

DAFTAR PUSTAKA

- Adiana, B. E., Soesanti, I., & Permanasari, A. E. (2018). Analisis Segmentasi Pelanggan Menggunakan Kombinasi RFM Model dan Teknik Clustering. *JUTEI*, (2). Retrieved from <https://doi.org/10.21460/jutei.2017.21.76>
- Anitha, P., & Patil, M. M. (2022). RFM model for customer purchase behavior using K-Means algorithm. *Journal of King Saud University - Computer and Information Sciences*, 34(5), 1785–1792. Retrieved from <https://doi.org/10.1016/j.jksuci.2019.12.011>
- Ayu, D., Dewi, I. C., & Pramita, K. (2019). Analisis Perbandingan Metode Elbow dan Silhouette pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali. *JURNAL MATRIX* (Vol. 9).
- Ayu Rospricilia, T., Ayu Ithriah, S., & Anjani Arifiyanti, A. (2020). SEGMENTASI PELANGGAN MENGGUNAKAN METODE K-MEANS CLUSTERING BERDASARKAN MODEL RFM PADA CV TITA JAYA. *Jurnal Informatika dan Sistem Informasi (JIFoSI)* (Vol. 1).
- Cahyo Wiguno, T., & Nataliani, Y. (2022). Penerapan k-Means Clustering Berdasarkan Analisis RFM Terhadap Segmentasi Pembeli untuk Meningkatkan Strategi CRM. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 6, 1871–1881. Retrieved from <https://doi.org/10.30865/mib.v6i4.4472>
- Carudin. (2021). PEMANFAATAN DATA TRANSAKSI UNTUK DASAR MEMBANGUN STRATEGI BERDASARKAN KARAKTERISTIK PELANGGAN DENGAN ALGORITMA K-MEANS CLUSTERING DAN MODEL RFM. *Jurnal Teknologi Terpadu*, 7(1), 7–14. Retrieved from <https://journal.nurulfikri.ac.id/index.php/jtt>
- Chen, A. H. L., Liang, Y. C., Chang, W. J., Siau, H. Y., & Minanda, V. (2022). RFM Model and K -Means Clustering Analysis of Transit Traveller Profiles: A Case Study. *Journal of Advanced Transportation*, 2022. Retrieved from <https://doi.org/10.1155/2022/1108105>
- Febriani, A., & Putri, S. A. (2020). Segmentasi Konsumen Berdasarkan Model Recency, Frequency, Monetary dengan Metode K-Means. *Jiems (Journal of Industrial Engineering and Management Systems)*, 13(2). Retrieved from <https://doi.org/10.30813/jiems.v13i2.2274>
- Firmansah, R. Y., Dedy Irawan, J., & Vendyansyah, N. (2021). ANALISIS RFM (RECENCY, FREQUENCY AND MONETARY) PRODUK MENGGUNAKAN METODE K-MEANS. *Jurnal Mahasiswa Teknik Informatika* (Vol. 5).
- Fithriyah, M., Yaqin, M. A., & Zaman, S. (2021). K-Means Clustering Untuk Segmentasi Produk Berdasarkan Analisis Recency, Frequency, Monetary (RFM) Pada Data Transaksi Penjualan. *ILKOMNIKA: Journal of Computer*

- Science and Applied Informatics*, 3(2), 151–164. Retrieved from <https://doi.org/10.28926/ilkomnika.v3i2.284>
- Gustientiedina, G., Adiya, M. H., & Desnelita, Y. (2019). Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 5(1), 17–24. Retrieved from <https://doi.org/10.25077/teknosi.v5i1.2019.17-24>
- Gustriansyah, R., Suhandi, N., & Antony, F. (2019). Clustering optimization in RFM analysis based on k-means. *Indonesian Journal of Electrical Engineering and Computer Science*, 18(1), 470–477. Retrieved from <https://doi.org/10.11591/ijeecs.v18.i1.pp470-477>
- Handoko, S., Fauziah, F., & Handayani, E. T. E. (2020). IMPLEMENTASI DATA MINING UNTUK MENENTUKAN TINGKAT PENJUALAN PAKET DATA TELKOMSEL MENGGUNAKAN METODE K-MEANS CLUSTERING. *Jurnal Ilmiah Teknologi Dan Rekayasa*, 25(1), 76–88. Retrieved from <https://doi.org/10.35760/tr.2020.v25i1.2677>
- Hartanti, N. T. (2020). Metode Elbow dan K-Means Guna Mengukur Kesiapan Siswa SMK Dalam Ujian Nasional. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 6(2), 82–89. Retrieved from <https://doi.org/10.25077/teknosi.v6i2.2020.82-89>
- Ilham, A., Setiawan, N. Y., & Afirianto, T. (2020). Analisis Segmentasi Pelanggan Kartu Prabayar Kabupaten Malang dengan RFM Model Menggunakan Metode Fuzzy C-Means Clustering (Studi Kasus : PT. XYZ) (Vol. 4). Retrieved from <http://j-ptiik.ub.ac.id>
- Larasati, A., Maren, R., & Wulandari, R. (2021). Utilizing Elbow Method for Text Clustering Optimization in Analyzing Social Media Marketing Content of Indonesian e-Commerce. *Jurnal Teknik Industri*, 23(2), 111–120. Retrieved from <https://doi.org/10.9744/jti.23.2.111-120>
- Monalisa, S. (2018). Segmentasi Perilaku Pembelian Pelanggan Berdasarkan Model RFM dengan Metode K-Means. *Jurnal Sistem Informasi*, 1.
- Muningsih, E., Maryani, I., & Handayani, V. R. (2021). Penerapan Metode K-Means dan Optimasi Jumlah Cluster dengan Index Davies Bouldin untuk Clustering Propinsi Berdasarkan Potensi Desa. *Jurnal Sains Dan Manajemen*, 9(1). Retrieved from www.bps.go.id
- Musyawah, I. Y., & Idayanti, D. (2022). Analisis Strategi Pemasaran Untuk Meningkatkan Penjualan Pada Usaha Ibu Bagas di Kecamatan Mamuju (online). *Jurnal Ilmiah Ilmu Manajemen*, 1(1), 2656–6265.
- Nofitri, R., & Irawati, N. (2019). ANALISIS DATA HASIL KEUNTUNGAN MENGGUNAKAN SOFTWARE RAPIDMINER. *JURTEKSI (Jurnal Teknologi Dan Sistem Informasi)*, 5(2), 199–204. Retrieved from <https://doi.org/10.33330/jurteksi.v5i2.365>

- Paembonan, S., Abduh, H., & Kunci, K. (2021). *Penerapan Metode Silhouette Coeficient Untuk Evaluasi Clustering Obat Clustering; K-means; Silhouette coeficient* (Vol. 6). Retrieved from <https://ojs.unanda.ac.id/index.php/jiit/index>
- Priyatman, H., Sajid, F., & Haldivany, D. (2019). Klasterisasi Menggunakan Algoritma K-Means Clustering untuk Memprediksi Waktu Kelulusan Mahasiswa. *JEPIN (Jurnal Edukasi Dan Penelitian Informatika)* .
- Puspa Wildyaksanjani, J., & Dadang Sugiana, dan. (2018). Strategi Customer Relationship Management (CRM) PT Angkasa Pura II (Persero). *Jurnal Kajian Komunikasi*, 6(1), 10–23.
- Rizki, B., Ginasta, N. G., Tamrin, M. A., & Rahman, A. (2020). Customer Loyalty Segmentation on Point of Sale System Using Recency-Frequency-Monetary (RFM) and K-Means. *Jurnal Online Informatika*, 5(2), 130. Retrieved from <https://doi.org/10.15575/join.v5i2.511>
- Rumiarti, C. D., & Budi, I. (2017). Customer Segmentation for Customer Relationship Management on Retail Company: Case Study PT Gramedia Asri Media. *Jurnal Sistem Informasi*, 13(1), 1. Retrieved from <https://doi.org/10.21609/jsi.v13i1.525>
- Sarasvananda, I. B. G., Wardoyo, R., & Sari, A. K. (2019). The K-Means Clustering Algorithm With Semantic Similarity To Estimate The Cost of Hospitalization. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 13(4), 313. Retrieved from <https://doi.org/10.22146/ijccs.45093>
- Savitri, A. D., Abdurrachman Bachtiar, F., & Setiawan, N. Y. (2018). *Segmentasi Pelanggan Menggunakan Metode K-Means Clustering Berdasarkan Model RFM Pada Klinik Kecantikan (Studi Kasus : Belle Crown Malang)* (Vol. 2). Retrieved from <http://j-ptiik.ub.ac.id>
- Sembinging Brahmana, R. W., Mohammed, F. A., & Chairuang, K. (2020). Customer Segmentation Based on RFM Model Using K-Means, K-Medoids, and DBSCAN Methods. *Lontar Komputer : Jurnal Ilmiah Teknologi Informasi*, 11(1), 32. Retrieved from <https://doi.org/10.24843/lkjiti.2020.v11.i01.p04>
- Shi, C., Wei, B., Wei, S., Wang, W., Liu, H., & Liu, J. (2021). A quantitative discriminant method of elbow point for the optimal number of clusters in clustering algorithm. *Eurasip Journal on Wireless Communications and Networking*, 2021(1). Retrieved from <https://doi.org/10.1186/s13638-021-01910-w>
- Shutaywi, M., & Kachouie, N. N. (2021). Silhouette analysis for performance evaluation in machine learning with applications to clustering. *Entropy*, 23(6). Retrieved from <https://doi.org/10.3390/e23060759>

- Sugiharto, A., Sari, B. N., & Padilah, T. N. (2021). ANALISIS CLUSTER SEBARAN COVID-19 MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING (STUDI KASUS: PROVINSI JAWA BARAT). *Journal of Information Technology and Computer Science (INTECOMS)*, 4(2), 291–301. Retrieved from <https://pikobar.jabarprov.go.id/>.
- Winarta, A., & Kurniawan, W. J. (2021). OPTIMASI CLUSTER K-MEANS MENGGUNAKAN METODE ELBOW PADA DATA PENGGUNA NARKOBA DENGAN PEMROGRAMAN PYTHON. *Jurnal Teknik Informatika Kaputama (JTIK)*, 5(1).
- Yanto, D., & Komputerisasi Akuntansi -AMIK Taruna Probolinggo Jl Raya Leces, D. (2019). Analisis RFM dan Algoritma K-Means untuk Clustering Loyalitas Customer, 9(1).



Skripsi Ganjil 22/23

by Aji Setiono



Submission date: 22-Dec-2022 09:49AM (UTC+0700)

Submission ID: 1985712012

File name: SKRIPSI_-_AJI_SETIONO_-_197064416019_-_TEST_TURNITIN.docx (2.32M)

Word count: 6564

Character count: 40441

Skripsi Ganjil 22/23

ORIGINALITY REPORT

18%

SIMILARITY INDEX

16%

INTERNET SOURCES

5%

PUBLICATIONS

7%

STUDENT PAPERS

PRIMARY SOURCES

1	etheses.uin-malang.ac.id Internet Source	2%
2	medium.com Internet Source	1%
3	ojs.uajy.ac.id Internet Source	1%
4	docplayer.info Internet Source	1%
5	Submitted to Universitas Brawijaya Student Paper	1%
6	repository.usd.ac.id Internet Source	1%
7	dspace.uui.ac.id Internet Source	1%
8	ejournal.uin-suska.ac.id Internet Source	1%
9	repository.mercubuana.ac.id Internet Source	<1%

Analisis RFM (*Recency Frequency Monetary*) dan K-Means *Clustering* pada Klinik Gigi untuk Menentukan Segmentasi Pasien

Aji Setiono¹, Agung Triayudi², Endah Tri Esti Handayani³

^{1,2,3} Jurusan Informatika Fakultas Teknologi Komunikasi dan Informatika Universitas Nasional
Jl. Sawo Manila Pejaten Pasar Minggu – Jakarta Selatan

¹aji.setiono16@gmail.com

²agung.triayudi@gmail.com

³endahtriesti@civitas.unas.ac.id

Abstrak

Palapa Dentists merupakan bidang usaha Klinik Gigi yang terletak di Jakarta Selatan. Agar mendapatkan pasien lebih banyak dan kepuasan pelayanan terhadap pasien, maka perusahaan harus mempunyai strategi. Palapa Dentists belum mengadopsi strategi CRM (*Customer Relationship Management*) masih memperlakukan semua pasien dengan pendekatan yang sama. Berdasarkan permasalahan tersebut maka diperlukan data *mining* menggunakan teknik *cluster* untuk mengetahui karakteristik setiap pasien. Penelitian ini menggunakan metode RFM (*Recency Frequency Monetary*) dan K-Means *Clustering* dengan tujuan menentukan segmentasi pasien dan memilih kelompok pasien mana yang paling menguntungkan bagi perusahaan. Penentuan jumlah *cluster* menggunakan *elbow method* yang menghasilkan jumlah *cluster* terbaik adalah 2. *Silhouette score* menghasilkan jumlah 2 *cluster* dengan *score* 0.6014345457538962. Sedangkan *hasil davies-bouldin score* menunjukkan *cluster* optimal dengan 3 *cluster* tapi skornya 0.7500785223208264 masih jauh dari 0. *Cluster* 1 memiliki 17.413 anggota dan *cluster* 2 memiliki 2.068 anggota. *Cluster* 1 memiliki nilai rata-rata *recency* 641,63, *frequency* 3,21, dan *monetary* Rp. 2.424.251,98. Sedangkan *cluster* 2 memiliki nilai rata-rata *recency* 286,87, *frequency* 19,32, dan *monetary* Rp. 20.087.467,49. Dapat disimpulkan *cluster* 2 adalah kelompok pasien yang lebih menguntungkan dibandingkan *cluster* 1.

Kata kunci: *Customer Relationship Management*, Segmentasi, RFM, K-Means *Clustering*, *Cluster*

I. PENDAHULUAN

Di era *modern* dengan semua kemudahannya dan teknologi yang terus berkembang, strategi setiap perusahaan dalam melakukan teknik pemasaran juga semakin berkembang. Dalam strategi pemasaran, sebelum melakukan berbagai macam promosi kepada pelanggan, hendaknya perusahaan terlebih dahulu membidik segmen secara jelas dan tepat. Dengan menerapkan strategi pemasaran yang tepat maka perusahaan akan lebih menghemat biaya pengeluaran untuk promosi. Sering kali perusahaan tidak dapat secara efektif melakukan promosi kepada pelanggan karena belum memiliki data yang akurat. Persaingan yang ketat antar perusahaan tentunya menuntut setiap perusahaan untuk beradaptasi dengan perkembangan dunia bisnis [1].

Dengan semakin banyaknya Klinik Gigi di Jakarta, maka persaingan untuk mendapatkan dan mempertahankan pasien akan semakin ketat. Setiap Klinik Gigi tentu memiliki strategi tersendiri dalam melakukan teknik pemasaran. CRM (*Customer Relationship Management*) merupakan perpaduan proses dan teknologi dalam memahami pelanggan [2]. Tujuan dari CRM (*Customer Relationship Management*) untuk mendapatkan pelanggan yang berkomitmen terhadap jasa dan

produk sebuah perusahaan dan menjaga hubungan baik antara perusahaan dengan pelanggan [3]. Palapa Dentists belum mengadopsi strategi CRM (*Customer Relationship Management*) masih memperlakukan semua pasien dengan pendekatan yang sama.

Data dengan jumlah yang besar membutuhkan peranan teknologi untuk mengolahnya [4]. Penggunaan data *mining* dalam proses CRM (*Customer Relationship Management*) dapat memberikan pemahaman lebih mendalam tentang pelanggan [5]. Istilah data *mining* sering dikenal dengan data *analysis* dan KDD (*Knowledge Discovery in Databases*). KDD merupakan metode atau cara untuk mendapatkan pengetahuan dari sebuah data [6]. Proses data *mining* mencakup pengumpulan, pemakaian data perusahaan untuk menemukan keteraturan, pola hubungan dalam himpunan data [7].

Segmentasi pelanggan dapat membagi pelanggan menjadi beberapa kelompok untuk dapat mempermudah perlakuan kepada setiap pelanggan atau membagi pasar agar lebih mudah ditangani oleh perusahaan [8]. Segmentasi pelanggan membutuhkan variabel deskriptif untuk mengidentifikasi pola perilaku pelanggan [9]. Tujuan segmentasi pelanggan adalah untuk mengetahui karakteristik pelanggan agar perusahaan dapat memilah pelanggan mana yang paling menguntungkan

atau kurang menguntungkan bagi perusahaan [10]. Dengan metode RFM (*Recency Frequency Monetary*) pelanggan dapat dikelompokkan berdasarkan waktu kunjungan terakhir, frekuensi seberapa banyak kunjungan, dan total uang yang dikeluarkan pelanggan [11].

Dalam mengelompokkan data pelanggan dalam jumlah besar diperlukan teknik *clustering*. *Clustering* merupakan salah satu teknik dalam data mining dan bersifat *unsupervised* yaitu memiliki arti bahwa atribut-atribut dari suatu kelompok atau *cluster* yang memiliki karakteristik yang sama dapat dikelompokkan dalam satu *cluster* [12], [13]. *Clustering* juga dapat diartikan sebagai proses pembentukan *cluster* atau kelompok data dari himpunan data yang tidak diketahui kelompoknya, kemudian mengelompokkan data tersebut ke dalam satu *cluster* yang memiliki kemiripan data [14], [15]. Salah satu metode *clustering* paling terkenal dengan kesederhanaan dan kecepatannya adalah K-Means *Clustering* [16]. Metode K-Means *Clustering* adalah metode mengelompokkan data yang terkenal dengan kesederhanaan algoritma, dan kecepatan memilih pusat *cluster* [17]. Simbol K pada K-means clustering diartikan sebagai jumlah *cluster* yang digunakan.

Penelitian ini menggunakan metode RFM (*Recency Frequency Monetary*) dan K-Means *Clustering*. Untuk menentukan jumlah *cluster* diperlukan metode untuk optimasi jumlah *cluster* terbaik. *Elbow method* adalah prosedur untuk mengoptimalkan algoritma K-Means dengan mengevaluasi *cluster* optimal melalui estimasi *sum square error* (SSE) di setiap rentang *cluster* yang ditentukan [18]. *Silhouette Coefficient* adalah metode evaluasi *cluster* yang menggabungkan metode *cohesion* dan *separation* [19]. Nilai *silhouette* yang dihasilkan terletak pada rentang nilai -1 sampai dengan 1. Semakin besar nilai koefisien *silhouette* mendekati nilai 1, semakin baik pengelompokan data dalam satu *cluster*. Sebaliknya, jika koefisien *silhouette* mendekati nilai -1, maka pengelompokan data dalam satu *cluster* akan semakin buruk. *Davies-Bouldin Index* adalah metode yang digunakan untuk mengukur validitas atau jumlah *cluster* paling optimal pada suatu metode pengelompokan dimana kohesi didefinisikan sebagai jumlah dari kedekatan data terhadap titik pusat *cluster* dari *cluster* yang diikuti [20]. Tujuan penelitian ini untuk menentukan segmentasi pasien dan memilih kelompok pasien mana yang paling menguntungkan bagi perusahaan.

II. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Tahapan pada penelitian ini meliputi:



Gambar 1. Tahapan Penelitian

A. Business Understanding

Melakukan analisis bisnis Palapa Dentists dan masalah segmentasi pasien. Pemahaman bisnis berfokus pada pemahaman tujuan dan persyaratan penelitian.

B. Data Understanding

Melakukan pemahaman tentang data yang ada di Palapa Dentists. Pada tahap ini peneliti melakukan identifikasi data dan penentuan sumber data. Pengumpulan data dilakukan di Palapa Dentists. Jenis data adalah transaksi pasien dari 26 Desember 2016 sampai 06 April 2022 yang berjumlah 259.657 transaksi.

C. Data Preparation

Menyiapkan data akhir untuk proses segmentasi. Data diolah menggunakan Google Colab, bahasa pemrograman python, dan menggunakan beberapa *library* data science dari python. Pada tahap ini data meliputi memilih data atau kolom apa saja yang akan digunakan, membersihkan data dari *missing value*, mencari atau mendapatkan atribut baru dari sebuah data, menggabungkan beberapa data untuk membuat atribut data baru, dan mengubah data sesuai kebutuhan analisis.

D. Modeling

Pada tahap ini dilakukan proses RFM (*Recency Frequency Monetary*) dari atribut data yang telah dipilih. Nilai *recency* diperoleh dari seberapa lama pasien melakukan kunjungan yaitu tanggal perhitungan *recency* dikurangi tanggal terakhir kunjungan pasien. Tanggal terakhir analisis adalah 7 April 2022 karena dataset terakhir adalah 6 April 2022. Nilai *frequency* diperoleh dengan menjumlahkan atribut kode transaksi pasien. Nilai *monetary* diperoleh dengan menjumlahkan total biaya yang dikeluarkan pasien. Kemudian data hasil RFM (*Recency Frequency Monetary*) dilakukan *clustering* menggunakan metode K-Means *Clustering*. Tahapan K-Means *Clustering* pada penelitian ini yaitu:

1. Mengecek data *outlier* pada dataset menggunakan boxplot.
2. Penanganan data *outlier* menggunakan IQR (*Interquartile Range*), nilai *outlier* adalah nilai data yang letaknya lebih dari 1.5 x panjang kotak IQR.
3. Standarisasi dataset.
4. Memilih nilai K *cluster*.

E. Evaluation

Pada tahap *evaluation* menggunakan tiga metode yaitu *Elbow Method*, *Silhouette Coefficient*, dan *Davies-Bouldin Index* untuk menentukan jumlah *cluster* terbaik. Gambaran akhir dari *elbow method* adalah grafik konsistensi *cluster* terbaik dengan memplot nilai SSE. Penurunan paling ekstrim dan paling berbentuk siku dianggap sebagai jumlah *cluster* paling optimal. Untuk menghitung nilai *sum square error* (SSE) dengan Rumus 1.

$$SSE = \sum_{i=1}^k \sum_{j \in C_i} \|X_j - C_i\|^2 \quad (1)$$

Keterangan :

k : nomor *cluster* yang ditentukan

x_j : data j pada *cluster* i

C_i : inialisasi *centroid* secara acak

Untuk menghitung *silhouette score* dengan Rumus 2.

$$Silhouette\ score = \frac{p-q}{\max(p,q)} \quad (2)$$

Keterangan:

p : jarak rata-rata ke titik-titik di *cluster* terdekat

q : jarak rata-rata intra *cluster* ke semua titik di *cluster* nya sendiri.

Davies-Bouldin Score didefinisikan sebagai rata-rata ukuran kesamaan dari setiap *cluster* dengan *cluster* yang paling mirip, dimana kesamaan adalah rasio jarak dalam *cluster* terhadap jarak antar *cluster*. Nilai *Davies-Bouldin Score* adalah 0-1, nilai *cluster* yang mendekati nilai 0 maka akan semakin baik pengelompokkannya.

F. Deployment

Pada tahap ini label jumlah *cluster* digabungkan dengan dataset, sehingga setiap data memiliki label *cluster*.

III. HASIL DAN PEMBAHASAN

A. RFM (Recency Frequency Monetary)

Dataset awal untuk proses RFM (*Recency Frequency Monetary*) berjumlah 259.601 baris. Atribut yang digunakan dalam proses RFM adalah Tanggal Transaksi, Kode Pasien, Kode Transaksi, dan Total. Terdapat 171 pasien dengan *monetary* sama dengan 0, pasien tersebut adalah pasien yang melakukan tindakan tetapi biayanya gratisan oleh pihak Palapa Dentists, kemudian ada juga pasien yang memang sudah membayar paket perawatan di awal jadi nilai *monetary* nya sama dengan 0. Dengan pertimbangan dan pengamatan pada saat *business understanding* pada Palapa Dentists, pasien dengan nilai *monetary* sama dengan 0 tidak akan berpengaruh

ke hasil *clustering*, dan juga ada beberapa pasien dengan nilai *monetary* kurang dari Rp. 150.000,00 yaitu pasien yang tidak melakukan tindakan tetapi hanya membeli produk, pasien tersebut kurang memberikan keuntungan bagi perusahaan, maka dataset pasien dengan *monetary* kurang dari sama dengan Rp. 150.000,00 akan dihapus.

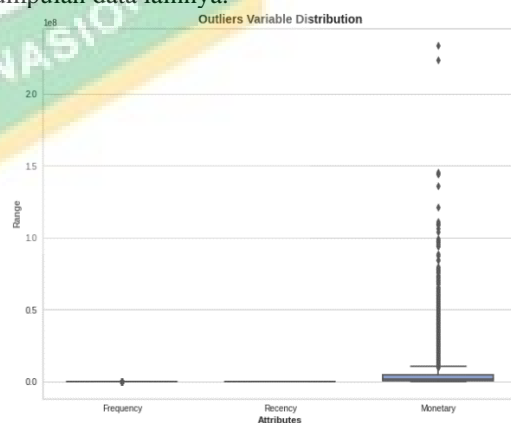
Tabel 1. Hasil RFM (Recency Frequency Monetary)

Kode Pasien	Recency	Frequency	Monetary
40681.0	62	75.0	233723125.0
412991.0	39	52.0	223585000.0
49740.0	4	42.0	145295000.0
47465.0	917	39.0	144105000.0
410053.0	587	58.0	144046500.0
48839.0	12	61.0	136178875.0
47666.0	71	82.0	121147000.0
....
422875.0	180	2.0	160000.0
423023.0	171	2.0	160000.0
422636.0	167	2.0	156000.0
420457.0	175	2.0	156000.0
422745.0	219	2.0	156000.0
422581.0	243	2.0	152000.0
422575.0	233	2.0	152000.0
422719.0	236	2.0	152000.0

Tabel diatas menunjukkan hasil akhir RFM (*Recency Frequency Monetary*). Setelah proses RFM (*Recency Frequency Monetary*) kemudian data hasilnya akan di proses *clustering* menggunakan metode K-Means *Clustering*.

B. K-Means Clustering

Pada tahap *clustering* menggunakan algoritma K-Means, dataset akan dilihat terlebih dahulu variabel *outlier* dari nilai RFM. *Outlier* adalah titik data yang secara signifikan berbeda dari kumpulan data lainnya.



Gambar 2. Distribusi Data Outlier

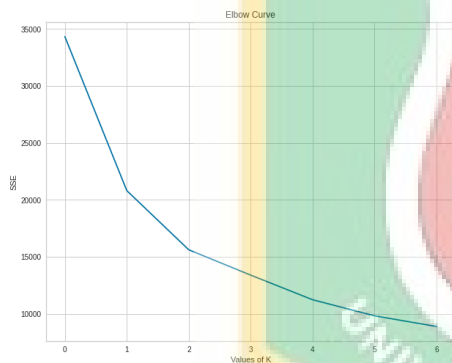
Setelah diketahui jumlah data outlier pada atribut *frequency* dan *monetary*, maka data *outlier* tersebut akan dihapus. Tahap selanjutnya adalah dilakukan *rescaling* terhadap atribut-atribut yang RFM. *Rescaling* atribut adalah salah satu tahap yang penting agar masing-masing atribut memiliki skala yang sebanding. Pada tahap *rescaling* menggunakan *library* StandarScaler.

Tabel 2. Rescaling RFM

Recency	Frequency	Monetary
1.925.160.428.24	5.986.364.625.74	11.454.179.622.3
1.490	9.000	85.900
19.089.121.265.1	5.501.995.057.58	0.5975836323010
67.900	6.860	582
		-
1.457.987.618.56	11.426.689.441.2	0.4905991813764
8.690	76.100	923
		-
14.466.725.197.1	19.499.515.577.3	0.8895995463915
21.700	11.800	941

C. Optimasi Jumlah Cluster

Optimasi jumlah *cluster* merupakan proses penting untuk menentukan berapa jumlah *cluster* terbaik dalam penelitian ini. Tiga metode untuk menentukan *cluster* paling optimal dalam penelitian ini menggunakan *Elbow Method*, *Silhouette Coefficient*, dan *Davies-Bouldin Index*.



Gambar 3. Hasil Elbow Curved

Pada gambar diatas menunjukkan bahwa jumlah *cluster* terbaik ditunjukkan dengan grafik presentase penurunan paling curam pada nilai *cluster* 2. Jadi untuk jumlah *cluster* terbaik dengan *elbow method* adalah dengan jumlah 2 *cluster*.

Tabel 3. Hasil Silhouette Score

Jumlah Cluster	Silhouette Score
2	0.6014345457538962
3	0.4795062977517168
4	0.48582949943797765
5	0.47608477887200473
6	0.4164932850705542
7	0.40748124769926497
8	0.4131848077156416

Dari pengujian menggunakan *silhouette score* pada tabel diatas menggunakan jumlah *cluster* 2 sampai 8 menunjukkan bahwa *cluster* 2 memiliki nilai *score* tertinggi dengan nilai 0.6014345457538962.

Tabel 4. Hasil Davies-Bouldin Score

Jumlah Cluster	Davies-Bouldin Score
2	0.8132733920961062
3	0.7500785223208264
4	0.811833607084641
5	0.8605487194764665
6	0.8516430173033278
7	0.9417882159850456
8	0.8902092017969905

Pada tabel diatas menunjukkan jumlah *cluster* dari 2 sampai 8 menggunakan *Davies-Bouldin Score*. Hasilnya menunjukkan jumlah 3 *cluster* memiliki *score* terendah yaitu 0.7500785223208264, sementara jumlah 2 *cluster* memiliki *score* 0.8132733920961062.

D. Hasil RFM dan K-Means Clustering

Berdasarkan 2 dari 3 metode optimasi menghasilkan jumlah 2 *cluster* adalah paling optimal, maka penelitian ini menggunakan jumlah 2 *cluster*. *Cluster* 1 memiliki 17.413 anggota, dan *cluster* 2 memiliki 2.068 anggota.

Tabel 5. Hasil RFM dan K-Means Clustering

Kode Pasien	Recency	Frequency	Monetary	Cluster Id
			14529500	
49740.0	4	42.0	0.0	2
47465.0	917	39.0	0.0	2
418420.0	347	12.0	0.0	2
41612.0	138	17.0	0.0	2
416837.0	72	20.0	0	2
46695.0	12	39.0	0	2
410282.0	130	32.0	0	2
...
423136.0	202	2.0	160000.0	1
422636.0	167	2.0	156000.0	1
420457.0	175	2.0	156000.0	1
422745.0	219	2.0	156000.0	1
422581.0	243	2.0	152000.0	1
422575.0	233	2.0	152000.0	1
422719.0	236	2.0	152000.0	1

Hasil nilai rata-rata setiap *cluster* adalah *cluster* 1 memiliki nilai rata-rata *recency* 641,63, *frequency* 3,21, dan *monetary* Rp. 2.424.251,98. Sedangkan *cluster* 2 memiliki nilai rata-rata *recency* 286,87, *frequency* 19,32, dan *monetary* Rp. 20.087.467,49.

IV. KESIMPULAN

Metode RFM (*Recency Frequency Monetary*) dapat membagi segmen pasien berdasarkan berapa lama terakhir kunjungan pasien, seberapa banyak frekuensi kunjungan pasien, dan total uang yang dikeluarkan pasien. Sedangkan K-Means *Clustering* dapat membagi hasil dari RFM (*Recency Frequency Monetary*) kedalam 2 *cluster*. Jumlah *cluster* optimal adalah menggunakan 2 *cluster* karena pada *elbow method* dan *silhouette coefficient* menunjukkan 2 *cluster*. Sedangkan hasil *davies-bouldin score* menunjukkan *cluster* optimal dengan 3 *cluster* tapi skornya 0.7500785223208264 masih jauh dari 0. Kelompok pasien yang menguntungkan adalah pasien dengan *frequency* lebih sering dan nilai *monetary* lebih tinggi. Sementara nilai *recency* yang tinggi menunjukkan pasien tersebut sudah lama tidak melakukan transaksi. *Cluster* 2 memiliki nilai *frequency* dan *monetary* lebih tinggi dibandingkan *cluster* 1, dan juga *cluster* 2 memiliki nilai rata-rata *recency* lebih rendah dibandingkan *cluster* 1. Dengan begitu *cluster* 2 merupakan kelompok pasien yang paling menguntungkan bagi Palapa Dentists

V. SARAN

Diperlukan penelitian selanjutnya menggunakan metode *clustering* lain untuk mengetahui metode mana yang lebih baik untuk menentukan segmentasi pasien di Palapa Dentists. Diperlukan juga penelitian tentang metode terbaik antara *elbow method*, *silhouette coefficient*, dan *davies-bouldin index* untuk optimasi jumlah *cluster* pada metode K-Means *Clustering*.

REFERENSI

- [1] I. Y. Musyawah and D. Idayanti, "Analisis Strategi Pemasaran Untuk Meningkatkan Penjualan Pada Usaha Ibu Bagas di Kecamatan Mamuju (online)," *Jurnal Ilmiah Ilmu Manajemen*, vol. 1, no. 1, pp. 2656–6265, 2022.
- [2] A. D. Savitri, F. Abdurrachman Bachtiar, and N. Y. Setiawan, "Segmentasi Pelanggan Menggunakan Metode K-Means Clustering Berdasarkan Model RFM Pada Klinik Kecantikan (Studi Kasus : Belle Crown Malang)," 2018. [Online]. Available: <http://j-ptiik.ub.ac.id>
- [3] J. Puspa Wildyaksanani and dan Dadang Sugiana, "Strategi Customer Relationship Management (CRM) PT Angkasa Pura II (Persero)," *Jurnal Kajian Komunikasi*, vol. 6, no. 1, pp. 10–23, 2018.
- [4] R. Nofitri and N. Irawati, "ANALISIS DATA HASIL KEUNTUNGAN MENGGUNAKAN SOFTWARE RAPIDMINER," *JURTEKSI (Jurnal Teknologi dan Sistem Informasi)*, vol. 5, no. 2, pp. 199–204, Jul. 2019, doi: 10.33330/jurtekxi.v5i2.365.
- [5] C. D. Rumiarti and I. Budi, "Customer Segmentation for Customer Relationship Management on Retail Company: Case Study PT Gramedia Asri Media," *Jurnal Sistem Informasi*, vol. 13, no. 1, p. 1, May 2017, doi: 10.21609/jsi.v13i1.525.
- [6] G. Gustientiedina, M. H. Adiya, and Y. Desnelita, "Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 5, no. 1, pp. 17–24, Apr. 2019, doi: 10.25077/teknosi.v5i1.2019.17-24.
- [7] S. Handoko, F. Fauziah, and E. T. E. Handayani, "IMPLEMENTASI DATA MINING UNTUK MENENTUKAN TINGKAT PENJUALAN PAKET DATA TELKOMSEL MENGGUNAKAN METODE K-MEANS CLUSTERING," *Jurnal Ilmiah Teknologi dan Rekayasa*, vol. 25, no. 1, pp. 76–88, 2020, doi: 10.35760/tr.2020.v25i1.2677.
- [8] S. Monalisa, "Segmentasi Perilaku Pembelian Pelanggan Berdasarkan Model RFM dengan Metode K-Means," *Jurnal Sistem Informasi*, p. 1, 2018.
- [9] P. Anitha and M. M. Patil, "RFM model for customer purchase behavior using K-Means algorithm," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 5, pp. 1785–1792, May 2022, doi: 10.1016/j.jksuci.2019.12.011.
- [10] A. Febriani and S. A. Putri, "Segmentasi Konsumen Berdasarkan Model Recency, Frequency, Monetary dengan Metode K-Means," *JIEMS (Journal of Industrial Engineering and Management Systems)*, vol. 13, no. 2, Sep. 2020, doi: 10.30813/jiems.v13i2.2274.
- [11] T. Ayu Rospricilia, S. Ayu Ithriah, and A. Anjani Arifiyanti, "SEGMENTASI PELANGGAN MENGGUNAKAN METODE K-MEANS CLUSTERING BERDASARKAN MODEL RFM PADA CV TITA JAYA," 2020.
- [12] R. Y. Firmansah, J. Dedy Irawan, and N. Vendyansyah, "ANALISIS RFM (RECENCY, FREQUENCY AND MONETARY) PRODUK MENGGUNAKAN METODE K-MEANS," 2021.
- [13] A. Sugiharto, B. N. Sari, and T. N. Padilah, "ANALISIS CLUSTER SEBARAN COVID-19 MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING (STUDI KASUS: PROVINSI JAWA BARAT)," *Journal of Information Technology and Computer Science (INTECOMS)*, vol. 4, no. 2, pp. 291–301, 2021, [Online]. Available: <https://pikobar.jabarprov.go.id/>.
- [14] H. Priyatman, F. Sajid, and D. Haldivany, "Klasterisasi Menggunakan Algoritma K-Means Clustering untuk Memprediksi Waktu Kelulusan Mahasiswa," *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, 2019.
- [15] M. M. Khairunnisa, A. Triayudi, and E. T. E. Handayani, "Application of K-Means Clustering on the Performance Evaluation of Lecturers Based on Student Questionnaire: Application of K-Means Clustering on the Performance Evaluation of Lecturers Based on Student Questionnaire," *Mantik*, vol. 4, no. 1, pp. 760-766, May 2020.
- [16] M. Shutaywi and N. N. Kachouie, "Silhouette analysis for performance evaluation in machine learning with

- applications to clustering.” *Entropy*, vol. 23, no. 6, Jun. 2021, doi: 10.3390/e23060759.
- [17] R. Gustriansyah, N. Suhandi, and F. Antony, “Clustering optimization in RFM analysis based on k-means,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 18, no. 1, pp. 470–477, 2019, doi: 10.11591/ijeecs.v18.i1.pp470-477.
- [18] A. Larasati, R. Maren, and R. Wulandari, “Utilizing Elbow Method for Text Clustering Optimization in Analyzing Social Media Marketing Content of Indonesian e-Commerce,” *Jurnal Teknik Industri*, vol. 23, no. 2, pp. 111–120, Dec. 2021, doi: 10.9744/jti.23.2.111-120.
- [19] S. Paembonan, H. Abduh, and K. Kunci, “Penerapan Metode Silhouette Coeficient Untuk Evaluasi Clustering Obat Clustering; K-means; Silhouette coeficient,” 2021. [Online]. Available: <https://ojs.unanda.ac.id/index.php/jiit/index>
- [20] E. Muningsih, I. Maryani, and V. R. Handayani, “Penerapan Metode K-Means dan Optimasi Jumlah Cluster dengan Index Davies Bouldin untuk Clustering Propinsi Berdasarkan Potensi Desa,” *Jurnal Sains dan Manajemen*, vol. 9, no. 1, 2021, [Online]. Available: www.bps.go.id



LETTER OF ACCEPTANCE (LoA)

No. 001/02.02/UNSERA/JSiI-LOA/I/2023

Dear Authors,

On the behalf of the Jurnal Sistem Informasi (JSiI)'s manager, we are pleased to inform that your paper, entitled:

“Analisis Recency Frequency Monetary dan K-Means Clustering pada Klinik Gigi untuk Menentukan Segmentasi Pasien”

Written by

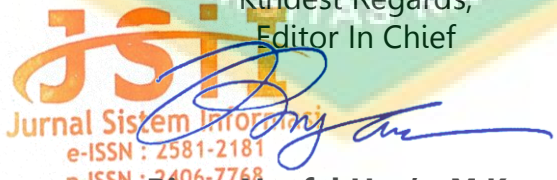
1. Aji Setiono
2. Agung Triayudi
3. Endah Tri Esti Handayani

Corresponding Author: Agung Triayudi

Has been **ACCEPTED** and will be proceed to be published in Jurnal Sistem Informasi (JSiI) Volume 10 Issue 1 March 2023.

We congratulate for your achievement. The technical issues about the publication will be informed later. Thank you very much for participating in our journal.

Kindest Regards,
Editor In Chief


Jurnal Sistem Informasi
e-ISSN : 2581-2181
p-ISSN : 2406-7768
Riyan Naufal Hay's, M.Kom.
NIK. 231208340