

Optimasi K-Means Menggunakan Algoritma Firefly Untuk Segmentasi Pelanggan pada E-commerce

Dwi Tatang Warianta¹, Paramesti Astagina², Richy Julianto³, Florentina Yuni Arini⁴
^{1,2,3,4}Teknik Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Semarang
¹tatangwarianta@students.unnes.ac.id *, ²astagina30@students.unnes.ac.id, ³richyjuli03@students.unnes.ac.id,
⁴floyuna@mail.unnes.ac.id

Abstract

Customer segmentation is a crucial component in e-commerce for enabling personalized offers, enhancing customer loyalty, and supporting strategic decision-making. This study proposes the optimization of the K-Means algorithm using the Firefly algorithm to improve customer segmentation accuracy based on the Recency, Frequency, and Monetary (RFM) model. The K-Means algorithm is chosen for its efficiency in processing large-scale data, but it often struggles with determining the global optimal solution due to sensitivity to random centroid initialization. The Firefly algorithm is implemented to overcome this limitation through a broader exploration of the solution space and the ability to avoid local optima. The dataset used in this study consists of over 500,000 online retail transactions. The research stages include data cleaning, exploratory data analysis (EDA), and clustering using K-Means, Firefly, and their combination. Experimental results demonstrate that the combination of K-Means and Firefly achieves a consistently high Silhouette Score above 0.9 across 50 iterations, with better stability compared to individual algorithms. The segmentation results yield five clusters reflecting distinct customer characteristics such as transaction activity, frequency, and monetary contribution. This hybrid approach not only improves cluster quality but also provides more stable and structured outcomes. In conclusion, the combination of K-Means and Firefly offers an effective method for customer segmentation in e-commerce, providing a robust foundation for targeted marketing decisions and strategies to enhance customer loyalty.

Keywords: K-Means, Firefly, Customer Segmentation, E-Commerce, RFM

Abstrak

Segmentasi pelanggan menjadi komponen krusial dalam e-commerce untuk mendukung personalisasi penawaran, meningkatkan loyalitas pelanggan, dan mendukung keputusan strategis. Penelitian ini mengusulkan optimasi algoritma K-Means dengan algoritma Firefly untuk meningkatkan akurasi segmentasi pelanggan berdasarkan model Recency, Frequency, dan Monetary (RFM). Algoritma K-Means dipilih karena efisiensinya dalam memproses data berskala besar, namun sering menghadapi keterbatasan dalam menentukan solusi optimal global akibat sensitivitas terhadap inisialisasi centroid. Algoritma Firefly diimplementasikan untuk mengatasi kelemahan tersebut melalui eksplorasi ruang solusi yang lebih luas dan kemampuan menghindari jebakan solusi lokal optimal. Data yang digunakan dalam penelitian ini adalah dataset transaksi retail daring yang mencakup lebih dari 500.000 entri. Tahapan penelitian meliputi pembersihan data, analisis eksplorasi data (EDA), klusterisasi dengan K-Means, Firefly, dan kombinasi keduanya. Hasil eksperimen menunjukkan bahwa kombinasi K-Means dan Firefly menghasilkan nilai Silhouette Score yang konsisten di atas 0,9 sepanjang 50 iterasi, dengan stabilitas yang lebih baik dibandingkan algoritma individu. Segmentasi menghasilkan lima kluster yang mencerminkan karakteristik pelanggan yang unik, seperti aktivitas transaksi, frekuensi, dan kontribusi moneter. Metode hibrida ini tidak hanya memperbaiki kualitas kluster tetapi juga memberikan hasil yang lebih stabil dan terstruktur. Kesimpulannya, kombinasi K-Means dan Firefly menawarkan pendekatan efektif dalam segmentasi pelanggan e-commerce, memberikan landasan yang lebih kuat untuk pengambilan keputusan pemasaran yang terarah dan strategi peningkatan loyalitas pelanggan.

Kata kunci: K-Means, Firefly, Segmentasi Pelanggan, E-Commerce, RFM

©This work is licensed under a Creative Commons Attribution - ShareAlike 4.0 International License

1. Pendahuluan

Popularitas *e-commerce* semakin meningkat dari hari ke hari [1]. Seiring dengan perkembangan pesat teknologi digital, *e-commerce* kini telah menjadi salah satu pilar utama dalam perdagangan global. Platform *e-commerce* telah muncul sebagai media utama di mana berbagai perusahaan memfasilitasi penjualan produk dan layanan secara daring [2]. Pemahaman mendalam terhadap pelanggan menjadi kunci utama untuk mencapai keberhasilan. Dengan meningkatnya jumlah konsumen yang beralih ke platform daring untuk

memenuhi kebutuhan, perusahaan-perusahaan *e-commerce* dihadapkan pada tantangan untuk memahami perilaku dan preferensi pelanggan. Salah satu strategi yang efektif untuk memahami perilaku pelanggan dan meningkatkan kinerja pemasaran adalah dengan menerapkan segmentasi pelanggan dan penargetan pembeli [3].

Segmentasi pelanggan, proses mengelompokkan pelanggan berdasarkan karakteristik atau pola kesamaan pelanggan untuk mengidentifikasi kelompok yang memiliki karakteristik serupa agar mempermudah

pemasaran yang lebih efektif [4, 5]. Dengan mengelompokkan pelanggan ke dalam kelompok yang berbeda berdasarkan karakteristik yang sama, seperti minat, kebiasaan, dan pola pembelian, dapat membantu perusahaan lebih memahami kebutuhan dan perilaku pelanggan toko, sehingga dapat mengambil strategi pemasaran yang ditargetkan untuk meningkatkan kepuasan pelanggan dan penjualan [6, 7]. Segmentasi pelanggan dilakukan untuk menemukan kelompok pelanggan yang potensial dan paling menguntungkan di antara semua pelanggan [8].

Klasterisasi sebagai salah satu teknik yang populer digunakan dalam segmentasi pelanggan, dengan cara mengelompokkan objek ke dalam kelompok (klaster) berdasarkan kemiripan antar objek. Teknik klasterisasi termasuk metode yang paling tepat yang memungkinkan bisnis dan perusahaan untuk mengidentifikasi segmen atau kelompok pelanggan guna menargetkan basis pengguna potensial [9]. Dalam konteks *e-commerce*, segmentasi pelanggan dapat dilakukan berdasarkan model RFM (*Recency, Frequency, Monetary*), yang mengukur waktu terakhir pembelian (*Recency*), frekuensi pembelian (*Frequency*), dan jumlah uang yang dibelanjakan oleh pelanggan (*Monetary*).

Berdasarkan hasil penelitian sebelumnya, algoritma *K-Means* terbukti menjadi salah satu metode yang populer untuk melakukan pengelompokan data dalam skala besar dengan efisiensi tinggi, seperti pada kasus analisis data vaksinasi *COVID-19* di Indonesia. Dalam penelitian tersebut, algoritma ini menghasilkan klaster dengan nilai indeks *Davies Bouldin* terkecil sebesar 0,487, menunjukkan kualitas pengelompokan yang baik pada data vaksinasi dosis pertama, kedua, dan ketiga di berbagai kabupaten/kota di Indonesia [10]. Di sisi lain, penelitian lain menunjukkan bahwa algoritma *K-Medoids* memiliki keunggulan dibandingkan *K-Means*, terutama dalam mengatasi kelemahan terhadap *outlier* dan *noise* pada data kompleks, seperti dalam analisis bulan rawan kabut asap di Pekanbaru. Penelitian ini menggunakan *K-Medoids* untuk mengelompokkan data polusi udara berdasarkan PM10 dengan hasil yang divalidasi menggunakan *Silhouette Coefficient*, memberikan nilai 0,638 yang mencerminkan struktur data yang baik [11].

Metode *K-Means Clustering*, *Hierarchical Clustering*, dan *DBSCAN* memungkinkan analisis data yang kompleks untuk mengidentifikasi pola perilaku konsumen yang mungkin tidak terlihat sebelumnya. Di antara metode tersebut, Algoritma *K-Means* menjadi salah satu algoritma pengelompokan yang populer dan sering digunakan karena kesederhanaan dan efisiensinya dalam memproses data berskala besar [12, 13]. Algoritma ini bekerja sebagai algoritma optimasi, yang meminimalkan fungsi objektif sebagai kriteria optimasi dan memilih kombinasi optimal sebagai skema klasterisasi yang optimal [14]. Namun, algoritma ini memiliki beberapa kelemahan, terutama dalam penentuan jumlah klaster optimal dan

sensitivitas terhadap inialisasi centroid yang acak. Kelemahan ini dapat menyebabkan hasil segmentasi yang kurang akurat dan tidak konsisten. Untuk mengatasi masalah tersebut, beberapa algoritma metaheuristik telah digabungkan dengan algoritma *K-means* untuk meningkatkan kinerjanya [15].

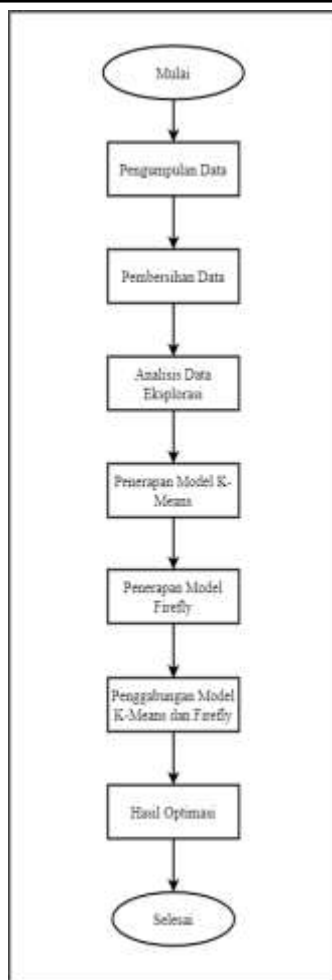
Untuk lebih meningkatkan performa algoritma *K-Means*, beberapa pendekatan telah diusulkan, salah satunya dengan menggabungkannya dengan algoritma optimasi metaheuristik seperti *Firefly*. Algoritma *Firefly* terinspirasi dari perilaku cahaya yang dihasilkan oleh kunang-kunang, di mana setiap kunang-kunang bergerak menuju cahaya yang lebih terang [16]. Dalam konteks optimasi, algoritma *Firefly* dapat membantu memperbaiki hasil pengelompokan dengan menjelajahi ruang solusi secara lebih luas dan menghindari terjebak dalam solusi lokal optimal [17].

Kelebihan dari algoritma *Firefly* dalam mengoptimalkan *K-Means* terletak pada kemampuannya untuk mencari solusi optimal secara lebih efisien dibandingkan algoritma tradisional lainnya [17]. *Firefly* memiliki keunggulan dalam menghindari solusi lokal optimal melalui mekanisme pencarian yang adaptif, di mana kunang-kunang bergerak menuju individu dengan intensitas cahaya lebih tinggi (solusi yang lebih baik). Hal ini memungkinkan algoritma *Firefly* untuk secara dinamis mengeksplorasi ruang solusi yang lebih luas, menghasilkan kelompok yang lebih stabil dan akurat dalam waktu yang relatif singkat.

Penelitian ini berfokus pada optimasi algoritma *K-Means* dengan memanfaatkan teknik *Firefly* untuk segmentasi pelanggan pada *e-commerce* berbasis model RFM. Algoritma *Firefly* terkenal karena kemampuannya dalam optimasi global, sementara *K-Means Clustering* efektif untuk pengelompokan data non-linear [18, 19]. Kinerja algoritma hibrida lebih unggul dibandingkan algoritma individu dalam hal ketahanan, efisiensi, dan akurasi [20]. Dengan pendekatan ini, diharapkan dapat diperoleh hasil pengelompokan pelanggan yang lebih baik, yang akan mendukung pengambilan keputusan bisnis dan strategi pemasaran yang lebih tepat sasaran.

2. Metode Penelitian

Pada penelitian ini ada beberapa tahapan yang dilakukan. Untuk mendapatkan pemahaman yang lebih detail mengenai proses penelitian yang dilakukan, dapat dilihat pada Gambar 1 yang menggambarkan setiap tahapan dalam proses penelitian.



Gambar 1. Flowchart Metode Penelitian

2.1. Pengumpulan Data

Data yang digunakan dalam penelitian ini berasal dari Kaggle. Dataset tersebut dapat diakses melalui Kaggle dengan tautan <https://www.kaggle.com/datasets/mashlyn/online-retail-ii-uci>. Dataset yang digunakan dalam penelitian ini berisi data transaksi penjualan yang terjadi di sebuah perusahaan retail online di Inggris. Data ini mencakup lebih dari 500.000 transaksi yang terjadi antara tanggal 01/12/2009 hingga 09/12/2011. Dataset ini terdiri dari 541.909 baris dan 8 kolom.

2.2. Pembersihan Data

Tahap pembersihan data diperlukan untuk memastikan data dalam *DataFrame* bersih, konsisten, dan valid sebelum dianalisis atau digunakan dalam model prediktif. Proses ini mencakup penghapusan nilai kosong (*missing values*) untuk mencegah bias yang dapat memengaruhi akurasi analisis, serta penghapusan baris duplikat agar hasil analisis tetap valid dan distribusi data tidak terganggu.

2.3. Analisis Data Eksplorasi

Tahap *Exploratory Data Analysis* (EDA) ini bertujuan untuk memahami lebih dalam karakteristik dan pola dalam data, khususnya terkait perilaku pelanggan.

Langkah-langkah dalam EDA ini dimulai dengan mengubah data tanggal menjadi format yang lebih mudah dianalisis, diikuti dengan pembuatan tiga metrik utama: *Recency* (seberapa lama sejak pembelian terakhir pelanggan), *Frequency* (jumlah transaksi yang dilakukan pelanggan), dan *Monetary* (total pengeluaran pelanggan). Metrik-metrik ini membantu dalam memahami kapan, seberapa sering, dan seberapa banyak pelanggan berbelanja.

2.4. Algoritma K-Means

K-Means sebagai salah satu algoritma klusterisasi yang banyak digunakan dalam analisis data, termasuk untuk segmentasi pelanggan di industri *e-commerce*, bertujuan untuk membagi data ke dalam beberapa kelompok (kluster) berdasarkan kemiripan fitur antar data dalam satu kelompok. Prinsip dasar *K-Means* dilakukan pengelompokan data dengan cara meminimalkan variasi data dalam setiap kluster dan memaksimalkan perbedaan antara kluster yang berbeda. Pada penelitian ini algoritma *K-Means* diterapkan untuk mengelompokkan data RFM (*Recency, Frequency, Monetary*) setelah melalui proses normalisasi.

Proses *K-Means* dimulai dengan inisialisasi sejumlah *k* *centroid* (titik pusat kluster) secara acak. Setiap titik data kemudian dikaitkan dengan *centroid* terdekat berdasarkan jarak *Euclidean*, yang dihitung menggunakan persamaan (1) [21].

$$d(Z, W) = \sqrt{\sum_{k=1}^d (Z_k - Z_w)^2} \quad (1)$$

Dimana *d* menunjukkan dimensi, *Z* menunjukkan titik data ke-*p* pada komponen ke-*k*, dan *W* menunjukkan *centroid* dari kluster ke-*j*.

Setelah semua data dikelompokkan, *centroid* baru dihitung dengan mengambil rata-rata dari data di dalam setiap kluster menggunakan persamaan (2) [21].

$$M = \frac{Z_1 + Z_2 + Z_3 + \dots + Z_n}{n} \quad (2)$$

Dimana *M* menunjukkan nilai rata-rata (*centroid*) dari kluster, *n* menunjukkan jumlah data dalam kluster. *Z* menunjukkan nilai dari data ke-*i* dalam kluster, dengan *i* = 1 sampai *n*.

Data tersebut akan ditugaskan ke kluster dengan jarak minimum, dihitung berdasarkan jarak terdekat ke *centroid*. Langkah ini diulang hingga posisi *centroid* tidak lagi berubah atau perubahan antara iterasi sudah sangat kecil. Langkah tersebut ditunjukkan dalam *pseudocode* pada Program 1.

Program 1. Algoritma K-Means

Inisialisasi centroid secara acak
Pilih *k* titik data secara acak sebagai *centroid* awal

Iterasi K-Means

```

while iterasi < MaksIterasi dan
KriteriaBerhenti = false do
    for setiap titik data x_i di dataset do
    
```

```
    Hitung jarak antara  $x_i$  dengan setiap centroid
    Tentukan centroid terdekat dan tetapkan  $x_i$  ke klaster tersebut
end for

perubahan = 0

for setiap klaster  $j = 1$  sampai  $k$  do
    Hitung ulang posisi centroid sebagai rata-rata dari semua titik data yang ditugaskan ke klaster tersebut
    Hitung perubahan posisi centroid
    if perubahan posisi centroid > perubahan then
        perubahan = perubahan posisi centroid
    end if
end for

if perubahan < toleransi then
    KriteriaBerhenti = true
end if

iterasi = iterasi + 1
end while
```

Tampilkan hasil (centroid dan klaster)

Berdasarkan Program 1 algoritma *K-Means* bekerja dengan memulai tahap inialisasi *centroid*, di mana k titik data dipilih secara acak dari dataset untuk dijadikan sebagai *centroid* awal. Setelah inialisasi, algoritma memasuki fase iterasi. Pada setiap iterasi, jarak antara setiap titik data dalam dataset dan setiap *centroid* dihitung menggunakan jarak *Euclidean*. Berdasarkan perhitungan jarak ini, setiap titik data kemudian dihubungkan dengan *centroid* terdekat dan ditetapkan ke klaster yang bersesuaian.

Setelah proses pengelompokan data, *centroid* diperbarui dengan menghitung posisi rata-rata dari seluruh titik data yang berada dalam klaster tersebut. Perubahan posisi *centroid* dari iterasi sebelumnya dihitung, dan jika perubahan ini melebihi nilai toleransi yang telah ditentukan, algoritma akan terus berjalan. Iterasi berlanjut hingga perubahan posisi *centroid* menjadi lebih kecil dari toleransi atau hingga jumlah iterasi mencapai batas maksimum.

2.5. Algoritma *Firefly*

Algoritma *Firefly* termasuk ke dalam algoritma metaheuristik yang terinspirasi dari perilaku kunang-kunang [22]. *Firefly* didasarkan pada perilaku daya tarik kunang-kunang tropis dan pola kilatan cahaya dari perilaku ideal kunang-kunang [23]. Algoritma ini termasuk dalam kategori algoritma *swarm intelligence*, yang mengandalkan interaksi dan kolaborasi antara individu dalam kelompok untuk mencapai tujuan bersama. Algoritma ini digunakan untuk mencari solusi optimal pada masalah optimasi. Setiap kunang-kunang dalam algoritma mewakili solusi potensial, dan kunang-kunang tertarik ke solusi-solusi yang lebih baik berdasarkan intensitas cahayanya.

Pada algoritma *Firefly*, sekelompok kunang-kunang diinisialisasi secara acak, dan setiap kunang-kunang menunjukkan satu solusi awal. Sebuah nilai kebugaran (*fitness*) dihitung berdasarkan fungsi objektif dari

setiap kunang-kunang, yang kemudian ditetapkan sebagai intensitas cahaya. Kunang-kunang dengan intensitas cahaya lebih rendah akan bergerak menuju kunang-kunang dengan intensitas cahaya lebih kuat yang berada di sekitarnya.

Pseudocode Algoritma *Firefly* ditunjukkan pada Program 2.

Program 2. Algoritma *Firefly*

```
Inialisasi parameter
Jumlah firefly ( $n$ )
Jumlah klaster ( $k$ )
Koefisien reduksi intensitas cahaya ( $\gamma$ )
Tingkat serapan cahaya ( $B$ )
Faktor randomisasi ( $a$ )
Iterasi maksimum ( $iterMax$ )
```

Inialisasi *firefly* secara acak di ruang pencarian

Evaluasi intensitas setiap *firefly* berdasarkan fungsi objektif klasterisasi

Perulangan untuk setiap iterasi ($iter = 1$ sampai $iterMax$)

```
for setiap firefly  $i$  ( $i = 1$  sampai  $n$ ) do
    for setiap firefly  $j$  ( $j = 1$  sampai  $n$ ) do
        if firefly  $j$  lebih terang dari firefly  $i$  then
            Hitung jarak antara firefly  $i$  dan  $j$ 
            Hitung perubahan intensitas cahaya berdasarkan jarak ( $\gamma$ )
            Perbarui posisi firefly  $i$  menuju firefly  $j$  berdasarkan  $B$  dan jarak
            Tambahkan randomisasi pada pergerakan menggunakan  $a$ 
        end if
        Evaluasi kembali intensitas firefly  $i$  setelah perpindahan
    end for
end for
```

Simpan *firefly* dengan intensitas terbaik

Cek kondisi penghentian ($iterMax$ tercapai)

Output

```
Firefly dengan intensitas terbaik sebagai solusi klastering (pusat klaster)
```

Pada Program 2 algoritma dimulai dengan proses pengambilan data RFM (*Recency, Frequency, Monetary*), sebuah model yang umum digunakan untuk menganalisis perilaku pelanggan dalam konteks pemasaran. Setelah data RFM diambil, dilakukan proses normalisasi untuk memastikan bahwa setiap fitur memiliki skala yang seimbang, sehingga algoritma dapat bekerja lebih efektif. Setelah data dipersiapkan, dilakukan inialisasi parameter seperti jumlah *firefly*, jumlah klaster yang diinginkan, koefisien pengurangan intensitas cahaya (γ), faktor randomisasi (a), dan jumlah iterasi maksimal yang ditentukan oleh pengguna. Kemudian, algoritma menginisialisasi posisi awal *firefly* (pusat klaster) secara acak di ruang pencarian. Langkah selanjutnya dilakukan evaluasi *firefly*, di mana setiap *firefly* dievaluasi berdasarkan intensitasnya. Update nilai kecocokan (*fitness*) dilakukan pada *firefly* yang kurang terang menuju *firefly* yang lebih terang. Posisi *firefly* kemudian diperbarui dalam proses perbandingan tersebut, dan

setelah setiap iterasi, *firefly* yang paling cocok (dengan intensitas terbaik) diperingkat ulang dan diperbarui posisinya. Proses ini diulang hingga mencapai iterasi maksimal (kondisi stop), yang memeriksa apakah algoritma telah menyelesaikan iterasi yang diizinkan. Jika iterasi maksimum tercapai, algoritma menghasilkan hasil optimal, dengan *firefly* intensitas terbaik yang mewakili solusi klusterisasi.

3. Hasil dan Pembahasan

Penelitian ini dilakukan untuk menganalisis segmentasi pelanggan menggunakan model RFM (*Recency, Frequency, Monetary*) dengan tiga pendekatan algoritma klusterisasi yang berbeda: *K-Means*, Algoritma *Firefly*, dan kombinasi antara keduanya. Analisis dilakukan dengan mengelompokkan pelanggan ke dalam segmen-segmen tertentu dengan karakteristik yang mirip antar pelanggan di tiap segmennya dalam setiap dimensi dari RFM. Hasil analisis nantinya dapat digunakan untuk mengambil keputusan bisnis yang terkait dengan pelanggan, misalnya pemasaran yang lebih terarah ke segmen tertentu.

3.1. Pengumpulan Data

Dataset ini terdiri dari 541.909 baris dan 8 kolom, serta terdapat 8 (delapan) fitur dalam dataset seperti Gambar 2.

ID	Invoice	Description	Quantity	InvoiceDate	Price	CustomerID	Country
0	460004	WINE CHRISTMAS (SANGRE DE LOS REYES)	12	2000-12-28 07:45:00	0.94	1	United Kingdom
1	460004	WINE CHRISTMAS (SANGRE DE LOS REYES)	12	2000-12-28 07:45:00	0.94	1	United Kingdom
2	460004	WINE CHRISTMAS (SANGRE DE LOS REYES)	12	2000-12-28 07:45:00	0.94	1	United Kingdom
3	460004	WINE CHRISTMAS (SANGRE DE LOS REYES)	12	2000-12-28 07:45:00	0.94	1	United Kingdom
4	460004	WINE CHRISTMAS (SANGRE DE LOS REYES)	12	2000-12-28 07:45:00	0.94	1	United Kingdom

Gambar 2. Sampel Dataset

Dengan detail sebagai berikut:

- Invoice: Nomor faktur transaksi.
- StockCode: Kode produk.
- Description: Nama produk.
- Quantity: Jumlah produk yang dibeli.
- InvoiceDate: Tanggal transaksi.
- Price: Harga per unit produk.
- CustomerID: ID unik pelanggan.
- Country: Negara asal pelanggan.

3.2. Pembersihan Data

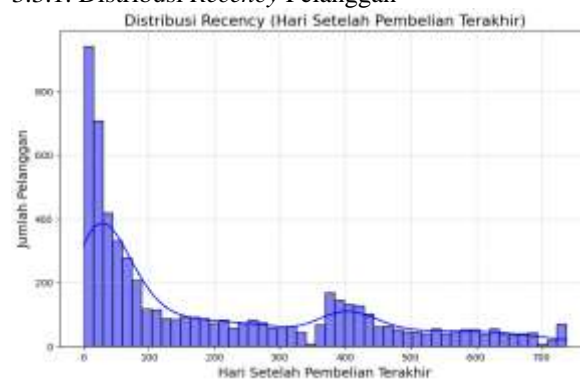
Tahap ini digunakan untuk membersihkan dan memproses data dalam sebuah *DataFrame*. *Data cleaning* bertujuan memastikan bahwa data yang akan digunakan dalam analisis atau model prediktif bersih, konsisten, dan valid. Menghapus nilai kosong (*missing values*) penting untuk menghindari kesalahan analisis yang disebabkan oleh data yang tidak lengkap. Jika ada nilai yang kosong dalam dataset, analisis atau model prediktif yang dihasilkan bisa menjadi bias atau kurang akurat. Selain itu, penghapusan baris duplikat diperlukan karena baris yang terduplikasi dapat membuat hasil analisis tidak valid, dengan menghitung observasi lebih dari sekali yang dapat mengganggu distribusi data dan hasil model. Validasi kolom

'Quantity' dan 'Price' juga diperlukan untuk memastikan hanya transaksi yang valid digunakan dalam analisis. Baris dengan nilai 'Quantity' atau 'Price' yang kurang dari atau sama dengan nol dihapus karena nilai-nilai tersebut tidak sesuai dengan konsep penjualan yang rasional, dimana kedua kolom ini harus bernilai positif. Penambahan fitur 'TotalPrice', yang dihitung sebagai hasil perkalian antara 'Quantity' dan 'Price', sangat berguna untuk analisis lanjutan, terutama untuk memahami total pengeluaran setiap transaksi, sebagai aspek penting dalam mengevaluasi perilaku pelanggan.

3.3. Analisis Data Eksplorasi

Setelah melakukan pembersihan data, selanjutnya masuk ke tahap analisis data eksplorasi, untuk mengetahui variasi signifikan dalam keterlibatan dan kontribusi pelanggan, dilakukan analisis distribusi terhadap setiap dimensi dari RFM. Hasil analisis ditunjukkan pada gambar 3, 4, dan 5 yang secara berurutan menunjukkan histogram distribusi *Recency, Frequency, Monetary*.

3.3.1. Distribusi *Recency* Pelanggan

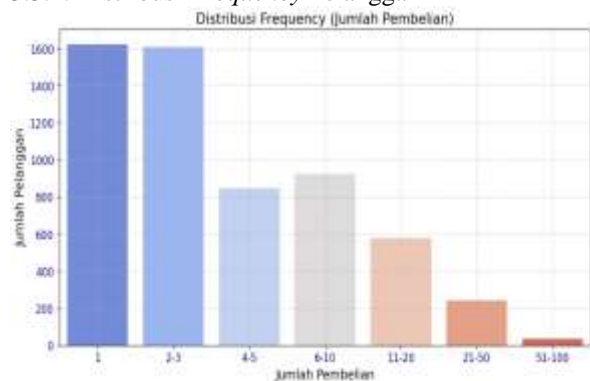


Gambar 3. Histogram Distribusi *Recency* Pelanggan

Gambar 3 menunjukkan histogram distribusi pelanggan berdasarkan *Recency*. Histogram ini menunjukkan bahwa terdapat kecenderungan besar pelanggan membeli dalam waktu dekat setelah pembelian terakhir. Mayoritas pelanggan membeli dalam rentang yang lebih kecil sekitar 0-100 hari. Oleh karena itu, histogram menunjukkan frekuensi tinggi pada bagian kiri. Setelah 100 hari pembelian, frekuensi menurun dan terlihat pada puncak kedua di sekitar 400 hari yang menunjukkan pelanggan lama yang kembali membeli setelah jangka waktu yang cukup lama. Dari histogram, terlihat bahwa sebagian besar pelanggan melakukan pembelian dalam waktu yang relatif dekat, dengan puncak frekuensi pada kategori 0 hingga 30 hari, di mana jumlah pelanggan hampir mencapai 1000. Seiring bertambahnya jumlah hari setelah pembelian terakhir, frekuensi pelanggan cenderung menurun, menunjukkan bahwa semakin lama waktu berlalu, semakin sedikit pelanggan yang kembali bertransaksi. Kategori antara 30 hingga 200 hari masih menunjukkan jumlah pelanggan yang signifikan, tetapi frekuensinya menurun secara bertahap. Setelah 200 hari, jumlah pelanggan yang melakukan pembelian semakin sedikit,

dengan frekuensi yang sangat rendah pada rentang 600 hingga 700 hari. Histogram ini juga dilengkapi dengan kurva distribusi yang menurun, mencerminkan pola yang sama dan mengindikasikan bahwa pelanggan yang bertransaksi baru-baru ini cenderung memiliki frekuensi yang lebih tinggi. Temuan ini memberikan wawasan penting bagi pemasar untuk mengidentifikasi pelanggan yang mungkin perlu ditargetkan dengan penawaran khusus guna mendorong pembelian kembali, serta membantu dalam merancang strategi retensi pelanggan yang lebih efektif.

3.3.2. Distribusi *Frequency* Pelanggan



Gambar 4. Histogram Distribusi Frequency Pelanggan

Gambar 4 menunjukkan histogram distribusi frekuensi jumlah pembelian. Pada sumbu horizontal terdapat jumlah pembelian yang dibagi menjadi 7 kelompok dengan intervalnya masing-masing. Sementara pada sumbu vertikal terdapat jumlah pelanggan yang melakukan pembelian pada setiap interval. Dari histogram tersebut, terlihat bahwa kategori 1 memiliki frekuensi tertinggi dengan lebih dari 1600 pelanggan, menunjukkan bahwa banyak pelanggan hanya melakukan satu kali pembelian. Kategori 2-3 juga menunjukkan frekuensi yang cukup tinggi, tetapi menurun tajam pada kategori-kategori selanjutnya, seperti 6-10, 11-20, 21-50, dan 51-100, yang menunjukkan jumlah pelanggan yang semakin sedikit. Hal ini mengindikasikan bahwa mayoritas pelanggan cenderung melakukan sedikit pembelian, dengan sebagian besar diantaranya melakukan satu kali pembelian saja. Mayoritas pelanggan hanya membeli sekali atau 2-3 kali, dengan sekitar 1600 pelanggan pada masing-masing kelompok. Setelahnya mengalami penurunan frekuensi pembelian signifikan sekitar 50% dan terus menurun secara signifikan.

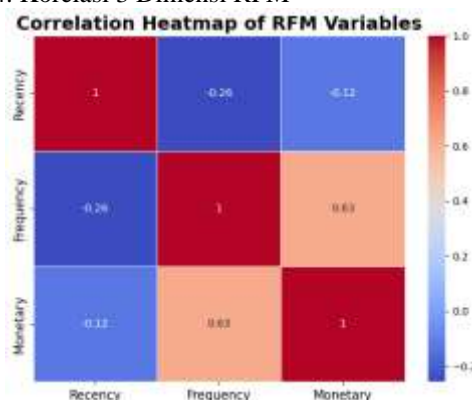
3.3.3. Distribusi *Monetary* Pelanggan



Gambar 5. Histogram Distribusi Monetary Pelanggan

Gambar 5 menunjukkan histogram distribusi moneter total pengeluaran pelanggan. Pada sumbu x terdapat total pengeluaran pelanggan yang dibagi dengan beberapa interval. Sementara pada sumbu y terdapat jumlah pelanggan dengan rentang 0-1600. Berdasarkan histogram tersebut, terdapat puncak pada interval 1001-3000 total pengeluaran dengan jumlah sekitar 1500 pelanggan. Namun, setelahnya mengalami penurunan kuantitas pelanggan dengan signifikan pada interval total pengeluaran di atas 3001. Lalu pada kelompok interval 0-1000, histogram menunjukkan jumlah yang stabil dan konsisten dengan rata-rata 900 pelanggan. Dengan histogram tersebut dapat diketahui bahwa kelompok pelanggan dengan jumlah yang masif terdapat pada 1001-3000 total pengeluaran. Temuan ini mengindikasikan bahwa mayoritas pelanggan cenderung memiliki total pengeluaran yang relatif rendah hingga sedang, dengan puncak frekuensi pada kategori menengah. Hal ini memberikan wawasan bagi pemasar untuk merancang strategi yang mendorong pelanggan agar meningkatkan pengeluaran, serta mengidentifikasi segmen pelanggan yang potensial untuk ditargetkan dengan penawaran khusus guna meningkatkan loyalitas dan frekuensi pembelian.

3.3.4. Korelasi 3 Dimensi RFM



Gambar 6. Correlation Heatmap 3 Dimensi RFM

Gambar 6 menunjukkan *Correlation Heatmap* dari tiga variabel RFM (*Recency*, *Frequency*, *Monetary*). *Correlation Heatmap* menggambarkan representasi grafis yang menunjukkan hubungan antar variabel numerik dalam bentuk matriks. Dalam *heatmap*,

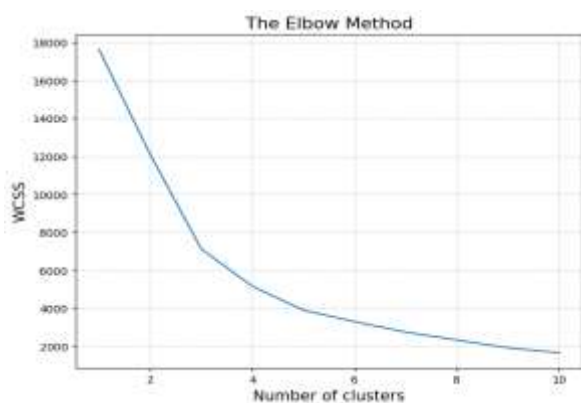
korelasi diwakili oleh warna, dimana warna merah tua menunjukkan korelasi positif yang kuat (mendekati 1), sedangkan warna biru menunjukkan korelasi negatif yang kuat (mendekati -1). Semakin terang atau mendekati putih warnanya, semakin lemah korelasi antara dua variabel tersebut.

Korelasi berkaitan dengan ukuran statistik yang menunjukkan sejauh mana dua variabel saling berkaitan secara linear. Dimana semakin mendekati merah atau 1 diindikasikan bahwa terdapat hubungan yang kuat antara dua variabel. Pada gambar diatas diketahui bahwa beberapa variabel memiliki korelasi yang cukup kuat sehingga menunjukkan adanya keterkaitan antar variabel. Terlihat hubungan sesama variabel terindikasi 1 yang artinya memiliki keterkaitan yang kuat, lalu pada *Recency* kurang memiliki keterkaitan yang kuat pada variabel lain. Sementara pada frekuensi memiliki keterkaitan pada moneter sehingga keduanya memiliki korelasi yang cukup kuat, dimana semakin tinggi jumlah pembelian pelanggan (*Frequency*), akan semakin tinggi juga total pengeluaran pelanggan (*Monetary*).

Pada heatmap yang menampilkan dimensi RFM (*Recency, Frequency, Monetary*), bisa dilihat hubungan antar variabel-variabel tersebut. Misalnya, dalam analisis ini, korelasi antara *Recency* dan *Frequency* menunjukkan nilai negatif (-0.26), artinya semakin sering pelanggan bertransaksi, semakin baru (lebih rendah) *Recency*-nya. Korelasi antara *Recency* dan *Monetary* sangat rendah (-0.12), menunjukkan bahwa waktu terakhir pelanggan bertransaksi tidak terlalu terkait dengan total nilai transaksi. Sebaliknya, terdapat korelasi positif yang cukup signifikan (0.63) antara *Frequency* dan *Monetary*, yang menunjukkan bahwa pelanggan yang lebih sering bertransaksi cenderung menghasilkan nilai transaksi yang lebih tinggi.

3.4. Segmentasi Menggunakan *K-Means*

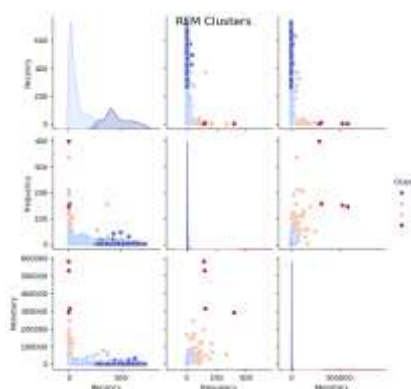
Pada penelitian ini, *K-Means* sebagai salah satu metode klusterisasi dalam melakukan segmentasi pelanggan menggunakan model RFM. Untuk menentukan jumlah kluster yang optimal, digunakan Metode *Elbow* yang ditunjukkan pada Gambar 7.



Gambar 7. Metode Elbow K-Means

Gambar 7 menunjukkan penggunaan Metode *Elbow* untuk menentukan jumlah kluster optimal pada algoritma *K-Means*. Sumbu x mewakili jumlah kluster dan sumbu y menunjukkan *Within-Cluster Sum of Squares* (WCSS) yang menggambarkan jarak titik data dengan *centroid*. Nilai WCSS akan menurun seiring bertambahnya kluster dan akan membentuk *elbow*. Pada gambar 7, titik *elbow* muncul pada kluster 3 atau 4 yang menunjukkan bahwa kluster optimal untuk data ini terdapat pada 3 atau 4 kluster. Penggunaan Metode *Elbow* penting untuk menyeimbangkan kompleksitas model dan kualitas klusterisasi.

Setelah menentukan jumlah kluster optimal, hasil klusterisasi divisualisasikan dalam pairplot yang memperlihatkan hubungan antara dimensi *Recency, Frequency, dan Monetary* sebagaimana ditunjukkan pada Gambar 8.



Gambar 8. Pairplot Segmentasi Menggunakan *K-Means*

Pada Gambar 8 menunjukkan distribusi pelanggan dalam empat kluster berdasarkan analisis RFM (*Recency, Frequency, Monetary*). Setiap kluster disimbolkan dengan warna yang berbeda untuk membedakan karakteristik pelanggan. Kluster 0 terdiri dari pelanggan dengan *recency* tinggi dan frekuensi serta *monetary* rendah, menunjukkan pelanggan yang kurang aktif. Kluster 1 memiliki pelanggan dengan *recency* rendah hingga sedang dan *monetary* sedang, mengindikasikan pelanggan yang lebih aktif belakangan ini. Kluster 2 diisi oleh pelanggan dengan frekuensi tinggi dan *monetary* sedang hingga tinggi, menunjukkan kelompok yang lebih bernilai dan aktif. Terakhir, kluster 3 mencakup pelanggan dengan *monetary* tertinggi dan frekuensi cukup tinggi, yang menunjukkan pelanggan bernilai tinggi dengan aktivitas transaksi yang konsisten. Hasil lengkap segmentasi pelanggan menggunakan *K-Means* ditunjukkan pada Tabel 1.

Tabel 1. Hasil Segmentasi Menggunakan *K-Means*

	Recency	Frequency	Monetary
0	462.03	2.21	748.54
1	66.05	7.27	2921.08
2	23.26	100.13	77728.27
3	2.50	212.50	428611.00

Pada Tabel 1 menunjukkan hasil segmentasi pelanggan menggunakan algoritma *K-Means* berdasarkan tiga variabel: *Recency* (seberapa lama pelanggan melakukan pembelian terakhir), *Frequency* (seberapa sering pelanggan melakukan pembelian), dan *Monetary* (jumlah uang yang dihabiskan oleh pelanggan). Dari tabel tersebut didapatkan empat kluster pelanggan dengan karakteristik yang berbeda beda.

- a. Kluster 0: Pelanggan Terlupakan
Recency tinggi (462.03), menunjukkan pelanggan jarang melakukan transaksi. *Frequency* rendah (2.21), berarti pelanggan melakukan pembelian sangat jarang. *Monetary* rendah (748.54), artinya nilai transaksi yang dilakukan oleh pelanggan ini relatif kecil.
- b. Kluster 1: Pelanggan Pasif
Recency relatif tinggi (66.05), menunjukkan bahwa pelanggan tidak terlalu sering melakukan transaksi. *Frequency* relatif rendah (7.27), menunjukkan pelanggan jarang melakukan transaksi. *Monetary* relatif rendah (2921.08), berarti nilai transaksi dari pelanggan ini juga tidak signifikan.
- c. Kluster 2: Pelanggan Setia
Recency sedang (23.26), artinya pelanggan ini cukup aktif bertransaksi. *Frequency* sedang (100.13), menunjukkan pelanggan melakukan transaksi dengan frekuensi sedang. *Monetary* sedang (77728.27), pelanggan ini memberikan kontribusi nilai transaksi yang sedang.
- d. Kluster 3: Pelanggan VIP
Recency rendah (2.50), menunjukkan pelanggan ini sangat sering melakukan pembelian. *Frequency* tinggi (212.50), berarti pelanggan ini sangat aktif melakukan transaksi. *Monetary* tinggi (428611.00), pelanggan ini menghasilkan nilai transaksi yang sangat besar.

Secara keseluruhan, pelanggan dalam Kluster 3 menunjukkan pelanggan yang paling aktif, loyal, dan bernilai tinggi bagi bisnis, yang menjadikannya kluster yang paling penting untuk dipertahankan dan dijaga loyalitasnya. Bisnis sebaiknya fokus memberikan perhatian lebih pada pelanggan di kluster ini karena pelanggan memberikan dampak paling signifikan terhadap pendapatan.

3.5. Segmentasi Menggunakan Algoritma *Firefly*

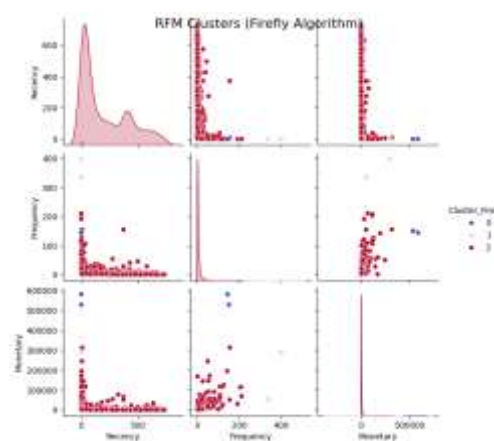
Setelah melakukan segmentasi menggunakan *K-Means*, dilakukan segmentasi dengan algoritma *Firefly*, sebuah algoritma berbasis *swarm* yang mampu melakukan klusterisasi secara efektif pada data yang kompleks. Untuk menentukan jumlah kluster optimal, *silhouette score* digunakan yang hasilnya terdapat pada Tabel 2.

Tabel 2. Silhouette Score Algoritma *Firefly*

Cluster	Silhouette Score
2	0.962
3	0.963
4	0.958
5	0.950
6	0.956

Tabel 2 menunjukkan *silhouette score* untuk masing-masing kluster dari algoritma *Firefly* dari kluster 2-6. Berdasarkan tabel tersebut, diperoleh *silhouette score* tertinggi pada kluster 3 dengan angka 0.963 yang menunjukkan bahwa kluster tersebut menunjukkan kluster paling optimal. Oleh karena itu, kluster 3 akan digunakan dalam analisis menggunakan algoritma *Firefly*.

Setelah kluster optimal ditemukan, hasil dari klusterisasi divisualisasikan melalui pairplot sesuai dengan gambar 9.



Gambar 9. Pairplot Segmentasi Menggunakan Algoritma *Firefly*

Gambar 9 menampilkan pairplot segmentasi pelanggan berdasarkan algoritma *Firefly* untuk analisis RFM (*Recency*, *Frequency*, *Monetary*). Dimensi *Recency* menunjukkan seberapa lama sejak terakhir kali pelanggan melakukan transaksi. Pada kluster 0 (titik merah), terlihat bahwa beberapa pelanggan memiliki nilai *Recency* yang tinggi, artinya pelanggan sudah lama tidak bertransaksi. Namun, dimensi lainnya tetap relevan. Dimensi *Frequency* menggambarkan seberapa sering pelanggan bertransaksi selama periode tertentu, dan pada kluster 0, terdapat pelanggan dengan nilai *Frequency* yang rendah hingga sedang, menandakan bahwa pelanggan tidak sering bertransaksi. Meski demikian, pada dimensi *Monetary*, kluster 0 menunjukkan banyak pelanggan dengan nilai transaksi total yang tinggi (lebih dari 500.000), menandakan kontribusi besar terhadap pendapatan, meskipun frekuensi transaksi pelanggan rendah atau tidak aktif lagi.

Hasil lengkap segmentasi pelanggan menggunakan algoritma *Firefly* ditunjukkan pada Tabel 3.

Tabel 3. Hasil Segmentasi Menggunakan Algoritma Firefly

	Recency	Frequency	Monetary
0	6096.78	125.09	512226.95
1	3569.24	165.72	145445.35
2	811.92	26.48	123246.03

Tabel 3 menunjukkan hasil segmentasi pelanggan menggunakan algoritma *Firefly* dengan tiga kluster berdasarkan dimensi *Recency*, *Frequency*, dan *Monetary*.

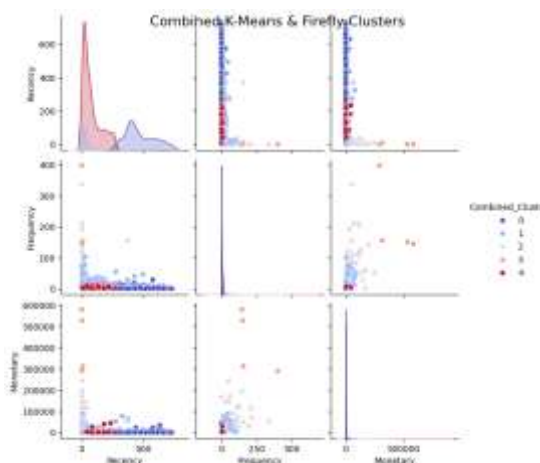
- a. Kluster 0: Pelanggan Elite Terlupakan
Recency sangat tinggi (6096,78), yang berarti pelanggan sudah lama tidak bertransaksi, tetapi dengan *Frequency* sedang (125,09) dan *Monetary* sangat tinggi (512.226,95). Pelanggan ini memberikan kontribusi besar pada pendapatan, meskipun sudah tidak aktif, sehingga perusahaan perlu melakukan strategi *re-engagement* melalui program loyalitas atau penawaran eksklusif untuk menghidupkan kembali transaksi pelanggan.
- b. Kluster 1: Pelanggan Kontributor Besar
Recency sangat tinggi (3569,24), *Frequency* tinggi (165,72), dan *Monetary* sedang (145.445,35). Pelanggan ini cukup sering bertransaksi dan memberikan kontribusi yang signifikan, sehingga perlu diprioritaskan dengan program loyalitas untuk menjaga keterlibatan pelanggan.
- c. Kluster 2: Pelanggan Berpengaruh
Recency tinggi (811,92), *Frequency* sedang (26,48), dan *Monetary* sedang (123.246,03). Kluster ini terdiri dari pelanggan yang masih terhubung dengan perusahaan, meski dengan frekuensi dan nilai transaksi yang belum optimal. Mereka berpotensi berkembang menjadi pelanggan bernilai tinggi jika diberikan promosi dan program *up-selling* yang tepat.

Dengan memahami pola transaksi setiap kluster, perusahaan dapat menyusun strategi pemasaran yang lebih efektif, seperti mengaktifkan kembali pelanggan bernilai tinggi di kluster 0, mempertahankan keterlibatan pelanggan kontributor di kluster 1, dan meningkatkan frekuensi serta nilai transaksi pelanggan di kluster 2. Segmentasi ini membantu perusahaan mengalokasikan sumber daya secara tepat untuk mengoptimalkan pertumbuhan dan loyalitas pelanggan.

3.6. Segmentasi Menggunakan Kombinasi *K-Means* dan Algoritma *Firefly*

Setelah melakukan segmentasi menggunakan *K-Means* dan algoritma *Firefly*, kedua algoritma tersebut dikombinasikan untuk memperoleh kecepatan dan akurasi dari *K-Means* dengan kemampuan optimasi dari algoritma *Firefly*. Tujuannya untuk mendapatkan hasil yang lebih stabil dengan akurasi yang baik.

Hasil analisis menggunakan kombinasi *K-Means* dan algoritma *Firefly* divisualisasikan melalui pairplot sesuai dengan gambar 10.



Gambar 10. Pairplot Segmentasi Menggunakan Kombinasi *K-Means* dan Algoritma *Firefly*

Berdasarkan Gambar 10 dapat diketahui bahwa setiap kluster memiliki karakteristik yang unik berdasarkan dimensi *Recency*, *Frequency*, dan *Monetary*. Pelanggan dengan kluster 0 menunjukkan pelanggan yang sudah lama tidak aktif dengan *Recency* yang tinggi. Sementara pelanggan kluster 3 menunjukkan pelanggan yang sangat aktif dengan *Recency* yang rendah dan *Monetary* yang tinggi. Visualisasi tersebut menunjukkan adanya lima kluster (0, 1, 2, 3, dan 4) yang masing-masing memiliki karakteristik unik berdasarkan tiga dimensi utama: *Recency*, *Frequency*, dan *Monetary*. Kluster 0 (warna abu-abu) terdiri dari pelanggan dengan *Recency* tinggi dan *Frequency* serta *Monetary* rendah, menandakan bahwa pelanggan sudah lama tidak bertransaksi dan memiliki kontribusi nilai yang kecil. Pelanggan dalam kluster ini cenderung kurang relevan dan membutuhkan strategi *re-engagement* jika masih ingin dipertahankan.

Hasil lengkap segmentasi, ditunjukkan pada Tabel 4.

Tabel 4. Hasil Segmentasi Menggunakan Kombinasi *K-Means* dan Algoritma *Firefly*

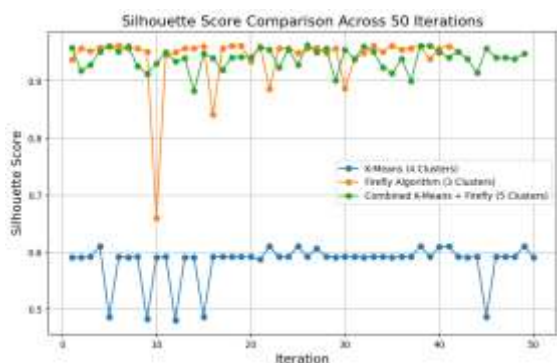
	Recency	Frequency	Monetary
0	475.41	2.23	1111.34
1	23.27	28.58	12801.73
2	32.32	119.34	98740.15
3	9.71	212.80	427960.32
4	74.81	5.04	1923.49

Tabel 4 menampilkan hasil segmentasi pelanggan menggunakan kombinasi *K-Means* dan algoritma *Firefly* dengan klusterisasi seperti berikut.

- a. Kluster 0: Pelanggan Terlupakan
Terdiri dari pelanggan yang sudah lama tidak bertransaksi, jarang melakukan pembelian,

- dan mengeluarkan sedikit uang. Pelanggan cenderung tidak aktif dan jarang berbelanja.
- b. Kluster 1: Pelanggan Reguler
Sering melakukan transaksi, baru-baru ini berbelanja, dan menghabiskan jumlah uang yang lebih besar dibandingkan dengan Kluster 0. Pelanggan pada kluster 1 menunjukkan pelanggan yang aktif dan memberikan kontribusi yang stabil.
 - c. Kluster 2: Pelanggan Berpotensi
Menunjukkan perilaku yang positif dengan frekuensi pembelian yang tinggi dan pengeluaran yang signifikan. Pelanggan pada kluster 2 menunjukkan pelanggan yang baru saja bertransaksi dan memiliki potensi untuk menjadi pelanggan yang lebih bernilai di masa depan.
 - d. Kluster 3: Pelanggan Premium
Kelompok pelanggan paling berharga. Pelanggan pada kluster 3 sering melakukan pembelian, baru-baru ini bertransaksi, dan menghabiskan jumlah uang yang sangat besar, menjadikannya segmen yang sangat penting bagi perusahaan.
 - e. Kluster 4: Pelanggan Pasif
sudah lama tidak berbelanja, jarang melakukan pembelian, dan mengeluarkan uang dalam jumlah kecil. Pelanggan pada kluster 4 menunjukkan pelanggan yang kurang aktif dan berisiko tidak lagi bertransaksi.

Setelah melakukan segmentasi menggunakan 3 algoritma yang berbeda, untuk mengetahui performa dari masing-masing algoritma yang digunakan, dilakukan perbandingan antara ketiga algoritma: *K-Means*, algoritma *Firefly*, dan kombinasi antara *K-Means* dan algoritma *Firefly*. *Silhouette score* digunakan sebagai salah satu parameter untuk mengukur performa suatu algoritma dalam melakukan segmentasi. Semakin tinggi *silhouette score* menunjukkan performa algoritma yang semakin. Perbandingan performa ketiga algoritma ditunjukkan melalui gambar 11.



Gambar 11. Perbandingan Silhouette Score

Gambar 11 menunjukkan perbandingan ketiga algoritma klusterisasi: *K-Means*, algoritma *Firefly*, dan kombinasi antara *K-Means* dan algoritma *Firefly*,

menggunakan *silhouette score* sebagai parameter. *Silhouette Score* digunakan sebagai metrik untuk mengukur seberapa baik sebuah klusterisasi pada data. Metrik ini mengevaluasi konsistensi dalam kluster dengan membandingkan jarak rata-rata antara data dalam satu kluster dan jarak rata-rata dengan kluster terdekat lainnya. Proses ini dilakukan melalui iterasi sebanyak 50, untuk memberikan gambaran yang jelas berkaitan dengan stabilitas dan kualitas klusterisasi masing-masing algoritma.

Dari hasil visualisasi tersebut, terlihat bahwa *K-Means* menunjukkan *silhouette score* yang konsisten di bawah kedua algoritma lain, dengan rata-rata di bawah angka 0.6. Selain nilai yang rendah, *K-Means* juga menunjukkan adanya ketidakstabilan yang signifikan, terlihat pada beberapa iterasi yang nilainya turun cukup tajam di bawah angka 0.5. Hal ini mengindikasikan bahwa *K-Means* tidak cukup optimal untuk memisahkan kluster yang berbeda.

Berbeda dengan *K-Means*, algoritma *Firefly* menunjukkan *silhouette score* yang jauh lebih tinggi dari *K-Means* dengan rata-rata nilai 0.9. Hasil ini menunjukkan algoritma *Firefly* memiliki kemampuan yang lebih baik daripada *K-Means* dalam melakukan klusterisasi. Meskipun algoritma *Firefly* memiliki *silhouette score* yang tinggi, ada beberapa iterasi yang menunjukkan algoritma ini menunjukkan ketidakkonsistenan dan menyebabkan *silhouette score* mengalami penurunan yang sangat signifikan hingga di bawah angka 0.7.

Adapun pendekatan dengan menggunakan kombinasi *K-Means* dan algoritma *Firefly* menunjukkan nilai yang paling optimal dan stabilitas yang terjaga dengan sangat baik. *Silhouette score* menunjukkan nilai yang cukup konsisten di atas 0.9 sepanjang iterasi, tanpa mengalami penurunan signifikan seperti pada algoritma *Firefly*. Kombinasi ini memanfaatkan keunggulan inisialisasi *centroid* dari *K-Means* dan kemampuan algoritma *Firefly* dalam mengoptimalkan pemisahan kluster, menghasilkan algoritma yang tidak hanya akurat, namun juga stabil dari iterasi awal hingga akhir.

Secara keseluruhan kombinasi dari *K-Means* dan algoritma *Firefly* menunjukkan hasil yang sangat superior dibandingkan dengan dua algoritma lainnya dalam melakukan klusterisasi. Dengan *silhouette score* yang tinggi dan performa yang sangat stabil, algoritma ini memberikan hasil segmentasi yang lebih akurat dan dapat diandalkan untuk pengambilan keputusan bisnis.

4. Kesimpulan

Kesimpulan dari penelitian ini menunjukkan bahwa pengoptimalan algoritma *K-Means* dengan algoritma *Firefly* pada segmentasi pelanggan di *e-commerce* berbasis model RFM menghasilkan performa yang lebih unggul. Penggunaan *K-Means* sendiri memberikan *silhouette score* rata-rata di bawah 0.6, dengan ketidakstabilan yang signifikan, terlihat dari

penurunan tajam hingga di bawah 0.5 pada beberapa iterasi. Sebaliknya, algoritma *Firefly* menunjukkan performa lebih baik dengan *silhouette score* rata-rata 0.9, meskipun ada beberapa penurunan hingga di bawah 0.7 pada iterasi tertentu. Kombinasi *K-Means* dan *Firefly* memberikan hasil paling optimal dan stabil, dengan *silhouette score* di atas 0.9 yang konsisten di seluruh iterasi. Pendekatan kombinasi ini memanfaatkan inialisasi *centroid* dari *K-Means* dan optimasi *Firefly* untuk pemisahan kluster yang lebih akurat. Hasil ini menghasilkan segmentasi yang presisi dan stabil, sehingga lebih andal untuk pengambilan keputusan bisnis berbasis data, seperti pengembangan strategi pemasaran dan peningkatan loyalitas pelanggan melalui segmentasi yang tepat.

Daftar Rujukan

- [1] V. Urbancokova, M. Kompan, Z. Trebulova, and M. Bielikova, "BEHAVIOR-BASED CUSTOMER DEMOGRAPHY PREDICTION IN E-COMMERCE." [Online]. Available: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L:2016:119:TOC>
- [2] Y. J. Purnomo, "Digital Marketing Strategy to Increase Sales Conversion on E-commerce Platforms," *Journal of Contemporary Administration and Management (ADMAN)*, vol. 1, no. 2, pp. 54–62, Aug. 2023, doi: 10.61100/adman.v1i2.23.
- [3] T. Jiang and A. Tuzhilin, "Improving Personalization Solutions through Optimal Segmentation of Customer Bases," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 3, pp. 305–320, March 2009, doi: 10.1109/TKDE.2008.163.
- [4] V. Vijilesh, A. Harini, M. H. Dharshini, and R. Priyadharshini, "CUSTOMER SEGMENTATION USING MACHINE LEARNING," *International Research Journal of Engineering and Technology*, 2021, [Online]. Available: www.irjet.net
- [5] K. Tabianan, S. Velu, and V. Ravi, "K-Means Clustering Approach for Intelligent Customer Segmentation Using Customer Purchase Behavior Data," *Sustainability (Switzerland)*, vol. 14, no. 12, Jun. 2022, doi: 10.3390/su14127243.
- [6] B. Jamalpur, D. Singh, B. S. Kumar, A. Nagpal, S. Pawar, and D. Banerjee, "Applications of Deep Learning in Marketing Analytics: Predictive Modeling and Segmenting Customers," in *Proceedings of International Conference on Communication, Computer Sciences and Engineering, IC3SE 2024*, Institute of Electrical and Electronics Engineers Inc., 2024, pp. 1473–1477. doi: 10.1109/IC3SE62002.2024.10593465.
- [7] X. Wang, B. Cai, and K. Salleh, "Research on the application of cluster analysis in cross-border e-commerce customer segmentation and market strategy," *Applied Mathematics and Nonlinear Sciences*, vol. 9, no. 1, Jan. 2024, doi: 10.2478/amns-2024-2568.
- [8] V. Dawane, P. Waghodekar, and J. Pagare, "International Conference on Smart Data Intelligence (ICSMDI 2021) RFM Analysis Using K-Means Clustering to Improve Revenue and Customer Retention." [Online]. Available: <https://ssrn.com/abstract=3852887>
- [9] A. Abdulhafedh, "Incorporating K-means, Hierarchical Clustering and PCA in Customer Segmentation," *Journal of City and Development*, vol. 3, no. 1, pp. 12–30, 2021, doi: 10.12691/jcd-3-1-3.
- [10] W. T. Saputro, M. Murhadi, and H. M. Jumasa, "Menemukan Pola Sebaran Vaksinasi Data Covid-19 di Indonesia Menggunakan Algoritma K-Means," *Jurnal Fasilkom*, vol. 13, no. 2, pp. 244–250, Aug. 2023. [Online]. Available: <https://doi.org/10.37859/jf.v13i02.5551>
- [11] M. P. A. Budiman and D. Winarso, "Penerapan Algoritma K-Medoids Clustering untuk Pengelompokan Bulan Rawan Bencana Kabut Asap di Kota Pekanbaru," *Jurnal Fasilkom*, vol. 14, no. 1, pp. 1–8, Apr. 2024. [Online]. Available: <https://doi.org/10.37859/jf.v14i1.6858>
- [12] G. Lefait and T. Kechadi, "Customer segmentation architecture based on clustering techniques," in *4th International Conference on Digital Society, ICDS 2010, Includes CYBERLAWS 2010: The 1st International Conference on Technical and Legal Aspects of the e-Society*, 2010, pp. 243–248. doi: 10.1109/ICDS.2010.47.
- [13] A. M. Ikotun, A. E. Ezugwu, L. Abualigah, B. Abuhaija, and J. Heming, "K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data," *Inf Sci (N Y)*, vol. 622, pp. 178–210, Apr. 2023, doi: 10.1016/j.ins.2022.11.139.
- [14] Y. Deng and Q. Gao, "A study on e-commerce customer segmentation management based on improved K-means algorithm," Dec. 01, 2020, Springer Science and Business Media Deutschland GmbH. doi: 10.1007/s10257-018-0381-3.
- [15] A. M. Ikotun, M. S. Almutari, and A. E. Ezugwu, "K-means-based nature-inspired metaheuristic algorithms for automatic data clustering problems: Recent advances and future directions," Dec. 01, 2021, MDPI. doi: 10.3390/app11231246.
- [16] Maazalahi, M., Hosseini, S. K-means and meta-heuristic algorithms for intrusion detection systems. *Cluster Comput* **27**, 10377–10419 (2024). doi: 10.1007/s10586-024-04510-7
- [17] S. Mehrotra, A. Sharan, and N. Varish, "Improving search result clustering using nature inspired approach," *Multimed Tools Appl*, vol. 83, no. 23, pp. 62971–62988, Jul. 2024, doi: 10.1007/s11042-023-18067-x.
- [18] H. Xie et al., "Improving K-means clustering with enhanced Firefly Algorithms," *Applied Soft Computing Journal*, vol. 84, Nov. 2019, doi: 10.1016/j.asoc.2019.105763.
- [19] V. Arul, A. Kumar, and A. Agarwal, "Segmenting Mall Customers Data to Improve Business into Higher Target using K-Means Clustering," in **2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)**, Dec. 2021, pp. 1602–1604. IEEE.
- [20] S. J. Nanda and G. Panda, "A survey on nature inspired metaheuristic algorithms for partitioned clustering," 2014, Elsevier. doi: 10.1016/j.swevo.2013.11.003.
- [21] Singh, R. V., & Bhatia, M. P. S. (2011). Data clustering with modified K-means algorithm. **IEEE International Conference on Recent Trends in Information Technology (ICRTIT)**, 717–721. Anna University, Chennai. <https://doi.org/10.1109/ICRTIT.2011.5972391>
- [22] M. Ghasemi, S. kadhoda Mohammadi, M. Zare, S. Mirjalili, M. Gil, and R. Hemmati, "A new firefly algorithm with improved global exploration and convergence with application to engineering optimization," *Decision Analytics Journal*, vol. 5, Dec. 2022, doi: 10.1016/j.dajour.2022.100125.
- [23] A. Alam and M. K. Ahamad, "K-Means Hybridization with Enhanced Firefly Algorithm for High-Dimension Automatic Clustering," *Journal of Advanced Research in Applied Sciences and Engineering Technology*, vol. 33, no. 3, pp. 137–153, Feb. 2024, doi: 10.37934/araset.33.3.137153.