

# ANALISIS *RECENCY FREQUENCY MONETARY* DAN *K-MEANS CLUSTERING* PADA KLINIK GIGI UNTUK MENENTUKAN SEGMENTASI PASIEN

Aji Setiono<sup>1</sup>, Agung Triayudi<sup>2\*</sup>, Endah Tri Esti Handayani<sup>3</sup>

<sup>1,2,3</sup> Jurusan Informatika Fakultas Teknologi Komunikasi dan Informatika Universitas Nasional  
Jl. Sawo Manila Pejaten Pasar Minggu – Jakarta Selatan

<sup>1</sup>aji.setiono16@gmail.com

<sup>2</sup>agungtriayudi@civitas.unas.ac.id

<sup>3</sup>endahtriesti@civitas.unas.ac.id

## Abstrak

Dengan semakin berkembangnya persaingan bisnis, 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*) adalah kombinasi proses dan teknologi dalam memahami pelanggan [2]. CRM (*Customer Relationship Management*) memiliki tujuan 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

\*Corresponding author: Agung Triayudi

merupakan metode untuk mendapatkan pengetahuan dari sebuah data [6]. Proses data *mining* meliputi pengumpulan data, dan pemakaian data perusahaan untuk mendapatkan 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* adalah teknik dalam data *mining* dan bersifat *unsupervised* yaitu berarti bahwa atribut-atribut dari suatu kelompok atau *cluster* yang memiliki persamaan karakteristik yang dapat dikelompokkan dalam satu kelompok atau *cluster* [12], [13]. *Clustering* juga dapat diartikan sebagai sebuah proses pembentukan kelompok atau *cluster* data dari kelompok data yang tidak diketahui kelompoknya, kemudian mengelompokkan data tersebut ke dalam satu *cluster* yang memiliki kesamaan data [14], [15]. Salah satu metode *clustering* paling terkenal dengan kesederhanaan dan kecepatannya adalah K-Means *Clustering* [16]. K-Means *Clustering* adalah metode mengelompokkan data yang terkenal dengan kesederhanaan algoritma, dan kecepatan memilih pusat *cluster* [17]. Simbol K pada K-means *clustering* memiliki arti sebagai jumlah *cluster* yang digunakan.

Savitri, Abdurrachman Bachtiar, & Setiawan, 2018 dalam penelitiannya menggunakan model RFM dan metode K-Means *Clustering* dengan menggunakan data riwayat transaksi pelanggan. Pada penelitian tersebut didapatkan 2 dan 3 segmen, dan hasil RFM menunjukkan pelanggan yang paling menguntungkan masuk dalam peringkat pertama [2]. Fithriyah, Yaqin, & Zaman, 2021 dalam penelitiannya yang bertujuan melakukan segmentasi produk guna meminimalisir kekurangan atau kelebihan stok [17]. Randi Rian dan Cendra Wadisman, 2018 dalam penelitiannya menggunakan algoritma K-Means untuk yang diimplementasikan dalam pemilihan pelanggan potensial di MC Laundry [18]. Elly Muningsih, Ina Maryani, dan Vembria Rose Handayani, 2021 dalam penelitiannya menggunakan metode K-Means untuk mengelompokkan propinsi berdasar potensi desa dengan banyaknya jenis industri yang dimiliki wilayahnya, dan menggunakan *davies-bouldin index* sebagai penentu jumlah *clusternya* [19]. Andre Winarta, dan Wahyu Joni Kurniawan, 2021 dalam penelitiannya menggunakan metode K-Means dan *elbow method* untuk menentukan *cluster* optimal dalam *clustering* penyebaran pengguna narkoba [20].

Penelitian ini menggunakan dua metode yaitu 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 score* yang mendekati nilai 1 berarti pengelompokan data *cluster* itu semakin baik. Jika koefisien *silhouette score* mendekati nilai -1, maka pengelompokan data dalam satu *cluster* akan semakin tidak baik atau buruk. *Davies-Bouldin Index* adalah salah satu metode untuk mengukur jumlah *cluster* paling baik atau optimal pada suatu *cluster* dimana kohesi diartikan sebagai jumlah dari kedekatan data terhadap titik pusat *cluster* dari *cluster* yang diikuti [20],[21]. Tujuan penelitian ini untuk menentukan segmentasi pasien dan memilih kelompok pasien mana yang paling menguntungkan bagi perusahaan.

## II. METODOLOGI PENELITIAN

Metodologi pada penelitian ini menggunakan CRISP-DM atau disebut *The Cross Industry Standard Process for Data Mining* adalah salah satu standard dalam data mining yang terpopuler.



Gambar 1. Tahapan Penelitian (Sumber: [www.djkn.kemenkeu.go.id](http://www.djkn.kemenkeu.go.id))

### 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

$C_i$  : inisialisasi *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 gratis 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*)

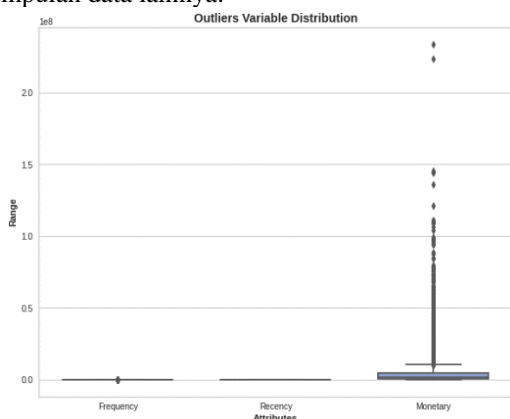
Kode Pasien	Recency	Frequency	Monetary
40681.0	62	75.0	233723125.0
412991.0	39	52.0	223585000.0
49740.0	4	42.0	145295000.0
47465.0	917	39.0	144105000.0
410053.0	587	58.0	144046500.0
48839.0	12	61.0	136178875.0
47666.0	71	82.0	121147000.0

....	...	...	...
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

Data awal sebelum dilakukan penanganan *outlier* berjumlah 19.595, Terdapat data *outlier* pada *frequency* dan *monetary* sebanyak 114 data, maka data itu akan dihapus. Data bersih tanpa *outlier* adalah sebanyak 19.481 data. Tahap selanjutnya adalah dilakukan *rescaling* terhadap atribut-atribut yang RFM. *Rescaling* bertujuan agar masing-masing atribut memiliki skala yang sebanding. Pada tahap *rescaling* menggunakan *library StandarScaler*.

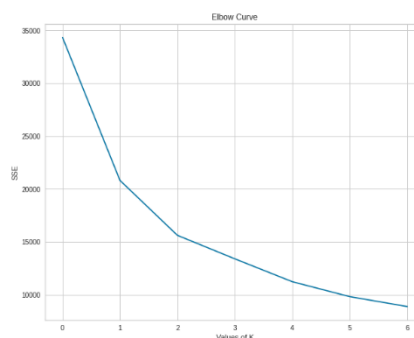
Tabel 2. Rescaling RFM

Recency	Frequency	Monetary
1.925.160.428.24	5.986.364.625.74	-
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	-
8.690	76.100	923
14.466.725.197.1	19.499.515.577.3	-
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
			14410500	
47465.0	917	39.0	0.0	2
			11108000	
418420.0	347	12.0	0.0	2
			11025130	
41612.0	138	17.0	0.0	2
			99246000.	
416837.0	72	20.0	0	2
			98022000.	
46695.0	12	39.0	0	2
			96020000.	
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.

Dari hasil *cluster* yang terbagi menjadi 2 *cluster*, *cluster* 1 adalah segmen pasien-pasien dengan *recency* atau terakhir kunjungan lebih lama dibandingkan *cluster* 2, nilai *frequency* atau seringnya berkunjung pada pasien *cluster* 1 juga lebih sedikit dibandingkan *cluster* 2, dan juga nilai *monetary* atau

uang yang dikeluarkan pasien pada *cluster* 1 lebih kecil dibandingkan *cluster* 2. Dengan hasil itu maka pada kelompok pasien di *cluster* 1 perlu promosi atau *recall* untuk membuat pasien tersebut lebih sering melakukan kunjungan ke Klinik. Sementara pada *cluster* 2 Klinik perlu terus meningkatkan pelayanan kepuasan agar pasien tersebut tetap atau meningkat kunjungannya.

#### 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. *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. Kekurangan penelitian ini adalah *cluster* yang terbagi hanya menjadi 2 *cluster*. 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. Musyawarah 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 Wildyaksanjani 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/jurteksi.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. 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.
- [16] 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.
- [17] M. Fithriyah, M. A. Yaqin, and S. Zaman, "K-Means Clustering Untuk Segmentasi Produk Berdasarkan Analisis Recency, Frequency, Monetary (RFM) Pada Data Transaksi Penjualan," *ILKOMNIKA: Journal of Computer Science and Applied Informatics*, vol. 3, no. 2, pp. 151–164, Aug. 2021, doi: 10.28926/ilkomnika.v3i2.284.
- [18] R. R. Putra and C. Wadisman, "Implementasi Data Mining Pemilihan Pelanggan Potensial Menggunakan Algoritma K-Means," *Journal of Information Technology and Computer Science (INTECOMS)*, vol. 1, no. 1, 2018.
- [19] 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](http://www.bps.go.id)
- [20] A. Winarta and W. J. Kurniawan, "Optimasi Cluster K-Means Menggunakan Metode Elbow Pada Data Pengguna Narkoba Dengan Pemrograman Python," *Jurnal Teknik Informatika Kaputama (JTik)*, vol. 5, no. 1, 2021.
- [21] 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.