Segmentasi Pelanggan Berdasarkan Recency, Frequency, dan Monetary dengan K-Means Clustering: Studi Kasus Toko Pakaian Almost Famous

Firman Arifianto[1)*)], Jonlisen Hasudungan[2)], Adamara Muzaky[3)], Harry T.Y Achsan [4)]

[1)2)3)4)] Program StudiTeknik Informatika, Fakultas Ilmu Rekayasa, Universitas Paramadina
[*)]Correspondence Author: firman.arifianto@students.paramadina.ac.id, Jakarta, Indonesia
DOI: https://doi.org/10.37012/jtik.v10i1.2096

Abstrak

Penelitian ini bertujuan untuk menganalisis loyalitas pelanggan dalam konteks bisnis distro pakaian dengan menggunakan model RFM (*Recency, Frequency, Monetary*) dan algoritma K-Means. Data yang dianalisis berasal dari basis data membership *Distro Almost Famous Clothing Store* yang mencakup tiga cabang di Beji, Jagakarsa, dan Kelapa Dua. Pengumpulan data melibatkan informasi penting mengenai pelanggan terdaftar, kunjungan terakhir pelanggan, dan jumlah pembelian selama menjadi anggota membership. Setelah melalui proses pra-pemrosesan data, dilakukan segmentasi pelanggan menggunakan model RFM untuk membagi pelanggan menjadi kelompok berdasarkan tingkat *recency, frequency,* dan *monetary value*. Selanjutnya, algoritma K-Means digunakan untuk memetakan kelompok pelanggan yang serupa dengan menggunakan metode *Elbow Curved*, *Silhouette Coefficient*, dan *Davies-Bouldin Index* untuk menentukan jumlah cluster yang optimal. Hasil penelitian menunjukkan adanya tiga kelompok pelanggan dengan tingkat loyalitas yang berbeda: Cluster 0 (loyalitas tinggi) dengan 3225 pelanggan, Cluster 1 (loyalitas sedang) dengan 3.119 pelanggan, dan Cluster 2 (loyalitas rendah) dengan 1258 pelanggan. Implikasi dari penelitian ini adalah memberikan panduan kepada perusahaan dalam merancang strategi yang sesuai dengan karakteristik masing-masing kelompok pelanggan untuk meningkatkan retensi pelanggan dan pertumbuhan bisnis secara keseluruhan. Bagi pelanggan dengan loyalitas rendah, disarankan perusahaan untuk menyelenggarakan potongan harga atau promosi khusus, meningkatkan kualitas produk atau layanan, serta menawarkan program loyalitas guna mendorong kembali kegiatan berbelanja. Bagi pelanggan dengan loyalitas sedang, perusahaan dapat meningkatkan daya tarik program loyalitas, memperluas portofolio produk atau layanan yang relevan, dan menjalankan strategi pemasaran yang dapat meningkatkan frekuensi pembelian. Bagi pelanggan dengan loyalitas tinggi, disarankan perusahaan memberikan penghargaan tambahan, meningkatkan pengalaman pelanggan melalui personalisasi, dan terus mengembangkan produk atau layanan baru.

Kata Kunci: Loyalitas Pelanggan, Segmentasi Pelanggan, Analisis RFM, K-Means

Abstract

This research aims to analyze customer loyalty in the context of the clothing distribution business using the RFM (Recency, Frequency, Monetary) model and the K-Means algorithm. The data analyzed comes from the Almost Famous Clothing Store Distro membership database which includes three branches in Beji, Jagakarsa and Kelapa Dua. Data collection involves important information regarding registered customers, the customer's last visit, and the number of purchases during membership. After going through the data pre-processing process, customer segmentation is carried out using the RFM model to divide customers into groups based on the level of recency, frequency and monetary value. Next, the K-Means algorithm is used to map groups of similar customers using the Elbow Curved, Silhouette Coefficient, and Davies-Bouldin Index methods to determine the optimal number of clusters. The research results show that there are three groups of customers with different levels of loyalty: Cluster 0 (high loyalty) with 3225 customers, Cluster 1 (medium loyalty) with 3,119 customers, and Cluster 2 (low loyalty) with 1258 customers. The implication of this research is to provide guidance to companies in designing strategies that suit the characteristics of each customer group to increase customer retention and overall business growth. For customers with low loyalty, it is recommended that companies hold discounts or special promotions, improve product or service quality, and offer loyalty programs to encourage returning shopping activities. For customers with moderate loyalty, companies can increase the attractiveness of loyalty programs, expand their portfolio of relevant products or services, and implement marketing strategies that can increase purchasing frequency. For customers with high loyalty, it is

recommended that companies provide additional rewards, improve customer experience through personalization, and continue to develop new products or services.

Keywords: *Customer Loyalty, Customer Segmentation, RFM Analysis, K-Means*

PENDAHULUAN

Memahami dan mempertahankan loyalitas pelanggan sangat penting untuk kesuksesan bisnis dalam era bisnis yang sangat kompetitif saat ini. Analisis loyalitas pelanggan sangat penting bagi perusahaan yang ingin tetap bersaing di pasar yang dinamis karena pelanggan yang setia tidak hanya menjadi sumber pendapatan yang stabil, tetapi juga dapat memberikan dampak positif dengan berbagi pengalaman positif mereka, baik melalui media sosial, ulasan online, atau rekomendasi pribadi (Kotler dan Keller, 2016)

Menggunakan teknologi untuk mengumpulkan data transaksi telah menjadi komponen penting dari strategi bisnis modern. Dalam era komputer dan internet saat ini, setiap transaksi pelanggan meninggalkan jejak digital yang dapat diproses dan digunakan untuk mengumpulkan informasi berharga. Pengumpulan data transaksi dengan teknologi ini dimaksudkan untuk membangun strategi bisnis yang lebih tepat, responsif, dan terarah.

Distro, atau *distribution store,* adalah bisnis yang menjual pakaian dan aksesoris yang dititipkan oleh produsen pakaian independen atau diproduksi sendiri oleh pemilik toko. Bidang usaha ini memiliki peluang yang besar, mengingat minat masyarakat terhadap produk pakaian yang terus meningkat. Hal ini didukung oleh kecendrungan *fashion* yang berkembang, yang semakin variatif dan berubah dengan cepat. Selain itu, pakaian merupakan kebutuhan pokok masyarakat. Oleh karena itu, analisis loyalitas pelanggan adalah salah satu strategi penting untuk kemajuan bisnis, dan membuat perusahaan dapat bersaing dengan perusahaan lain di industri yang sama.

Analisis loyalitas pelanggan melibatkan pemahaman mendalam terhadap sejumlah faktor kritis, termasuk frekuensi pembelian, jumlah belanja, dan kebaruan. Meskipun metode ini memberikan wawasan berharga, keberhasilannya seringkali dapat ditingkatkan melalui pendekatan yang lebih canggih. Salah satu metode yang semakin populer dalam meningkatkan ketepatan analisis loyalitas pelanggan adalah penggunaan teknik *clustering*, seperti *K-Means.*

K-Means clustering merupakan salah satu algoritma pembelajaran mesin tanpa pengawasan yang populer untuk mengelompokkan data. Algoritma ini menggunakan

pendekatan berbasis *centroid* dan memisahkan kumpulan data yang tidak memiliki label ke dalam beberapa klaster berdasarkan kesamaan karakteristiknya. Pengelompokan atau *clustering* memungkinkan perusahaan untuk mengelompokkan pelanggan ke dalam segmen-segmen berdasarkan kesamaan karakteristik, seperti pola pembelian atau preferensi produk tertentu.

Dalam konteks ini, *K-Means*, sebagai salah satu metode *clustering* yang efektif, digunakan untuk memetakan kelompok pelanggan yang serupa. Kombinasi *RFM* dan *K-Means*, yang merupakan dua teknik pengelompokan data untuk segmentasi pelanggan, memberikan pandangan yang kuat tentang perilaku pelanggan. *RFM* adalah metode yang efektif untuk menilai pelanggan berdasarkan riwayat pembelian mereka. Metode ini menggunakan tiga dimensi, yaitu *recency* (keterdapatan waktu), *frequency* (frekuensi), dan *monetary* value (nilai uang). Analisis *RFM* menjadi metrik kunci untuk menganalisis perilaku pelanggan dan riwayat transaksi. Dengan menggabungkan *RFM* dan *K-Means*, perusahaan dapat memahami secara lebih baik kebutuhan dan preferensi pelanggan, serta menyesuaikan layanan mereka sesuai dengan karakteristik masing-masing segmen.

METODE

Pada penelitian ini, data yang dianalisis berasal dari basis data membership *Distro Almost Famous Clothing Store* yang mencakup tiga cabang, yaitu di Beji, Jagakarsa, dan Kelapa Dua. Data yang dikumpulkan melibatkan informasi-informasi penting mengenai pelanggan yang sudah terdaftar sebagai anggota *membership*, dintaranya adalah data pelanggan terdaftar, kunjungan terakhir pelanggan, dan jumlah pembelian yang dilakukan selama menjadi anggota *membership*.

Data mentah yang diperoleh dari *Distro Almost Famous Clothing Store* perlu melalui proses pembersihan untuk menghilangkan nilai atau atribut yang tidak relevan. Tujuannya adalah memastikan bahwa data yang digunakan dalam analisis memiliki kualitas yang baik. Setelah proses pembersihan data selesai, dilakukan pemilihan atribut yang relevan dari delapan atribut yang terdapat pada data mentah tersebut.

Model *RFM* (*Recency*, *Frequency*, *Monetary*) digunakan untuk segmentasi pelanggan berdasarkan sejarah dan perilaku loyalitas pada *Distro Almost Famous Clothing*

Store. Data yang digunakan meliputi tanggal terakhir kunjungan, total pesanan, dan jumlah transaksi selama menjadi anggota *membership*. Model *RFM* membagi pelanggan menjadi kelompok-kelompok berdasarkan 3 kriteria, yaitu:

1. *Recency*: Menunjukkan kebaruan aktivitas atau pembelian pelanggan. Dalam konteks data ini, dimensi *recency* dihitung berdasarkan tanggal terakhir kunjungan pelanggan dan tanggal terbaru yang terdapat dalam dataset sebagai titik acuan untuk mengukur tingkat kebaruan.
2. *Frequency*: Menghitung jumlah interaksi pelanggan dalam aktivitas tertentu. Data yang dihitung adalah total pembelian yang dilakukan oleh pelanggan selama periode menjadi anggota *membership*.
3. *Monetary*: Mengukur total pengeluaran pelanggan dalam periode tertentu. Data yang dihitung mencakup jumlah nilai yang dihabiskan oleh pelanggan selama menjadi anggota *membership*.

Algoritma *K-means* dikenal sebagai metode pengelompokan yang terkenal karena sifatnya yang sederhana, efisien, dan tangguh. Keunggulan tersebut menjadikannya pilihan utama dalam berbagai konteks, seperti segmentasi pelanggan, segmentasi gambar, dan pengelompokan dokumen. (Jain, 2010).

Menurut penjelasan Hastie et al. (2009) dalam bukunya "*The Elements of Statistical Learning*," algoritma *K-means* merupakan suatu metode *clustering* yang sederhana dan mudah dipahami. Algoritma ini melibatkan dua langkah dasar, yakni inisialisasi *centroid-cluster* secara acak atau strategis, dan alokasi data *point* ke *cluster* terdekat. Keefisienan *K-means* juga menjadi sorotan, di mana algoritma ini mampu dengan cepat mengelompokkan data dalam jumlah besar.

HASIL DAN PEMBAHASAN

Data yang menjadi subjek penelitian ini dikumpulkan dari database keanggotaan *membership Distro Almost Famous Clothing Store*, sebuah perusahaan yang berfokus pada distribusi pakaian di wilayah Depok dan Jakarta. Dataset ini mencakup delapan atribut, yaitu: nama, tanggal bergabung, tanggal terakhir kunjungan, jumlah pesanan, jumlah pengeluaran bulan ini, jumlah pengeluaran tahun ini, jumlah pengeluaran seumur hidup, dan rata-rata pengeluaran per transaksi. Rentang waktu pengambilan data tersebut melibatkan periode mulai dari tanggal terlama dalam atribut tanggal bergabung hingga tanggal terbaru dalam atribut tanggal terakhir kunjungan, yaitu dari 01-21-2019 hingga 10-22-2023. Dataset ini terdiri dari 7.602 *record*.

Tabel 1. Data Membership Almost Famous

No	Nama Atribut	Keterangan
1	Name	Nama pelanggan yang terdaftar sebagai membership
2	Customer Since	Tanggal pelanggan tersebut mendaftar menjadi member.
3	Last Visit	Tanggal terakhir pelanggan tersebut melakukan kunjungan atau transaksi.
4	Total # of orders	Jumlah total pesanan yang pernah dilakukan oleh pelanggan.
5	Amount This Month	Jumlah total pengeluaran pelanggan pada bulan ini.
6	Amount This Year	Jumlah total pengeluaran pelanggan pada tahun ini.
7	Amount Lifetime	Jumlah total pengeluaran pelanggan selama menjadi member.
8	Amount Average	Rata-rata pengeluaran pelanggan per transaksi.

Tahapan Pra pemrosesan data dimulai dengan pengumpulan data keanggotaan Almost Famous Clothing Store dari tanggal 21 Januari 2019 hingga 22 Oktober 2023. Data mentah yang dikumpulkan kemudian dilakukan pra-pemrosesan untuk membersihkan, menata, dan mengonversi informasi sehingga dapat diolah dengan lebih efisien. Dalam proses ini, juga ditambahkan atribut baru berupa kolom *unique_id* pada setiap data *record* untuk menghindari duplikasi data jika menggunakan nama sebagai kunci pengidentifikasi.

```
import pandas as pd
import numpy as np
import os

# Check if 'unique_id' column already exists in the DataFrame
if 'unique_id' not in data.columns:
    # Generate random unique_id only for rows where it's not already present
    data['unique_id'] = np.random.randint(100000000, 999999999, size=len(data))

# Reorder columns to have "unique_id" on the left of "name"
columns_order = ['unique_id'] + [col for col in data.columns if col != 'unique_id']
data = data[columns_order]

# Rename columns
data.columns = ['unique_id', 'name', 'customer_since', 'last_visit', 'total_orders', 'amount_this_month', 'amount_this_year', 'a

# Replace NaN or empty values in 'amount_lifetime' with default value
default_amount_lifetime = 0
data['amount_lifetime'].fillna(default_amount_lifetime, inplace=True)

# Replace NaN values in 'total_orders' with 0
data['total_orders'].fillna(0, inplace=True)

# Gunakan raw string untuk menghindari interpretasi karakter khusus dalam string
export_file_name = r'almost_prepo_dataset_evaluated1.csv'

# Check if the export file already exists
if os.path.exists(export_file_name):
    # Load existing data from CSV
    existing_data = pd.read_csv(export_file_name)

    # Check if the columns order matches the desired order
    if list(existing_data.columns) != list(data.columns):
        # Update the existing file with the new DataFrame
        data.to_csv(export_file_name, index=False)
else:
    # Export the DataFrame to CSV if the file doesn't exist
    data.to_csv(export_file_name, index=False)
```

```
df2.head()
```

	unique_id	name	customer_since	last_visit	total_orders	amount_this_month	amount_this_year	amount_lifetime	amount_average
0	870799801	0121 ida	26-01-2022	26-01-2022	2	NaN	NaN	491000	245500
1	715308711	0122 alfi ryandha	30-01-2022	30-01-2022	2	NaN	NaN	609500	304750
2	343124529	0122 ambar	10/1/2022	10/1/2022	1	NaN	NaN	280000	280000
3	460254008	0122 dani	28-01-2022	28-01-2022	1	NaN	NaN	449970	449970
4	610532576	0122 erika	9/1/2022	9/1/2022	1	NaN	NaN	458250	458250

Gambar 1. Kode Dan Output Pra-Pemrosesan Data

Pada tahap berikutnya, dilakukan proses segmentasi pelanggan berdasarkan *RFM* (*Recency, Frequency, dan Monetary*) dari atribut data yang telah dipilih. Nilai *recency* dihitung berdasarkan selisih antara tanggal terakhir kunjungan pelanggan dan tanggal terbaru yang terdapat dalam dataset, yaitu 22 Oktober 2023. Nilai *frequency* dihitung berdasarkan jumlah total pesanan yang dilakukan oleh pelanggan selama periode menjadi anggota *membership.* Nilai *monetary* dihitung berdasarkan jumlah total pengeluaran pelanggan selama menjadi anggota *membership*.

```python
import pandas as pd

# Baca DataFrame dari file CSV
df2 = pd.read_csv('almost_prepo_dataset_evaluated1.csv')

# Konversi kolom 'last_visit' dan 'customer_since' menjadi tipe data datetime
df2['last_visit'] = pd.to_datetime(df2['last_visit'], infer_datetime_format=True)
df2['customer_since'] = pd.to_datetime(df2['customer_since'], infer_datetime_format=True)

# Ambil tanggal terbaru dalam data
tanggal_terbaru_di_dataset = df2['last_visit'].max()

# Isi nilai kosong pada 'last_visit' dengan nilai 'customer_since'
df2['last_visit'].fillna(df2['customer_since'], inplace=True)

# Hitung kembali Recency ke tanggal terbaru di dataset
df2['Recency'] = (tanggal_terbaru_di_dataset - df2['last_visit']).dt.days

df2['Frequency'] = df2['total_orders']
df2['Monetary'] = df2['amount_lifetime']

# Pilih kolom yang diperlukan untuk analisis RFM
rfm_data = df2[['unique_id', 'name', 'Recency', 'Frequency', 'Monetary']]

# Tampilkan hasil RFM
rfm_data.head()
```

Gambar 2. Modeling RFM

Setelah melakukan proses pemilihan atribut dan perhitungan menggunakan *Python* pada dataset awal, diperoleh nilai-nilai untuk *recency*, *frequency*, dan *monetary*, sebagaimana terlihat cuplikan data pada gambar 3.

	unique_id	name	Recency	Frequency	Monetary
0	870799801	0121 ida	683	2	491000
1	715308711	0122 alfi ryandha	679	2	609500
2	343124529	0122 ambar	435	1	280000
3	460254008	0122 dani	681	1	449970
4	610532576	0122 erika	465	1	458250

Gambar 3. Output Data Model RFM

Setelah mendapatkan nilai *RFM*, yang mencakup nilai *recency*, *frequency*, dan *monetary* dari setiap pelanggan, selanjutnya akan dicari nilai minimum, maksimum, dan rata-rata dari masing-masing variabel *recency*, *frequency* dan *monetary*. Gambar 4 adalah

baris kode yang digunakan untuk melakukan pencarian nilai minimum, maksimum dan rata-rata dataset yang digunakan untuk melakukan normalisasi data.

```
# Pilih kolom yang diperlukan untuk analisis RFM
rfm_data = df2[['Recency', 'Frequency', 'Monetary']]

# Mencari nilai minimal, maksimal, dan rata-rata dari masing-masing variabel
min_values = rfm_data.min()
max_values = rfm_data.max()
avg_values = rfm_data.mean()


# Simpan hasil dalam DataFrame
result_df = pd.DataFrame({
    'Variable': ['Recency', 'Frequency', 'Monetary'],
    'Min': min_values.values,
    'Max': max_values.values,
    'Rata-Rata': avg_values.values
})

# Menampilkan hasil dalam bentuk matriks
print(result_df)
```

Gambar 4. Kode Nilai Minimum, Maksimum Dan Rata-Rata

Berikut hasil dari pencarian nilai minimum, maksimum, dan rata-rata dari variabel *recency*, *frequency* dan *monetary* data yang terlihat pada gambar 5.

	Variable	Min	Max	Rata-Rata
0	Recency	0.0	1801.0	664.300184
1	Frequency	0.0	198.0	2.167587
2	Monetary	0.0	72116450.0	713645.206522

Gambar 5. Cuplikan Data RFM

Setelah mendapatkan nilai *RFM* dan mencari nilai minimum, maksimum dan rata-rata, yang mencakup nilai *recency*, *frequency*, dan *monetary* dari setiap pelanggan, perlu dilakukan normalisasi untuk menormalkan skala nilai. Langkah ini diperlukan karena terkadang terdapat perbedaan yang signifikan dalam nilai *RFM* antar pelanggan. Gambar 6 adalah baris kode yang digunakan untuk melakukan normalisasi data.

```
#Data Normalization
normalized_rfm = (rfm_data[['Recency', 'Frequency', 'Monetary']] - min_values) / (max_values - min_values)
```

Gambar 6. Kode Untuk Melakukan Normalisasi Data

Setelah melakukan proses normalisasi data, selanjutnya dilakukan *clustering* untuk mengelompokkan pelanggan berdasarkan nilai-nilai *RFM* yang telah dinormalisasi. Proses *clustering* bertujuan untuk mengidentifikasi pola atau kelompok pelanggan yang memiliki karakteristik serupa dalam hal *recency*, *frequency*, dan *monetary*. Berikut cuplikan hasil dari normalisasi data yang terlihat pada gambar 7.

Cuplikan Data yang Sudah Dinormalisasi:

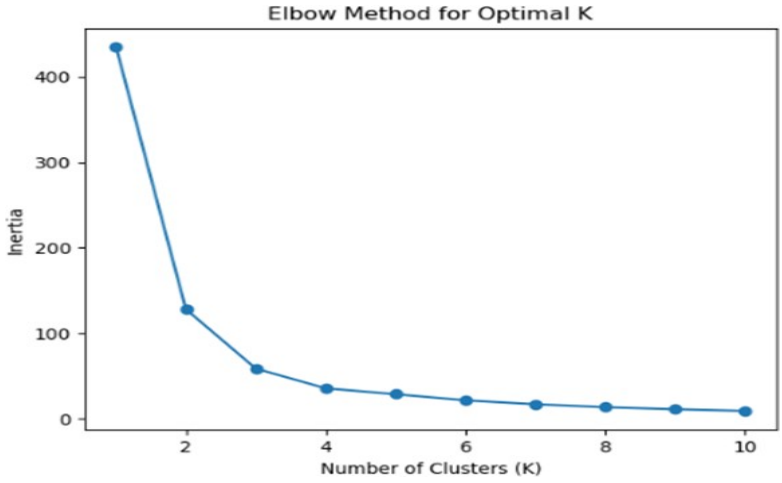
	unique_id	Recency	Frequency	Monetary
0	870799801	0.379234	0.010101	0.006808
1	715308711	0.377013	0.010101	0.008452
2	343124529	0.241532	0.005051	0.003883
3	460254008	0.378123	0.005051	0.006239
4	610532576	0.258190	0.005051	0.006354

Gambar 7. Cuplikan Data RFM Yang Sudah Dinormalisasi

Dalam langkah pengelompokan ini, algoritma *K-Means* pertama-tama digunakan untuk menentukan jumlah k, yang mewakili jumlah segmen data yang diinginkan. Dalam konteks ini, k dipilih dengan beberapa metode yakni *Elbow Curved*, *Silhouette Coefficient*, dan *Davies-Bouldin Index* untuk mencari jumlah *cluster* yang optimal dengan memperhatikan keseimbangan antara penurunan variasi, kecocokan antar kluster, dan pemisahan antar kluster.

1. *Elbow Curved*

   Jumlah cluster optimal ditunjukkan oleh grafik yang menunjukkan penurunan persentase yang paling curam, yang sesuai dengan nilai cluster 2. Oleh karena itu, menurut metode *elbow curved*, jumlah cluster optimal adalah 2.



Gambar 8. Hasil *Elbow Curved*

2. *Silhouette Coefficient*

Dengan rentang nilai dari -1 hingga 1, *silhouette coefficient* mengukur sejauh mana suatu objek cocok dalam klasternya dibandingkan dengan klaster lainnya, di mana nilai positif mendekati 1 menunjukkan pembagian klaster yang baik. *Cluster* 2 dan 3 memiliki skor siluet tertinggi (0.6074 dan 0.6071), yang menunjukkan kualitas pembagian klaster yang baik yang dapat dilihat pada Gambar 9.

```
Cluster: 2, Silhouette Coefficient: 0.6074518948404121
Cluster: 3, Silhouette Coefficient: 0.6070934340565861
Cluster: 4, Silhouette Coefficient: 0.581045751659248
Cluster: 5, Silhouette Coefficient: 0.5389911970501797
Cluster: 6, Silhouette Coefficient: 0.5209118256519807
Cluster: 7, Silhouette Coefficient: 0.517302284535673
Cluster: 8, Silhouette Coefficient: 0.5190528965537542
Cluster: 9, Silhouette Coefficient: 0.5146310903125237
Cluster: 10, Silhouette Coefficient: 0.5112459738623657
```

Gambar 9. Hasil Perhitungan *Silhouette Coefficient*

3. *Davies-Bouldin Index* (DBI)

Cluster 2, 3, dan 4, yang menunjukkan nilai DBI terendah (0.5116, 0.4939, dan 0.4457), mengindikasikan pemisahan antar-klaster yang baik dan kepadatan internal yang tinggi, sesuai dengan parameter penilaian *Davies-Bouldin Index* (DBI) di mana semakin rendah nilai DBI, semakin baik kepadatan klaster dan pemisahan antar klaster.

```
Cluster (K): 2, Davies-Bouldin Index: 0.5347279570440112
Cluster (K): 3, Davies-Bouldin Index: 0.49426418868752325
Cluster (K): 4, Davies-Bouldin Index: 0.5034526359887543
Cluster (K): 5, Davies-Bouldin Index: 0.5755586047547178
Cluster (K): 6, Davies-Bouldin Index: 0.604336548754056
Cluster (K): 7, Davies-Bouldin Index: 0.6010583558779287
Cluster (K): 8, Davies-Bouldin Index: 0.559726274933879
Cluster (K): 9, Davies-Bouldin Index: 0.552676076295818
Cluster (K): 10, Davies-Bouldin Index: 0.5547534406928865
```

Gambar 10. Hasil Perhitungan *Davies-Bouldin Index*

Dengan mempertimbangkan perbedaan signifikan antara hasil evaluasi dari metode *Elbow Curved*, *Silhouette Coefficient*, dan *Davies-Bouldin Index*, pendekatan yang seimbang adalah membagi jumlah cluster menjadi 3.

```
from sklearn.impute import SimpleImputer
from sklearn.cluster import KMeans
import os

# Pilih kolom yang diperlukan untuk analisis RFM
rfm_data = df2[['unique_id', 'name', 'Recency', 'Frequency', 'Monetary']]

# Handle missing values using mean imputation
imputer = SimpleImputer(strategy='mean')
rfm_data[['Recency', 'Frequency', 'Monetary']] = imputer.fit_transform(rfm_data[['Recency', 'Frequency', 'Monetary']])

# Data Normalization
normalized_rfm = (rfm_data[['Recency', 'Frequency', 'Monetary']] - min_values) / (max_values - min_values)

# Menentukan jumlah klaster
num_clusters = 3

# Melakukan klastering menggunakan K-means
kmeans = KMeans(n_clusters=num_clusters, random_state=42)
rfm_data['Cluster_ID'] = kmeans.fit_predict(normalized_rfm)
```

Gambar 11. Kode Untuk Melakukan Proses *K-Means Clustering*

Setelah melakukan perhitungan nilai RFM dari masing-masing cluster dengan menghitung nilai centroid, yang selanjutnya digunakan untuk menentukan kluster poin data. Proses ini membantu dalam memahami karakteristik masing-masing kelompok dan memfasilitasi analisis lebih lanjut terhadap data keanggotaan.

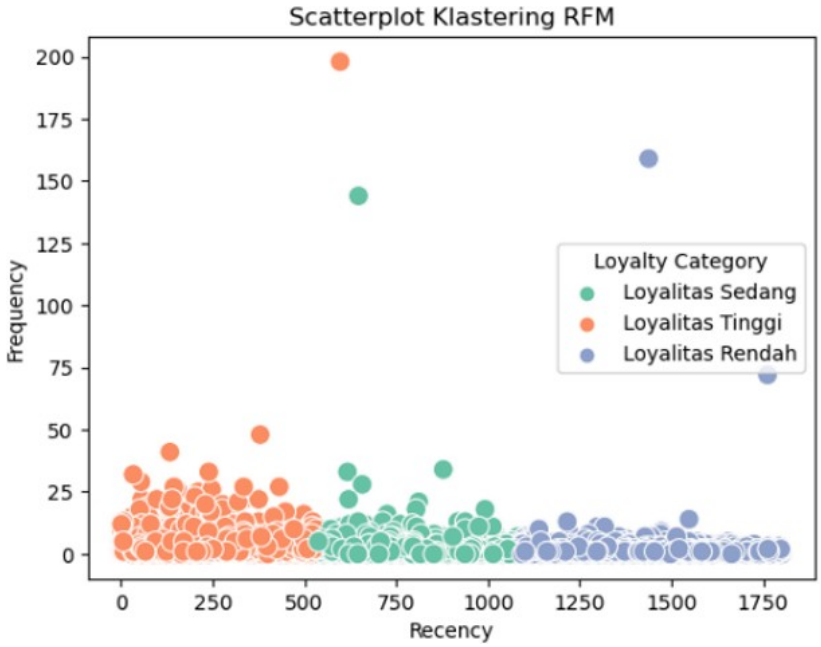
```
# Compute cluster centroids
centroid_table = kmeans.cluster_centers_

# Create a DataFrame to represent the centroid values
centroid_df = pd.DataFrame(centroid_table, columns=['Recency', 'Frequency', 'Monetary'])

# Display the centroid values
print("Centroid Table:")
print(centroid_df)
```

```
Centroid Table:
    Recency  Frequency  Monetary
0  0.752303   0.005413  0.000350
1  0.122874   0.010077  0.001128
2  0.437405   0.005200  0.000929
```

Gambar 12. Kode dan Output *Centroid Table*

Berdasarkan tabel *centroid* pada Gambar 12, loyalitas pelanggan dapat diklasifikasikan berdasarkan nilai *recency*, *frequency*, dan *monetary*. *Cluster* 0 menunjukkan kemungkinan loyalitas tinggi dengan pembelian terbaru, frekuensi pembelian dan pengeluaran yang lebih tinggi dibandingkan cluster lain. *Cluster* 1 menunjukkan loyalitas sedang dengan pembelian yang mungkin kurang baru, frekuensi dan pengeluaran berada di kisaran menengah. *Cluster* 2 menunjukkan loyalitas rendah dengan pembelian yang paling tidak baru, frekuensi dan pengeluaran yang rendah.

Dengan melakukan visualisasi dengan *scatterplot* maka bisa dilihat hubungan antara *recency* dan *frequency* karena dapat memberikan gambaran visual yang jelas terkait pola distribusi atau kumpulan data pelanggan dalam ruang dua dimensi. Seperti yang terlihat pada Gambar 13.



Gambar 13. Visualisasi Scatterplot

Jumlah anggota yang terdapat pada masing-masing cluster, seperti yang ditunjukkan pada Gambar 14, adalah sebagai berikut: *Cluster* 0 memiliki 3225 pelanggan dengan loyalitas tinggi, *Cluster* 1 memiliki 3119 pelanggan dengan loyalitas sedang, dan *Cluster* 2 memiliki 1258 pelanggan dengan loyalitas rendah.

Gambar 14. Visualisasi Jumlah Data Pada Setiap Cluster

KESIMPULAN DAN REKOMENDASI

Berdasarkan hasil penelitian, dapat disimpulkan bahwa terdapat tiga kelompok pelanggan dalam *Distro Almost Famous Clothing Store*, masing-masing dengan tingkat loyalitas yang berbeda: *Cluster* 0 (loyalitas tinggi) dengan 3225 pelanggan, *Cluster* 1 (loyalitas sedang) dengan 3.119 pelanggan, *Cluster* 2 (loyalitas rendah) dengan 1258 pelanggan dengan total seluruh data yakni 7602 *record*. Untuk meningkatkan loyalitas pelanggan, perusahaan dapat mengimplementasikan strategi yang sesuai dengan karakteristik masing-masing kelompok.

Bagi pelanggan dengan loyalitas rendah, disarankan perusahaan untuk menyelenggarakan potongan harga atau promosi khusus, meningkatkan kualitas produk atau layanan, serta menawarkan program loyalitas guna mendorong kembali kegiatan berbelanja.

Bagi pelanggan dengan loyalitas sedang, perusahaan dapat meningkatkan daya tarik program loyalitas, memperluas portofolio produk atau layanan yang relevan, dan menjalankan strategi pemasaran yang dapat meningkatkan frekuensi pembelian.

Bagi pelanggan dengan loyalitas tinggi, disarankan perusahaan memberikan penghargaan tambahan, meningkatkan pengalaman pelanggan melalui personalisasi, dan terus mengembangkan produk atau layanan baru yang memenuhi kebutuhan mereka. Dengan demikian, implementasi strategi yang sesuai dengan karakteristik setiap kelompok pelanggan diharapkan dapat membawa dampak positif terhadap retensi pelanggan dan pertumbuhan bisnis secara keseluruhan.

REFERENSI

Allegue, S., Abdellatif, T., & Bannour, K. (2020). RFMC: a spending-category segmentation. In 2020 IEEE 17th International Conference on Web and Information Technologies (WETICE) (pp. 165-170). IEEE.

Anitha, P., & Patil, M. M. (2022). RFM model for customer purchase behavior using K-Means algorithm. Journal of King Saud University - Computer and Information Sciences, 34(5).

Ashari, F., Ilham, Nugroho, E., Baraku, R., Yanda, I., & Liwardana, R. (2023). Analysis of Elbow, Silhouette, Davies-Bouldin, Calinski-Harabasz, and Rand-Index Evaluation on K-Means Algorithm for Classifying Flood-Affected Areas in Jakarta. Journal of

Applied Informatics and Computing, 7, 89-97. https://doi.org/10.30871/jaic.v7i1.4947

Basri, F. M., Gata, W., & Risnandar. (2020). Analisis loyalitas pelanggan berbasis model recency, frequency, dan monetary (RFM).

Christy, A. J., Umamakeswari, A., Priyatharsini, L., & Neyaa, A. (2021). RFM ranking – An effective approach to customer segmentation. Journal of King Saud University - Computer and Information Sciences, 33(10), 1251-1257.

Febriani, A., & Putri, S. A. (2020). Consumer Segmentation Based on Recency, Frequency, Monetary Models with the K-Means Method. Journal of Industrial Engineering and Management Systems, 13(2), 52-57.

Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). The elements of statistical learning: Data mining, inference, and prediction. Springer.

Hossain, MZ, Akhtar, MN, Ahmad, RB, & ... (2019). A dynamic K-means clustering for data mining. Indonesian Journal of …, squ.elsevierpure.com, https://squ.elsevierpure.com/ar/publications/a-dynamic-k-means-clustering-for-data-mining

Ikotun, A. M., Ezugwu, A. E., Abualigah, L., Abuhaija, B., Heming, J., & Abualigah, L. (2023). K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. Information Sciences, 622, 178-210.

Jain, A. K. (2010). Data clustering: 50 years beyond k-means. Pattern Recognition Letters, 31(8), 651-666.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning (Vol. 112). New York, NY: Springer.

Kotler, P., & Keller, K. L. (2016). Marketing management (15th ed.). New Jersey: Pearson Education.

Rokach, L., & Maimon, O. (Eds.). (2005). Data mining and knowledge discovery handbook. Berlin, Germany: Springer Science & Business Media.

Setiono, A., Triayudi, A., & Handayani, E. T. E. (2023). Analysis of Recency Frequency Monetary and K-Means Clustering in Dental Clinics to Determine Patient Segmentation. JSiI | Jurnal Sistem Informasi, 10(1), 1-6.

Sinaga, K. P., & Yang, M.-S. (2020). Unsupervised K-Means Clustering Algorithm. IEEE Access, 8, 80716-80727.