

Penerapan Algoritma K-Means Clustering Untuk Segmentasi Pelanggan B-Clean Laundry Banyuwangi

Dwi Yulian RL¹, Sony Panca Budiarto²

^{1,2}Teknik Informatika

Sekolah Tinggi Ilmu Komputer PGRI Banyuwangi

Banyuwangi, Indonesia

e-mail: ¹lingkeku@gmail.com, ²sonystikombanyuwangi@gmail.com

Abstrak

Dalam menjalankan roda bisnisnya, B-Clean Laundry juga membutuhkan suatu strategi marketing dengan tujuan meningkatkan pelayanan kepada pelanggannya. Karena itu kebutuhan informasi yang berkaitan dengan faktor perilaku dari pelanggan sangat diperlukan sebagai pengetahuan bisnis untuk bersaing dari kompetitor lainnya. Tema penelitian yang diangkat bertujuan untuk mengetahui perilaku pelanggan dan melakukan clustering terhadap data pelanggan tersebut sehingga informasi yang didapatkan dapat membantu dalam pengambilan keputusan manajemen serta dapat dijadikan rujukan dalam pemberian reward kepada pelanggan setia. Metode clustering yang digunakan adalah metode K-Means, sedangkan data yang digunakan adalah data transaksi selama bulan Mei 2022 dengan jumlah record transaksi sebanyak 442 data dari jumlah pelanggan sebanyak 240 orang. Untuk pengujian cluster digunakan metode Rasio BCV/WCV. Setelah dilakukan analisis perbandingan berapa jumlah cluster yang ideal didapatkan bahwa 2 cluster memiliki hasil lebih baik jika dibandingkan dengan pembagian 3 dan 4 cluster. Hasil yang didapatkan menunjukkan cluster 1 beranggotakan 237 pelanggan, sedangkan cluster 2 beranggotakan 3 pelanggan, dengan nilai rasio sebesar 0,009.

Kata kunci: K-Means, Clustering, Segmentasi Pelanggan, Data Mining.

Abstract

When running its business, B-Clean Laundry also needs a marketing strategy with the aim of improving service to its customers. Therefore, information related to behavioral factors from customers is needed as business knowledge to compete with other competitors. The research theme raised aims to determine customer behavior and classify customer data so that the information obtained can assist in management decision making and can be used as a reference in providing rewards to loyal customers. The clustering method used is the K-Means method, while the data used is transaction data for May 2022 with a total of 442 transaction records from a total of 240 customers. For cluster testing, the BCV/WCV ratio method is used. After analyzing the comparison of the ideal number of clusters, it was found that 2 clusters had better results when compared to the division of 3 and 4 clusters. The results obtained indicate that cluster 1 consists of 237 customers, while cluster 2 consists of 3 customers, with a ratio value of 0.009.

Keywords: K-Means, Clustering, Customer Segmentation, Data Mining.

1. Pendahuluan

B-Clean Laundry merupakan UMKM yang bergerak dibidang pelayanan jasa laundry, berlokasi di kota Banyuwangi Jawa Timur. B-Clean Laundry telah beroperasi tahun 2007 dan telah menerapkan sistem informasi komputer dalam pelayanannya. Tetapi data warehouse yang dimiliki belum bisa dimanfaatkan secara maksimal untuk pengambilan keputusan manajemen dan strategi dalam pemasaran. Hal ini sangat disayangkan karena bisnis ini termasuk dalam kategori bisnis dengan perputaran yang cepat, maksudnya rentang waktu permintaan pelanggan antara permintaan pertama dan permintaan selanjutnya pada jasa ini yang memakan waktu relatif singkat[1]. Salah satu strategi marketing yang dapat digunakan oleh B-Clean Laundry adalah adanya informasi karakteristik dari pelanggan dimana pelanggan dapat dikelompokkan menurut karakteristik tertentu. Pelanggan merupakan aset perusahaan yang perlu dijaga didalam suatu bisnis, sehingga perusahaan harus memiliki cara supaya aset tersebut bisa lebih berkembang[2].

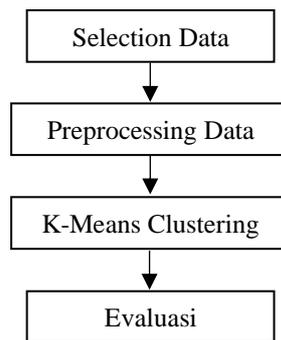
Algoritma *K-Means* pertama kali diperkenalkan oleh J.B. MacQueen pada tahun 1976. Metode *K-Means* ini memilah data kedalam *cluster* (kelompok) sehingga data yang memiliki karakteristik yang mirip akan masuk ke dalam *cluster* yang sama dan data yang mempunyai karakteristik berbeda akan masuk kedalam kelompok yang lain[3].

Beberapa penelitian yang menerapkan metode *K-Means* pernah dilakukan. Metode *K-Means* clustering digunakan dalam analisis pengaruh penilaian asesor terhadap kinerja guru mata pelajaran. Hasil akhir dari penelitian ini adalah terbentuk tiga kelompok dari 40 data yang diterapkan dalam dua kategori data penilaian, yaitu penilaian kinerja guru oleh asesor dan penilaian hasil perhitungan angka kredit dalam satu tahun. Data penilaian tersebut menghasilkan tiga kelompok yaitu; sangat baik, baik, dan cukup baik. Adanya pembagian cluster ini pihak sekolah dapat mengetahui guru-guru mana yang layak untuk mendapatkan apresiasi[4].

Metode *K-Means* juga digunakan dalam segmentasi pelanggan sebagai strategi pemasaran dengan studi kasus PT Coversuper Indonesia Global. Proses segmentasi digunakan untuk mengetahui karakteristik dari pelanggan, menghasilkan 4 cluster dari keseluruhan data pelanggan sebanyak 736 data, yaitu kelompok pelanggan Consumers 226, Ordinary 186, Big Consumers 299 dan Top Class 25[5]. Di penelitian lain juga disebutkan bahwa metode *K-Means* dapat digunakan dalam menilai loyalitas customer yang dikombinasikan dengan metode RFM (*Recency, Frquency, Monerity*). Pada penelitian tersebut telah ditetapkan 5 jenis kategori customer. Hasil akhir yang didapat dari 503 data, Customer dengan status Sangat Loyal sebanyak 14, Loyal 44, Biasa 219, Kurang Loyal 192 dan Tidak Loyal 34[6]. Dari penjabaran tersebut diatas dapat diketahui bahwa metode *K-Means* dapat digunakan untuk pengelompokan suatu data.

2. Metode Penelitian

Tahapan yang dilakukan dalam penelitian ini dibagi empat tahapan : *selection* data, *preprocessing* data, *K-Means* clustering dan evaluasi[6]. Tahapan penelitian disajikan dalam gambar 1.



Gambar 6. Tahapan dari Penelitian

2.1. Selection Data

Tahapan dari selection data disajikan dalam diagram seperti pada gambar 2.

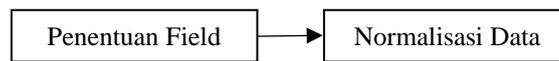


Gambar 2. Tahapan Selection Data

Data sampel sebagai dataset diambil dari transaksi pelanggan laundry yang dikumpulkan selama satu bulan di bulan Mei 2022. Selanjutnya dilakukan selection dengan menghilangkan *record* yang tidak relevan dengan kebutuhan. Hasil dari proses selection merupakan dataset yang akan menjadi input untuk proses preproccessing.

2.2. Preprocessing Data

Tahapan Preprocessing Data disajikan dalam diagram seperti pada gambar 3. Dari dataset yang telah didapat, tahap berikutnya adalah proses penentuan field, data yang dipilih hanyalah data nama pelanggan, jumlah kunjungan dan *monetary*. Jumlah kunjungan didapatkan dari jumlah nota yang ada dari nama pelanggan yang sama dalam periode satu bulan. Monetary didapatkan dari akumulasi dari tiap jumlah besaran rupiah dari nota-nota yang ada sesuai nama pelanggan.



Gambar 3. Tahapan Preprocessing Data

Tahapan berikutnya adalah tahapan normalisasi data atau *Descritization (Equal Width)* agar skala data tidak terlalu jauh[7]. Untuk proses normalisasi data digunakan metode MinMax dengan menggunakan skala min = 1 dan maks = 5. Rumus persamaan MinMax :

$$x_{baru} = \frac{(x_{lama} - \min x_{lama})}{(\max x_{lama} - \min x_{lama})} (\max x_{baru} - \min x_{baru}) + \min x_{baru} \quad (1)$$

X_{baru} adalah nilai data hasil normalisasi. X_{lama} adalah nilai data sebelumnya dinormalisasi. $Max_{x_{lama}}$ adalah nilai terbesar dari data sebelum normalisasi. $Min_{x_{lama}}$ adalah nilai terkecil dari data sebelum normalisasi. $Max_{x_{baru}}$ adalah nilai terbesar dari skala yang digunakan, sedangkan $min_{x_{baru}}$ adalah nilai terkecil dari skala yang digunakan [8].

2.3. K-Means Clustering

Algoritma K-Means merupakan algoritma clustering yang mengelompokkan data berdasarkan titik pusat cluster (*centroid*). Berikut merupakan algoritma K-Means[9]:

1. Tentukan jumlah cluster
2. Hitung nilai centroid awal secara acak
3. Menghitung jarak antara titik centroid dengan tiap data
4. Alokasikan data ke jarak centroid terdekat
5. Ulangi langkah 2 s/d 4 hingga nilai centroid tetap dan anggota cluster tidak berpindah ke cluster lain

Pada penelitian ini akan dilakukan perbandingan jumlah cluster antara 2 sampai 4 dan dipilih jumlah cluster mana yang terbaik dengan menggunakan perhitungan DBI pada tahap evaluasi.

2.4. Evaluasi

Metode yang digunakan untuk evaluasi hasil clustering adalah dengan perhitungan rasio antara BetweenClass Variation (BCV) dan Within-Class Variation (WCV), yaitu rasio = BCV/WCV. BCV merupakan nilai jarak dari seluruh centroid yang terbentuk, sedangkan WCV adalah nilai keseluruhan jarak terkecil antara data dengan centroid pada masing-masing cluster. Pengujian ini dilakukan untuk menentukan kriteria penilaian hasil dari perhitungan clustering. Semakin besar rasio hasil perhitungan yang diperoleh, maka semakin baik tingkat kualitas clustering tersebut[4], [10].

Rumus perhitungan BCV :

$$BCV = \frac{1}{n_k} \sum_{k \neq k'} d(m_k, m_{k'}) ; k, k' = 1, \dots, K \quad (2)$$

dengan

- K = jumlah cluster
 m_k = pusat cluster ke- k
 n_k = jumlah data di cluster ke- k
 $d(m_k, m_{k'})$ = $\|m_k - m_{k'}\|^2$

Rumus perhitungan WCV :

$$WCV = \sum_{k=1}^K \sum_{p \in C_k} d(p, m_k) \quad (3)$$

Keterangan:

- $p \in C_k$ = data di cluster ke- k
 m_k = pusat cluster ke- k
 $d(p, m_k)$ = $\|p - m_k\|^2$

3. Hasil dan Pembahasan

3.1. Selection Data

Pada tahap ini, data sampel transaksi yang digunakan adalah transaksi dalam satu bulan dibulan Mei 2022. Dari proses pengumpulan data didapatkan 442 data transaksi. Struktur tabel database dari data mentah yang didapatkan, disajikan dalam tabel 1.

Tabel 1. Struktur Tabel Transaksi

NO	Nama Atribut	Deskripsi
1	No Nota	Id nota
2	Customer	Nama pelanggan
3	No Telp Customer	No telp pelanggan
4	Progress Pengerjaan	Persentase dari pengerjaan laundry
5	Outlet	Nama dari outlet yang mengerjakan
6	Tgl Terima	Tanggal penerimaan laundry
7	Tgl Selesai	Tanggal terselesaikannya pengerjaan
8	Subtotal	Subtotal tagihan
9	Tambahan Express	Besarnya layanan tambahan express
10	Diskon	Besarnya diskon yang diberikan
11	Pajak	Besarnya pajak
12	Total Tagihan	Total dengan layanan express, pajak dan dikurangi diskon
13	Jenis	Jenis layanan laundry
14	Pembayaran	Status dari pembayaran
15	Pengambilan	Status dari pengambilan
16	Pembuat Nota	Nama karyawan pembuat nota
17	Keterangan Nota	Keterangan tambahan untuk nota
18	Detail Layanan	Rincian dari layanan yang dipilih pelanggan

3.2. Preprocessing Data

Berikutnya dilakukan reduksi field dengan menghilangkan semua data yang dianggap tidak valid. Langkah ini bertujuan untuk menentukan field mana yang digunakan dalam mining. Atribut yang dihasilkan adalah customer untuk mencatat nama pelanggan, jumlah nota tiap pelanggan dan total yang dibayar tiap pelanggan. Dari proses reduksi didapat 240 pelanggan dengan record transaksinya. Berikut adalah bentuk data transaksi yang telah dilakukan reduksi field atau penghapusan field yang tidak dibutuhkan, terlihat pada tabel 2.

Tabel 2. Hasil Reduksi Data Transaksi

Customer	Jumlah Nota	Total
Abi	1	30.000
Adel Bu	8	1.447.000
Adli	5	347.500
Afif	1	21.200
Agnes	1	152.400
...
Zila	4	316.600

Dari data di tabel 2 tersebut selanjutnya dilakukan normalisasi data menggunakan MinMax pada persamaan 1 pembahasan sebelumnya. Hasil normalisasi data terlihat pada tabel 3.

Tabel 3. Hasil Normalisasi Data

Customer	Jumlah Nota	Total
Abi	1,00	1,04
Adel Bu	3,80	4,86
Adli	2,60	1,89
Afif	1,00	1,01
Agnes	1,00	1,37
...
Zila	2,20	1,81

3.3. K-Means Clustering

Output dari K-Means clustering berupa informasi yang menampilkan jumlah anggota setiap cluster dan titik pusat atau centroid. Dari hasil perhitungan didapat bahwa untuk jumlah cluster sebanyak 2, iterasi dilakukan sebanyak 3 kali, sedangkan untuk jumlah cluster 3 dan 4 dilakukan iterasi sebanyak 2 kali. Hasil proses clustering terlihat pada tabel 4, tabel 5 dan tabel 6. Jumlah anggota masing-masing cluster adalah untuk cluster 1 beranggotakan 237 pelanggan dan cluster 2 beranggotakan 3 pelanggan.

Tabel 4. Hasil K-Means 2 Cluster

Customer	Jumlah Nota	Total	C1	C2	Keanggotaan
Abi	1,00	1,04	0,28	2,72	C1
Adel Bu	3,80	4,86	4,58	3,82	C2
Adli	2,60	1,89	1,58	1,41	C1
Afif	1,00	1,01	0,28	2,72	C1
Agnes	1,00	1,37	0,43	2,74	C1
...
Zila	2,20	1,81	1,21	1,71	C1

Tabel 5. Hasil K-Means 3 Cluster

Customer	Jumlah Nota	Total	C1	C2	C3	Keanggotaan
Abi	1,00	1,04	0,10	1,08	3,07	C1
Adel Bu	3,80	4,86	4,68	4,19	3,83	C3
Adli	2,60	1,89	1,73	1,00	1,70	C2
Afif	1,00	1,01	0,11	1,08	3,07	C1
Agnes	1,00	1,37	0,35	1,13	3,08	C1
...
Zila	2,20	1,81	1,34	0,78	2,02	C2

Dari hasil clustering dengan 3 cluster didapatkan jumlah anggota masing-masing cluster adalah untuk cluster 1 beranggotakan 193 pelanggan, cluster 2 beranggotakan 43 pelanggan dan cluster 3 beranggotakan 4 pelanggan. Untuk jumlah cluster 4 didapatkan jumlah anggota masing-masing cluster adalah untuk cluster 1 beranggotakan 193 pelanggan, cluster 2 beranggotakan 42 pelanggan, cluster 3 beranggotakan 3 pelanggan dan cluster 4 beranggotakan 2 pelanggan.

Tabel 6. Hasil K-Means 4 Cluster

Customer	Jumlah Nota	Total	C1	C2	C3	C4	Keanggotaan
Abi	1,00	1,04	0,10	1,03	2,30	2,60	C1
Adel Bu	3,80	4,86	4,68	4,21	3,86	3,83	C4
Adli	2,60	1,89	1,73	1,03	1,11	1,32	C2
Afif	1,00	1,01	0,11	1,03	2,30	2,60	C1
Agnes	1,00	1,37	0,35	1,08	2,32	2,62	C1
...
Zila	2,20	1,81	1,34	0,79	1,34	1,60	C2

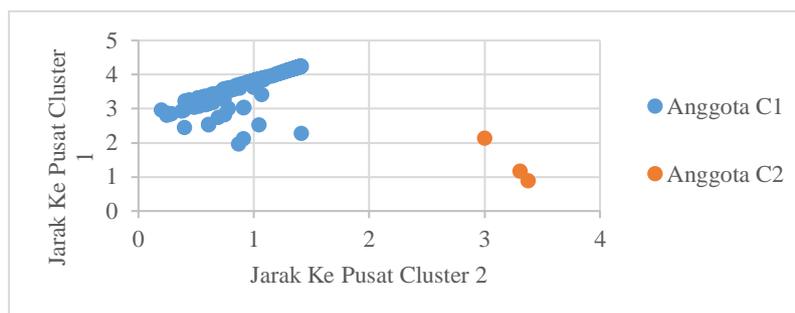
3.4. Evaluasi

Hasil evaluasi terhadap perhitungan nilai rasio ditunjukkan dalam tabel 7.

Tabel 7. Hasil Evaluasi WCV, BCV dan Rasio

Jumlah Cluster	Iterasi	WCV	BCV	Rasio
2	1	329,509	2,828	0,009
	2	110,007	2,775	0,025
	3	105,849	2,444	0,023
3	1	50,744	11,314	0,223
	2	76,110	5,928	0,078
4	1	50,422	18,385	0,365
	2	76,796	8,760	0,114

Dari tabel 7 terlihat bahwa nilai rasio terkecil ada pada jumlah cluster 2 pada iterasi ke 1 yaitu sebesar 0,009 yang menandakan bahwa nilai centroid dan cluster yang dihasilkan memiliki kualitas yang terbaik. Grafik visualisasi hasil clustering terlihat pada gambar 4, dimana C1 menandakan anggota dari cluster 1 dengan jumlah 3 atau 1% dan C2 adalah anggota dari cluster 2 dengan jumlah 237 atau 99%.



Gambar 4. Hasil Klustering K-Means

4. Kesimpulan

Dari hasil pengambilan data transaksi laundry pada bulan Mei 2022 yang kemudian dilakukan proses selection data, preprocessing data, K-Means clustering dan evaluasi, didapatkan bahwa jumlah cluster terbaik ada sebanyak 2 cluster. Jumlah anggota dari cluster 1 adalah sebanyak 237 pelanggan atau 99% dan cluster ke 2 adalah sebanyak 3 pelanggan atau 1%. Pelanggan pada cluster ke 1 tergolong pelanggan regular, dan pelanggan pada cluster ke 2 tergolong pelanggan premium. Hasil dari penelitian ini dapat digunakan oleh pihak B-Clean Laundry untuk acuan dalam strategi marketing, pemberian reward serta promo menarik kepada para pelanggannya sehingga dapat meningkatkan pendapatan usaha.

Daftar Pustaka

- [1] B. Mulyadi, Jaroji, and A. T, "Aplikasi Sistem Pemesanan Jasa Laundry (E-Laundry) Berbasis Android," *Zo. J. Sist. Inf.*, vol. 1, no. 1, pp. 48–57, 2019, doi: 10.31849/zn.v1i1.2386.
- [2] B. Basri, W. Gata, and R. Risnandar, "Analisis Loyalitas Pelanggan Berbasis Model Recency, Frequency, dan Monetary (RFM) dan Decision Tree pada PT. Solo," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 5, p. 943, 2020, doi: 10.25126/jtiik.2020752284.
- [3] K. Auliasari and M. Kertaningtyas, "Penerapan Algoritma K-Means untuk Segmentasi Konsumen Menggunakan R," *J. Teknol. dan Manaj. Inform.*, vol. 5, no. 2, 2019, doi: 10.26905/jtmi.v5i2.3644.
- [4] M. F. Tri and Y. Nataliani, "Analisis Pengaruh Penilaian Asesor terhadap Kinerja Guru Mata Pelajaran dengan k-Means Clustering," *Indones. J. Comput. Model.*, 2021, [Online]. Available: <https://ejournal.uksw.edu/icm/article/view/5063%0Ahttps://ejournal.uksw.edu/icm/article/download/5063/1817>
- [5] A. T. Widiyanto and A. Witanti, "Segmentasi Pelanggan Berdasarkan Analisis RFM Menggunakan Algoritma K-Means Sebagai Dasar Strategi Pemasaran (Studi Kasus PT Coversuper Indonesia Global)," *KONSTELASI Konvergensi Teknol. dan Sist. Inf.*, vol. 1, no. 1, pp. 204–215, 2021, doi: 10.24002/konstelasi.v1i1.4293.
- [6] J. Jamal and D. Yanto, "Analisis RFM dan Algoritma K-Means untuk Clustering Loyalitas Customer," *Energy*, vol. 9, no. 1, pp. 1–8, 2019.
- [7] S. Monalisa and I. Erza, "Analisis Loyalitas Agen Biasa dan Agenstok Menggunakan Model RFM (Recency, Frequency, Monetary) dan Algoritma K-Medoids pada BC 4 HPAI Pekanbaru," *Techno.Com*, vol. 20, no. 1, pp. 109–121, 2021, doi: 10.33633/tc.v20i1.4219.
- [8] D. Methods, R. Wati, B. Sembiring, F. A. Mohammed, and K. Chairuang, "Customer Segmentation Based on RFM Model Using," *LONTAR Komput.*, vol. 11, no. 1, pp. 32–43, 2020.
- [9] Z. G. Prastyawan, M. Ridho Bagaskara, and D. Fitriati, "Segmentasi Pelanggan Restoran Menggunakan Metode Clustering Simple K-Means (Studi Kasus Xyz)," *Inform. Dan Keamanan Siber Jakarta-Indonesia*, pp. 198–203, 2018.
- [10] R. T. S. Muhammad Hariyanto, "Clustering pada Data Mining untuk Mengetahui Potensi Penyebaran Penyakit DBD Menggunakan Metode Algoritma K-Means dan Metode Perhitungan Jarak Euclidean Distance," *Sist. Komput. dan Tek. Inform.*, vol. 1, no. 1, pp. 117–122, 2018.