

Penerapan Metode Clustering Dalam Segmentasi Pelanggan Perusahaan Logistik

Nisa Hanum¹, Cahyo Prianto², Woro Isti Rahayu³, Hanan Destiarin Kishendrian⁴

¹²⁴Teknik Informatika, Fakultas Sekolah Vokasi, Universitas Logistik & Bisnis Internasional

³Sains Data, Universitas Logistik & Bisnis Internasional

Jl. Sari Asih No.54, Sarijadi, Kec. Sukasari, Kota Bandung, Jawa Barat, Indonesia

e-mail: nisa@ulbi.ac.id¹, chprianto@gmail.com², wistirahayu@yahoo.com³,
hananksndrian@gmail.com⁴

Received : October, 2023

Accepted : November, 2023

Published : December, 2023

Abstract

Marketing is important in business to compete and maintain market share. The development of technology brings major changes in the industry. In addition to product development as well as required services, and customer segmentation becomes a factor to consider in marketing strategies. Clustering, such as the K-Means method, is used in customer segmentation to divide data into groups based on similarities. This technique helps in useful pattern recognition and customer segmentation. By applying Clustering techniques in Data mining, companies can understand customer behavior, recognize similar customer groups, and plan marketing strategies accordingly. The results showed that the best cluster was generated with a k value of 4, and the data was normalized using the Min-Max Normalization method. Grouping customers in the form of clusters can enable the identification of consumer profiles to guide companies in decision making.

Keywords: clustering, customer segmentation, RFM, data mining, K-Means, CRISP-DM

Abstrak

Pemasaran penting dalam bisnis untuk bersaing dan mempertahankan pangsa pasar. Perkembangan teknologi membawa perubahan besar dalam industri ini. Selain pengembangan produk serta jasa yang dibutuhkan, dan segmentasi pelanggan menjadi faktor yang perlu dipertimbangkan dalam strategi pemasaran. Clustering, seperti metode K-Means, digunakan dalam segmentasi pelanggan untuk membagi data menjadi kelompok berdasarkan kesamaan. Teknik ini membantu dalam pengenalan pola yang bermanfaat dan segmentasi pelanggan. Dengan menerapkan teknik Clustering dalam Data mining, perusahaan dapat memahami perilaku pelanggan, mengenali kelompok pelanggan yang serupa, dan merencanakan strategi pemasaran yang sesuai. Hasil penelitian menunjukkan bahwa cluster terbaik dihasilkan dengan nilai k 4, serta data dinormalisasi menggunakan metode Min-Max Normalization. Pengelompokan pelanggan dalam bentuk cluster dapat memungkinkan identifikasi profil konsumen guna memberikan panduan pada perusahaan dalam pengambilan keputusan.

Kata Kunci: clustering, segmentasi pelanggan, RFM, data mining, K-Means, CRISP-DM

1. PENDAHULUAN

Di setiap bisnis, baik di sektor jasa ataupun manufaktur, pemasaran memainkan peran penting dalam kemajuan perusahaan. Pemasaran memiliki pengaruh besar terhadap kegiatan perusahaan dalam berkompetisi dengan pesaing lainnya untuk menguasai dan mempertahankan pangsa pasar yang berpotensi. Selain pengembangan produk serta jasa yang dibutuhkan, dan segmentasi pelanggan [1] juga menjadi faktor yang perlu dipertimbangkan dalam strategi pemasaran. Perusahaan ekspedisi adalah entitas bisnis yang menyediakan jasa pengiriman barang dan logistik.[2]. Perkembangan teknologi dan globalisasi telah membawa perubahan besar dalam industri ekspedisi. Banyak perusahaan ekspedisi telah mengadopsi solusi digital dan teknologi canggih untuk meningkatkan efisiensi operasional dan memberikan visibilitas [3] yang lebih baik kepada pelanggan. Sistem pelacakan online, integrasi dengan platform e-commerce, dan aplikasi mobile adalah contoh dari inovasi yang digunakan untuk meningkatkan pengalaman pelanggan.[4]. Pelanggan harus mempertimbangkan beberapa faktor, termasuk reliabilitas, kecepatan pengiriman, jangkauan geografis, layanan tambahan yang disediakan, harga, dan dukungan pelanggan.

Saat ini, perusahaan ekspedisi di Indonesia tengah menghadirkan banyak inovasi dalam bisnis mereka dalam tingkat persaingan yang semakin tinggi, dikarenakan pelanggan memiliki hak untuk memilih dari banyaknya penyedia layanan tersedia. Persaingan tersebut menjadi penting untuk keberlangsungan perusahaan karena memiliki dampak pada pendapatan perusahaan. Akibat persaingan di pasar, terjadi perpindahan pelanggan ke perusahaan telekomunikasi lain, yang biasa disebut sebagai Churn. Churn merupakan proses di mana pelanggan memutuskan jasa yang diberikan oleh suatu perusahaan dan beralih ke layanan dari pesaing [5]. Dalam persaingan pasar ini, pengalaman menunjukkan bahwa tingkat churn mencapai sekitar 30-35% setiap tahunnya [6]. Hal ini menyebabkan perusahaan harus melakukan upaya dan pengeluaran sebanyak 5-10 kali lebih besar untuk mendapatkan pelanggan baru dibandingkan [7] dengan mempertahankan pelanggan yang sudah ada. Oleh karena itu, dalam industri ekspedisi, penting untuk fokus

pada upaya mempertahankan pelanggan yang ada. Saat ini, PT. Ekspedisi Indonesia sedang melaksanakan strategi pemasaran menggunakan publisitas produk di media sosial, periklanan serta personal selling. Meskipun demikian, efektivitas dari kegiatan pemasaran tersebut masih dirasa kurang, terutama karena terdapat beberapa barang yang kurang diminati oleh pelanggan. Fokus terpenting perusahaan adalah menghadapi persaingan dengan kompetitor yaitu pelanggan. Namun, saat ini belum tersedia informasi yang cukup mengenai karakteristik pelanggan PT. Ekspedisi Indonesia. Oleh karena itu, sebagai langkah awal, diperlukan analisis karakteristik pelanggan PT. Ekspedisi Indonesia sebagai dasar untuk menetapkan segmentasi dan menentukan target pasar yang akan dilayani.

Segmentasi pelanggan melibatkan pengelompokan pelanggan ke dalam beberapa kelompok atau segmen yang memiliki karakteristik serupa[8], [9]. Konsep ini juga dapat diartikan sebagai proses Clustering. Analisis Clustering bertujuan untuk [10], [11],[12] membentuk kelompok-kelompok yang baik ketika setiap anggota dalam kelompok memiliki tingkat keasaman yang tinggi, menciptakan homogenitas internal. Tujuan dari segmentasi pelanggan yaitu untuk mengimplementasikan strategi pemasaran yang sesuai serta memahami perilaku pelanggan, dengan tujuan mendapatkan keuntungan bagi perusahaan. [13], [14]Proses segmentasi pelanggan ini dilakukan menggunakan teknik Data mining untuk mengelompokkan pelanggan ke dalam segmen-segmen yang relevan.

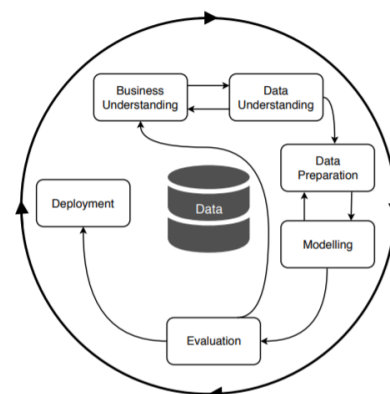
Segmentasi pelanggan merupakan sebuah tahap membagi pelanggan menjadi beberapa kategori pelanggan yang memiliki kemiripan antara satu sama lain yang sangat relevan, seperti minat serta kebiasaan belanja, jenis kelamin ataupun metode transaksi yang digunakan. [15]. RFM merupakan kerangka kerja yang sederhana guna mengukur tingkah laku pelanggan. RFM merupakan Recency, Frequency dan Monetary.[16], [17]. Recency merupakan kapan terakhir kali pelanggan melakukan pembelian? Nilai recency merupakan jumlah hari yang dibutuhkan pelanggan antara dua pembelian. Nilai sebuah recency yang kecil mengartikan bahwa pelanggan melakukan pembelian secara

berulang dalam kurun waktu yang cukup singkat. Sedangkan nilai sebuah recency yang besar mengartikan bahwa pelanggan kurang melakukan aktivitas pembelian dalam kurun waktu yang singkat. [15]. Frequency merupakan berapa kali pelanggan melakukan pembelian? Frequency dapat adalah jumlah pembelian yang telah dilakukan oleh pelanggan dalam suatu periode tertentu. Semakin besar nilai frequency maka semakin menunjukkan bahwa pelanggan setia terhadap Perusahaan. [18]. Monetary merupakan berapa banyak uang yang dihabiskan oleh pelanggan? Monetary merupakan jumlah uang yang dihabiskan oleh pelanggan di kurun waktu tertentu. Semakin besar jumlah uang yang dihabiskan oleh pelanggan, maka semakin banyak pendapatan yang didapatkan oleh Perusahaan. [18].

Data mining merupakan tahap penemuan pola menarik dari kumpulan data besar menggunakan algoritma untuk mengungkapkan informasi yang tersembunyi. [19] Dalam konteks segmentasi pelanggan, Clustering memiliki peran penting dalam Data mining [20]. K-Means merupakan satu diantara metode non-hirarki dalam Clustering data bertujuan untuk membagi data ke dalam satu atau lebih Cluster. [21], Metode ini [22] melakukan pemisahan data ke dalam kelompok-kelompok sehingga data dengan karakteristik yang serupa dikelompokkan dalam satu Cluster yang sama. Melalui pemanfaatan teknik Clustering dalam Data mining, perusahaan dapat mendapatkan pemahaman yang berharga tentang perilaku pelanggan, mengenali kelompok pelanggan yang memiliki kesamaan, dan merencanakan strategi pemasaran yang lebih terarah dan sesuai dengan kebutuhan mereka. [23], [24] Data akan diolah untuk menentukan segmentasi pelanggan yang sesuai dengan melakukan Clustering data pelanggan menggunakan metode K-means. [25], [26]. Hasil dari pemodelan data yang telah dilakukan akan mengelompokkan pelanggan ke dalam beberapa Cluster serta mengidentifikasi customer profiling (profil konsumen). Pengelompokan ini menghasilkan karakteristik yang berbeda dari setiap pelanggan dalam setiap Cluster, yang bisa digunakan sebagai panduan guna pengambilan keputusan perusahaan.

2. METODE PENELITIAN

Metodologi penelitian yang digunakan yaitu CRISP-DM (Cross Standart Industries for Data Mining). Metodologi CRISP-DM merupakan standarisasi data mining yang dikembangkan oleh tiga pionir pasar data mining yaitu Daimler Chrysler (Daimler Benz), SPSS, NCR. Kemudian dikembangkan di berbagai workshop (1997-1999). Lebih dari 300 organisasi berpartisipasi dalam proses pemodelan ini dan CRISP-DM akhirnya diterbitkan pada 1999 [27]. Terdapat beberapa langkah dalam CRISP-DM (Cross Industries Process for Data Mining) yaitu *business understanding, data understanding, data preparation, modeling, evaluation, deployment*. Namun langkah yang dilakukan pada penelitian ini hanya sampai tahapan evaluation karena tujuan dari penelitian ini adalah untuk mengetahui beberapa segmen data pelanggan yang dibagi serta untuk memperoleh informasi mengenai profil pelanggan berdasarkan segmentasi [28].



Gambar 1. Diagram Alur CRISP-DM [Sumber: Martínez-Plumed, 2019]

2.1 Business Understanding

Tujuan bisnis yang dilakukan pada penelitian ini yaitu bagaimana cara melakukan customer profiling serta segmentasi data pelanggan menggunakan metode RFM dan menggunakan algoritma K-Means. [29] Pembagian segmentasi dilakukan berdasarkan transaksi pada penggunaan pospay dengan atribut yang digunakan yaitu recency, frequency, dan monetary.

2.2 Data Understanding

Pada tahap ini dikenal juga sebagai tahap pemahaman terhadap data yang diperoleh kemudian data awal yang dikumpulkan melalui observasi langsung, dengan menganalisa berkas

data pelanggan pada objek penelitian. Kajian dilakukan sedemikian rupa sehingga dapat mengetahui data yang akan digunakan. Penelitian ini menggunakan data yang berjumlah 684 record dimana data tersebut diambil dari periode bulan januari sampai dengan bulan desember, berdasarkan data yang ditunjukkan oleh objek penelitian, serta data mentah diolah kembali dalam kategori RFM dari setiap record. Semakin banyak data yang di peroleh, maka hasil akurasi datanya juga semakin maksimal.

2.3 Data Preparation

Fase ini merupakan fase pengolahan data atau bisa disebut juga dengan fase pengolahan data. Atribut yang digunakan dibagi menjadi 3 kelompok atau 3 bagian yaitu recency, frequency, dan monetary. Untuk memperoleh 3 kelompok tersebut data mentah atau data asli yang dimiliki berupa data transaksi dari 57 perusahaan dengan periode transaksi dari bulan januari sampai desember, data tersebut dibagi menjadi 3 buah bagian berdasarkan data asli sehingga dihasilkan recency, frequency, dan monetary dari setiap record data. Selanjutnya data baru yang telah dihasilkan menghasilkan kesenjangan yang cukup tinggi antara setiap atribut nya, sehingga dilakukan normalisasi data menggunakan 3 metode normalisasi yaitu Min-Max Normalization [30], Z-Score Normalization, serta Decimal Scaling. Setelah dilakukan normalisasi menghasilkan data yang memiliki skala atribut range antara -1 sampai 1 atau 0 sampai 1. Data yang dihasilkan tidak lagi memiliki kesenjangan yang tinggi. [31]

Untuk menormalisasi data menggunakan metode min-max normalization menggunakan code berikut pada google collab :

```
df['Recency'] = (df['Recency'] -
df['Recency'].min()) / (df['Recency'].max() -
df['Recency'].min())
```

```
df['Frequency'] = (df['Frequency'] -
df['Frequency'].min()) /
(df['Frequency'].max() -
df['Frequency'].min())
```

```
df['Monetary'] = (df['Monetary'] -
df['Monetary'].min()) / (df['Monetary'].max()
- df['Monetary'].min())
```

Gambar 2. Code Min-Max Normalization

Untuk menormalisasi data menggunakan metode Z-Score menggunakan code berikut pada google collab :

```
for col in ['Recency', 'Frequency',
'Monetary'] :
df[col] = (df[col] - df[col].mean()) /
df[col].std()
```

Gambar 3. Code Z-Score

Untuk menormalisasi data menggunakan metode Decimal Scaling menggunakan code berikut pada google collab :

```
d = 2
df['Recency'] = df['Recency'] / (10 ** d)
```

```
d = 2
df['Frequency'] = df['Frequency'] / (10 ** d)
```

```
d = 12
df['Monetary'] = df['Monetary'] / (10 ** d)
```

Gambar 4. Code Decimal Scaling

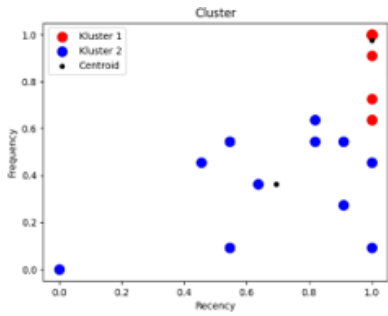
2.4 Modeling

Tahap pemodelan dilakukan dengan menggunakan jupyter notebook, dan menyertakan metode k-means. Pada tahap persiapan data, clustering menggunakan data dari atribut yang dipilih yaitu recency, frequency, dan monetary sebagai parameter. Untuk menentukan nilai k yang akan digunakan pada metode k-means menggunakan metode elbow [32], [33],[34]. berdasarkan atribut yang telah dipilih, sehingga menghasilkan nilai k [35] yang akan digunakan. Nilai k yang dihasilkan tersaji dalam tabel 1

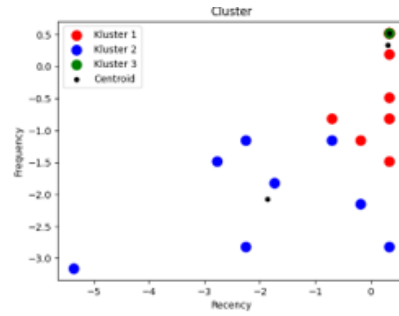
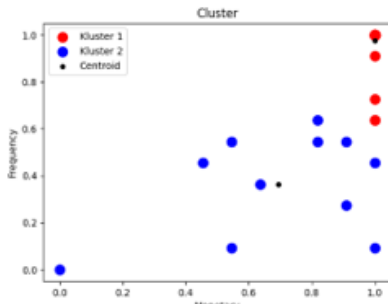
Tabel 1: Tabel Hasil Nilai K

Metode Normalisasi	Min-Max Normalization	Z-Score	Decimal Scaling
Nilai k	2 dan 4	3 dan 4	3 dan 4

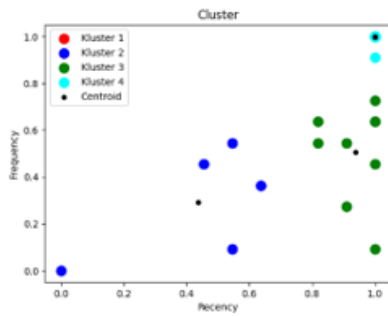
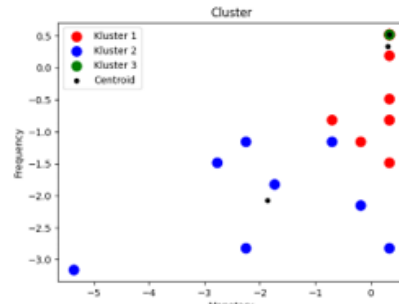
Dari nilai k yang sudah diperoleh menghasilkan sebuah cluster, untuk cluster yang dihasilkan dari Min-Max Normalization sebagai berikut :



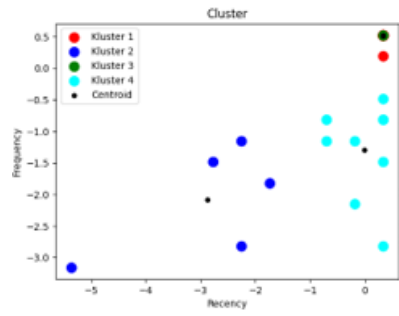
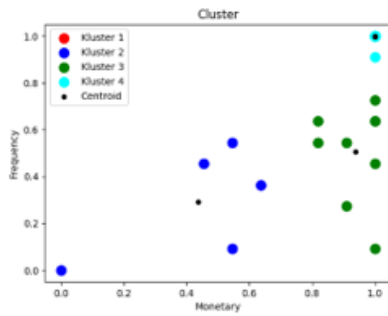
Gambar 5. Hasil Cluster K=2 metode Min-Max Normalization



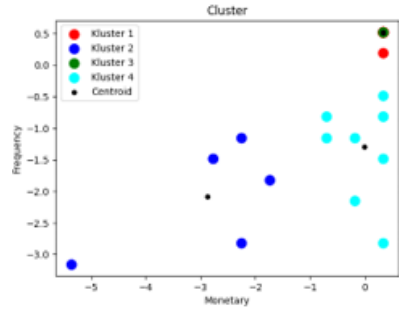
Gambar 7. Hasil Cluster K=3 metode Z-Score



Gambar 6. Hasil Cluster K=4 metode Min-Max Normalization

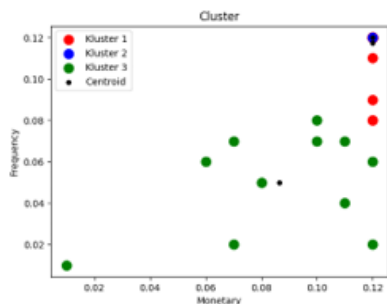
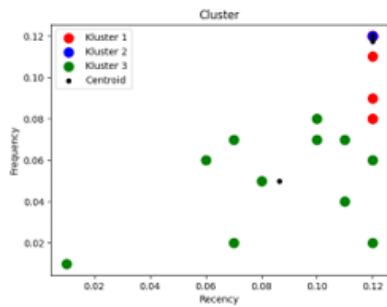


Gambar 8. Hasil Cluster K=4 metode Z-Score

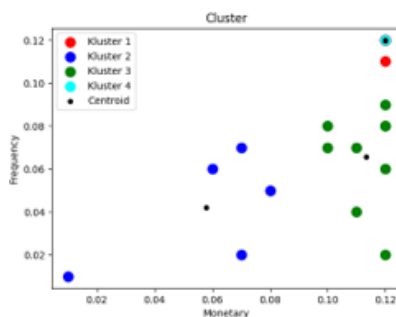
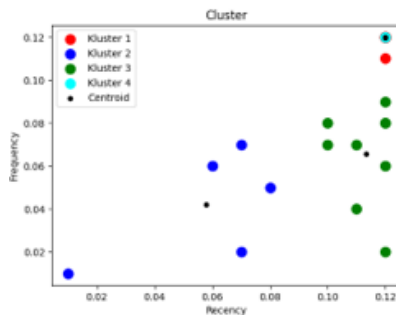


Sedangkan untuk cluster yang dihasilkan dari Z-Score terdapat pada gambar 7 dan 8.

Sedangkan untuk cluster yang dihasilkan dari Decimal Scaling terdapat pada gambar 9 dan 10.



Gambar 9. Hasil Cluster K=3 metode Decimal Scaling



Gambar 10. Hasil Cluster K=4 metode Decimal Scaling

2.5 Evaluation

Langkah ini merupakan tahapan analisis yang merupakan hasil dari Langkah pengolahan sebelumnya yaitu menginterpretasikan data yang kemudian diperoleh segmentasi pelanggan serta profiling pelanggan. Evaluation pada penelitian ini menggunakan metode Cluster yaitu Davies-Bouldin Index dan Calinski Harabasz Index. Pada metode Davies-Bouldin Index jika nilai yang dihasilkan rendah

menunjukkan klasterisasi yang lebih unggul, begitupun sebaliknya jika nilai yang dihasilkan lebih besar maka menunjukkan klasterisasi yang kurang baik.[36]. Sedangkan pada Calinski Harabasz Index jika nilai yang dihasilkan nilai yang lebih tinggi menunjukkan klasterisasi lebih unggul, sebaliknya jika nilai yang dihasilkan lebih kecil menunjukkan klasterisasi yang kurang baik. [37], [36], [38].

Pada data yang dinormalisasi menggunakan metode Min-Max Normalization dengan Cluster 2 dilakukan evaluation menggunakan metode Davies-Bouldin Index menghasilkan nilai score rata-rata 13,16. sedangkan jika menggunakan metode Calinski Harabasz Index menghasilkan nilai score rata-rata 83. Pada Cluster 4 jika dilakukan evaluation menggunakan metode Davies-Bouldin Index menghasilkan nilai score rata-rata 13,16. Sedangkan jika menggunakan metode Calinski Harabasz Index menghasilkan nilai score rata-rata 100. Jika dibandingkan pada saat menggunakan metode Davies-Bouldin Index antara Cluster 2 dan 4 menghasilkan nilai score rata-rata yang sama. Sedangkan jika dibandingkan pada saat menggunakan metode Calinski Harabasz Index antara Cluster 2 dan 4 menghasilkan nilai score rata-rata 83 dan 100 dapat diartikan Cluster 4 lebih unggul dari pada Cluster 2.

Pada data yang dinormalisasi menggunakan metode Z-Score Normalization dengan Cluster 3 dilakukan evaluation menggunakan metode Davies-Bouldin Index menghasilkan nilai score rata-rata 17,28. sedangkan jika menggunakan metode Calinski Harabasz Index menghasilkan nilai score rata-rata 65,25. Pada Cluster 4 jika dilakukan evaluation menggunakan metode Davies-Bouldin Index menghasilkan nilai score rata-rata 13,36. Sedangkan jika menggunakan metode Calinski Harabasz Index menghasilkan nilai score rata-rata 94,64. Jika dibandingkan pada saat menggunakan metode Davies-Bouldin Index antara Cluster 3 dan 4 menghasilkan nilai score rata-rata 17,28 dan 13,36 dapat diartikan Cluster 3 lebih unggul dari pada Cluster 4. Sedangkan jika dibandingkan pada saat menggunakan metode Calinski Harabasz Index antara Cluster 3 dan 4 menghasilkan nilai score rata-rata 65,25 dan 94,64 dapat diartikan Cluster 4 lebih unggul dari pada Cluster 3.

Pada data yang dinormalisasi menggunakan metode Decimal Scaling dengan Cluster 3 dilakukan evaluation menggunakan metode Davies-Bouldin Index menghasilkan nilai score rata-rata 17,47. sedangkan jika menggunakan metode Calinski Harabasz Index menghasilkan nilai score rata-rata 83,28. Pada Cluster 4 jika dilakukan evaluation menggunakan metode Davies-Bouldin Index menghasilkan nilai score rata-rata 13,25. Sedangkan jika menggunakan metode Calinski Harabasz Index menghasilkan nilai score rata-rata 98,64. Jika dibandingkan pada saat menggunakan metode Davies-Bouldin Index antara Cluster 3 dan 4 menghasilkan nilai score rata-rata 17,47 dan 13,25 dapat diartikan Cluster 4 lebih unggul dari pada Cluster 3. Sedangkan jika dibandingkan pada saat menggunakan metode Calinski Harabasz Index antara Cluster 3 dan 4 menghasilkan nilai score rata-rata 83,28 dan 98,64 dapat diartikan Cluster 4 lebih unggul dari pada Cluster 3.

3. HASIL DAN PEMBAHASAN

3.1 Deskripsi Data

Data yang dijadikan bahan untuk melakukan analisis merupakan data pelanggan perusahaan ekspedisi di wilayah kota Bandung. Data yang digunakan pada penelitian ini berupa data transaksi dari 57 perusahaan dan berjumlah 684 record sebagai populasi dimana data tersebut diambil dari periode bulan Januari sampai bulan Desember.

3.2 Pembahasan

Pada tabel 2 merupakan hasil dari penelitian, dimana jika menggunakan metode normalisasi Min-Max Normalization menghasilkan nilai k 2 dan 4, serta dilakukan evaluation pada nilai k tersebut jika evaluation menggunakan metode Davies-Bouldin Index untuk nilai k 2 menghasilkan score 13,16, dan untuk nilai k 4 menghasilkan score 13,16. Sedangkan jika evaluation menggunakan metode Calinski Harabasz Index untuk nilai k 2 menghasilkan score 83 dan untuk nilai k 4 menghasilkan score 100.

Jika metode normalisasi yang digunakan yaitu Z-Score menghasilkan nilai k 3 dan 4, kemudian dilakukan evaluation menggunakan metode Davies-Bouldin Index untuk nilai k 3 menghasilkan score 17,28 dan untuk nilai k 4 menghasilkan score 13,36. Sedangkan jika

evaluation menggunakan metode Calinski Harabasz Index untuk nilai k 3 menghasilkan score 65,25 dan untuk nilai k 4 menghasilkan score 94,64.

Sedangkan jika normalisasi data dilakukan menggunakan metode Decimal Scaling menghasilkan nilai k 3 dan 4, kemudian dilakukan evaluation menggunakan metode Davies-Bouldin Index untuk nilai k 3 menghasilkan score 17,47 dan untuk nilai k 4 menghasilkan score 13,25. Sedangkan jika evaluation menggunakan metode Calinski Harabasz Index untuk nilai k 3 menghasilkan score 83,28 dan untuk nilai k 4 menghasilkan score 98,64.

Tabel 2: Tabel Hasil Penelitian

	Metode Normalisasi					
	Min-Max Normalization		Z-Score		Decimal Scaling	
Nilai k	2	4	3	4	3	4
	Evaluation					
Davies-Bouldin Index	13,16	13,16	17,28	13,36	17,47	13,25
Calinski Harabasz Index	83	100	65,25	94,64	83,28	98,64

4. KESIMPULAN

Dari penelitian yang telah dilakukan pada Penerapan metode Clustering dalam Segmentasi Pelanggan Perusahaan Logistik, penulis telah mengumpulkan data-data yang dibutuhkan sehingga menjadi sebuah dataset yang kemudian diolah dan dilakukan normalisasi, metode normalisasi dilakukan dengan 3 metode untuk dijadikan perbandingan yaitu Min-Max Normalization, Z-Score, dan Decimal Scaling Normalization, setelah dilakukan normalisasi tersebut dihasilkan nilai k yang dilakukan menggunakan metode elbow sebagai berikut :

1. Metode Min-Max Normalization menghasilkan 2 dan 4 cluster.
2. Metode Z-Score menghasilkan 3 dan 4 cluster.
3. Metode Decimal Scaling menghasilkan 3 dan 4 cluster.

Pada penelitian ini juga menggunakan 2 metode evaluasi yaitu :

1. Pada metode Davies-Bouldin Index jika nilai yang dihasilkan rendah menunjukkan klasterisasi yang lebih unggul, yang paling unggul adalah cluster 2 dan 4 dengan metode Min-Max Normalization dengan presentase 13,16%.

2. Sedangkan pada Calinski Harabasz Index jika nilai yang dihasilkan nilai yang lebih tinggi menunjukkan klasterisasi lebih unggul. menggunakan metode Calinski-Harabasz Index hasil yang paling unggul adalah cluster 4 dengan metode Min-Max Normalization dengan presentase 100%.

Untuk penelitian selanjutnya alangkah baiknya menggunakan dataset yang lebih mendukung agar menghasilkan Cluster yang lebih baik lagi.

DAFTAR PUSTAKA

- [1] A. Johansson, Wikström, and Jonas, "Customer segmentation using machine learning," 2021.
- [2] C. Harini, "Entrepreneurial Marketing for UMKM : a Literatur Review," *J. Mantik*, vol. 4, no. 2, pp. 10–19, 2020, [Online]. Available: <http://iocscience.org/ejournal/index.php/mantik/article/view/882/595>
- [3] K. A. Barnhill *et al.*, "Ship-to-shore training for active deep-sea capacity development," *ICES J. Mar. Sci.*, vol. 80, no. 6, pp. 1619–1628, 2023, doi: 10.1093/icesjms/fsad088.
- [4] S. E. J. Vähä-Karvia, "How to adapt the requirements of an expedition to meet customer competences," 2021.
- [5] H. Ribeiro, B. Barbosa, A. C. Moreira, and R. G. Rodrigues, *Determinants of churn in telecommunication services: a systematic literature review*. 2023. doi: 10.1007/s11301-023-00335-7.
- [6] Y. Li, B. Hou, Y. Wu, D. Zhao, A. Xie, and P. Zou, "Giant fight: Customer churn prediction in traditional broadcast industry," *J. Bus. Res.*, vol. 131, no. January, pp. 630–639, 2021, doi: 10.1016/j.jbusres.2021.01.022.
- [7] S. R. Baker, B. Baugh, and M. C. Sammon, "NBER WORKING PAPER SERIES MEASURING CUSTOMER CHURN AND INTERCONNECTEDNESS," no. <http://www.nber.org/papers/w7707>, 2020.
- [8] J. Salminen, M. Mustak, M. Sufyan, and B. J. Jansen, "How can algorithms help in segmenting users and customers? A systematic review and research agenda for algorithmic customer segmentation," *J. Mark. Anal.*, no. 0123456789, 2023, doi: 10.1057/s41270-023-00235-5.
- [9] A. Agrawal, P. Kaur, and M. Singh, "Customer Segmentation Model using K-means Clustering on E-commerce," *Int. Conf. Sustain. Comput. Data Commun. Syst. (ICSCDS), Erode, India*, pp. 1–6, 2023.
- [10] H. Bu, "Deep Clustering Based on Contractive Autoencoder and Self-paced Learning," *Int. Conf. Comput. Vision, Image Deep Learn.*, pp. 458–462, 2023.
- [11] I. Masich, N. Rezova, G. Shkaberina, S. Mironov, M. Bartosh, and L. Kazakovtsev, "Subgroup Discovery in Machine Learning Problems with Formal Concepts Analysis and Test Theory Algorithms," *Algorithms*, vol. 16, no. 5, 2023, doi: 10.3390/a16050246.
- [12] P. Grabusts, "Evaluation of Clustering Results in the Aspect of Information Theory," *2020 61st Int. Sci. Conf. Inf. Technol. Manag. Sci. Riga Tech. Univ. ITMS 2020 - Proc.*, 2020, doi: 10.1109/ITMS51158.2020.9259227.
- [13] S. I. R. Rodríguez and F. de A. T. de Carvalho, "Soft subspace clustering of interval-valued data with regularizations," *Knowledge-Based Syst.*, vol. 227, p. 107191, 2021, doi: 10.1016/j.knosys.2021.107191.
- [14] V. Shinde, N. Ransing, S. Ransing, S. Chitranshi, and P. A. S. Shinde, "An Implementation on User Centered Website Using Customer Segmentation," *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 11, no. 5, pp. 5212–5217, 2023, doi: 10.22214/ijraset.2023.52864.
- [15] F. Asmat, K. Suryadi, and R. Govindaraju, "Data mining framework for the identification of profitable customer based on recency, frequency, monetary (RFM)," *Nucleation Atmos. Aerosols*, 2023.
- [16] S. H. Shihab, S. Afroge, and S. Z. Mishu, "RFM Based Market Segmentation Approach Using Advanced K-means and Agglomerative Clustering : A Comparative Study," *2019 Int. Conf. Electr. Comput. Commun. Eng.*, pp. 1–4, 2019.

- [17] G. WIBOWO, "Segmentasi Pelanggan Berdasarkan Recency FREQUENCY MONETARY (RFM) dan USER EVENT TRACKING MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING (Budi Luhur Tesis)," 2022.
- [18] A. J. Christy, A. Umamakeswari, L. Priyatharsini, and A. Neyaa, "RFM ranking – An effective approach to customer segmentation," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 33, no. 10, pp. 1251–1257, 2021, doi: 10.1016/j.jksuci.2018.09.004.
- [19] D. D and J. P. Prathap, "A Brief Study of Privacy-Preserving Practices (PPP) in Data Mining," *TEST Engineering Manag.*, vol. 82, no. 7611, pp. 7611–7622, 2020.
- [20] F. Yoseph *et al.*, "The impact of big data market segmentation using data mining and clustering techniques," *J. Intell. Fuzzy Syst.*, vol. 38(5), no. <https://doi.org/10.3233/JIFS-179698>, pp. 6159–6173, 2020.
- [21] L. Rajput and S. N. Singh, "Customer Segmentation of E-commerce data using K-means Clustering Algorithm," *Int. Conf. Cloud Comput. Data Sci. Eng.*, pp. 658–664, 2023.
- [22] S. Mohamed, K. Sethom, A. Namoun, A. Tufail, K. H. Kim, and H. Almoamari, "Customer Profiling Using Internet of Things Based Recommendations," *Sustain.*, vol. 14, no. 18, pp. 1–21, 2022, doi: 10.3390/su141811200.
- [23] M. S. Kasem, M. Hamada, and I. Taj-Eddin, "Customer Profiling, Segmentation, and Sales Prediction using AI in Direct Marketing," 2023, [Online]. Available: <http://arxiv.org/abs/2302.01786>
- [24] Luh Putu Wiwien Widhyastuti, I. N. Sukajaya, and Kadek Yota Ernanda Aryanto, "The Customer Profiling berdasarkan Model RFM dengan Metode K-Means pada Institusi Pendidikan untuk menunjang Strategi Bisnis di Masa Pandemi Covid-19," *JTIM J. Teknol. Inf. dan Multimed.*, vol. 4, no. 2, pp. 94–108, 2022, doi: 10.35746/jtim.v4i2.232.
- [25] G. ASLANTAŞ, M. GENÇGÜL, M. RUMELLİ, M. ÖZSARAÇ, and G. BAKIRLI, "Customer Segmentation Using K-Means Clustering Algorithm and RFM Model," *Deu Muhendis. FDokuz Eylul Univ. Fac. Eng. J. Sci. Eng.*, vol. 25, no. 74, pp. 491–503, 2023, doi: 10.21205/deufmd.2023257418.
- [26] F. Nie, Z. Li, R. Wang, and X. Li, "An Effective and Efficient Algorithm for K-Means Clustering With New Formulation," *IEEE Trans. Knowl. Data Eng.*, vol. 35, pp. 3433–3443, 2023.
- [27] C. Wulandari, Y. Ansori, and K. F. H. H, "13760-50057-1-Pb," *MATICS J. Ilmu Komput. dan Teknol. Inf.*, vol. 14, no. 2, pp. 35–40, 2022.
- [28] A. F. Khairati, A. . Adlina, G. . Hertono, and B. . Handari, "Kajian Indeks Validitas pada Algoritma K-Means Enhanced dan K-Means MMCA," *Prism. Pros. Semin. Nas. Mat.*, vol. 2, pp. 161–170, 2019, [Online]. Available: <https://journal.unnes.ac.id/sju/index.php/prisma/article/view/28906>
- [29] S. Monalisa, Y. Juniarti, E. Saputra, F. Muttakin, and T. K. Ahsyar, "Customer segmentation with RFM models and demographic variable using DBSCAN algorithm," *Telkomnika (Telecommunication Comput. Electron. Control.*, vol. 21, no. 4, pp. 742–749, 2023, doi: 10.12928/TELKOMNIKA.v21i4.22759.
- [30] H. Henderi, "Comparison of Min-Max normalization and Z-Score Normalization in the K-nearest neighbor (kNN) Algorithm to Test the Accuracy of Types of Breast Cancer," *IJIS Int. J. Informatics Inf. Syst.*, vol. 4, no. 1, pp. 13–20, 2021, doi: 10.47738/ijis.v4i1.73.
- [31] D. Singh and B. Singh, "Investigating the impact of data normalization on classification performance," *Appl. Soft Comput.*, vol. 97, no. xxx, p. 105524, 2020, doi: 10.1016/j.asoc.2019.105524.
- [32] F. Liu and Y. Deng, "Determine the Number of Unknown Targets in Open World Based on Elbow Method," *IEEE Trans. Fuzzy Syst.*, vol. 29, no. 5, pp. 986–995, 2021, doi: 10.1109/TFUZZ.2020.2966182.
- [33] R. Ünlü and P. Xanthopoulos, "Estimating the number of clusters in a dataset via consensus clustering," *Expert Syst. Appl.*, vol. 125, pp. 33–39, 2019, doi: 10.1016/j.eswa.2019.01.074.
- [34] M. Al Ghifari and W. T. Harsanti Putri, "Clustering Courses Based On Student

- Grades Using K-Means Algorithm With Elbow Method For Centroid Determination," *Inf. J. Ilm. Bid. Teknol. Inf. dan Komun.*, vol. 8, no. 1, pp. 42–46, 2023, doi: 10.25139/inform.v8i1.4519.
- [35] J. Chen, L. Zhao, M. Zhou, Y. Liu, and S. Qin, "An Approach to Determine the Optimal k-Value of K-means Clustering in Adaptive Random Testing," *Proc. - 2020 IEEE 20th Int. Conf. Softw. Qual. Reliab. Secur. QRS 2020*, pp. 160–167, 2020, doi: 10.1109/QRS51102.2020.00032.
- [36] A. Idrus, N. Tarihoran, U. Supriatna, A. Tohir, S. Suwarni, and R. Rahim, "Distance Analysis Measuring for Clustering using K-Means and Davies Bouldin Index Algorithm," *TEM J.*, vol. 11, no. 4, pp. 1871–1876, 2022, doi: 10.18421/TEM114-55.
- [37] X. Wang and Y. Xu, "An improved index for clustering validation based on Silhouette index and Calinski-Harabasz index," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 569, no. 5, 2019, doi: 10.1088/1757-899X/569/5/052024.
- [38] W. Zhang *et al.*, "Modulation format identification using the Calinski-Harabasz index," *Appl. Opt.*, vol. 61, no. 3, pp. 851–857, 2022.