

Behavioral Segmentation of Ride-Hailing Users Using the K-Means Clustering Algorithm

¹Fajar Adi Pratama, ²Gusmelia Testiana

^{1,2}UIN Raden Fatah Palembang

Jl. Prof. K. H. Zainal Abidin Fikri No.3, Kota Palembang, Sumatera Selatan, Indonesia

email : 23021450016_uin@radenfatah.ac.id, gusmeliatestiana_uin@radenfatah.ac.id

ABSTRACT

The rapid growth of the ride-hailing industry has generated massive volumes of transactional data (*Big Data*), containing valuable information regarding user behavior patterns. One of the major challenges faced by companies is the difficulty of mapping user characteristics at an individual level, resulting in marketing strategies that are often generalized and insufficiently targeted. This study aims to segment ride-hailing service users using the K-Means Clustering algorithm to identify groups of user behavior based on the *Frequency*, *Monetary*, and *Rating* attributes. The research methodology follows the stages of the *Knowledge Discovery in Databases* (KDD) process, including data identification, data preprocessing, and result interpretation using Orange Data Mining software. The results indicate that four optimal user clusters were formed. Profile analysis identified four user segments: *Loyal Customers* (frequent users with high satisfaction), *Thrifty Customers* (cost-conscious yet satisfied users), *Risky Customers* (frequent users who are dissatisfied with the service), and *Passive Customers* (users with relatively low transaction activity). This study recommends segment-specific *Customer Relationship Management* (CRM) strategies, with particular emphasis on service recovery for high-risk users and retention programs for loyal users to maximize the company's profitability.

Keywords - Data Mining, K-Means Clustering, Ride-Hailing, CRM.

1. Introduction

Di era teknologi informasi modern saat ini, banyak perubahan telah terjadi, salah satunya pada bisnis transportasi, yang mengacu pada hadirnya layanan *ride-hailing*. Ribuan bahkan jutaan transaksi yang terjadi setiap saat, menghasilkan volume data yang besar, yang menyimpan berbagai informasi mengenai perilaku pengguna layanan *ride-hailing*. Mulai dari pola perjalanan, pemesanan, pengeluaran, dan pemberian rating. Informasi tersebut menjadi sebuah data, data yang di mana semakin lama semakin besar dan tidak akan memiliki nilai strategis jika tidak diolah menjadi sebuah informasi yang bermanfaat dalam pengambilan keputusan bisnis [1].

Permasalahan yang sering dihadapi penyedia layanan *ride-hailing* saat ini adalah bagaimana cara memahami karakteristik setiap pengguna secara personal, ditambah dengan jumlah data yang semakin bertambah dan semakin besar. Hal ini menyebabkan pihak perusahaan kesusahan dalam menerapkan strategi pemasaran yang bersifat umum kepada seluruh pengguna. Menurut prinsip manajemen pemasaran, setiap pengguna memiliki preferensi dan nilai (*value*) yang berbeda [2]. Sehingga, perlakuan yang sama terhadap pengguna royal dan pengguna pasif mengakibatkan inefisiensi biaya promosi dan risiko kehilangan pelanggan. Menurut Buttle dan Maklan [11], manajemen hubungan pelanggan yang efektif dapat meningkatkan profit perusahaan melalui pemahaman terhadap penilaian pengguna. Oleh karena itu, dibutuhkan pendekatan teknis untuk melakukan segmentasi guna mengelompokkan pengguna berdasarkan perilaku mereka.

Solusi yang ditawarkan untuk menangani permasalahan tersebut adalah melakukan penerapan *Data Mining* dengan teknik *Clustering*. Teknik *Clustering* mampu mengelompokkan data tanpa *label* ke dalam beberapa klaster berdasarkan karakteristiknya, dengan algoritma yang umum digunakan adalah *K-Means Clustering*. Algoritma ini memiliki efisiensi tinggi dalam mengelola data besar dan sering bertambah, sehingga mudah diimplementasikan untuk kasus segmentasi bisnis [3].

Sejumlah penelitian terdahulu telah membuktikan, bahwa algoritma *K-Means Clustering* mampu melakukan segmentasi pengguna. Awalina dan Rahayu [4] menerapkan *K-Means* pada data transaksi ritel *online* dan berhasil mengidentifikasi kelompok pengguna untuk optimalisasi strategi pemasaran. Penelitian serupa oleh Kiat et al. [5], memanfaatkan model *RFM (Recency, Frequency, Monetary)* dan metode *Elbow* untuk menentukan jumlah klaster terbaik dalam segmentasi pelanggan. Didukung dengan pengembangan variabel yang dilakukan Kristanti et al. [6] yang tidak hanya menggunakan *RFM*, tetapi juga mengintegrasikan variabel demografis seperti usia dan pendapatan pada studi kasus toko *offline*.

Penelitian-penelitian tersebut berhasil melakukan segmentasi pelanggan menggunakan *K-Means*, namun mayoritas objek penelitian berfokus pada sektor ritel. Belum banyak penelitian yang menerapkan segmentasi pada sektor transportasi (*ride-hailing*) dengan melibatkan variabel *user rating* sebagai karakteristik perilaku. Oleh karena itu, penelitian hadir untuk mengisi celah dengan mengkombinasikan variabel transaksi dan variabel *rating* untuk menghasilkan profil pelanggan yang lebih komprehensif.

2. Research Method

Penelitian dilakukan dengan mengikuti tahapan *Knowledge Discovery in Database (KDD)* sebagaimana didefinisikan oleh Fayyad et al. [7], yang disesuaikan untuk kebutuhan studi kasus. Tahapan dimulai dari identifikasi masalah, pengumpulan data, pra-pemrosesan data, penerapan algoritma *K-Means*, hingga analisis hasil klaster. Alat bantu yang digunakan dalam penelitian ini adalah menggunakan perangkat lunak *Orange Data Mining* [8] versi terbaru.

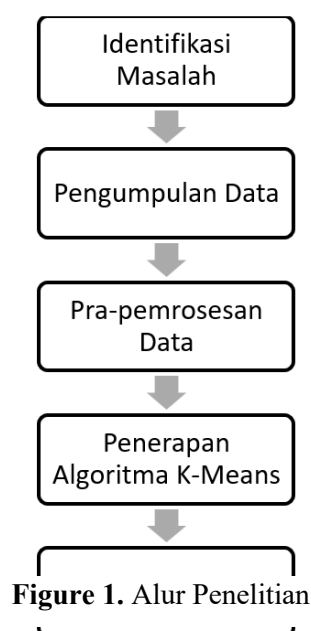


Figure 1. Alur Penelitian

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari repositori publik (Kaggle) oleh Zahraa [9], yang menyajikan riwayat transaksi pengguna layanan *ride-hailing* sebanyak 500 data. Dataset ini merupakan data simulasi yang mencerminkan pola perilaku pengguna nyata.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
Customer ID	Request Time	Pickup Location	Latitude Pickup	Longitude Pickup	Dropoff Location	Latitude Dropoff	Longitude Dropoff	Ride Distance (in miles)	Fare Amount (in \$)	Payment Method	Driver ID	Vehicle Type	Traffic Condition	Peak Hours	Day of Week	Public Holiday	User Rating
1	9/15/22 17:53	66.634374,135.003249	17.581381	177.70173	66.648223,-50.335487	-10.335329	-166.554534	1826.455308	202.65	Apple Pay	M33439	SUV	High	No	Sunday	Yes	1
2	1/21/22 10:42	66.664754,106.248391	-47.788693	-147.386595	53.172444,-50.446042	-18.388815	53.525353	4398.808844	478.88	Cash	K59550	Bus	High	Yes	Wednesday	No	5
3	10/15/22 7:12	24.332165,176.686135	56.88338	-12.596775	48.801493,-17.366678	-91.338857	107.38132	9032.200396	925.23	Debit Card	AP1181	Bus	Low	No	Wednesday	Yes	3
4	3/8/22 22:43	60.132341,-11.059698	47.285989	-135.238759	-54.835591,53.004335	87.448327	-174.855587	8032.320088	828.23	Cash	SE3390	SUV	High	Yes	Monday	No	1
5	10/12/22 7:07	1.339379,93.042777	-81.399148	-23.00383	55.832426,-53.814087	-13.594046	39.548566	4286.318575	448.63	Debit Card	GN1232	SUV	High	No	Friday	Yes	4
6	2/8/24 7:58	6.431558,-96.561204	77.184723	87.577561	-44.065556,61.428412	-14.735454	-60.087879	1306.741375	150.67	Visa	ER0243	Motorcycle	Low	Yes	Sunday	No	5
7	5/13/22 18:56	47.378213,-46.257985	7.885064	137.480873	54.900097,-110.115135	25.173386	-148.841857	2418.119623	261.81	Visa	K25916	SUV	High	No	Sunday	Yes	5
8	10/6/24 22:15	68.538838,58.780868	37.227124	146.133699	-51.131304,-75.284749	27.307668	-141.444983	672.81514621	87.28	Apple Card	JS6436	Sedan	Low	Yes	Thursday	No	1
9	9/24/24 14:01	10.5918,151.859583	-78.079822	61.157787	0.358334,66.480144	-46.801064	51.889387	10612.90879	1081.29	Debit Card	KS4211	Bus	Medium	Yes	Thursday	Yes	5
10	1/18/23 18:28	45.306872,-40.02319	7.813131	-61.789328	-21.124508,-171.483735	25.338965	-175.725118	7394.343183	749.43	GoPay	CU1785	SUV	Medium	No	Wednesday	No	5
11	3/28/22 5:53	-18.80184,106.515997	55.567192	-172.062302	74.069179,129.148891	-65.652866	80.273174	5050.61179	525.06	Apple Pay	K23633	Sedan	Low	Yes	Thursday	No	5
12	9/12/22 10:48	66.664854,-107.842129	-15.18148	140.098387	-81.281014,-10.013396	13.321448	-75.56732	7388.673988	756.87	Visa	YP4485	SUV	High	Yes	Thursday	No	5
13	1/21/24 3:39	46.288208,-145.665648	-3.30448	-100.080495	-2.98105,-33.827858	-63.791489	36.681597	5841.640889	584.16	Debit Card	M83274	Motorcycle	Low	Yes	Friday	Yes	5
14	1/21/22 1:35	1.456765,68.853317	84.811125	87.806597	29.585419,-51.585557	23.748088	130.865138	428.521897	62.85	Debit Card	QE3681	Bus	High	Yes	Tuesday	Yes	1
15	3/31/23 14:11	64.860713,-71.335816	-4.78167	-66.705666	15.802014,54.583388	18.257147	-128.861313	7238.414448	743.84	Cash	CR0071	SUV	Low	Yes	Tuesday	No	4
16	6/17/23 12:02	47.190894,-12.602471	38.612863	162.310388	-55.835777,87.802028	46.686455	-52.158623	4959.284891	515.93	GoPay	TE8907	Sedan	High	Yes	Thursday	Yes	3
17	3/28/24 5:55	78.880376,-17.85841	28.362779	143.488815	65.147288,-48.387011	-75.42723	-45.315187	7591.897026	779.19	Debit Card	W25952	Motorcycle	High	No	Wednesday	No	4
18	3/14/24 16:19	79.000116,145.348821	-43.968345	-78.838451	44.101915,38.810109	31.882318	51.869629	2472.958151	287.40	Cash	R86848	Sedan	Low	Yes	Tuesday	Yes	4
19	5/18/24 6:53	-47.33823,58.895214	78.634844	0.804648	-75.340857,-135.632616	-28.755421	150.756211	8948.45174	714.84	GoPay	B84505	Motorcycle	High	Yes	Monday	No	3
20	8/8/22 18:34	45.899816,68.435641	77.514382	175.706664	17.000232,40.136515	-12.412378	-26.321858	6762.944388	686.26	Debit Card	DC0128	SUV	Low	Yes	Saturday	Yes	4
21	8/6/23 18:56	17.732121,-80.810044	-17.093388	-64.563896	52.15414,66.688328	-68.238967	-167.557626	5075.440765	527.54	Visa	GG1390	Bus	High	Yes	Friday	Yes	5
22	1/23/23 12:58	68.783779,-17.636056	-54.089376	105.050802	-56.865381,-130.611344	-15.740031	38.822533	10363.184521	1056.32	GoPay	H99881	SUV	Low	No	Monday	Yes	1
23	1/21/22 10:53	-70.634598,-177.482714	12.212812	-132.418665	-40.8083,-87.871488	-12.734722	-165.047332	4737.322197	489.73	Cash	Y19759	Motorcycle	High	No	Saturday	No	3
24	8/15/22 20:38	64.698416,-127.092286	0.152369	-172.321747	-40.879681,-12.106041	62.110051	64.835554	11883.38814	1208.34	GoPay	JP2404	SUV	Low	No	Tuesday	No	1
25	8/7/24 5:39	60.016391,136.738343	46.825474	-141.788754	-2.326755,-45.878341	15.078933	101.308625	9235.705051	842.87	Apple Pay	PA4422	Motorcycle	High	No	Thursday	Yes	2

Figure 2. Dataset Ride-Sharing

Data yang diperoleh di awal pada **Figure 2**. Dataset Ride-Sharing berbentuk data transaksi. Agar dapat diolah menggunakan K-Means, dilakukan proses *cleaning*, *aggregation*, dan *selection*. Garcia et al. [12] menekankan bahwa data nyata sering tidak konsisten dan mengandung *noise*, sehingga digunakan teknik pembersihan agar data mutlak sebelum pemodelan dilakukan. Atribut atau variabel utama yang dipilih untuk proses klusterisasi meliputi:

- Frequency*, berisi data jumlah pesanan yang dilakukan pengguna selama satu periode.
- Monetary*, berisi data total pengeluaran yang dikeluarkan pengguna.
- Total Distance*, berisi data total jarak tempuh perjalanan pengguna.
- Avg_Rating*, berisi data penilaian yang diberikan oleh pengguna.

Hasil pra-pemrosesan data ditunjukkan pada **Figure 3**. Hasil Pra-pemrosesan Data

	A	B	C	D	E
1	Row Labels	Frequency	Total_Distance	Monetary	Avg_Rating
2	1	1	1826,46	202,65	1
3	2	1	4588,81	478,88	5
4	3	1	9052,28	925,23	3
5	4	1	8082,33	828,23	1
6	5	1	4286,32	448,63	4
7	6	1	1306,74	150,67	5
8	7	1	2419,12	261,91	5
9	8	1	672,82	87,28	1
10	9	1	10612,91	1081,29	5
11	10	1	7294,34	749,43	5
12	11	1	5050,63	525,06	5
13	12	1	7368,67	756,87	5
14	13	1	5641,64	584,16	5
15	14	1	428,52	62,85	1
16	15	1	7238,41	743,84	4
17	16	1	4959,28	515,93	3
18	17	1	7591,90	779,19	4
19	18	1	2473,96	267,40	4
20	19	1	6948,43	714,84	3

Figure 3. Hasil Pra-pemrosesan Data

Metode yang digunakan untuk segmentasi adalah K-Means *Clustering*, karena algoritma ini diakui sebagai salah satu dari sepuluh algoritma terbaik dalam penambahan data [13]. Algoritma ini merupakan

metode *unsupervised learning* yang mempartisipasi data ke dalam k kelompok (*cluster*). Algoritma K-Means juga merupakan metode *clustering* berbasis jarak yang membagi data ke dalam sejumlah *Cluster* dan hanya bekerja pada atribut numerik. Perhitungan jarak menggunakan rumus *Euclidean Distance* [10] sebagai berikut:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

Di mana:

- $d(x, y)$: jarak antara titik data x dan pusat *cluster* y .
 x_i : nilai atribut ke- i pada data.
 y_i : nilai atribut ke- i pada *centroid*.

Kemudian, proses *clustering* dilakukan dengan menggunakan *workflow* pada *Orange Data Mining*, ditunjukkan pada **Figure 4**. Workflow K-Means pada *Orange Data Mining*

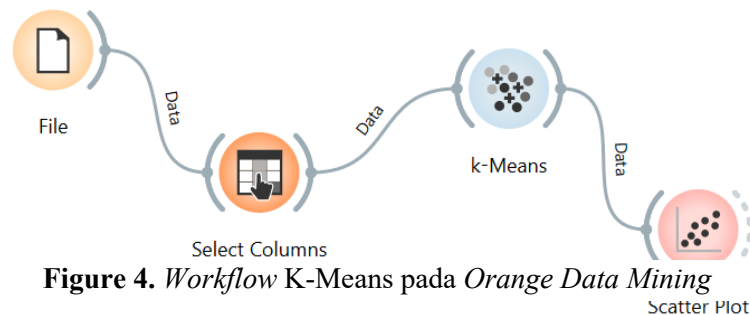


Figure 4. Workflow K-Means pada *Orange Data Mining*

3. Result and Analysis

Berdasarkan percobaan dengan melakukan penerapan Algoritma K-Means menggunakan perangkat lunak *Orange Data Mining*, jumlah *cluster* (k) yang dipilih sebanyak 4 *cluster*.

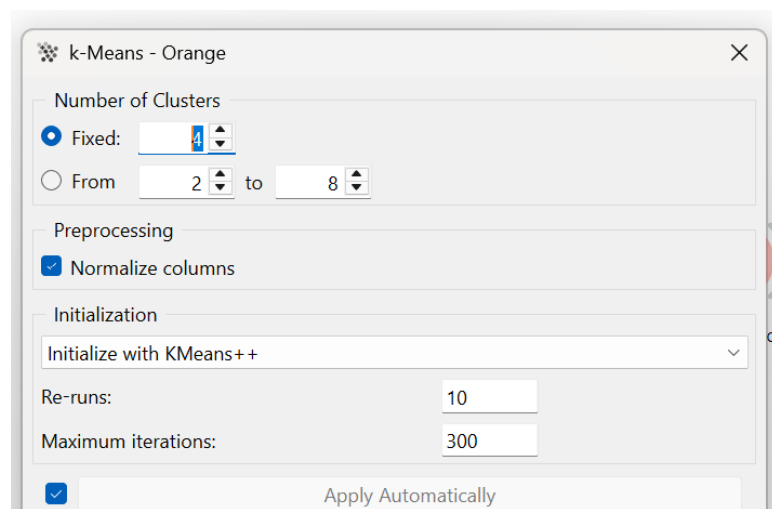


Figure 5. Pemilihan Clustering

Pemilihan $k = 4$ didasarkan pada keterpisahan data yang jelas dalam mempresentasikan kelompok perilaku pengguna yang berbeda secara signifikan. Hasil visualisasi dari *scatter plot*, persebaran data dapat dilihat pada **Figure 6**. Visualisasi Sebaran Cluster Pelanggan (Sumbu X: Monetary, Sumbu Y: Rating) Sumbu X merepresentasikan variabel *Monetary* (Total Pengeluaran), sedangkan sumbu Y merepresentasikan variabel *Avg_Rating* (Tingkat Kepuasan). Dari hasil visualisasi tersebut, algoritma K-Means berhasil mempartisipasi pengguna ke dalam empat kuadran yang berbeda, berdasarkan pola

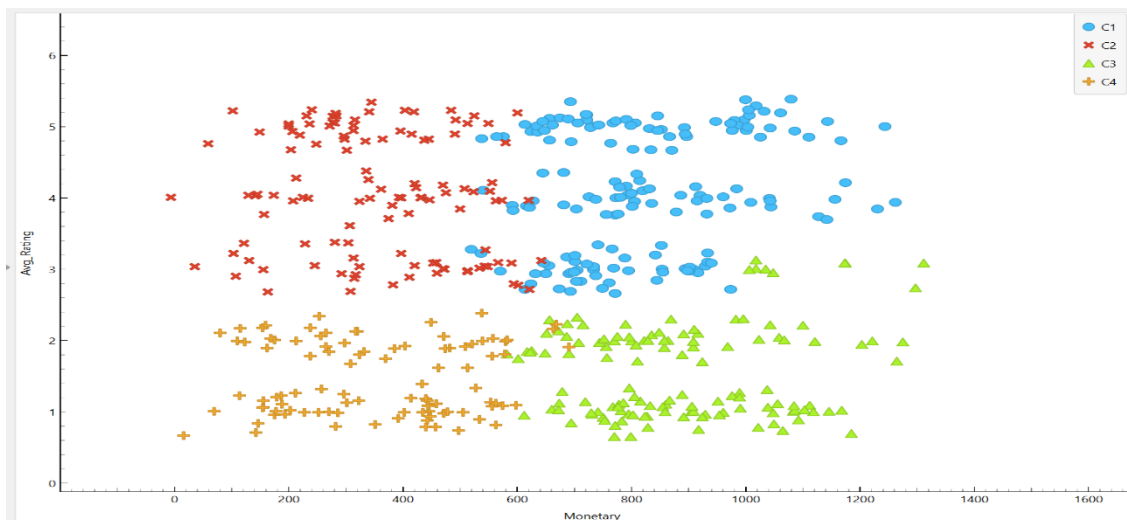


Figure 6. Visualisasi Sebaran *Cluster* Pelanggan (Sumbu X: *Monetary*, Sumbu Y: *Rating*)

pengeluaran dan tingkat kepuasannya.

Dalam penentuan jumlah *cluster* yang optimal, diperlukan hasil yang valid untuk memastikan *cluster* tersebut penting. Metode validasi internal seperti *Silhouette Coefficient* sering digunakan untuk mengukur seberapa baik suatu objek ditempatkan dalam *cluster*-nya sendiri dibandingkan *cluster* lain [14].

a. *Cluster 1* (Warna Biru) - *Loyal Customers*

Karakteristik kelompok ini adalah dengan tingkat pengeluaran (*Monetary*) tertinggi dan pemberian *rating* di atas rata-rata (mendekati 5.0). Dengan analisa, segmen pelanggan paling berharga, karena mereka sering menggunakan dan mengeluarkan uang yang banyak untuk menggunakan layanan, dengan tingkat kepuasan yang tinggi. Mereka bisa disebut sebagai aset loyalitas perusahaan.

b. *Cluster 2* (Warna Merah) – *Thrifty Customers*

Karakteristik kelompok ini adalah kelompok dengan tingkat pengeluaran (*Monetary*) rendah, tetapi pemberian *rating* tinggi. Dengan analisa, segmen pelanggan puas dengan layanan, karena mereka menggunakan layanan untuk perjalanan jarak pendek dan pengeluaran hemat, namun menyukai layanan yang diberikan.

c. *Cluster 3* (Warna Hijau) – *Risky Customers*

Karakteristik kelompok ini adalah dengan tingkat pengeluaran (*Monetary*) tertinggi, tetapi pemberian *rating* terendah. Dengan Analisa, segmen pelanggan paling kritikan, karena mereka mengeluarkan uang yang banyak untuk layanan, namun sering mengalami pengalaman yang buruk. Jika tidak segera diatasi, kelompok ini akan berhenti layanan dan menyebabkan kerugian pendapatan perusahaan.

d. *Cluster 4* (Warna Jingga) – *Passive Customer*

Karakteristik kelompok ini adalah dengan tingkat pengeluaran (*Monetary*) dan pemberian *rating* terendah. Dengan analisa, segmen pelanggan yang jarang menggunakan layanan dengan

kepuasan yang cenderung rendah. Perekonomian mereka bagi perusahaan relatif kecil dibandingkan kelompok lainnya.

Dengan hasil segmentasi dan analisis pengguna yang telah dilakukan sebelumnya, perusahaan direkomendasikan untuk menerapkan strategi CRM (*Customer Relationship Management*) yang berbeda untuk setiap segmen. Strategi ini dimuat dan dirangkum dalam **Table 1**. Rekomendasi Strategi Bisnis per Cluster

Table 1. Rekomendasi Strategi Bisnis per *Cluster*

<i>Cluster</i>	<i>Label Segment</i>	<i>Strategi</i>	<i>Rekomendasi</i>
<i>Cluster 1</i>	<i>Loyal Customers</i>	Pertahankan	Memberikan status prioritas, melalui akses layanan eksklusif agar merasa dihargai dan tidak berpindah ke kompetitor lain.
<i>Cluster 2</i>	<i>Thrifty Customers</i>	Tingkatkan	Memberikan potongan harga, untuk mendorong menggunakan layanan dengan pengeluaran yang banyak.
<i>Cluster 3</i>	<i>Risky Customers</i>	Pemulihan	Melakukan kontak personal, untuk menanyakan permasalahan, kemudian memberikan kompensasi layanan yang menguntungkan.
<i>Cluster 4</i>	<i>Pasive Customer</i>	Efisiensi	Melakukan promosi standar, untuk menarik kembali perhatian tanpa ada pemasaran yang terlalu besar.

Khusus untuk segmen 3, penerapan strategi pemulihan menjadi prioritas. Michel et al. [15] menyatakan bahwa penanganan keluhan yang cepat, mempermudah dalam meningkatkan kepercayaan pengguna terhadap pelayanan ke tingkat yang lebih tinggi dibandingkan sebelum masalah terjadi.

4. Conclusion

Penerapan algoritma K-Means *Clustering* terbukti dalam memetakan profil pengguna layanan *ride-hailing*, berdasarkan pola *Monetary* dan *Rating*. Berdasarkan percobaan model, terbentuk 4 *cluster* yang merepresentasikan segmen perilaku pengguna yang berbeda-beda. Keempat *cluster* yang terbentuk memiliki karakteristik yang unik di setiap segmennya. *Cluster 1* merupakan segmen yang paling ideal dalam perloyalan yang tinggi, *Cluster 2* yang segmen dengan harga murah namun kepuasan sesuai layanan, *Cluster 3* dengan segmen yang memiliki nilai pengeluaran tinggi namun tingkat kepuasan rendah, serta *Cluster 4* yang memiliki kontribusi minim bagi perusahaan. Prioritas penanganan utama yang perlu dilakukan, ditujukan pada segmen *Risky Customers* melalui strategi pemulihan untuk mencegah perpindahan pelanggan ke kompetitor lainnya, serta menjaga segmen *Loyal Customers*.

Keterbatasan penelitian ini hanya menggunakan variabel *Frequency*, *Monetary*, dan *Rating*. Penelitian ini dapat dikembangkan dan disarankan menambahkan variabel lainnya, seperti dimensi waktu atau jenis layanan yang disediakan, untuk mendapatkan profil pelaku pengguna yang lebih mendetail. Untuk perbandingan, disarankan menggunakan kinerja Algoritma K-Means dengan Algoritma *clustering* lain seperti K-Medoids atau DBSCAN, untuk melihat metode mana yang menghasilkan pengelompokan terbaik pada dataset transportasi. erisikan kesimpulan berdasarkan hasil-hasil penelitian yang diperoleh, serta kelebihan dan kekurangan dari penelitian yang dilaksanakan. Dapat pula ditambahkan beberapa saran pengembangan untuk penelitian selanjutnya.

References

- [1] E. Turban, R. Sharda and D. Delen, *Decision Support and Business Intelligence Systems*, 4 ed., Pearson, 2011, pp. 1-696.
- [2] P. Kotler and K. L. Keller, *Marketing Management*, 15 ed., Pearson, 2016, pp. 1-692.
- [3] J. Han, M. Kamber and J. Pei, *Data Mining: Concepts and Techniques*, 3 ed., Elsevier, 2012, pp. 1-740.
- [4] E. F. L. Awalina and W. I. Rahayu, "Optimalisasi Strategi Pemasaran dengan Segmentasi Pelanggan Menggunakan Penerapan Penerapan K-Means Clustering pada Transaksi Online Retail," *Jurnal Teknologi dan Informasi*, vol. 13, pp. 122-137, 2023.
- [5] A. B. H. Kiat, Y. Azhar and V. Rahmayanti, "Penerapan Metode K-Means Dengan Metode Elbow Untuk Segmentasi Pelanggan Menggunakan Model RFM (Recency, Frequency & Monetary)," *Repositor*, vol. 2, no. 7, pp. 945-952, 2020.
- [6] B. T. Kristanti, A. Junaidi and E. P. Mandyartha, "Implementasi K-Means Clustering Dalam Segmentasi Pelanggan Berdasarkan Usia, Pendapatan, Dan Model Rfm (Studi Kasus: Lantikya Store Jombang)," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 3, pp. 2099-2112, 2024.
- [7] U. Fayyad, G. Shapiro-Piatetsky and P. Smyth, "From Data Mining to Knowledge Discovery in Databse," *AI Magazine*, vol. 17, no. 3, pp. 37-54, 1996.
- [8] J. Demsar, T. Curk, A. Erjavec, C. Gorup, T. Hocevar, M. Milutinovic, M. Mozina, M. Polajnar, M. Toplak, A. Staric, M. Stajdohar, L. Umek, L. Zagar, J. Zbontar, M. Zitnik and B. Zupan, "Orange: Data Mining Toolbox in Python," *Machine Learning Research*, vol. 14, pp. 2349-2353, 2013.
- [9] Z. Al-Touq, "Ride-Sharing Dataset," 2024. <https://www.kaggle.com/datasets/zahraatouq/ride-sharing-dataset>.
- [10] A. K. Jain, "Data clustering: 50 years beyond K-means," *Elseiver*, vol. 31, no. 8, pp. 651-666, 2010.
- [11] F. Buttle and S. Maklan, *Customer Relationship Management*, 4 ed., London: Routledge, 2019, p. 468.
- [12] S. Garcia, J. Luenego and F. Herrera, *Data Preprocessing in Data Mining*, Springer, 2015.
- [13] X. Wu, V. Kumar, J. R. Quinlan, J. Ghosh, Q. Yang, H. Motoda, G. J. McLachlan, A. Ng, B. Liu, P. S. Yu, M. Steinbach, D. J. Hand and D. Steinberg, "Top 10 algorithms in data mining," *Knowledge and Information Systems*, vol. 14, pp. 1-37, 2008.
- [14] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *Elseiver*, vol. 20, pp. 53-65, 1987.
- [15] S. Michel, D. Bowen and R. Johnston, "Why service recovery fails: Tensions among customer, employee, and process perspectives," *Service Management*, vol. 20, no. 3, pp. 253-273, 2009.