

Segmentasi Pelanggan Berdasarkan Analisis *Recency*, *Frequency*, *Monetary* Menggunakan Algoritma K-Means Pada CV. Toedjoe Sinar Group

Sri Wahyuni*¹, Tina Tri Wulansari², Fahrullah³

^{1,2}Jl. Pahlawan No.2A Samarinda Kalimantan Timur, Telp. 0541-741864 | Fax. 0541-735984

³Program Studi Sistem Informasi, Universitas Mulia PSDKU Samarinda, Indonesia

e-mail: *¹wahyuni.sri@students.universitasmulia.ac.id, ²tina@universitasmulia.ac.id,

³fahrullah@universitasmulia.ac.id

Abstrak

CV. Toedjoe Sinar Group merupakan salah satu percetakan di Samarinda yang bergerak di bidang digital printing dengan beragam produk. Percetakan ini tidak hanya fokus pada strategi yang mengutamakan produk, tetapi juga melakukan strategi yang mengutamakan pelanggan. Hal ini diperlukan untuk mengetahui perilaku pelanggan sehingga akan membantu dalam penerapan strategi pemasaran yang tepat untuk meningkatkan pendapatan perusahaan. Sebanyak 8.020 data transaksi pelanggan dari Bulan Januari hingga Maret tahun 2022 digunakan dalam penelitian ini. Pemodelan data yang digunakan dalam pengelompokan pelanggan adalah *Recency*, *Frequency*, *Monetary* (RFM) dan metode clustering K-Means melalui teknik data mining. Penentuan jumlah cluster berdasarkan nilai k menggunakan metode elbow. Penelitian ini menghasilkan nilai $k=7$ sehingga cluster yang terbentuk sebanyak 7 cluster. Kelompok terbaik yaitu cluster 5 dengan nilai rata-rata *Recency* sebesar 0,0, *Frequency* sebesar 140,33 serta rata-rata *monetary* sebesar 4.3328.833.

Kata kunci—Algoritma K-Means, RFM, Segmentasi Pelanggan

1. PENDAHULUAN

CV. Toedjoe Sinar Group merupakan salah satu percetakan yang ada di Samarinda dan bergerak di bidang digital printing dengan berbagai macam produk seperti banner, baliho, stiker, poster, kartu nama, brosur, id card, stempel dan lain-lain. Saat ini, CV. Toedjoe Sinar Group tidak hanya fokus pada strategi yang mengutamakan produk, tetapi juga melakukan strategi yang mengutamakan pelanggan. Perusahaan ini menyadari akan pentingnya hubungan antara pelanggan yang loyal dengan keberhasilan bisnis. Hal ini tentu penting dilakukan untuk perkembangan perusahaan dengan menciptakan hubungan baik antara perusahaan dan pelanggan. Dalam situasi ekonomi yang semakin berubah membuat persaingan industri percetakan semakin ketat dan ditandai dengan munculnya perusahaan percetakan baru.

Segmentasi pelanggan adalah suatu cara untuk mengelompokkan pelanggan ke dalam beberapa cluster (kelompok) dan setiap *cluster* memiliki beberapa anggota dengan karakteristik yang sama [1]. Segmentasi pelanggan melakukan pengelompokan karakteristik, kebutuhan, dan minat konsumen/pelanggan yang sama ke dalam satu kelompok. Tujuannya untuk menerapkan model bisnis yang mampu memenuhi setiap kebutuhan dari segmen-segmen tersebut agar strategi pemasaran dari perusahaan lebih tepat sasaran dan efisien. Dengan dilakukannya segmentasi, pelaku usaha dapat memahami bahwa pelanggan pun memiliki karakteristik khusus saat berinteraksi dengan produk [2]. Model *Recency*, *Frequency*, *Monetary* (RFM) merupakan model untuk menentukan segmentasi konsumen berdasarkan kapan transaksi terakhir dilakukan,

frequency yaitu tingkat keseringan pelanggan bertransaksi dan *monetary* ialah besarnya nilai transaksi yang dilakukan [3]. Analisa RFM adalah proses menganalisis perilaku pelanggan. Hal ini umumnya digunakan dalam pemasaran database dan pemasaran langsung. Analisis RFM merupakan suatu model perhitungan yang terdiri dari tiga atribut domain yang memperhatikan transaksi pelanggan berdasarkan transaksi terakhir/keterkinian (*Recency*), jumlah transaksi (*Frequency*), serta nominal dari transaksi (*Monetary*) [4]. Tujuan dari RFM adalah untuk meramalkan perilaku pelanggan di masa depan agar dapat mengarahkan keputusan segmentasi yang lebih baik. Model RFM ini merupakan metode yang sudah lama dan populer untuk mengukur hubungan dengan pelanggan [5].

Penggunaan teknik data mining merupakan salah satu solusi untuk persoalan segmentasi. Data mining merupakan gabungan sejumlah disiplin ilmu komputer yang didefinisikan sebagai proses penemuan pola-pola baru dari kumpulan-kumpulan data sangat besar, meliputi metode-metode yang merupakan irisan dari *artificial intelligence*, *machine learning*, *statistics*, dan *database systems* [6]. Data mining merupakan bidang ilmu yang digunakan untuk menangani masalah pengambilan informasi dari *database* yang besar dengan menggabungkan teknik dari statistik, pembelajaran mesin, visualisasi data, pengenalan pola, dan *database* [7].

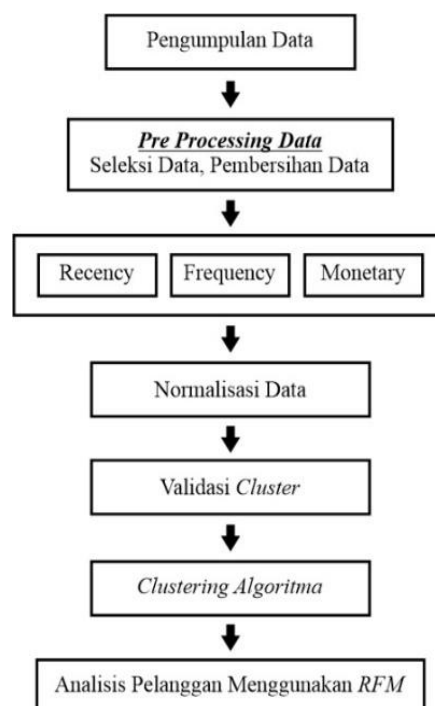
2. METODE PENELITIAN

2.1. Objek Penelitian

Penelitian ini dilakukan di CV. Toedjoe Sinar Group yang beralamat di Jalan K.H. Wahid Hasyim I Nomor 40, Sempaja Selatan, Kecamatan Samarinda Utara, Kota Samarinda, Kalimantan Timur. Penelitian ini dilaksanakan selama 3 bulan yaitu pada bulan Januari 2022 sampai dengan bulan Maret 2022. Scope penelitian pada skripsi ini adalah pelanggan Toedjoe Sinar Group berdasarkan kelompok transaksi pelanggan.

2.2. Kerangka Pikir

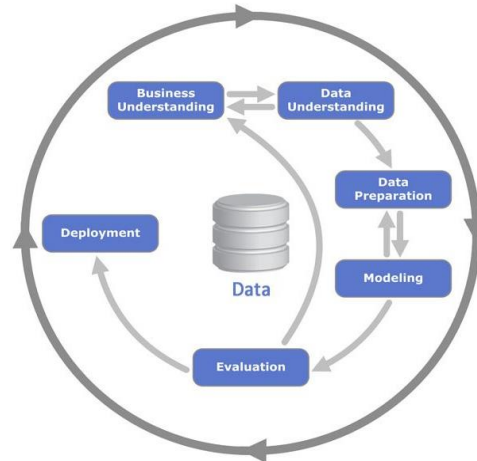
Kerangka pikir adalah tahapan yang harus dilakukan dalam melakukan penelitian ini, diawali dari langkah pertama menentukan business understanding sampai dengan menghasilkan pengetahuan yang tersaji pada Gambar 1.



Gambar 1. Kerangka Pikir Penelitian

2.3. Tahapan Penelitian

Alur penelitian penerapan algoritma K-Means ini dengan metode analisis data menggunakan *Cross-Industry Standard Process for Data Mining (CRISP-DM)* dalam segmentasi pelanggan berdasarkan analisis *Recency, Frequency, Monetary* menggunakan algoritma K-Means yang dapat dilihat pada Gambar 2.



Gambar 2. *Crisp-DM* Tahapan Penelitian

Tahapan penelitian untuk segmentasi pelanggan menggunakan algoritma K-Means pada Gambar 2 sebagai berikut:

1) *Business Understanding*

Pada fase ini berfokus pada pemahaman dan perspektif bisnis proses dari suatu sistem atau penelitian. Yaitu penentuan tujuan penelitian, menentukan tujuan bisnis, dan menilai situasi, dan menentukan tujuan data mining

2) *Data Understanding*

Pada tahap ini yaitu berfokus pada pembelajaran data yang sudah ada, pengumpulan dan penyeleksian data.

3) *Data Preparation*

Pada fase persiapan data terdiri dari pemilihan data, pembersihan data, mengintegrasikan data, dan transformasi data agar dapat dilanjutkan kedalam tahap pemodelan. Oleh sebab itu, dibutuhkan teknik data preprocessing untuk mempersiapkan data yang akan diproses.

4) *Modeling*

fase ini proses yang terjadi adalah pemilihan model yang sesuai agar mengoptimalkan hasil. Model clustering dimulai dari *dataset* akan dilakukan pemodelan dengan algoritma K-Means sehingga dihasilkan model clustering dan memunculkan parameter evaluasi

5) *Evaluation*

Pada fase ini akan dilakukan proses evaluasi dari phase sebelumnya. Fase evaluasi ini akan dilakukan uji validasi dengan *Sum of Square Error (SEE)*. *SEE* digunakan untuk menganalisis variasi pada jarak antar pusat kluster yang terbentuk

6) *Deployment*

Pada fase ini proses yang terjadi adalah penyusunan laporan atau presentasi dari pengetahuan yang didapat dari analisis masing-masing atribut RFM.

2.4. Metode Pengambilan Data

penelitian skripsi pada CV. Toedjoe Sinar Group diperlukan metode pengumpulan data yang merupakan salah satu aspek yang berperan dalam kelancaran dan keberhasilan dalam suatu penelitian. Dalam penelitian ini metode pengumpulan data yang digunakan sebagai berikut:

- a. Wawancara

Wawancara adalah teknik pengambilan data yang dilakukan untuk kebutuhan penelitian, wawancara dilakukan dengan narasumber yaitu Bapak Naufal selaku pimpinan CV. Toedjoe Sinar Group.

b. Data *Primer*

Data *Primer* diperoleh dari database aplikasi di CV. Toedjoe Sinar Group yang berupa data *excel*. Data transaksi pelanggan yang diambil selama 3 Bulan.

2.5. K-Means Clustering

Clustering adalah teknik menemukan sekelompok data dari pemecahan atau pemisahan sekumpulan data menurut karakteristik tertentu yang telah ditentukan. Dalam pengelompokan tersebut nilai label nya belum diketahui label dari data tersebut. Metode *clustering* juga sering disebut tahapan awal sebelum melakukan metode lain seperti klasifikasi [8]. *Clustering* merupakan kumpulan objek data yang memiliki kemiripan antara satu dengan yang lain dalam kelompok. *Clustering* atau lebih dikenal dengan analisis *cluster* merupakan proses pengelompokkan satu set benda fisik ataupun abstrak kedalam satu kelas objek yang sama [9]. Semakin kecil jumlah *cluster* yang digunakan, maka akan mempermudah pemahaman terhadap struktur data di dalamnya, tetapi pola informasi penting yang ada akan terabaikan. Sebaliknya dengan memberikan jumlah cluster akhir yang lebih besar, maka kesamaan yang dimiliki antar kelompok akan semakin meningkat pula dan mengabaikan struktur data yang ada.

Salah satu metode *clustering* yang sangat populer dan banyak dipelajari untuk meminimalkan kesalahan *clustering* untuk titik ruang *Euclidean* disebut K-Means clustering. K-Means adalah metode *clustering* berbasis jarak yang membagi data ke dalam sejumlah *cluster* dan algoritma ini hanya bekerja pada atribut numerik [10], [11]. Algoritma K-Means partitioning clustering yang memisahkan data ke K daerah yang terpisah. Algoritma K-Means sangat terkenal karena kemudahan dan kemampuannya untuk mengklaster data yang besar dan data yang outlir dengan sangat cepat. dalam algoritma K-Means, setiap data harus termasuk ke *cluster* tertentu dan bisa dimungkinkan bagi setiap data yang termasuk *cluster* tertentu pada suatu tahapan proses, pada tahapan berikutnya berpindah ke *cluster* lainnya [12]. Langkah-langkah dari algoritma K-Means sebagai berikut [9], [13]:

1. Tentukan berapa banyak jumlah cluster(c)
2. Tentukan secara acak pusat cluster awal (centroid)
3. Ukur jarak setiap data dengan pusat cluster(centroid) yaitu dengan rumus Euclidean Distance pada persamaan (1)

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

Keterangan :

x_i : objek x ke-i

y_i : titik pusat y ke-i

n : banyaknya objek

4. Kelompokkan setiap data ke dalam cluster berdasarkan jarak minimum
5. Lakukan proses iterasi dengan menentukan pusat klaser (centroid) baru menggunakan persamaan (2)

$$C_k = \frac{1}{N_k} (X_{1k} + X_{2k} + \dots + X_{nk}) \quad (2)$$

Keterangan :

C_k = Rata-rata baru

N_k = Jumlah training pattern pada cluster (k)

X_{nk} = Pola ke (n) yang menjadi bagian cluster (k).

- Ulangi langkah 3 hingga tidak terdapat perubahan cluster pada setiap data dari proses iterasi sebelumnya

2.6. Metode Elbow

Metode Elbow adalah untuk menghasilkan informasi dengan cara melihat perbandingan hasil antara jumlah *cluster* yang akan membentuk suatu titik terakhir, *cluster* nilai akan digunakan sebagai model data untuk *cluster* terbaik. Selain itu, hasil perhitungan akan digunakan untuk membandingkan jumlah *cluster*. Persamaan (3) Rumus yang digunakan untuk perhitungan SSE [14]:

$$SSE = \sum_{k=1}^k \sum_{x_i \in S_k} \|X_i - C_k\|^2 \quad (3)$$

Keterangan:

- X_i : nilai atribut dari data ke- i
 C_k : nilai atribut titik pusat *Cluster* ke- i

Metode Elbow untuk menentukan jumlah *cluster* yang paling optimum atau yang terbaik. Langkah-langkah metode Elbow [15], [16]:

- Awal nilai cluster ($k=2$)
- Menaikkan nilai *cluster* sampai jumlah *cluster* ($k=4$)
- Menghitung nilai SSE dari setiap *cluster*
- Melihat hasil SSE dari nilai *cluster* yang turun secara drastis atau nilai yang berubah signifikan.
- Menetapkan nilai *cluster* yang berbentuk siku.

3. HASIL DAN PEMBAHASAN

3.1. Pengumpulan Data

Objek dalam penelitian ini adalah pelanggan CV. Toedjoe sinar group, data yang diperoleh dari database aplikasi CV. Toedjoe Sinar Group sebanyak 8020 data transaksi pelanggan dimulai dari bulan Januari 2022 sampai dengan bulan Maret 2022.

3.2. Tampilan Data Model Recency, Frequency, Monetary

Setelah memperoleh masing-masing nilai *recency*, *frequency*, dan *monetary*. Selanjutnya menggabungkan tabel *recency*, *frequency*, dan *monetary* menjadi satu tampilan atau satu tabel seperti pada Gambar 3 berikut.

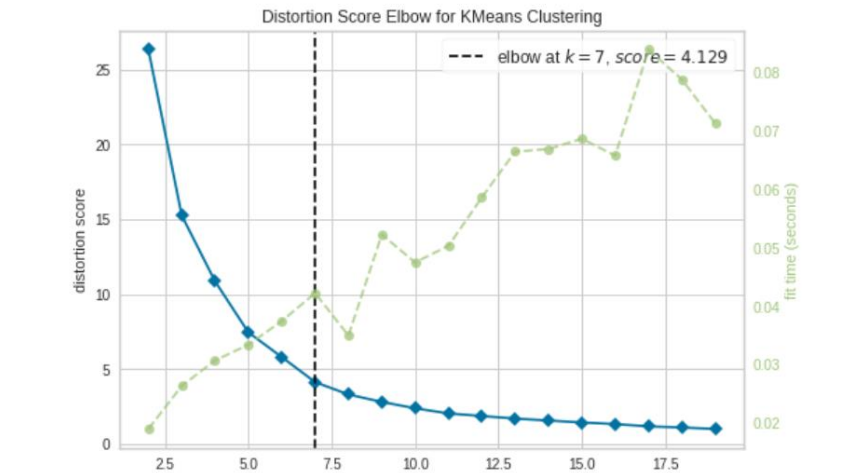
	Recency	Frequency	Monetary
id_pelanggan			
7	0	26	11524000
9	1	63	17307000
10	0	187	47883500
11	0	31	6544000
12	34	1	150000

Gambar 3. Tampilan Data Model RFM

Jumlah data pada hasil pembersihan data sebanyak 7895 data, dan dilakukan pemodelan RFM sehingga data berubah menjadi 843 data, dan data ini menjadi *dataset* yang akan digunakan dalam proses selanjutnya yaitu proses *clustering* ke algoritma K-Means.

3.3. Implementasi Algoritma K-Means

Penentuan jumlah *cluster* atau mencari nilai K dengan menggunakan metode *elbow*.



Gambar 4. Metode *Elbow* Dalam Penentuan Jumlah *Cluster*

Dapat dilihat pada Gambar 4 diatas, tampilan *grafik* untuk mencari nilai K dengan menggunakan metode *elbow*, dan hasilnya menunjukkan nilai K=7. Dengan demikian *cluster* yang akan digunakan adalah sebanyak 7 *cluster*.

3.4. Perbandingan Hasil Cluster Analisis Pelanggan

Setelah analisis pelanggan dari *cluster* 1 hingga 7 telah dilakukan, selanjutnya dilakukan perbandingan hasil *cluster* analisis pelanggan dengan menampilkan nilai RFM semua *cluster* yang dapat dilihat pada Tabel 1 berikut.

Tabel 1. Nilai RFM Setiap *cluster*

	<i>Recency</i>	<i>Frequency</i>	<i>Monetary</i>
<i>CLUSTER 1</i>	42,79	1,76	484533
<i>CLUSTER 2</i>	5,64	6,16	1266482
<i>CLUSTER 3</i>	76,42	1,16	225955
<i>CLUSTER 4</i>	22,43	2,28	456420
<i>CLUSTER 5</i>	0,0	140,33	43328833
<i>CLUSTER 6</i>	59,85	1,41	298318
<i>CLUSTER 7</i>	0,96	45,24	10510712

Hasil nilai RFM pada semua *cluster* terlihat *cluster* dengan nilai terbaik yaitu pada *cluster* 5, *cluster* terbaik kedua yaitu pada *cluster* 7 dan *cluster* dengan nilai RFM terendah yaitu *cluster* 3.

4. KESIMPULAN

Berdasarkan analisis yang dilakukan didapatkan kesimpulan bahwa segmentasi pelanggan pada CV. Toedjoe Sinar Group menghasilkan nilai K=7 dan *cluster* terbaik = 5 dengan nilai rata-rata *recency* 0,0, *frequency* 140,33 dan *monetary* 43328833

5. SARAN

Berdasarkan kesimpulan yang telah diuraikan, maka saran yang dapat disampaikan yaitu pada penelitian selanjutnya diharapkan dapat menambahkan atribut produk untuk mengidentifikasi dan mencari karakteristik pembeda agar setiap kelompok dapat diidentifikasi berdasarkan dengan produk yang paling laris. Selain itu, pada penelitian selanjutnya diharapkan mampu melakukan komparasi untuk mencari metode *clustering* terbaik dengan menggunakan algoritma *clustering* lainnya seperti *K-Medoids* dan *Hierarchical Clustering*.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Bapak Muhammad Yani, S.Kom., M.T.I. selaku Kepala Kantor Universitas Mulia Samarinda Program Studi di luar kampus Utama Samarinda serta pihak Toedjoe Sinar Group yang telah memberikan dukungan fasilitas dan ketersediaan data selama penelitian ini berlangsung.

DAFTAR PUSTAKA

- [1] N. H. Harani and F. A. Nugraha, *Segmentasi Pelanggan Menggunakan Python*. Kreatif, 2020.
- [2] R. P. Justitia, N. Hidayat, and E. Santoso, "Implementasi Metode Agglomerative Hierarchical Clustering Pada Segmentasi Pelanggan Barbershop (Studi Kasus: RichDjoe Barbershop Malang)," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 5, no. 3, pp. 1048–1054, 2021.
- [3] F. Hadi, M. Mustakim, D. O. Rahmadia, F. H. Nugraha, N. P. Bulan, and S. Monalisa, "Penerapan K-Means Clustering Berdasarkan RFM Mofek Sebagai Pemetaan dan Pendukung Strategi Pengelolaan Pelanggan (Studi Kasus: PT. Herbal Penawar Alwahidah Indonesia Pekanbaru)," *SITEKIN: Jurnal Sains, Teknologi dan Industri*, vol. 15, no. 1, pp. 69–76, 2017.
- [4] P. I. Pangestu, T. I. Hermanto, and D. Irmayanti, "Analisis Segmentasi Pelanggan Berbasis Model Recency Frequency Dan Monetary (RFM) Menggunakan Algoritma K-Means," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 11, no. 3s1, pp. 930–937, 2023.
- [5] B. E. Adiana, I. Soesanti, and A. E. Permanasari, "Analisis segmentasi pelanggan menggunakan kombinasi RFM model dan teknik clustering," *Jurnal Terapan Teknologi Informasi*, vol. 2, no. 1, pp. 23–32, 2018.
- [6] Suyanto, *Data Mining Untuk Klarifikasi dan Klasterisasi Data*, 1st ed. Bandung: INFORMATIKA, 2017.
- [7] I. Werdiningsih, M. Kom, B. Nuqoba, M. Kom, and S. S. Muhammadun, *Data Mining Menggunakan Android, Weka, dan SPSS*. Airlangga University Press, 2020.
- [8] V. Virtusena, A. Johar, and A. Wijanarko, "Pengelompokan Potensi Kelulusan Mahasiswa Fakultas Teknik Unib Menggunakan Algoritme K-Means (Studi Kasus: Fakultas Teknik Universitas Bengkulu)," *Rekursif: Jurnal Informatika*, vol. 9, no. 2, pp. 206–225, 2021.
- [9] E. Irfiani and S. S. Rani, "Algoritma K-Means Clustering untuk Menentukan Nilai Gizi Balita," *J. Sist. dan Teknol. Inf*, vol. 6, no. 4, p. 161, 2018.

- [10] M. Yoalifa, H. Havaluddin, M. Wati, N. Puspitasari, and U. Hairah, “Analisa Mutu Sekolah Pada Provinsi Kalimantan Timur Menggunakan Algoritma K-Means,” *Sains, Aplikasi, Komputasi dan Teknologi Informasi*, vol. 3, no. 2, pp. 53–60, 2023.
- [11] M. Wati, M. S. Noorlah, A. Tejawati, A. Septiarini, M. Jamil, and N. Puspitasari, “Implementation of The K-means Clustering for The Public Health Center Data,” in *2022 International Conference on Electrical Engineering, Computer and Information Technology (ICEECIT)*, IEEE, 2022, pp. 65–69.
- [12] M. Wahyudi, M. Masitha, R. Saragih, and S. Solikhun, *Data Mining: Penerapan Algoritma K-Means Clustering dan K-Medoids Clustering*. Yayasan Kita Menulis, 2020.
- [13] L. ‘Izzah1 and A. Jananto, “Penerapan Algoritma K-Means Clustering Untuk Perencanaan Kebutuhan Obat Di Klinik Citra Medika,” *Progresif: Jurnal Ilmiah Komputer*, vol. 18, no. 1, pp. 69–76, 2022.
- [14] A. P. Riani, A. Voutama, and T. Ridwan, “Penerapan K-Means Clustering Dalam Pengelompokan Hasil Belajar Peserta Didik Dengan Metode Elbow,” *Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD*, vol. 6, no. 1, pp. 164–172, 2023.
- [15] N. P. E. Merliana and A. J. Santoso, “Analisa Penentuan Jumlah Cluster Terbaik pada Metode K-Means Clustering,” 2015.
- [16] M. Al Ghifari and W. T. H. Putri, “Clustering Courses Based On Student Grades Using K-Means Algorithm With Elbow Method For Centroid Determination,” *Inform: Jurnal Ilmiah Bidang Teknologi Informasi dan Komunikasi*, vol. 8, no. 1, pp. 42–46, 2023.