

Analisis Segmentasi Frekuensi Transaksi ATM dengan K-Means Clustering pada Bank BJB Kantor Cabang Tasikmalaya

Transaction Frequency Segmentation Analysis Using K-Means Clustering at Bank BJB Tasikmalaya Branch Office

Robi Mustakim^{1*}, Dede Ruhimat², Satia Suhada³, Resti Yulistria⁴

Program Studi Sistem Informasi PSDKU Kota Sukabumi, Universitas Bina Sarana Informatika, Jakarta, Indonesia

*E-mail: robimustakimtobsi@gmail.com

Abstrak

Segmentasi nasabah merupakan salah satu strategi penting yang dapat membantu bank dalam memahami perilaku nasabah dan menyusun strategi pelayanan yang lebih efektif. Untuk mencapai tujuan tersebut, metode K-Means Clustering digunakan dalam penelitian ini guna mengelompokkan data transaksi ATM Bank BJB Kantor Cabang Tasikmalaya ke dalam empat klaster dengan karakteristik yang berbeda. Hasil analisis menunjukkan bahwa Cluster 0 memiliki 33 anggota, Cluster 1 terdiri dari 94 anggota, Cluster 2 berisi 136 anggota, dan Cluster 3 mencakup 171 anggota. Setiap klaster menunjukkan perbedaan signifikan dalam frekuensi transaksi, di mana Cluster 3 memiliki jumlah anggota terbesar. Evaluasi dan validasi clustering dilakukan menggunakan metode Elbow dan Silhouette Score, yang menunjukkan segmentasi berkualitas. Visualisasi dengan PCA digunakan untuk membantu interpretasi distribusi dan karakteristik setiap klaster. Hasil penelitian ini memberikan wawasan penting bagi Bank BJB dalam mengembangkan strategi pemasaran yang lebih tepat sasaran, program loyalitas, dan layanan yang lebih personal sesuai dengan karakteristik masing-masing klaster nasabah. Dengan penerapan K-Means Clustering, Bank BJB dapat meningkatkan efisiensi operasional dan kepuasan nasabah, serta memperkuat daya saing di era digital.

Kata kunci: Transaksi ATM, Nasabah, K-Means, Klasterisasi, Segmentasi.

Abstract

Customer segmentation is one of the key strategies that can help banks understand customer behavior and develop more effective service strategies. To achieve this goal, the K-Means Clustering method is used in this study to group ATM transaction data from Bank BJB Tasikmalaya Branch into four clusters with different characteristics. The analysis results show that Cluster 0 has 33 members, Cluster 1 consists of 94 members, Cluster 2 contains 136 members, and Cluster 3 includes 171 members. Each cluster shows significant differences in transaction frequency, with Cluster 3 having the largest number of members. Clustering evaluation and validation were carried out using the Elbow and Silhouette Score methods, which indicate high-quality segmentation. Visualization with PCA is used to help interpret the distribution and characteristics of each cluster. This research provides important insights for Bank BJB in developing more targeted marketing strategies, loyalty programs, and more personalized services according to the characteristics of each customer cluster. With the implementation of K-Means Clustering, Bank BJB can improve operational efficiency and customer satisfaction, as well as strengthen its competitiveness in the digital era.

Keywords: ATM Transactions, Customer, K-Means, Clustering, Segmentations.

Naskah diterima 02 Ags. 2024; direvisi 26 Sept. 2024; dipublikasikan 01 Apr. 2025.

JAMIKA is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.



I. PENDAHULUAN

Dalam era digital seperti saat ini, perbankan semakin mengandalkan teknologi untuk meningkatkan efisiensi dan pelayanan kepada nasabah [1]. Salah satu aspek teknologi yang sangat penting dalam operasional perbankan adalah pengelolaan dan analisis data transaksi nasabah, terutama transaksi ATM. Analisis ini tidak hanya membantu bank dalam memahami perilaku penggunaan layanan ATM, tetapi juga menjadi landasan untuk menyusun strategi pelayanan yang lebih baik dan personalisasi pengalaman nasabah [2].

Frekuensi transaksi ATM dapat memberikan wawasan yang berharga kepada bank mengenai kebiasaan penggunaan layanan perbankan oleh nasabah [3]. Dengan menganalisis pola transaksi ini, bank dapat mengidentifikasi segmentasi nasabah berdasarkan intensitas penggunaan ATM, memahami preferensi transaksi yang paling umum, serta mengeksplorasi tren atau pola yang mungkin tersembunyi di dalam data [4].

Salah satu metode data mining yang efektif untuk menganalisis segmentasi berdasarkan frekuensi transaksi ATM adalah dengan menggunakan algoritma K-Means *Clustering* [5]. K-Means *Clustering* adalah metode *Clustering* yang sangat populer dalam analisis data karena kemampuannya untuk mengelompokkan data berdasarkan kesamaan fitur tertentu, dalam hal ini adalah frekuensi transaksi ATM [4]. Dengan algoritma ini, data transaksi nasabah dapat dikelompokkan ke dalam beberapa kelompok atau klaster yang mewakili pola penggunaan ATM yang serupa, tanpa memerlukan label klasifikasi sebelumnya [6].

Penggunaan K-Means *Clustering* dalam konteks analisis frekuensi transaksi ATM oleh BANK BJB memberikan beberapa keuntungan signifikan. Pertama, algoritma ini memungkinkan identifikasi segmentasi nasabah secara objektif berdasarkan data transaksi yang tersedia, sehingga bank dapat menyesuaikan strategi pelayanan dan promosi yang lebih terfokus [7]. Kedua, K-Means *Clustering* efisien dalam menangani jumlah data besar yang umumnya dimiliki oleh bank, memungkinkan analisis yang cepat dan akurat. Ketiga, hasil dari *Clustering* dapat memberikan wawasan mendalam mengenai preferensi dan kebiasaan nasabah dalam menggunakan layanan ATM, yang menjadi landasan untuk pengambilan keputusan yang lebih baik dalam mengelola pelayanan nasabah [8].

Di BANK BJB, penggunaan K-Means *Clustering* dalam analisis frekuensi transaksi ATM bertujuan untuk meningkatkan pemahaman tentang perilaku nasabah serta untuk meningkatkan kepuasan nasabah melalui pelayanan yang lebih personal. Dengan menerapkan pendekatan ini, BANK BJB dapat mengoptimalkan operasional, meningkatkan efisiensi biaya, dan merancang layanan yang lebih sesuai dengan kebutuhan nasabah berdasarkan pola penggunaan ATM.

Meskipun K-Means *Clustering* menawarkan banyak keuntungan, tetap ada beberapa tantangan yang perlu diatasi. Salah satunya adalah pemilihan jumlah klaster yang optimal, yang memerlukan evaluasi dan validasi yang cermat agar hasil *Clustering* sesuai dengan karakteristik data [4]. Selain itu, perlu memperhatikan kualitas data yang digunakan dalam analisis untuk memastikan hasil yang akurat dan dapat diandalkan [5].

Penelitian ini bertujuan untuk menganalisis segmentasi nasabah Bank BJB Kantor Cabang Tasikmalaya berdasarkan frekuensi transaksi ATM menggunakan metode K-Means *Clustering*. Melalui analisis ini diharapkan diperoleh wawasan mendalam mengenai pola perilaku nasabah yang dapat dijadikan landasan dalam pengambilan keputusan strategis yang lebih akurat dan efektif di bidang pelayanan perbankan. Dengan pemahaman yang lebih baik tentang karakteristik dan kebutuhan nasabah melalui segmentasi terperinci, Bank BJB dapat mengembangkan layanan yang lebih personal, meningkatkan efisiensi operasional, serta memperkuat loyalitas nasabah. Hal ini, pada akhirnya akan mendukung pertumbuhan dan daya saing bank di pasar yang semakin kompetitif. Pemanfaatan K-Means *Clustering* dalam analisis frekuensi transaksi ATM memungkinkan Bank BJB khususnya Kantor Cabang Tasikmalaya, untuk meningkatkan pemahaman tentang nasabah, mengoptimalkan layanan, dan meningkatkan kepuasan nasabah secara keseluruhan. Analisis ini diharapkan memberikan kontribusi positif dalam pengembangan strategi pelayanan yang lebih adaptif dan responsif terhadap kebutuhan nasabah di era digital saat ini.

Penelitian terdahulu tentang segmentasi nasabah berbasis transaksi ATM telah menunjukkan penggunaan metode *Clustering* seperti K-Means. [4] mengelompokkan nasabah bank berdasarkan pengambilan kredit menggunakan K-Means *Clustering*, menghasilkan empat klaster utama dari 1000 data nasabah untuk membantu bank merumuskan strategi pemasaran yang lebih efektif.

Penelitian lainnya oleh [3] Menyoroti pentingnya segmentasi nasabah untuk mengoptimalkan strategi pelayanan di perbankan. Strategi segmentasi pasar di BMT Sidogiri Capem Sempu yang mencakup segmentasi geografik, demografik dan psikografik berhasil meningkatkan jumlah nasabah dan daya saing di pasar. Penelitian ini menggunakan algoritma Naive Bayes untuk memprediksi frekuensi penarikan uang tunai di ATM. Hasilnya menunjukkan akurasi hingga 81%, hasil ini membantu bank dalam meningkatkan layanan dan mengurangi uang yang tidak terpakai di mesin ATM [9]. Namun, terdapat celah yang bisa dieksplorasi lebih lanjut, terutama dalam hal penggunaan algoritma lain seperti K-Means *Clustering* untuk segmentasi nasabah yang lebih optimal dan analisis perilaku transaksi yang lebih kompleks.

Penelitian ini bertujuan untuk mengatasi permasalahan dalam hal segmentasi nasabah, yang selama ini masih dilakukan secara umum dan kurang terarah. Hal ini berdampak pada strategi pelayanan yang kurang efisien, terutama dalam memenuhi kebutuhan nasabah yang memiliki karakteristik berbeda. Dengan menggunakan metode K-Means *Clustering*, penelitian ini berusaha memberikan solusi dengan mengelompokkan nasabah berdasarkan kesamaan karakteristik transaksi mereka, sehingga pihak bank dapat menyesuaikan strategi pelayanannya secara lebih personal dan efektif. Segmentasi yang lebih akurat diharapkan dapat meningkatkan kepuasan nasabah dan mendukung efisiensi operasional bank. Meskipun demikian, penelitian ini memiliki keterbatasan pada jumlah data dan cakupan geografis yang terbatas pada satu

cabang bank, sehingga hasilnya mungkin tidak sepenuhnya dapat digeneralisasi untuk semua wilayah atau cabang yang menjadi area penting untuk eksplorasi lebih lanjut dalam penelitian mendatang.

II. METODE PENELITIAN

Metode penelitian adalah serangkaian prosedur atau langkah-langkah yang dilakukan secara sistematis untuk memperoleh pengetahuan ilmiah [10]. Proses ini bertujuan untuk menyusun dan mengembangkan ilmu pengetahuan dengan cara yang terstruktur dan terencana, sehingga hasil yang diperoleh dapat dipertanggungjawabkan secara akademis dan ilmiah [11]. Berikut adalah metodologi penelitian pada gambar 1.



Gambar 1. Metode Penelitian

Berikut adalah penjabaran dari metode yang ditampilkan pada Gambar 1:

Pengumpulan Dataset

Tahap awal dari penelitian ini adalah mengumpulkan dataset yang akan diteliti, yaitu data frekuensi transaksi mesin ATM di BANK BJB dalam hal ini Mengacu pada kumpulan data yang akan digunakan dalam penelitian [12]. Dataset ini menjadi input utama untuk seluruh proses. Sumber dataset ini merupakan data primer yang bersumber dari data transaksi ATM selama bulan Juni 2024 di kantor wilayah tiga Bank BJB khususnya Tasikmalaya. Total data yang dikumpulkan sebanyak 447 data dengan 6 atribut. Pada tahapan ini dilakukan serangkaian kegiatan meliputi:

1. Identifikasi kebutuhan data
2. Pengumpulan data transaksi dari sumber primer (sistem ATM Bank BJB)
3. Verifikasi dan validasi data untuk memastikan keakuratannya
4. Penyimpanan data dalam format yang sesuai untuk analisis lebih lanjut dalam hal ini dilakukan konversi bentuk data dari microsoft excel kedalam bentuk *comma delimited* (csv) supaya mudah dibaca oleh *tools*.

Preprocessing

Tahap awal pengolahan data yang bertujuan untuk mempersiapkan data mentah sebelum masuk ke proses lebih lanjut. Tahap ini dapat mencakup berbagai kegiatan seperti penghapusan duplikasi data, penanganan data yang hilang, dan normalisasi data [13].

Cleaning

Proses *cleaning* atau pembersihan merupakan tahapan pada bagian *preprocessing* dimana untuk memastikan data yang akan digunakan bebas dari kesalahan atau inkonsistensi [14]. *Cleaning* dapat mencakup kegiatan seperti penghapusan data yang tidak relevan, perbaikan kesalahan entri data, dan penanganan *missing values* [4]. Tahapan-tahapan yang dilakukan pada proses ini diantaranya:

1. Penghapusan data yang tidak relevan atau tidak sesuai dengan tujuan penelitian, misalnya menghapus data transaksi yang frekuensinya 0 atau tidak ada transaksi sehingga perlu di seleksi supaya tidak dibaca pada saat pemodelan.
2. Pemilihan atribut yang dianggap penting dalam mempengaruhi hasil penelitian, karena hasil yang dapat diproses pada *Kmeans Clustering* adalah data dengan type integer, maka dari itu atribut-atribut yang dianggap penting seperti kode cabang, cabang induk dan frekuensi adalah atribut paling relevan yang dapat mempengaruhi hasil penelitian.

TABEL 1
 DATA HASIL CLEANING

	Sebelum	Sesudah
Jumlah Dataset	447	434

	Sebelum	Sesudah
Atribut Terpilih	ID ATM, KD_CAB, CAB INDUK, KW, Frekuensi	KD_CAB, CAB INDUK, Frekuensi

Transformation

Proses transformasi data yang bertujuan untuk mengubah format atau struktur data agar sesuai dengan kebutuhan analisis selanjutnya. Transformasi bisa mencakup perubahan skala, pengkodean variabel kategori, atau pembentukan fitur baru dari data yang ada [13].

Clustering Menggunakan K-Means

Proses pengelompokan data menggunakan algoritma K-Means. Metode ini dipilih karena metode sederhana, efisien, dan sangat cocok untuk data yang memiliki jumlah kluster yang sudah ditentukan sebelumnya. Dalam penelitian ini, dataset yang digunakan menunjukkan kecenderungan kluster yang jelas dan terdefinisi, sehingga K-Means dapat mengelompokkan data dengan baik berdasarkan centroid terdekat [15]. Metode ini juga mudah diimplementasikan dan memberikan hasil yang dapat diinterpretasikan dengan jelas, di mana setiap kluster diwakili oleh centroid yang mewakili rata-rata dari titik data dalam kluster tersebut. Hal ini sangat sesuai dengan tujuan penelitian yang membutuhkan pembagian kluster yang terukur dan mudah dipahami [16].

Jika dibandingkan dengan metode lain seperti Gaussian Mixture Model (GMM) dan DBSCAN, K-Means lebih tepat karena data yang digunakan tidak memerlukan asumsi distribusi probabilitas seperti pada GMM, dan tidak mengandung banyak noise atau kluster dengan bentuk tidak beraturan yang lebih cocok untuk DBSCAN. Oleh karena itu, K-Means dipilih sebagai metode clustering yang paling sesuai untuk dataset ini karena mampu memberikan hasil yang optimal dengan struktur data yang jelas dan terkelompok secara alami [17].

Pada proses ini merupakan tahapan paling penting karena akan melakukan analisa terhadap dataset penelitian. Adapun tahapan-tahapan yang dilakukan dengan menggunakan metode K-Means *Clustering* meliputi:

1. Menentukan jumlah kluster yang optimal secara manual misalnya menentukan nilai $k=2$ dengan menggunakan syarat minimal k untuk *Clustering*.
2. Menentukan nilai tengah dengan memilih sample data nilai yang ada ditengah-tengah sebagai patokan mengklusterkan data lainnya.
3. Menghitung jarak objek ke centroid dengan persamaan:

$$De = De = \sqrt{\sum_{i=0}^n (xi - y1)^2} \quad (1)$$

Keterangan:

De = *Enclidean Distance*

x = data pusat kluster

y = data pada atribut

i = setiap data

n = jumlah data

xi = data pada pusat kluster ke i

yi = data pada setiap data ke i

4. Menentukan centroid baru dengan persamaan:

$$Ci = \frac{\sum_{i=1}^n xi \in si}{n} \quad (2)$$

Keterangan:

Ci = centroid baru ke i

si = objek ke i

xi = nilai pada objek ke i

n = jumlah data pada tiap kelompok

5. Menggunakan metode Elbow untuk mendapatkan nilai k paling optimal.

Analisa Hasil Cluster

Pada tahap ini, dilakukan evaluasi dan interpretasi hasil pengelompokan yang telah dilakukan. Tujuannya adalah untuk memahami karakteristik setiap kluster dan menarik kesimpulan yang relevan dengan tujuan penelitian [18].

III. HASIL DAN PEMBAHASAN

Analisis segmentasi frekuensi transaksi ATM menggunakan metode K-Means Clustering di Bank BJB Kantor Cabang Tasikmalaya mencakup berbagai tahapan, mulai dari persiapan dan eksplorasi data, pemilihan jumlah kluster, penerapan algoritma, hingga evaluasi hasil *Clustering*. Proses ini bertujuan untuk memahami pola transaksi nasabah dan mengidentifikasi segmentasi yang dapat digunakan untuk mengembangkan strategi pemasaran yang lebih efektif dan terarah. Hasil analisis ini diharapkan dapat memberikan wawasan yang signifikan bagi pengelolaan dan pengembangan layanan bank secara keseluruhan [19].

Tahap pertama adalah pengumpulan data transaksi ATM dari Bank BJB Kantor Cabang Tasikmalaya. Data ini mencakup berbagai informasi tentang frekuensi transaksi nasabah selama periode Juni 2024. Data yang terkumpul kemudian disiapkan untuk tahap analisis lebih lanjut. Dari hasil analisis di dapatkan data sejumlah 447 dengan 5 kolom utama. Berikut Tabel 2 sampel dari datasetnya.

TABEL 2
 DATA FREKUENASI TRANSAKSI ATM

No	ID ATM	KD_CAB	CAB INDUK	KW	Frekuensi
1	A019	0002	0002	Kanwil III	3045
2	A022	0017	0017	Kanwil III	358
3	A028	0002	0002	Kanwil III	2864
4	A029	0004	0004	Kanwil III	3388
5	A045	0016	0016	Kanwil III	4333
...
444	B952	0009	0067	Kanwil III	616
445	B953	0016	0016	Kanwil III	6823
446	B958	0016	0016	Kanwil III	1673
447	B959	0016	0016	Kanwil III	2945

Berikut penjelasan dari masing-masing atribut pada tabel 2 frekuensi transaksi di ATM:

1. ID ATM adalah kode unik yang diberikan untuk mengidentifikasi setiap mesin ATM secara individual.
2. KD_CAB adalah kode cabang bank tempat ATM tersebut berada. Kode ini digunakan untuk mengidentifikasi cabang bank secara spesifik.
3. CAB INDUK adalah kode atau nama cabang induk yang membawahi beberapa cabang termasuk yang memiliki ATM tersebut. Cabang induk mengatur beberapa cabang yang lebih kecil.
4. KW adalah singkatan dari Kantor Wilayah. Ini adalah kode yang menunjukkan wilayah atau area geografis di mana ATM tersebut berada.
5. Frekuensi adalah jumlah atau tingkat transaksi yang terjadi pada ATM tersebut dalam periode 1 bulan. Contoh: 50 (menunjukkan ada 50 transaksi), 100 (menunjukkan ada 100 transaksi).

Data yang telah dikumpulkan kemudian masuk ke tahap *preprocessing*, di mana data diolah untuk memastikan kebersihan dan kesiapan untuk analisis [20]. Proses ini meliputi pengisian nilai hilang, penghapusan duplikasi, normalisasi data, dan transformasi data dari format excel ke .csv untuk diolah menggunakan Google Colabs dan Python. Atribut data yang jumlah frekuensi transaksi 0 dihilangkan, dan kolom yang tidak berperan dalam pengklasifikasian dihapus untuk memudahkan proses *clustering*. Hasil dari tahapan ini didapatkan data sebanyak 434 data yang dianggap memenuhi untuk dilakukan *clustering* dan berdasarkan hasil ini kolom-kolom yang dianggap penting, yaitu pada tabel 3.

TABEL 3
 ATRIBUT CLUSTERING

KD_CAB	CAB INDUK	Frekuensi
2	2	3045
17	17	358

KD_CAB	CAB INDUK	Frekuensi
2	2	2864
4	4	3388
16	16	4333
...
17	17	3958
9	67	616
16	16	6823
16	16	1673
16	16	2945

Pada Tabel 3 Tabel 3 menunjukkan hasil pengelompokan atau *clustering* berdasarkan atribut KD_CAB (Kode Cabang), CAB INDUK (Cabang Induk), dan Frekuensi. Pengelompokan ini bertujuan untuk mengidentifikasi pola dan hubungan antara cabang-cabang dan cabang induk berdasarkan frekuensi tertentu.

Tahap selanjutnya merupakan tahapan paling penting dalam melakukan pengelompokan atau *clustering* terhadap dataset penelitian dimana *Clustering* merupakan proses pengelompokan yang menggunakan teknik *unsupervised learning*, dimana tidak diperlukan fase pembelajaran dan tidak menggunakan pelabelan pada setiap kelompok [21]. Metode *Clustering* membagi data menjadi kelompok-kelompok sehingga data yang memiliki karakteristik serupa dikelompokkan dalam satu klaster yang sama [22]. Data yang telah siap kemudian dikelompokkan menggunakan metode K-Means *Clustering*. Algoritma K-Means diterapkan untuk membagi data ke dalam beberapa klaster berdasarkan kesamaan pola transaksi [23]. Jumlah klaster ditentukan berdasarkan metode Elbow dan Silhouette Score yang menunjukkan jumlah klaster optimal untuk segmentasi [19]. Langkah-langkah Algoritma Metode K-Means meliputi:

Menentukan Jumlah Klaster

Langkah pertama dalam menggunakan algoritma k-means ini adalah menentukan jumlah klaster yang akan dibuat. jumlah klaster ini akan menjadi inisial bagi data-data yang telah dikumpulkan. Tidak ada aturan pasti dalam menentukan jumlah klaster k, tetapi jumlah klaster yang diinginkan dapat disesuaikan dengan kebutuhan subjektif masing-masing individu [24]. Pada tahap ini misalnya dari 434 data akan dibagi menjadi dua klaster, yaitu *cluster 0* (klaster ke-1) dan *cluster 1* (klaster ke-2).

Menentukan Nilai Awal Titik Tengah

Pada langkah ini adalah menentukan nilai tengah untuk menentukan data-data selain yang dipilih masuk ke *cluster 0* (klaster ke-1) atau *cluster 1* (klaster ke-2) [18], karena jumlah data ada 434 maka diambil 2 nilai yang ada di tengah-tengah dataset, yaitu data ke 222 dan 223 seperti yang ditunjukkan pada tabel 4.

TABEL 4
MENENTUKAN NILAI TENGAH

ID ATM	KD_CAB	CAB INDUK	Frekuensi	Keterangan
B192	9	9	1011	<i>Cluster 0</i>
B193	15	15	2893	<i>Cluster 1</i>

Menghitung Jarak objek ke Centroid

Untuk menghitung jarak objek ke centroid rumus yang digunakan dengan menggunakan persamaan (1) maka berdasarkan persamaan tersebut jika akan menentukan data ke 1 misalnya apakah masuk ke klaster ke-1 atau masuk ke klaster ke-2, maka jarak paling dekat ke *cluster* adalah yang menentukannya. Misalnya Data ke-1 pada tabel 5.

TABEL 5
DATASET KE-1

ID ATM	KD_CAB	CAB INDUK	Frekuensi
A019	2	2	3045

Hasil data ke-1 terhadap klaster ke-1

$$De = \sqrt{(2 - 9)^2 + (2 - 9)^2 + (3045 - 1011)^2}$$

$$De = \sqrt{49 + 49 + 4137156}$$

$$De = 4137254$$

Hasil data ke-1 terhadap klaster ke-2

$$De = \sqrt{(2 - 15)^2 + (2 - 15)^2 + (3045 - 2893)^2}$$

$$De = \sqrt{169 + 169 + 23104}$$

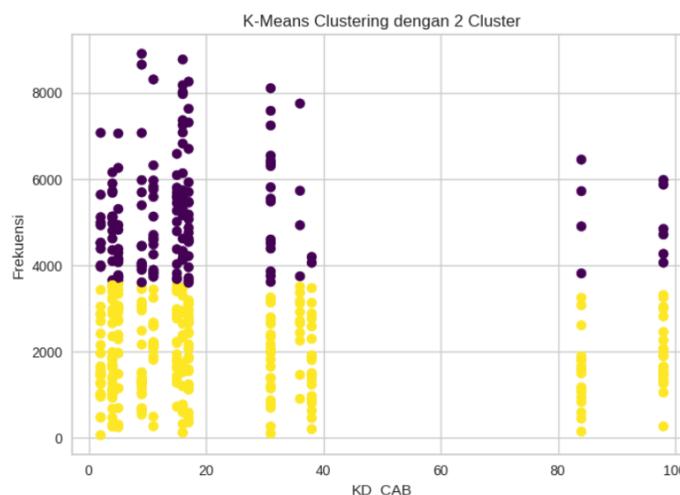
$$De = 23442$$

Berdasarkan hasil perhitungan dari dataset ke 1 terhadap klaster ke-1 (*cluster 0*) memiliki nilai jarak enclidean sebesar 4.137.254 dan terhadap klaster ke-2 (*cluster 1*) memiliki nilai jarak enclidean sebesar 23.442 maka dari itu data ke 1 dapat di kategorikan sebagai klaster ke-2 (*cluster 1*) karena memiliki jarak yang lebih kecil. Perhitungan ini dilakukan terhadap semua item dataset sehingga dapat ditentukan untuk semua data masuk kedalam klaster/kelompok yang mana. Dikarenakan jumlah data sebanyak 434 data untuk mempercepat perhitungan menggunakan tools google colabs dengan bahasa pemrograman python. Hasilnya pada tabel 6.

TABEL 6
 HASIL KLASTERISASI DENGAN PYTHON

	ID ATM	KD_CAB	CAB INDUK	Frekuensi	Cluster
0	A019	2	2	3045	1
1	A022	17	17	358	1
2	A028	2	2	2864	1
3	A029	4	4	3388	1
4	A045	16	16	4333	0
5	A046	9	9	5976	0
...
430	B927	36	206	3119	1
431	B932	38	38	1480	1
432	B933	11	293	5132	0
433	B934	98	98	1562	1
434	B935	98	98	2066	1

Pada tabel 6 semua dataset sudah dikelompokkan dan memiliki label baru dengan nama *cluster*, sehingga masing-masing klaster jika di definisikan memiliki jumlah anggota sebanyak 152 sebagai anggota *cluster 0* (klaster ke-1) dan 282 anggota untuk *cluster 1* (klaster ke-2). Jika di ilustrasikan menggunakan plot hasilnya pada gambar 2.



Gambar 2. Sebaran Data *Clustering* dengan k=2

Gambar 2 menunjukkan hasil visualisasi dari algoritma K-Means *Clustering* dengan k=2 pada dataset dengan kolom `KD_CAB` (kode cabang bank) dan `Frekuensi` (jumlah transaksi di ATM). Sumbu X mewakili

`KD_CAB`, sedangkan sumbu Y mewakili `Frekuensi`. Setiap titik pada plot menggambarkan satu baris data dengan kombinasi nilai `KD_CAB` dan `Frekuensi`, dengan warna titik menunjukkan kluster yang dihasilkan oleh K-Means. Dalam visualisasi ini, dua warna (ungu dan kuning) menunjukkan dua kluster yang berbeda. Kluster 0 (ungu) mungkin memiliki frekuensi transaksi tinggi, sementara kluster 1 (kuning) mungkin memiliki frekuensi transaksi lebih rendah. Pemisahan kluster ini membantu memahami distribusi frekuensi transaksi di berbagai cabang dan dapat digunakan untuk pengambilan keputusan bisnis berdasarkan pola yang teridentifikasi.

Menentukan Centroid Baru

Langkah selanjutnya setelah melakukan perhitungan jarak, kemudian hitung centroid baru dengan menggunakan nilai mean dari semua data dalam sel kluster tertentu dengan persamaan (2). Pada data sebelumnya *cluster 0* (kluster ke-1) sudah memiliki anggota kelompok data sejumlah 152 dan *cluster 1* (kluster ke-2) memiliki jumlah anggota 282. Untuk menentukan centroid baru berdasarkan rumus kita ambil sampel dataset yang dari *cluster 0* dan *cluster 1*.

Cluster 0 :

Nilai n = 152

Sample : 16, 16, 4333 (Data ke-4) dan 9, 9, 5976 (Date ke-5) => n=152

Cluster 1 :

nilai n = 282

sample : 2, 2, 3045 (Data ke-1) dan 17, 17, 358 (Data ke-2) => n=282

Menghitung centroid baru:

Cluster 0

$$\begin{aligned} C0_{KD_CAB} &= \frac{(16+9)}{2} &= \frac{(25)}{2} &= 12,5 \\ C0_{CAB_INDUK} &= \frac{(16+9)}{2} &= \frac{(25)}{2} &= 12,5 \\ C0_{Frekuensi} &= \frac{(4333+5976)}{2} &= \frac{(10309)}{2} &= 5154,5 \end{aligned}$$

Maka berdasarkan nilai centroid baru diatas mean atau rata-rata *cluster 0* berada pada nilai 12,5; 12,5 ; 5154,5

Cluster 1

$$\begin{aligned} C1_{KD_CAB} &= \frac{(2+17)}{2} &= \frac{(19)}{2} &= 9,5 \\ C1_{CAB_INDUK} &= \frac{(2+17)}{2} &= \frac{(19)}{2} &= 9,5 \\ C1_{Frekuensi} &= \frac{(3045+358)}{2} &= \frac{(3403)}{2} &= 1701,5 \end{aligned}$$

Maka berdasarkan nilai centroid baru diatas mean atau rata-rata *cluster 1* (kluster ke-2) berada pada nilai 9,5 ; 9,5 ; 1701,5.

Proses-proses ini dilakukan terhadap semua dataset sehingga di hasilkan nilai untuk masing-masing kluster adalah

1. Centroid baru untuk *Cluster 0*: (19.368, 126.782, 5249.638)
2. Centroid baru untuk *Cluster 1*: (26.936, 56.854, 1905.578)

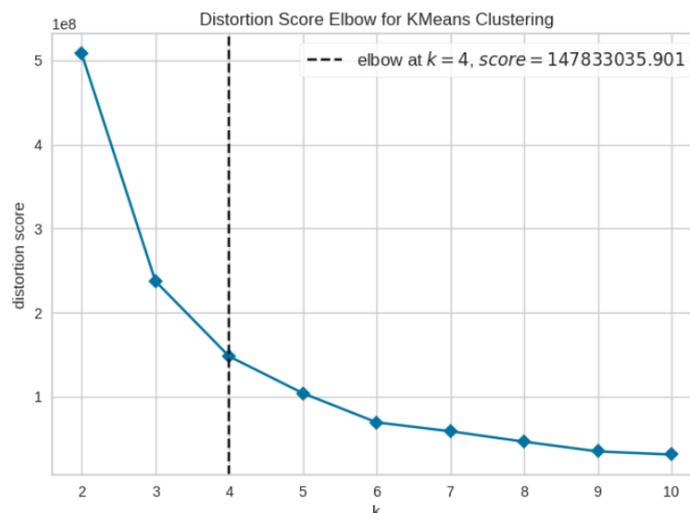
Evaluasi dan Validasi

Pada tahapan evaluasi dan validasi adalah dengan menentukan nilai kluster optimal adalah langkah penting dalam proses *Clustering* untuk memastikan bahwa data telah dikelompokkan dengan cara yang paling berarti. Evaluasi dilakukan dengan mengukur *Within-Cluster Sum of Squares* (WCSS) dan menggunakan *Davies-Bouldin Index*.

$$WCSS = \sum_{n=1}^{\dots} (X_i - Y_i)^2 \tag{3}$$

Di mana ' Y_i ' adalah centroid untuk observasi ' X_i ' dan ' n ' adalah total jumlah observasi.

Dari persamaan (3) tersebut, dapat diinterpretasikan bahwa ketika jumlah kluster meningkat, jarak antara titik dan centroid-nya berkurang, sehingga WCSS berkurang. Pemilihan jumlah kluster yang tepat meningkatkan interpretabilitas hasil dan memastikan struktur alami data teridentifikasi secara akurat. Penelitian ini menggunakan metode Elbow untuk menentukan nilai kluster optimal, yang bertujuan mengevaluasi dan memilih jumlah kluster yang memberikan keseimbangan terbaik antara kompleksitas model dan kemampuan menangkap struktur data sebenarnya. Hasil metode Elbow menentukan nilai k paling optimal untuk data yang diteliti, seperti yang ditunjukkan pada gambar 3.



Gambar 3. Hasil Clustering dengan Metode Elbow

Gambar 3 adalah hasil dari metode Elbow yang digunakan untuk menentukan jumlah kluster optimal dalam algoritma K-Means Clustering. Berikut adalah penjelasan lebih rinci mengenai hasil yang ditampilkan pada grafik:

1. Sumbu X (k). Mewakili jumlah kluster (k) yang diuji, mulai dari 1 hingga 10.
2. Sumbu Y (*Distortion Score*): Mewakili skor distorsi atau inerti, yang merupakan jumlah kuadrat dari jarak antara setiap titik data dan centroid terdekatnya dalam kluster tersebut.
3. *Distortion score* adalah ukuran seberapa baik titik data dalam kluster dikompresi oleh centroid. Nilai yang lebih rendah menunjukkan kluster yang lebih kompak dan lebih baik.
4. Kurva menunjukkan perubahan *distortion score* seiring dengan bertambahnya jumlah kluster (k). Kurva cenderung menurun saat jumlah kluster meningkat, karena menambahkan lebih banyak kluster akan cenderung mengurangi *distortion score*.
5. Titik Elbow ($k = 4$). Titik elbow adalah titik di mana penurunan *distortion score* mulai melambat secara signifikan.

Dalam grafik ini, titik elbow berada pada $k = 4$, dengan *distortion score* sekitar 147833035.901. Ini menunjukkan bahwa menggunakan 4 kluster adalah pilihan optimal, karena menambah lebih banyak kluster setelah titik ini memberikan pengurangan *distortion score* yang tidak signifikan.

Analisis Hasil

Setelah proses Clustering selesai, hasil dari setiap kluster dianalisis untuk memahami karakteristik masing-masing segmentasi. Kluster diidentifikasi berdasarkan rata-rata frekuensi transaksi dan karakteristik lainnya. Analisis ini memberikan wawasan tentang perilaku transaksi nasabah yang berbeda-beda. Pada hasil analisis ini melihat data-data kecenderungan setelah dihitung mengikuti tahap perhitungan pada sub bab sebelumnya, yaitu urutan Clustering dengan menggunakan algoritma Kmeans. Hasil-hasil yang ditampilkan setelah dibagi menjadi 4 kluster pada tabel 7.

TABEL 7
HASIL PENGELOMPOKKAN DENGAN MENGGUNAKAN KLASTER

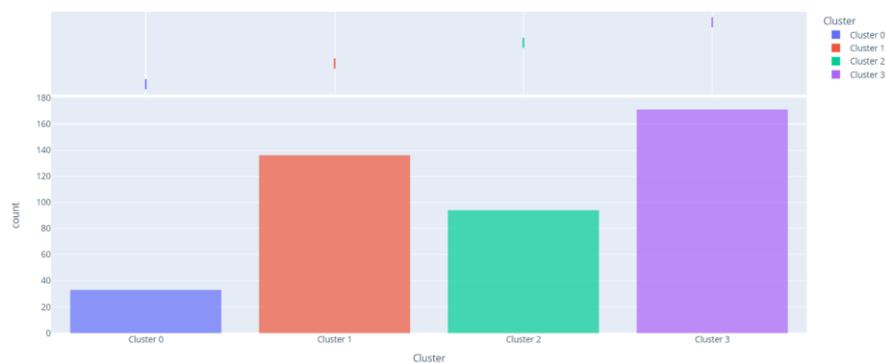
	KD_CAB	CAB INDUK	Frekuensi	Klaster
0	2	2	3045	1
1	17	17	358	3
2	2	2	2864	1
3	4	4	3388	1
4	16	16	4333	2
5	9	9	5976	2
...
430	36	206	3119	1
431	38	38	1480	3
432	11	293	5132	2
433	98	98	1562	3
434	98	98	2066	3

Hasil plotting dengan menggunakan PCA 2 dimensi pada gambar 4.



Gambar 4. Visualisasi dengan 4 Klaster

Plot PCA dua dimensi menunjukkan bahwa algoritma K-Means berhasil mengidentifikasi empat klaster dalam dataset dengan distribusi dan variasi yang berbeda. Meskipun ada beberapa tumpang tindih titik, klaster secara umum terpisah dengan baik, menunjukkan struktur data yang dapat dikenali. Plot PCA ini membantu memvisualisasikan dan memahami struktur data kompleks dalam bentuk yang lebih sederhana. Jumlah data pada masing-masing klaster divisualisasikan pada Gambar 5.



Gambar 5. Jumlah Anggota Klaster Dalam Diagram Batang

Hasil *clustering* dengan algoritma K-Means menunjukkan distribusi jumlah anggota dalam masing-masing kluster, dengan *Cluster 0* berisi 33 anggota, *Cluster 1* berisi 94 anggota, *Cluster 2* berisi 136 anggota, dan *Cluster 3* berisi 171 anggota. Kluster ini mengelompokkan data berdasarkan kesamaan karakteristik, di mana *Cluster 0* memiliki jumlah anggota terkecil dan *Cluster 3* memiliki jumlah anggota terbesar. Jumlah anggota dalam setiap kluster menunjukkan berapa banyak data points yang memiliki kesamaan karakteristik tertentu. Berdasarkan hasil *clustering* segmentasi berdasarkan frekuensi transaksi, data transaksi ATM di Bank BJB Kantor Cabang Tasikmalaya telah dikelompokkan ke dalam beberapa kluster, memberikan wawasan penting tentang pola transaksi nasabah diantaranya sebagai berikut:

1. *Cluster 0*/Kluster ke-1 (Frekuensi Sangat Tinggi)
 - Kluster ini mencakup cabang dengan frekuensi transaksi yang sangat tinggi dan mungkin memerlukan perhatian khusus dalam penempatan mesin ATM serta pengelolaan uang tunai.
 - Promosi eksklusif dan program loyalitas premium bisa menjadi pendekatan efektif untuk meningkatkan kepuasan nasabah di kluster ini.
2. *Cluster 1*/Kluster ke-2 (Frekuensi Rendah)
 - Kluster ini terdiri dari cabang dengan frekuensi transaksi yang rendah dan dapat diberikan program edukasi mengenai manfaat penggunaan ATM untuk meningkatkan frekuensi transaksi.
 - Promosi relevan dan insentif bisa dirancang untuk mendorong nasabah meningkatkan penggunaan layanan ATM.
3. *Cluster 2*/Kluster ke-3 (Frekuensi Tinggi)
 - Kluster ini terdiri dari cabang dengan frekuensi transaksi yang tinggi dan mungkin memerlukan jumlah mesin ATM yang lebih banyak untuk menghindari antrian.
 - Strategi pemasaran agresif dengan program loyalitas khusus dapat diterapkan untuk mempertahankan nasabah di kluster ini.
4. *Cluster 3*/Kluster ke-4 (Frekuensi Terendah)
 - Kluster ini terdiri dari cabang dengan frekuensi transaksi yang sangat rendah dan perlu diberikan perhatian khusus melalui program edukasi untuk meningkatkan penggunaan layanan ATM.
 - Promosi khusus dan insentif bisa dirancang untuk menarik nasabah meningkatkan frekuensi transaksi.

IV. KESIMPULAN

Penelitian ini menggunakan algoritma K-Means *clustering* untuk menganalisis segmentasi frekuensi transaksi ATM di Bank BJB Kantor Cabang Tasikmalaya dengan 447 dataset, menghasilkan empat kluster dengan karakteristik berbeda. Evaluasi menggunakan WCSS dan Davies-Bouldin Index menunjukkan kualitas *clustering* yang baik, memberikan wawasan untuk strategi pemasaran dan operasional yang lebih efisien. Bank BJB dapat menyesuaikan lokasi ATM untuk mengurangi antrian di kluster frekuensi tinggi, mengevaluasi kembali ATM di kluster frekuensi rendah, dan melakukan program edukasi untuk nasabah dengan frekuensi transaksi rendah. Dengan memahami pola transaksi ini, bank dapat mengoptimalkan pengelolaan uang tunai di setiap ATM untuk mengurangi biaya pengisian ulang dan meningkatkan efisiensi operasional. Data hasil *clustering* dapat digunakan untuk mengembangkan produk dan layanan baru yang sesuai dengan kebutuhan dan perilaku transaksi dari setiap kluster nasabah. Bank juga dapat merancang promosi dan insentif yang disesuaikan dengan karakteristik setiap kluster, seperti menawarkan hadiah atau poin loyalitas untuk nasabah yang meningkatkan frekuensi transaksi ATM, sehingga meningkatkan kepuasan nasabah, efisiensi operasional, dan daya saing di pasar.

Untuk memperluas hasil dan manfaat penelitian, maka dapat mengkombinasikan K-Means dengan algoritma lain untuk meningkatkan akurasi segmentasi. Selain itu, integrasi data tambahan seperti preferensi digital dan demografi nasabah dapat menghasilkan strategi yang lebih adaptif. Uji coba hasil *clustering* di cabang lain juga penting untuk menguji konsistensi dan efektivitas strategi, serta memperluas penerapan di seluruh jaringan perbankan.

UCAPAN TERIMA KASIH

Penulis mengucapkan banyak terimakasih atas segala bimbingan yang telah diberikan oleh dosen pembimbing sehingga penelitian ini dapat selesai sesuai dengan waktu yang telah ditentukan.

DAFTAR PUSTAKA

- [1] K. Kirman and F. Kurniawan, "Implementasi Metode Kruskal dan Open Street Map Pada Pencarian Lokasi Anjungan Tunai Mandiri (ATM) Kota Bengkulu," *J. Technopreneursh. Inf. Syst.*, vol. 6, no. 3, pp. 67–73, 2023.
- [2] N. Al Indiani, K. A. Santoso, and D. Anggraeni, "Penentuan Lokasi Atm Bank Syariah Indonesia Di Wilayah Jember Kota Menggunakan K-Means Clustering," *UNEJ e-Proceeding*, vol. 23, pp. 68–79, 2022.
- [3] M. Z. Alifudin and M. Y. Zamrozi, "Strategi Segmentasi Pasar Untuk Peningkatan Jumlah Nasabah Di Bmt Sidogiri Capem Sempu," *J. Ekon. Syariah Darussalam*, vol. 2, no. I, pp. 2745–8407, 2021.
- [4] N. Ahsina, F. Fatimah, and F. Rachmawati, "Analisis Segmentasi Pelanggan Bank Berdasarkan Pengambilan Kredit Dengan Menggunakan Metode K-Means Clustering," *J. Ilm. Teknol. Infomasi Terap.*, vol. 8, no. 3, 2022.
- [5] F. P. Dewi, P. S. Aryni, and Y. Umidah, "Implementasi Algoritma K-Means Clustering Seleksi Siswa Berprestasi Berdasarkan Keaktifan dalam Proses Pembelajaran," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 7, no. 2, pp. 111–121, 2022.
- [6] L. ' Izzah and A. Jananto, "Penerapan Algoritma K-Means Clustering Untuk Perencanaan Kebutuhan Obat Di Klinik Citra Medika," *Progresif J. Ilm. Komput.*, vol. 18, no. 1, p. 69, 2022.
- [7] C. N. Prabiantissa and G. E. Yuliasuti, "Prediksi Pergerakan Ikan Di Pesisir Pulau Madura Menggunakan Metode Gaussian Mixture Model Dan K-Means Clustering," *J. Teknol. Inf. dan Terap.*, vol. 8, no. 2, pp. 121–128, 2021.
- [8] M. M. Effendi and A. Siswandi, "Analysis Prediksi Wilayah Rawan Banjir dengan Algoritma K-Means," *J. Inf. Syst. Res.*, vol. 5, no. 2, pp. 697–703, 2024.
- [9] S. W. Mulia, S. Sujiharno, and A. Wibowo, "Implementasi Algoritma Naive Bayes Untuk Memprediksi Frekuensi Tunai Pada Mesin Atm Di Masa Transisi Pembatasan Sosial Berskala Besar (Psbb) Pandemi Covid-19," *SINTECH (Science Inf. Technol. J.)*, vol. 4, no. 1, pp. 47–52, 2021.
- [10] M. F. Al Halik and L. Septiana, "Analisa Data Untuk Prediksi Daerah Rawan Bencana Alam Di Jawa Barat Menggunakan Algoritma K-Means Clustering," *J. Inf. Syst. Applied, Manag. Account. Res.*, vol. 6, no. 4, pp. 856–870, 2022.
- [11] M. R. A. Fernanda, P. Sokibi, and R. Fahrudin, "Sistem Prediksi Ketepatan Kelulusan Mahasiswa Berdasarkan Data Akademik Dan Non Akademik Menggunakan Metode K-Means (Studi Kasus : Universitas Catur Insan Cendekia)," *J. Digit*, vol. 11, no. 1, p. 89, 2021.
- [12] R. M. Sagala, "Prediksi Kelulusan Mahasiswa Menggunakan Data Mining Algoritma K-Means," *TeIka*, vol. 11, no. 2, pp. 131–142, 2021.
- [13] W. Gata, S. Surohman, and H. M. Nawawi, "Twitter in analysis of policy sentiments of the omnibus law work creative design," *AIP Conf. Proc.*, vol. 2714, no. September 2011, 2023.
- [14] M. P. A. Ariawan, I. B. A. Peling, and G. B. Subiksa, "Prediksi Nilai Akhir Matakuliah Mahasiswa Menggunakan Metode K-Means Clustering (Studi Kasus : Matakuliah Pemrograman Dasar)," *J. Nas. Teknol. dan Sist. Inf.*, vol. 9, no. 2, pp. 122–131, 2023.
- [15] R. Sovia, E. P. W. Mandala, and S. Mardhiah, "Algoritma K-Means dalam Pemilihan Siswa Berprestasi dan Metode SAW untuk Prediksi Penerima Beasiswa Berprestasi," *J. Edukasi dan Penelit. Inform.*, vol. 6, no. 2, p. 181, 2020.
- [16] Z. Wahidah and D. T. Utari, "Comparison of K-Means and Gaussian Mixture Model in Profiling Areas By Poverty Indicators," *BAREKENG J. Ilmu Mat. dan Terap.*, vol. 17, no. 2, pp. 0717–0726, 2023.
- [17] R. Novia, S. S. Prasetyowati, and Y. Sibaroni, "Identify User Behavior Based on The Type of Tweet on Twitter Platform Using Gaussian Mixture Model Clustering," *J. Comput. Syst. Informatics*, vol. 3, no. 4, pp. 502–506, 2022.
- [18] F. A. Fithri and S. Wardhana, "The CLUSTER ANALYSIS OF SALES TRANSACTION DATA USING K-MEANS CLUSTERING AT TOKO USAHA MANDIRI," *J. Pilar Nusa Mandiri*, vol. 17, no. 2, pp. 113–118, 2021.
- [19] T. D. Andini and L. Farokhah, "Peningkatan Ketersediaan Darah Sesuai Segmentasi Umur Menggunakan K-Means Clustering," *J. Manaj. Inform.*, vol. 12, no. 2, pp. 126–136, 2022.
- [20] A. Fatunnisa and H. Marcos, "Prediksi Kelulusan Tepat Waktu Siswa SMK Teknik Komputer Menggunakan Algoritma Random Forest," *J. Manaj. Inform.*, vol. 14, no. 1, pp. 101–111, 2024.
- [21] A. Prasatya, R. R. A. Siregar, and R. Arianto, "Penerapan Metode K-Means Dan C4.5 Untuk Prediksi Penderita Diabetes," *Petir*, vol. 13, no. 1, pp. 86–100, 2020.
- [22] H. Priyatman, F. Sajid, and D. Haldivany, "Klasterisasi Menggunakan Algoritma K-Means Clustering

-
- untuk Memprediksi Waktu Kelulusan Mahasiswa,” *J. Edukasi dan Penelit. Inform.*, vol. 5, no. 1, pp. 62–66, 2019.
- [23] F. N. Dhewayani, D. Amelia, D. N. Alifah, B. N. Sari, and M. Jajuli, “Implementasi K-Means Clustering untuk Pengelompokan Daerah Rawan Bencana Kebakaran Menggunakan Model CRISP-DM,” *J. Teknol. dan Inf.*, vol. 12, no. 1, pp. 64–77, 2022.
- [24] E. F. L. Awalina and W. I. Rahayu, “Optimalisasi Strategi Pemasaran dengan Segmentasi Pelanggan Menggunakan Penerapan K-Means Clustering pada Transaksi Online Retail,” *J. Teknol. dan Inf.*, vol. 13, no. 2, pp. 122–137, 2023.