusahaan untuk dapat membantu pengembangan bisnisnya. Data yang demikian banyak dapat dianalisa sehingga ditemukan pola dan pengetahuan yang menarik dengan menggunakan *data mining* [1].

ABC (nama sesungguhnya disembunyikan) adalah sebuah startup yang memberikan layanan pengumpulan limbah rumah tangga yang dapat didaur ulang seperti plastik kertas dan logam. Agar dapat memberikan pelayanan yang baik kepada

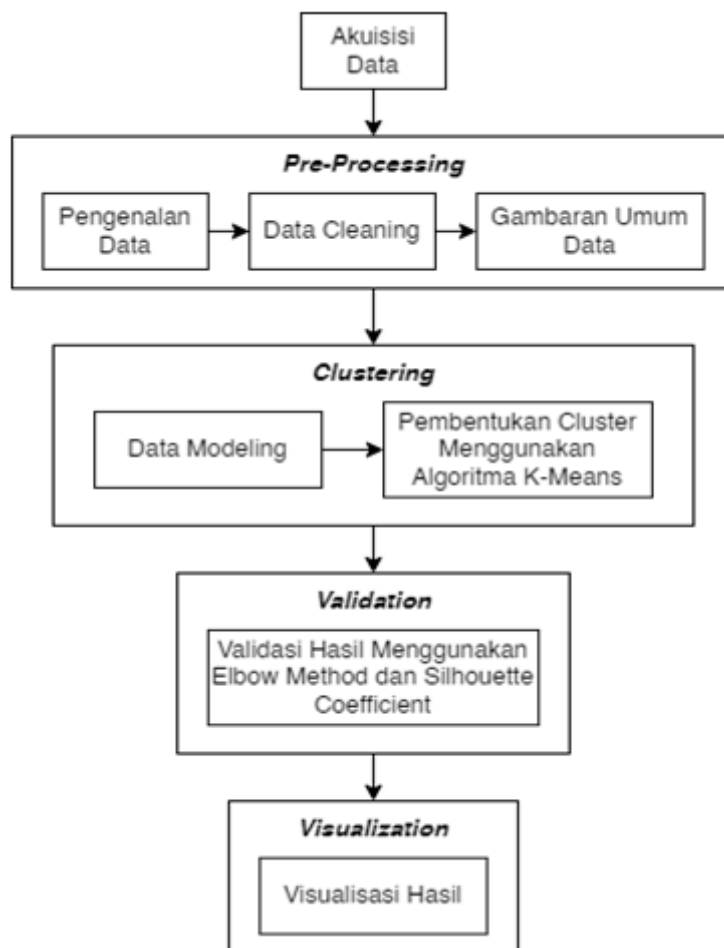
seluruh customernya tentu perusahaan pengembang ABC harus memiliki strategi yang dapat diterapkan pada proses bisnisnya. Untuk itu diperlukan suatu analisa terhadap pola atau kebiasaan *customer* sehingga dapat dijadikan acuan dalam menerapkan strategi bisnis yang maksimal. Informasi tentang segmen-segmen pelanggan digunakan untuk mengetahui perilaku dari pelanggan. Hal ini dapat dianalisa dengan menggunakan salah satu teknik data mining dengan pemodelan RFM (*Recency, Frequency, Monetary*). RFM digunakan untuk segmentasi pelanggan berdasarkan interval waktu kunjungan pelanggan, frekuensi kunjungan pelanggan dan besaran nilai yang dikeluarkan tiap transaksi. Dalam proses *clustering* digunakan algoritma *k-means*. Algoritma *k-means* merupakan salah satu algoritma yang sering digunakan untuk memecahkan masalah *clustering*.

Beberapa penelitian terdahulu terkait dengan segmentasi pelanggan menggunakan model RFM

dan algoritma *k-means* diberikan oleh: [2-7]. Untuk itu pada penelitian ini dilakukan segmentasi pelanggan dengan algoritma *k-means* dan model RFM untuk dapat mengelompokkan *customer* dan melakukan validasi hasil menggunakan metode validasi yang banyak digunakan pada proses data mining yaitu *elbow method* dan *silhouette coefficient*.

Metode Penelitian

Tahapan dari penelitian ini ditampilkan pada Gambar 1. Proses penelitian ini dilakukan menggunakan ide jupyter notebook dan bahasa pemrograman python.



Gambar 1: Tahapan Penelitian

Akuisisi Data

Data yang digunakan pada penelitian ini adalah data customer order yang tercatat pada database sistem ABC dari bulan Oktober 2020 sampai Oktober 2021. Data yang diambil merupakan data mentah atau data *raw* yang belum diolah dan dalam bentuk Microsoft Excel Format (.xlsx) file. Data sebelum dilakukan pre-processing berisi 11103 row dan 19 columns.

Pre-Processing

Tahapan *pre-processing* merupakan tahap awal dalam penelitian ini dimana data *raw* yang ada diolah dan diubah agar menjadi data yang bersih

dan sesuai untuk dilakukan proses *clustering*, pada tahap ini dilakukan beberapa proses diantaranya Pengenalan Data, *Data Cleaning* dan Gambaran Umum Data.

Pengenalan Data merupakan subtahapan pertama pada tahapan *pre-processing*. Pada subtahapan ini dataset berupa file .xlsx akan dibaca menggunakan *library* pandas. Seluruh elemen data akan dideskripsikan berdasarkan tipe data, jumlah data, jumlah *unique value* dan jumlah data null. Hal ini dapat lebih mempermudah dalam mengenali data untuk proses *data cleaning*.

Data Cleaning ialah sub tahapan selanjutnya setelah sub tahapan pengenalan data. Tujuan dari proses data cleaning adalah untuk menghilangkan *noise* pada data sehingga proses segmentasi yang

akan dilakukan dapat menghasilkan informasi yang *valid*. Pada proses ini dilakukan beberapa proses penghilangan *noise*, diantaranya adalah: menghilangkan kolom data yang tidak berpengaruh pada segmentasi, memberikan nilai pada kolom yang bernilai null, menghilangkan spasi pada kolom label dan mengubah format string menjadi datetime pada kolom Tanggal Selesai.

Gambaran Umum Data merupakan sub-tahapan analisa data yang bersifat umum. Tujuan dari proses ini adalah untuk memberikan gambaran terhadap data yang akan digunakan pada proses segmentasi. Pada tahap ini ditampilkan data diantaranya jumlah customer, jumlah item, total kotamadya dan total point per kotamadya.

Clustering

Setelah tahapan *pre-processing* selesai dilakukan maka tahap selanjutnya adalah tahap *clustering* dimana pada tahap ini akan dilakukan 2 proses yaitu *data modeling* dan *k-means clustering*. Data Modeling adalah proses *modeling* terhadap data order yang ada sesuai dengan model RFM dan algoritma *k-means*. Dimana data order akan dipecah menjadi beberapa variabel seperti variabel *Recency*, *Frequency* dan *Monetary*. Sehingga didapatkan hasil value distribution terhadap masing-masing variabel yang nantinya akan digunakan pada proses *clustering* menggunakan algoritma *k-means*. Pada tahap modeling juga dilakukan proses pembobotan dengan quartil dan normalisasi data menggunakan *MinMax Scaler*, proses normalisasi ini digunakan untuk menghilangkan *data outlier* atau data yang memiliki *range* perhitungan yang terlalu jauh atau berbeda skalanya. Pada proses normalisasi ini nilai data dari tiap variabel akan ditransformasi menjadi nilai interval 0 – 1. Perhitungan dari *min-max scaler* didefinisikan pada persamaan (1).

$$X_{scaled} = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (1)$$

Dimana X_{scaled} adalah nilai x setelah dilakukan normalisasi, x adalah nilai awal data, $\min(x)$ adalah nilai minimum pada *column*, $\max(x)$ adalah nilai maksimum dari *column*.

K-means clustering adalah proses utama pada penelitian ini, dimana pada proses ini dilakukan penentuan jumlah *cluster* optimal dengan menggunakan *Elbow Method* yang akan menghasilkan nilai *Sum of Square Error* [8]. Perhitungan nilai SSE dapat menggunakan persamaan (2) berikut ini.

$$SSE = \sum_{K=1}^K \sum_{X_i} |X_i - ck|^2 \quad (2)$$

Dimana K adalah cluster ke c , X_i adalah jarak data objek ke- i , ck adalah pusat cluster ke-1.

Proses selanjutnya adalah melakukan *clustering* menggunakan algoritma *k-means*. Algoritma *k-means* akan mengelompokkan item data dalam suatu

dataset ke suatu *cluster* berdasarkan jarak terdekat [9]. Proses *clustering* ini menggunakan *library data science* yang menyediakan modul *clustering* menggunakan algoritma *k-means* yaitu *sklearn*.

Validation

Tahap selanjutnya yang dilakukan adalah validasi, tahap ini merupakan proses validasi cluster yang telah terbentuk pada tahap clustering sebelumnya. Pada proses validasi ini digunakan metode *silhouette coefficient* untuk mengetahui kualitas dan kekuatan cluster yang terbentuk. Semakin tinggi nilai *silhouette coefficient* maka semakin baik kualitas dari cluster yang terbentuk. Proses perhitungan *silhouette coefficient* pada penelitian ini menggunakan fungsi yang sudah tersedia pada *library sklearn*, dimana nilai hasil clustering akan dimasukkan kedalam sebuah fungsi lalu akan menghasilkan output angka dengan interval 0 – 1. Semakin tinggi nilainya atau semakin mendekati 1 maka dapat dikatakan bahwa cluster tersebut baik atau cluster yang kuat keterikatan antar elemennya. Secara umum *silhouette* dapat dihitung menggunakan persamaan (3) berikut ini.

$$SC = \max_k SI(k) \quad (3)$$

Dimana SC adalah *silhouette coefficient*, SI adalah *silhouette index global*, k adalah jumlah *cluster*. Untuk menghitung nilai SI dari sebuah data ke- i , ada 2 komponen yaitu a_i dan b_i . Berikut persamaan untuk menghitung nilai SI dapat dilihat pada persamaan (4).

$$Si = \frac{(bi - ai)}{\max\{ai, bi\}} \quad (4)$$

Dimana Si adalah nilai *silhouette coefficient*, a_i adalah rata-rata jarak antara titik i dengan seluruh titik dalam A (cluster dimana titik i berada), b_i adalah rata-rata jarak antara titik i ke seluruh titik dalam cluster selain A. Kriteria subjektif pengukuran pengelompokkan berdasarkan *Silhouette Coefficient* (SC) [10] dapat dilihat pada Tabel 1.

Tabel 1: Kriteria pengukuran *Silhouette Coefficient*

Nilai SC	Kriteria
0,71 – 1,00	Struktur Kuat
0,51 – 0,70	Struktur Baik
0,26 – 0,50	Struktur Lemah
< 0,25	Struktur Buruk

5. Visualization Tahap ini merupakan tahap akhir penelitian dimana pada tahap ini dilakukan proses visualisasi hasil kedalam bentuk *chart* seperti *pie chart* dan juga *scatter plot* menggunakan *library matplotlib* dan interpretasi hasil menjadi sebuah informasi yang dapat digunakan sebagai bahan untuk menentukan strategi bisnis.

Hasil dan Pembahasan

Setelah tahapan proses penelitian dilakukan maka didapatkan hasil sebagai berikut. Hasil gambaran umum data berupa jumlah *customer*, transaksi, jenis item dan kotamadya dapat dilihat pada Tabel 2 serta top 5 point per kotamadya dapat dilihat pada Tabel 3.

Tabel 2: Gambaran Umum Data

Column	Quantity
Jenis Item	38
Transaksi	2896
Customer	1011
Kotamadya	83

Tabel 3: Gambaran Umum Data Lanjutan.

Kotamadya	Point
Jatiasih	2.148.088
Tapos	1.920.227
Cimanggis	1.310.925
Duren Sawit	1.238.739
Kelapa Gading	1.201.881

Tabel 4: Contoh 5 Data Raw

ID	Id User	Tanggal Selesai	Kotamadya
ORD-141021-00019	US-31577	2021-10-14	Jatiasih
NaN	NaN	NaN	NaN
NaN	NaN	NaN	NaN
ORD-230621-00017	US-18818	2021-06-28	Kebayoran Lama
ORD-130421-00002	US-04685	2021-04-20	Grogol Petamburan

Hasil dari serangkaian proses *pre-processing* yaitu tahap pengenalan data adalah sebuah tabel

informasi dari data diantaranya adalah informasi tentang *data types*, jumlah data, jumlah data unik, jumlah data *null*, dan nilai minimum dari data untuk masing-masing *column* yang ada pada data *raw*. Contoh 5 data *raw* dari *customer order* dapat dilihat pada Tabel 4. Setelah itu dilakukan proses *data cleaning* untuk menghilangkan noise yang terdapat pada data *raw*. Maka terbentuklah sebuah *dataset* baru yang dapat dilihat pada Tabel 5.

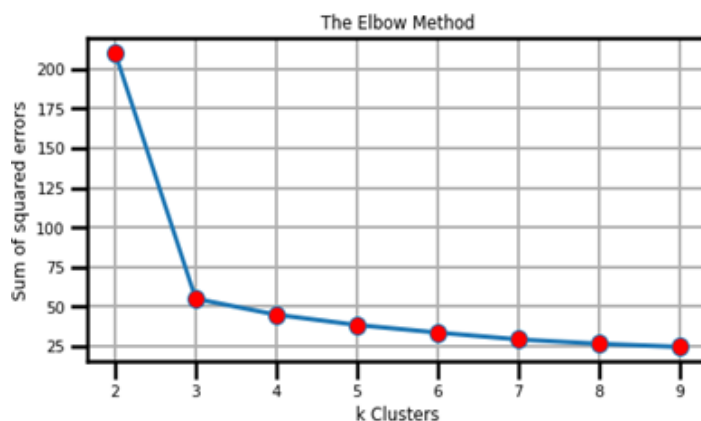
Tabel 5: Tabel Hasil Data Cleaning

ID	IdUser	Tanggal Selesai	Kotamadya
ORD-070621-00002	5641	2021-06-14	Kebayoran Lama
ORD-270621-00028	24618	2021-06-28	Makasar
ORD-240921-00006	2370	2021-09-27	Bekasi Barat
ORD-080321-00001	4248	2021-03-17	Pulo Gadung
ORD-011021-00015	25944	2021-10-04	Pamulang

Hasil dari tahapan *clustering* yaitu proses *data modeling* adalah sebuah nilai hasil perhitungan dari masing-masing variabel RFM dan juga pembobotan dari masing-masing variabelnya dengan menggunakan *quartil* yaitu pembobotan dengan nilai 1,2,3,4. Contoh hasil *data modeling* dari 5 *customer* pertama ditampilkan pada Tabel 6.

Tabel 6: Hasil Modeling dan Pembobotan

IDUser	R	F	M	Rq	Fq	Mq
3	23	6	96465	3	4	4
4	135	3	7456	1	3	2
5	106	6	28906	2	4	4
8	259	1	5500	1	1	2
9	14	3	8010	4	3	2

Gambar 2: . Grafik *Elbow Method*

Semakin tinggi nilai bobot RFM maka semakin baik customer tersebut untuk masing-masing variabelnya. Selanjutnya hasil *data modeling* akan digunakan dalam proses penentuan jumlah *cluster* optimal dengan menggunakan *elbow method* dan didapatkan hasil yaitu *cluster* optimal yang terbentuk adalah $k = 3$ karena penurunan nilai SSE tertinggi terjadi pada $k = 3$ dan penurunan tidak signifikan untuk *cluster* setelahnya. Hasil perhitungan *elbow method* dapat dilihat pada Gambar 2.

Nilai *Sum of Square Error* dapat dilihat pada Tabel 7.

Tabel 7: Nilai *Sum of Square Error*

k (cluster)	Sum of Square Error	Selisih SSE
2	210,40099	210,40099
3	54,99142	155,40957
4	44,94880	10,04262
5	38,45687	6,49193
6	33,65637	4,80050
7	29,34597	4,31040
8	26,60605	2,73992
9	24,63456	1,97149

Dengan asumsi nilai *cluster* optimal adalah $k = 3$. Maka dilakukan pengujian *clustering* dengan algoritma k-means menggunakan $k = 3$ dan menggunakan 11101 *row* data transaksi serta 1101 customer. 5 Sample data hasil *clustering* yang terbentuk dengan $k = 3$ dapat dilihat pada Tabel 8.

Tabel 8: Hasil *Clustering* Dengan $k = 3$

IDUser	R	F	M	Segment	Cluster
142	1	4	4	Lost Customers	1
2	4	4	4	Best Customer	1
9	4	3	2	Others	2
230	1	1	1	Lost Customers	0
142	1	1	1	Lost Customers	0

Pada *clustering* menggunakan $k = 3$ terbentuk 3 *cluster* yang diberi label *cluster* 0, *cluster* 1 dan *cluster* 2 dimana masing-masing *cluster* memiliki jumlah anggota *cluster* sebanyak 561 untuk *cluster* 0, 308 untuk *cluster* 1 dan 232 untuk *cluster* 2. Pada ketiga *cluster* yang terbentuk, *cluster* 1 merupakan *cluster* terbaik karena nilai rata-rata *recency* dari anggota *cluster* 1 tinggi artinya rentan waktu transaksi terakhirnya kecil, untuk nilai *frequency*-nya didominasi oleh customer dengan nilai tinggi, dan *monetary* tinggi. Untuk peringkat kedua adalah *cluster* 2 karena pada *cluster* ini nilai *recency*-nya sedang, *frequency* sedang dan *monetary*

tinggi. *cluster* terakhir adalah *cluster* 0 dimana anggota *cluster* 0 memiliki rata-rata nilai *recency* yang rendah yang berarti rentan waktu transaksi terakhirnya lama, untuk nilai *frequency* sedang dan *monetary* rendah.

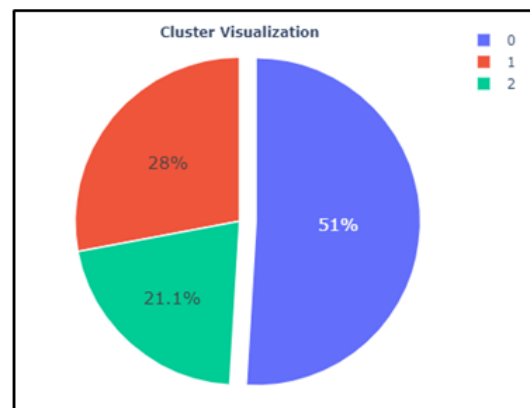
Setelah hasil proses *clustering* telah didapatkan maka selanjutnya dilakukan validasi hasil *cluster* dengan menggunakan *silhouette coefficient* untuk mengetahui apakah *cluster* yang terbentuk memiliki keterikatan antar anggota yang baik. Validasi dilakukan dengan *range cluster* dari 2 hingga 9. Hasilnya dapat dilihat pada Tabel 9.

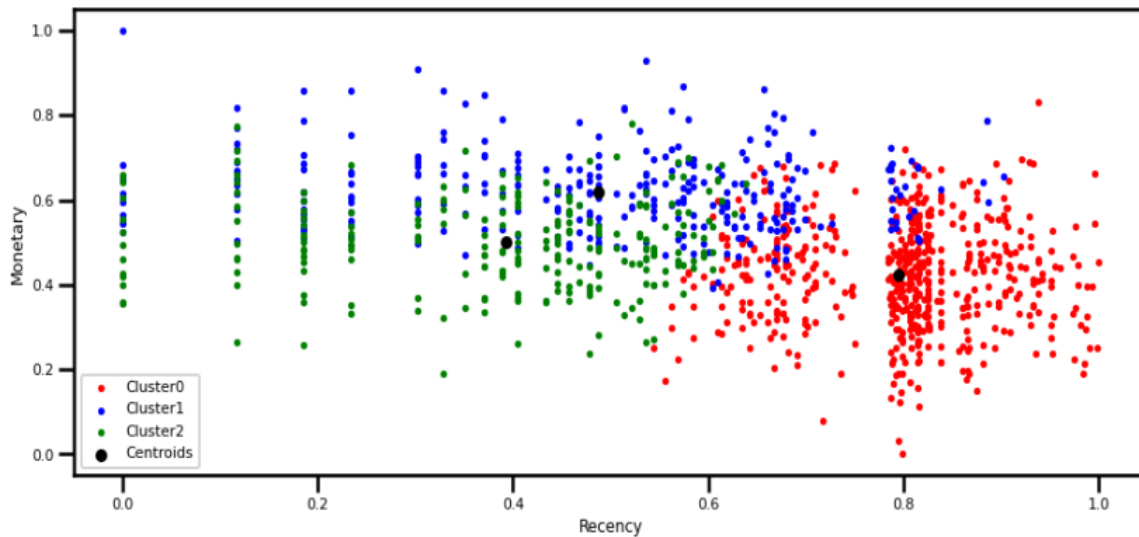
Tabel 9: Hasil Validasi *Silhouette Coefficient*

k (cluster)	Silhouette Score
2	0,6823993794315959
3	0,7576967818785827
4	0,6721915865126188
5	0,4399114367286028
6	0,4016695249366388
7	0,3886216016851118
8	0,3525769202504598
9	0,39797278737118463

Dari hasil perhitungan *silhouette score* diatas dapat diketahui bahwa *cluster* dengan nilai $k = 3$ merupakan *cluster* yang paling baik dengan nilai *silhouette score* paling tinggi dan mendekati 1 yaitu dengan nilai 0,7576967818785827.

Hasil *clustering* dan segmentasi yang masih berupa tabel dan angka selanjutnya divisualisasikan kedalam bentuk *chart* dan *plot* agar dapat mempermudah perusahaan ABC dalam menentukan strategi bisnisnya. Visualisasi hasil *cluster* dapat dilihat pada Gambar 3 hingga Gambar 7.

Gambar 3: Visualisasi Hasil *Cluster*



Gambar 4: Scatter Plot Hasil Clustering Dengan K-means

Berdasarkan hasil visualisasi *chart, scatter plot* pada Gambar 3 dan Gambar 4 akan diinterpretasikan kedalam sebuah informasi yaitu:

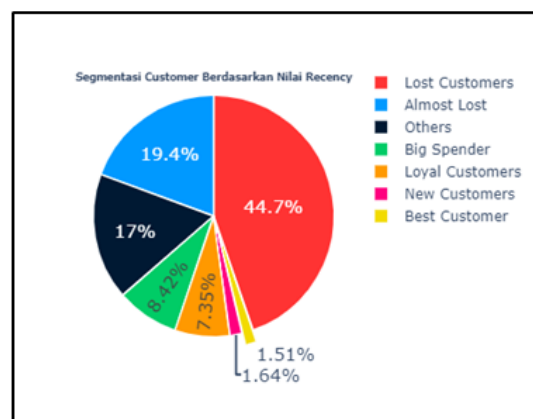
1. Anggota yang termasuk kedalam *cluster 1* merupakan *customer-customer* yang paling menguntungkan bagi perusahaan. Karenanya diperlukan suatu strategi bisnis yang dapat terus meningkatkan rasa percaya dan kepuasan terhadap pelayanan yang diberikan.
2. Anggota yang termasuk kedalam *cluster 2* merupakan *customer-customer* yang *acceptable* dengan pelayanan yang diberikan oleh ABC. Namun diperlukan suatu strategi bisnis yang dapat memenuhi kebutuhan *customer* pada *cluster 2* yang mungkin belum dimiliki oleh layanan saat ini.
3. Anggota yang termasuk kedalam *cluster 0* merupakan *customer-customer* yang belum memiliki tingkat kepercayaan dan kepuasan terhadap layanan yang diberikan sehingga perlu adanya strategi bisnis yang tepat dan menarik.

Selain itu, berdasarkan hasil segmentasi yang divisualisasikan pada Gambar 5, Gambar 6 dan Gambar 7. Didapatkan informasi sebagai berikut.

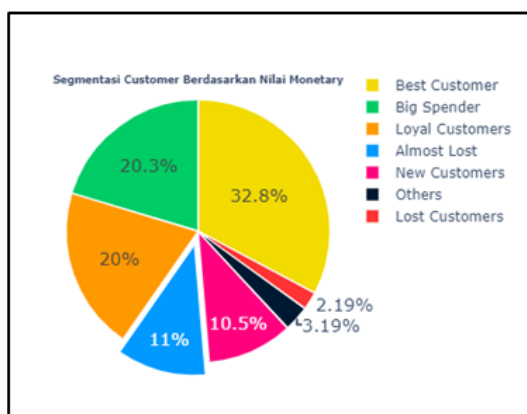
1. *Recency*: pada variabel *recency* hasil segmentasi yang terbentuk didominasi oleh segment *Lost Customer* dengan nilai persentase sebesar 44,7%, hal ini menandakan bahwa tingkat rentan waktu *customer* dalam bertransaksi tergolong cukup lama yang menandakan bahwa banyak *customer* yang sudah tidak menggunakan layanan lagi, hal ini perlu dapat digunakan oleh perusahaan ABC untuk dapat menentukan strategi bisnis yang bisa mengakuisisi *customer* baru.

2. *Frequency*: pada variabel *frequency* segmentasi tertinggi yang terbentuk adalah *segment Best Customer* dengan nilai persentase 34%, hal ini menandakan bahwa berdasarkan tingkat frekuensi transaksi cukup baik.

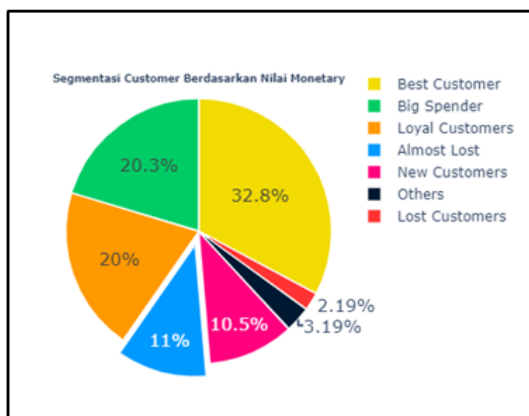
3. *Monetary*: pada variabel *monetary* segmentasi yang memiliki nilai persentase tertinggi adalah *segment Best Customer* dengan nilai persentase sebesar 32,8%. serupa dengan variabel *frequency*, dapat dikatakan bahwa tingkat *monetary* dari *customer* cukup baik



Gambar 5: Visualisasi Segmentasi Customer Berdasarkan Recency



Gambar 6: Visualisasi Segmentasi Customer Berdasarkan Frequency



Gambar 7: . Visualisasi Segmentasi Customer Berdasarkan Frequency

Penutup

Proses *clustering* terhadap customer ABC menggunakan algoritma *k-means* dan model RFM berhasil dilakukan dan memberikan hasil yang baik. Dimana berdasarkan *clustering* menggunakan *k-means* dihasilkan 3 buah *cluster* yaitu Cluster 0, Cluster 1 dan Cluster 2. Cluster 1 merupakan *cluster* terbaik dan paling menguntungkan bagi perusahaan dengan jumlah anggota *cluster* sebanyak 308 customer.

Pada segmentasi berdasarkan variabel RFM, diketahui bahwa variabel recency masih terbilang rendah karena didominasi oleh *segment Lost Customer* dengan nilai persentase 44,7% sedangkan untuk variabel *frequency* dan *monetary* dikategorikan cukup baik karena persentase tertinggi adalah *segment Best Customer* dengan persentase 34% dan 32,8%. Proses validasi hasil menggunakan *silhouette coefficient* terhadap cluster yang terbentuk menunjukkan bahwa jumlah *cluster* $k = 3$ merupakan

cluster paling baik dengan nilai *silhouette coefficient* 0,7576967818785827. Dengan hasil penelitian yang telah dilakukan, didapatkan beberapa informasi terkait pola customer ABC yang dapat digunakan sebagai bahan untuk menentukan strategi bisnisnya.

Daftar Pustaka

- [1] Jiawei Han, Micheline Kamber, Jian Pei, "Cluster Analysis: Basic Concepts and Methods, In the Morgan Kaufmann Series in Data Management Systems, Data Mining (Third Edition)", Morgan Kaufmann, Pages 443-495, <https://doi.org/10.1016/B978-0-12-381479-1.00010-1>. 2012.
- [2] Siti Monalisa, "Klasterisasi Customer Lifetime Value Dengan Model LRFM Menggunakan Algoritma K-means", Jurnal Teknologi Informasi dan Ilmu Komputer, Vol. 5, No. 2, hlm. 247-252, Mei 2018. DOI: 10.25126/jtiik.201852690
- [3] Jamal & Dwiyanto, "Analisis RFM dan Algoritma K-means untuk Clustering Loyalitas Customer", Jurnal Ilmiah Ilmu-Ilmu Teknik, Vol.9, No.1, Mei 2019.
- [4] Wiratama, Nanang & Fitra, "Analisis Segmentasi Pelanggan Dengan RFM Model Pada PT. Arthamas Citra Mandiri Menggunakan Metode Fuzzy C-Means Clustering", Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, Vol. 3, No. 2, hlm. 1986-1993, Februari 2019.
- [5] D.A. Indah & D.A Kadek, "Analisis Perbandingan Metode Elbow dan Silhouette pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali", Jurnal Matrix, Vol. 9, No. 3, November 2019.
- [6] N. Puspitasari, J. A. Widiand & N. B. Setiawan, "Segmentasi pelanggan menggunakan algoritme bisecting k-means berdasarkan model recency, frequency, dan monetary (RFM)", Jurnal Teknologi dan Sistem Komputer, vol. 8, no. 2, pp. 78-83, Apr 2020. <https://doi.org/10.14710/jtsiskom.8.2.2020.78-83>
- [7] Rahma, Fahd & Kankamol, "Customer Segmentation Based on RFM Model Using K-Means, K-Medoids, and DBSCAN Methods", Lontar Komputer:Jurnal Ilmiah Teknologi Informasi, Vol.11, No.1, p.32-43, April 2020.
- [8] Rena, "Improved the Performance of the K-Means Cluster Using the Sum of Squared Error (SSE) optimized by using the Elbow Method", Journal of Physics: Conference Series, Vol. 1361. 012015. 2019.

- [9] B. Bangoria, N. Mankad, N. & V. Pambhar, "A Survey on Efficient Enhanced K-Means Clustering Algorithm", International Journal for Scientific Research & Development, I (9), pp.1698-700, 2013.
- [10] L. Kaufmann and P. J. Rousseeuw, "Finding groups in data: An introduction to cluster analysis", USA: Wiley Series in Probability and Statistics, 2005.