

PENERAPAN *EXPLORATORY DATA ANALYSIS* (EDA) DAN ANALISIS *RECENCY, FREQUENCY, AND MONETARY* (RFM) UNTUK SEGMENTASI PELANGGAN E-COMMERCE

Fatia Amalia Maresti¹, Gabriella Mustika Anugraheni², Raditio Ananto Hargiyanto³,
Kiki Mustaqim⁴

S1 Sains Data, Fakultas Logistik Teknologi dan Bisnis, Universitas Logistik dan Bisnis Internasional¹
email: fatiaamalia@ulbi.ac.id¹

S1 Sains Data, Fakultas Logistik Teknologi dan Bisnis, Universitas Logistik dan Bisnis Internasional²
email: 184220028@std.ulbi.ac.id²

S1 Sains Data, Fakultas Logistik Teknologi dan Bisnis, Universitas Logistik dan Bisnis Internasional³
email: raditioananto2003@gmail.com³

S1 Sains Data, Fakultas Logistik Teknologi dan Bisnis, Universitas Logistik dan Bisnis Internasional⁴
email: kiki@ulbi.ac.id⁴

Abstrak

E-commerce menghasilkan volume data yang sangat besar setiap harinya yang dihasilkan oleh pelanggan saat bertransaksi. Pelanggan merupakan kelompok maupun individu yang memegang peranan penting dalam proses strategi bisnis dengan membeli produk atau jasa berdasarkan pada keputusan sendiri yang beralaskan oleh pertimbangan dari segi kebermanfaatan serta harga dari setiap produk atau jasa. Dalam penelitian ini pelanggan akan dilakukan segmentasi berdasarkan data transaksi penjualan pada e-commerce menggunakan metode *Exploratory Data Analysis* (EDA) dan analisis *Recency, Frequency, and Monetary* (RFM). Pemberlakuan segmentasi pelanggan bertujuan untuk memahami karakteristik pelanggan dan mengimplementasi strategi retensi agar meningkatkan kemajuan suatu perusahaan dengan membuat strategi pemasaran yang efektif dan memprediksi perilaku pelanggan di masa mendatang. Pada tahapan awal segmentasi pelanggan dilakukan investigasi untuk menemukan pola, mengevaluasi hipotesis, dan memverifikasi asumsi berkaitan dengan data pelanggan melalui rangkuman statistik menggunakan metode *Exploratory Data Analysis* (EDA). Tahapan selanjutnya dilakukan identifikasi untuk mengetahui pola perilaku pelanggan yang melakukan pembelian barang sehingga dapat dikelompokkan berdasarkan karakteristik dengan analisis *Recency, Frequency, and Monetary* (RFM). Terdapat 11 kategori yang dihasilkan dalam segmentasi pelanggan, di antaranya adalah *Champions, Loyal Customer, Promising Customer, New Customer, Need Attention, At Risk, Potential, About To Sleep, Old Potensial, Hibernating, dan Lost*.

Kata Kunci: Segmentasi Pelanggan, *Exploratory Data Analysis* (EDA), *Recency, Frequency, and Monetary* (RFM), E-Commerce.

Abstract

E-commerce generates a massive volume of data every day produced by customer transactions. Customers are groups or individuals who play an important role in the business strategy process by purchasing products or services based on their own decisions after considerations of the convenience and price of each product or service. In this study, customer segmentation will be carried out based on the transaction data of an e-commerce using the Exploratory Data Analysis (EDA) and Recency, Frequency, and Monetary (RFM) analysis methods. The implementation of customer segmentation aims to understand customer characteristics and implement retention strategies to improve the progress of a company by creating effective marketing strategies and predicting customer behaviors in the future. Initially, an investigation is carried out to find patterns, evaluate hypotheses, and verify assumptions related to customer data through statistical summaries using the Exploratory Data Analysis (EDA) method. Identification is then executed to determine the behavioral patterns of customers with purchases to then be grouped based on characteristics with Recency, Frequency, and Monetary (RFM) analysis. There are 11 categories generated in this customer segmentation which are Champions, Loyal

Customer, Promising Customer, New Customer, Need Attention, At Risk, Potential, About To Sleep, Old Potential, Hibernating, and Lost.

Keywords: *Customer Segmentation, Exploratory Data Analysis (EDA), Recency, Frequency, and Monetary (RFM), E-Commerce.*

1. PENDAHULUAN

Pelanggan merupakan kelompok maupun individu yang membeli produk atau jasa berdasarkan pada keputusan sendiri yang beralaskan oleh pertimbangan dari segi kebermanfaatannya serta harga dari setiap produk atau jasa dan pelanggan memegang peranan penting dalam proses strategi bisnis. Seiring berkembangnya persaingan industri, perusahaan harus mengenali keberagaman perilaku pelanggan melalui segmentasi. Segmentasi pelanggan merupakan upaya untuk memahami karakteristik pelanggan dan mengimplementasikan strategi retensi yang efektif agar meningkatkan kemajuan suatu perusahaan. Salah satu kunci yang dapat meningkatkan kemajuan suatu perusahaan adalah pemasaran, terlepas apapun bidang perusahaan tersebut (Harani, dkk., 2020).

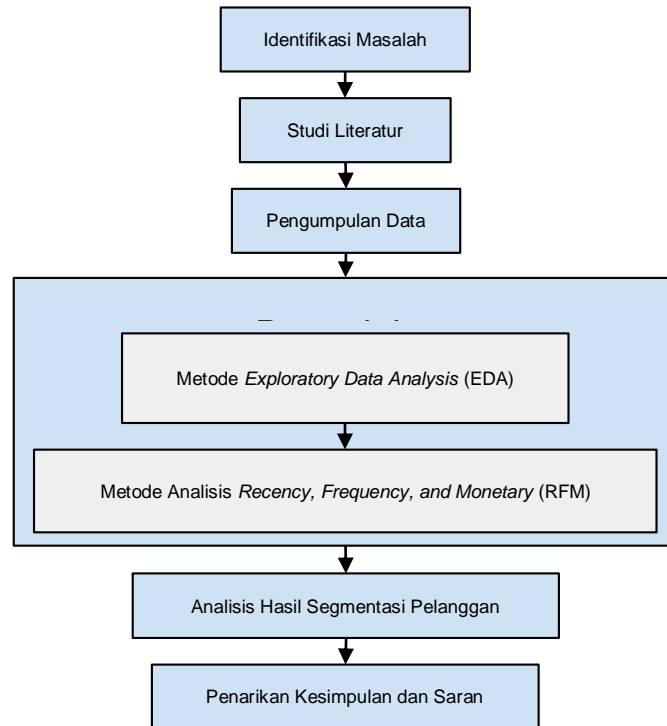
Exploratory Data Analysis (EDA) adalah tahapan awal dalam menginvestigasi atau menemukan pola, mengevaluasi hipotesis, memverifikasi asumsi berkaitan dengan data melalui rangkuman statistik pada visualisasi (Angela, J. B. dan Irsyad, A, 2023). Metode EDA merupakan tahapan awal yang sangat penting dalam pengolahan data agar data yang dianalisa dapat dipastikan berkualitas. *Recency, Frequency, and Monetary* (RFM) adalah metode untuk mengetahui pola perilaku pelanggan dalam melakukan pembelian barang (Tivano, D., Ruslianto, I., & Prawira, D, 2023). Metode RFM merupakan metode analisis yang digunakan untuk mengidentifikasi pelanggan yang memiliki potensi melakukan transaksi kembali. RFM digunakan untuk mengelompokkan pelanggan berdasarkan karakteristik mereka, seperti kepuasan, loyalitas, dan pelanggan potensial.

Dilansir melalui DataIndonesia.id, Bank Indonesia (BI) mencatat perdagangan elektronik atau e-commerce di Indonesia memiliki nilai transaksi sebesar Rp 476,3 triliun dari 3,49 miliar transaksi, meningkat 18,8% dibandingkan tahun sebelumnya sebesar Rp 410 triliun. Selain itu, DataIndonesia.id juga mengutip data dari Statista Market Insights, yang menyebutkan bahwa jumlah pengguna e-commerce di Indonesia mencapai 178,94 juta orang pada 2022 dan diproyeksikan meningkat menjadi 196,47 juta orang pada akhir 2023 (Monavia Ayu Rizaty & Ridhwan Mustajab, 2023). E-commerce menghasilkan volume data yang sangat besar yang disebabkan oleh aktivitas transaksi secara online setiap harinya. Data ini dapat digunakan untuk mengidentifikasi *trend* dan membantu pembuatan keputusan strategis dan tata kelola perusahaan dengan melakukan segmentasi pelanggan menggunakan metode *Exploratory Data Analysis* (EDA) dan analisis *Recency, Frequency, and Monetary* (RFM).

2. METODE PENELITIAN

2.1 Alur Penelitian

Metode penelitian digunakan sebagai panduan untuk penyusunan agar dapat dilakukan secara terarah dan terstruktur dalam memperoleh data demi memenuhi tujuan penelitian. Pada penelitian ini metode yang digunakan merupakan metode penelitian kuantitatif dengan data numerik dan statistik yang menjadi fokus utama dalam penelitian. Adapun alur pengerjaan penelitian ini pada **Gambar 1.** berikut.



Gambar 1. Alur Metode Penelitian

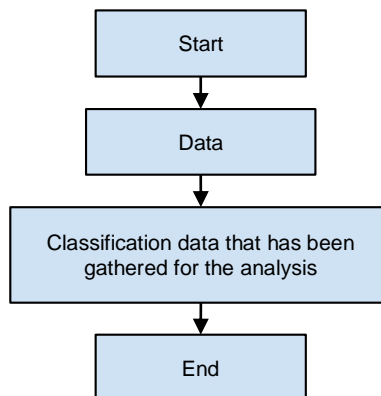
2.2 Tahapan-Tahapan Alur Penelitian

2.2.1 Pengumpulan Data

Tahap ini melakukan pengumpulan data yang dibutuhkan sebagai pendukung utama dalam pengerjaan penelitian. Terdapat dua *dataset* berbeda yang berisi tentang data transaksi penjualan e-commerce, meliputi.

1. Data tahunan yang dikumpulkan dimulai dari tahun 2009 hingga 2010 yang didapatkan dari kaggle.com dalam bentuk dokumen Excel. Data tahun tersebut disimpan ke dalam variabel *df1*.
2. Data tahunan yang dikumpulkan dimulai dari tahun 2010 hingga 2011 yang didapatkan dari kaggle.com dalam bentuk dokumen Excel yang akan disimpan pada *df2*.

Dari kedua *dataset* ini, transaksi pelanggan di United Kingdom dipilih sebagai fokus analisis karena mayoritas pelanggan berasal dari wilayah ini, yang sekaligus menjadi domisili e-commerce tersebut. Dominasi pelanggan United Kingdom menjadikannya pasar utama yang paling relevan untuk dianalisis.



Gambar 2. Alur Pengumpulan Data

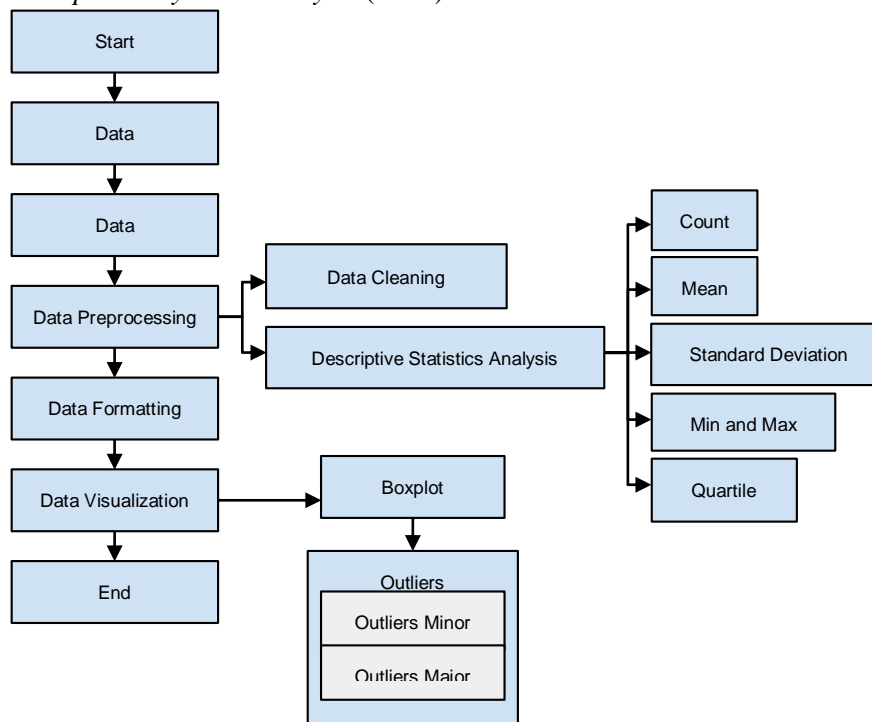
2.2.2 Segmentasi Pelanggan

Segmentasi pelanggan adalah proses pengidentifikasian dengan membagi basis pelanggan menjadi beberapa kelompok dalam satu karakteristik yang serupa dari perilaku pelanggan untuk mengembangkan strategi pemasaran yang lebih efektif dan tepat sasaran pada setiap kelompok pelanggan. Menurut Schiffman dan Kanuk (2000) mendefinisikan perilaku pelanggan sebagai berikut: “*The term consumer behavior refers to the behavior that consumers display in searching for, purchasing, using, evaluating, and disposing of products and services that they expect will satisfy their needs,*” yang berarti perilaku pelanggan merupakan perilaku yang diperlihatkan oleh pelanggan dalam mencari, membeli, menggunakan, mengevaluasi, serta menghabiskan produk dan jasa yang diharapkan dapat memuaskan kebutuhannya (Jefri Putri Nugraha, dkk, 2021). Melalui segmentasi, perusahaan dapat mengembangkan program loyalitas yang sesuai dengan karakteristik masing-masing segmen. Penelitian menunjukkan bahwa strategi retensi yang disesuaikan dengan segmen tertentu dapat meningkatkan loyalitas pelanggan dan profitabilitas (V. Katragadda, 2022). Segmentasi pelanggan yang efektif memungkinkan bisnis untuk menyesuaikan produk dan layanan mereka untuk memenuhi kebutuhan spesifik dari kelompok pelanggan yang berbeda. Personalisasi ini semakin penting dalam pasar yang kompetitif saat ini, strategi pemasaran yang umum menjadi kurang efektif. Pergeseran dari segmentasi berbasis demografis tradisional ke pendekatan yang lebih dinamis dan berbasis perilaku memungkinkan perusahaan untuk merespons perubahan preferensi konsumen secara *real-time* (R. Amutha, dkk, 2023).

2.2.3 Pengolahan Data

Pengolahan data ini mencakup metode *Exploratory Data Analysis* (EDA) dan metode analisis *Recency, Frequency, and Monetary* (RFM) menggunakan bahasa pemrograman Python.

2.2.3.1 Metode *Exploratory Data Analysis* (EDA)



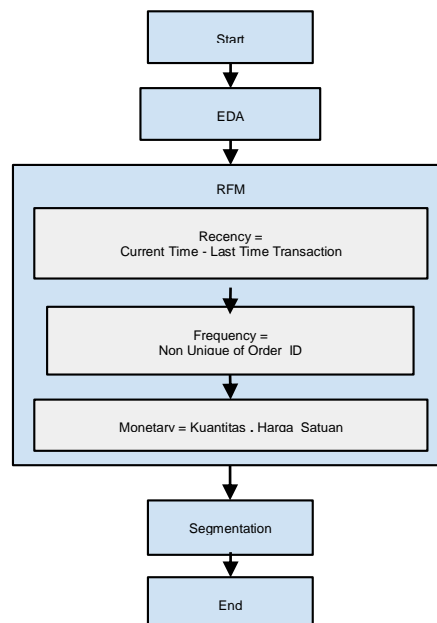
Gambar 3. Alur Metode EDA

Exploratory Data Analysis (EDA) adalah prosedur yang krusial di dalam data science untuk menganalisis dan mengetahui data sebelum dilakukan analisis yang lebih lanjut. EDA merupakan suatu cara yang dipakai untuk memahami pola suatu data, format data, dan hubungan antar data yang dituangkan ke dalam bentuk grafik supaya pengamat dapat mengetahui suatu data dengan mudah (Leni, D., dkk, 2023). EDA fokus pada pemahaman struktur data sebelum menggunakan metode statistik atau algoritma pembelajaran mesin yang lebih kompleks. EDA membantu mengidentifikasi pola, tren, dan anomali dalam data yang penting untuk analisis lebih lanjut dan pengembangan model prediktif yang efektif (V. Da Poian, dkk, 2023).

Menurut Angela, J. B. dan Irsyad, A. (2023), EDA melalui beberapa tahapan untuk memperoleh informasi apa saja yang dapat digali dari data, di antaranya.

1. *Data Understanding*, melakukan pemahaman data setelah data dikumpulkan.
2. *Data Preprocessing*, melakukan pembersihan data atau *data cleaning*, pengeliminasian duplikasi, nilai kosong, data yang tidak konsisten, dan kesalahan lainnya. *Data cleaning* adalah proses mengidentifikasi dan mengoreksi atau membuang data inkonsisten, keliru, duplikat, salah format, atau eror lain dari *dataset* yang mengganggu proses analisis data dan hasil analisis yang tidak akurat. Proses dalam *data cleaning* melibatkan penggantian, penghapusan, atau perbaikan data yang tidak relevan.
3. *Data Formatting*, melakukan perbaikan terhadap data yang tidak sejalan dengan topik dan memastikan tipe data yang tepat pada setiap kolom.
4. *Data Visualization*, melakukan penampilan visualisasi yang sesuai dengan kebutuhan.

2.2.3.2 Metode Analisis *Recency, Frequency, and Monetary* (RFM)



Gambar 4 Alur Metode Analisis RFM

RFM merupakan sebuah metode analisis marketing yang digunakan dalam segmentasi pelanggan dengan melakukan pengamatan perilaku pelanggan melalui pengumpulan data mengenai transaksi pelanggan, seperti waktu terakhir transaksi, jumlah transaksi, dan jumlah uang yang dikeluarkan (Iriawan, Nur dan Permadina Kanah, dkk, 2021). Segmentasi RFM adalah salah satu teknik yang digunakan untuk menilai sejauh mana loyalitas klien (Alzami, F., Sambasri, dkk, 2023). Analisis RFM bergantung pada parameter *Recency* (R), *Frequency* (F), dan *Monetary* (M). Analisis RFM dijalankan dengan menggunakan dataset transaksional untuk mengevaluasi pelanggan atas perilaku pembelian mereka (Anitha, P., & Patil, M. M., 2022). Pembuatan model dalam RFM digunakan sebagai peramalan perilaku konsumen dengan menerjemahkan perilaku konsumen ke dalam angka sehingga mengarahkan keputusan segmentasi yang lebih baik, menggunakan data yang telah melewati tahap pengumpulan dan pemrosesan, lalu dianalisis melalui proses klasifikasi (Supangat, S., & Mulyani, Y., 2023).

Menurut Wicaksana, Putu Aryasuta, dkk., (2022), RFM didasari oleh tiga aspek variabel utama dari perilaku pelanggan, yaitu *Recency*, *Frequency*, dan *Monetary*.

1. *Recency*

Nilai R atau *Recency* merupakan variabel untuk mengukur nilai pelanggan berdasarkan perilaku pelanggan yang mengacu pada interval waktu (tanggal, bulan, tahun) transaksi pembelian terakhir dengan waktu saat ini. Semakin kecil interval waktu maka nilai *recency* semakin besar. Nilai *recency* ini akan diambil dari kolom *Tgl_Pembayaran*.

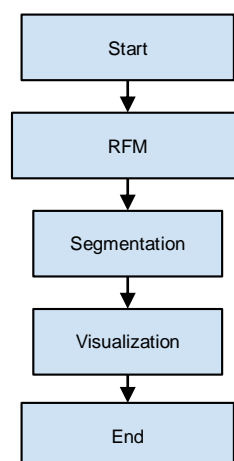
2. *Frequency*

Nilai F atau *Frequency* merupakan variabel untuk mengukur nilai pelanggan berdasarkan jumlah transaksi yang dilakukan pelanggan dalam satu periode waktu tertentu. Semakin banyak jumlah transaksi yang dilakukan maka nilai F semakin besar. Nilai *frequency* ini diambil dari kolom *Item_ID*.

3. *Monetary*

Nilai M atau *Monetary* merupakan variabel untuk mengukur nilai pelanggan berdasarkan jumlah besaran uang yang dihabiskan pelanggan untuk tiap transaksi pembelian dalam satu periode waktu tertentu. Semakin banyak jumlah besaran uang yang dihabiskan maka nilai M semakin besar. Nilai *monetary* ini diambil dari kolom *Jumlah* yang merupakan hasil akhir dari kolom *Quantity* dan *Price*.

2.2.4 Analisis Segmentasi Pelanggan



Gambar 5. Alur Hasil Segmentasi Pelanggan

Hasil analisis meliputi ciri khas setiap segmen yang diklasifikasi melalui perilaku pelanggan dan secara tidak langsung dapat menjelaskan perbedaan satu segmen dengan yang lainnya berupa visualisasi tulisan maupun grafik.

3. HASIL DAN PEMBAHASAN

3.1 Data Uji Coba

Dalam penyusunan laporan yang berjudul “Segmentasi Pelanggan dengan Menggunakan Metode *Exploratory Data Analysis* (EDA) dan Analisis *Recency, Frequency, and Monetary* (RFM) berdasarkan Data Transaksi Penjualan E-Commerce,” *dataset* yang menjadi fokus untuk melakukan analisis diperoleh dari E-Commerce yang berpusat di *United Kingdom* dengan judul “Online Retail II.” Terdapat beberapa kolom dalam *dataset* ini yang meliputi.

1. Invoice (Nomor faktur) berisi angka 6 digit yang ditetapkan secara unik untuk setiap transaksi. Jika kode ini dimulai dengan huruf 'C', ini mengindikasikan pembatalan transaksi.
2. StockCode berisi kode produk pada barang yang berisi angka nominal 5 digit yang ditetapkan secara unik untuk setiap produk yang berbeda.
3. Description berisi nama produk barang.
4. Quantity berisi jumlah setiap produk (item) per transaksi.
5. InvoiceDate (Tanggal faktur) berisi tanggal dan waktu pembuatan faktur meliputi hari dan waktu ketika transaksi dilakukan.
6. UnitPrice berisi harga satuan pada harga produk per unit dalam *pound sterling* (£).
7. CustomerID (Nomor pelanggan) berisi angka 5 digit yang ditetapkan secara unik untuk setiap pelanggan.
8. Country (Nama negara) berisi nama negara tempat tinggal pelanggan.

Dalam *dataset* “RFM Analysis Online Retail II,” pada tahun 2009–2010 memuat 525.461 data, sementara data tahun 2010–2011 memuat 541.910 data. Metode EDA akan dilakukan terlebih dahulu sebelum analisis RFM dilakukan guna mengetahui data apa saja yang dibutuhkan dan memastikan kebersihan data. Pada *dataset* tersebut terdapat 3 variabel yang akan digunakan untuk diproses dan diuji ke dalam analisis sesuai dengan komponen RFM, yakni InvoiceDate sebagai komponen waktu terakhir pembelian (*recency*), Invoice sebagai komponen frekuensi pembelian (*frequency*), dan Quantity dan Price sebagai komponen besarnya biaya pembelian (*monetary*).

3.2 Analisis Data

3.2.1 *Exploratory Data Analysis* (EDA)

1. *Data Understanding*

Tahapan EDA yang paling pertama adalah *data understanding* agar mengetahui kolom–kolom, tipe data, jumlah baris data, dan rentang pesan data pada *dataset*. Proses tersebut dapat dilakukan dengan menggunakan fungsi *info()* untuk mendapatkan informasi mengenai jumlah baris data, tipe data, dan sebagainya, serta fungsi *min()* dan *max()* untuk mendapatkan informasi mengenai rentang data.

2. *Data Preprocessing*

Tahapan selanjutnya adalah menghapus missing values, menghapus barang pembelian yang dibatalkan dan menghapus anomali data pada kolom numerik melalui analisis statistika deskriptif. Adapun penerapannya seperti berikut.

3. *Data Formatting*

Proses penghapusan bisa diabaikan dan dilanjutkan pada proses selanjutnya yaitu mengganti tipe data pada kolom *Customer ID* dengan tipe *float* menjadi *integer* agar menghilangkan nilai desimalnya pada kedua *dataset* (df1 dan df2)

4. *Data Visualization*

Pada bagian ini tahapan untuk memastikan bahwa tidak ada anomali data, terutama pada nilai outliers melalui visualisasi data yang disajikan dalam boxplot. Adapun penerapannya seperti berikut.

3.2.2 *Recency, Frequency, and Monetary* (RFM)

InvoiceDate sebagai komponen waktu terakhir pembelian (*recency*), Invoice sebagai komponen frekuensi pembelian (*frequency*), dan Quantity dan Price sebagai komponen besarnya biaya pembelian (*monetary*). Oleh karena itu, perkalian nilai Quantity dan Price untuk nilai *monetary* harus dihitung terlebih dahulu sebelum penghitungan skor RFM secara keseluruhan dilakukan. Adapun penerapannya akan dilakukan pada df1 (2009–2010) dan df2 (2010–2011).

3.3 Hasil dan Pembahasan

Metode EDA menemukan *missing value* dalam *dataset* United Kingdom pada kolom CustomerID dan Description. Di tahun 2009–2010, total *missing value* yang ditemukan dalam *dataset* pada kolom CustomerID dan Description sebesar **106.429** dan **2.928 data**. Sementara di tahun 2010–2011, total *missing value* yang ditemukan dalam *dataset* pada kolom CustomerID dan Description sebesar **133.600** dan **1.454 data**.

Anomali data lain yang ditemukan dalam *dataset* melalui metode EDA terdapat pada kolom Invoice dan Quantity. Diketahui nilai negatif yang berada pada Quantity memiliki karakter khusus dalam Invoice dengan ditandai adanya huruf “C” pada nomor faktur. Nilai ini disebabkan oleh pelanggan yang melakukan pembatalan pesanan. Dalam *dataset* United Kingdom jumlah karakter “C” yang ditemukan pada tahun 2009–2010 dan 2010–2011 sebesar **8.472 dari 379.423 data** dan **7.533 dari 361.878 data**.

Anomali data ini harus dibersihkan agar tidak mengganggu proses lanjutan penelitian. Total akhir data setelah dilakukan proses pembersihan data dalam metode EDA pada *dataset* United Kingdom tahun 2009–2010 dan 2010–2011 menjadi sebesar **370.951** dan **354.345 data**.

Setelah anomali data ditandai dan dibersihkan kembali, hasil akhir jumlah data dalam proses metode EDA mendapatkan total jumlah data pada *dataset* United Kingdom tahun 2009–2010 adalah **294.089 data** dan **271.313 data** pada *dataset* tahun 2010–2011.

Persebaran data dalam data transaksi penjualan tahun 2009–2010 dan 2010–2011 yang dimiliki e-commerce khususnya negara *United Kingdom* yang diteliti pada penelitian ini dapat diketahui pada proses EDA dengan melihat outlier pada boxplot. Jumlah outlier yang ditemukan pada data ini sangat besar. Hal ini menyebabkan terbaginya outlier menjadi 2 bagian, yakni *Outlier Minor* dan *Outlier Major*. Pada perhitungan dalam kolom Quantity dan Price jumlah kuantitas barang dan biaya dengan jumlah nilai yang paling besar yang dibeli dan dikeluarkan oleh pelanggan dapat dilihat dari *Outlier Major*. Sementara pada *Outlier Minor* merupakan jumlah kuantitas barang dan biaya dalam kolom Quantity dan Price dengan nilai yang relatif besar.

Penggabungan metode EDA dengan analisis RFM menyebabkan perubahan jumlah baris pada proses pembersihan dan pengelompokan data. Total jumlah data yang telah dilakukan pembersihan data melalui proses metode EDA pada *dataset* United Kingdom tahun 2009–2010 adalah **294.089 data**, sementara pada *dataset* tahun 2010–2011 adalah **271.313 data**. Setelah dilakukan pengelompokan data dengan analisis RFM *dataset* United Kingdom tahun 2009–2010 berjumlah **3.926 data** dan *dataset* tahun 2010–2011 berjumlah **3.887 data**.

Hal ini terjadi karena pada saat memasuki proses analisis data dengan RFM, data akan dikelompokkan sesuai dengan data CustomerID. Pelanggan yang melakukan pembelian secara besar atau grosir akan tercatat pada CustomerID secara berulang. Analisis RFM diperlukan untuk

menghindari adanya duplikasi data dan mengelompokkan pelanggan sesuai dengan karakteristik pelanggannya. Analisis RFM dapat menemukan karakteristik pelanggan dengan melihat dari skor masing-masing kategori yang telah ditentukan. Skor pada analisis RFM berpengaruh terhadap karakteristik pelanggan sehingga menghasilkan segmentasi pelanggan dengan karakteristik pelanggan sebagai berikut.

Tabel 1. Segmentasi Pelanggan berdasarkan Karakteristik Pelanggan

No	Segmen	Deskripsi	Skor <i>Recency</i>	Skor <i>Frequency</i>	Skor <i>Monetary</i>
1	<i>Champions</i>	Pelanggan dengan pembelian dekat atau terdekat, sering bertransaksi, dan jumlah uang yang sangat besar	4-5	3-5	4-5
2	<i>Loyal Customer</i>	Pelanggan dengan pembelian dekat atau terdekat, sering bertransaksi, dan jumlah uang yang cukup besar	4-5	3-5	1-3
3	<i>Promising Customer</i>	Pelanggan dengan pembelian dekat atau terdekat, jarang bertransaksi, dan jumlah uang yang besar	4-5	1-2	4-5
4	<i>New Customer</i>	Pelanggan dengan pembelian dekat atau terdekat, jarang bertransaksi, dan jumlah uang yang cukup besar	4-5	1-2	1-3
5	<i>Need Attention</i>	Pelanggan dengan pembelian cukup dekat, sering bertransaksi berapapun besaran jumlah uangnya	3	3-5	1-5
6	<i>At risk</i>	Pelanggan dengan pembelian cukup dekat, jarang bertransaksi berapapun besaran jumlah uangnya	3	1-2	1-5
7	<i>Potential</i>	Pelanggan dengan pembelian lama, sering bertransaksi berapapun jumlah besaran uangnya	2	3-5	1-5
8	<i>About to sleep</i>	Pelanggan dengan pembelian lama, jarang bertransaksi berapapun jumlah besaran uangnya	2	1-2	1-5
9	<i>Old Potential</i>	Pelanggan dengan pembelian terlama, sering bertransaksi dan jumlah uang yang besar	1	3-5	3-5
10	<i>Hibernating</i>	Pelanggan dengan pembelian terlama, berapapun jumlah transaksi dan jumlah yang uang sedikit	1	1-5	1-2
11	<i>Lost</i>	Pelanggan dengan pembelian terlama, jarang bertransaksi berapapun jumlah besaran uangnya	1	1-2	1-5

Pada **Tabel 1.** menunjukkan bahwa analisis RFM menghasilkan 11 kategori dalam segmentasi pelanggan dengan 7 di antaranya merupakan kategori yang telah mencakup ketiga nilai komponen pada RFM (*Recency*, *Frequency*, *Monetary*), yakni *Champions*, *Loyal Customer*, *Promising Customer*, *New Customer*, *Old Potential*, *Hibernating* dan *Lost* dengan 4 lainnya merupakan kategori yang tidak terlalu berpengaruh terhadap salah satu nilai komponen RFM (*Monetary*), yakni *Need Attention*, *At risk*, *Potential* dan *About to sleep*.

Setelah menganalisis karakteristik pelanggan untuk segmentasi, diperoleh perbandingan jumlah pelanggan berdasarkan segmen dari dua *dataset*: *df1* (2009–2010) dan *df2* (2010–2011), khusus untuk pelanggan di United Kingdom. Berikut hasil perbandingan yang didapatkan.

Tabel 2. Perbandingan Jumlah Pelanggan Segmentasi Pelanggan pada *df1* dan *df2*

No	Segmentasi Pelanggan	Jumlah Pelanggan Tahun 2009–2010	Jumlah Pelanggan Tahun 2010–2011
----	----------------------	----------------------------------	----------------------------------

1	<i>Champions</i>	918	937
2	<i>Loyal Customer</i>	329	357
3	<i>Promising Customer</i>	23	20
4	<i>New Customer</i>	301	259
5	<i>Need Attention</i>	477	454
6	<i>At risk</i>	320	320
7	<i>Potential</i>	435	403
8	<i>About to sleep</i>	344	360
9	<i>Old Potential</i>	126	101
10	<i>Hibernating</i>	70	80
11	<i>Lost</i>	583	596

Dalam hasil yang ditunjukkan oleh **Tabel 2.**, terdapat beberapa kategori hasil segmentasi pelanggan di negara *United Kingdom* mengalami penurunan dan peningkatan. Pada negara *United Kingdom* beberapa kategori yang mengalami peningkatan pelanggan adalah *Champions*, *Loyal Customer*, *About to sleep*, *Hibernating* dan *Lost*. Sementara kategori yang mengalami penurunan pelanggan adalah *Promising Customer*, *New Customer*, *Need Attention*, *At risk*, *Potential* dan *Old Potential*.

4. KESIMPULAN DAN SARAN

4.1 Kesimpulan

Berdasarkan hasil dan pembahasan dalam penelitian “Penerapan *Exploratory Data Analysis* (EDA) dan Analisis *Recency, Frequency, and Monetary* (RFM) untuk Segmentasi Pelanggan E-Commerce” didapatkan kesimpulan sebagai berikut.

1. Metode *Exploratory Data Analysis* (EDA) dapat menangani anomali data yang tersembunyi dengan menemukan nilai negatif sebanyak 8.472 (*df1 UK*) dan 7.533 (*df2 UK*) pada kolom *Quantity* yang disebabkan oleh pembatalan transaksi barang yang ditandai huruf “C” pada kolom *Invoice*.
2. Metode *Exploratory Data Analysis* (EDA) dapat dikolaborasikan dengan analisis *Recency, Frequency, and Monetary* (RFM). EDA meningkatkan kualitas data dengan memastikan data yang akan digunakan sudah bersih sebelum melakukan analisis RFM. Metode EDA memastikan tidak adanya tumpang tindih pada skor RFM satu pelanggan ke pelanggan lainnya saat dilakukan segmentasi pelanggan. Dalam proses EDA dilakukan pendalaman mengenai karakteristik setiap skor pada komponen RFM sehingga mempermudah segmentasi pelanggan.
3. Analisis RFM negara *United Kingdom* menghasilkan 11 kategori segmentasi pelanggan, yaitu *Champions*, *Loyal Customer*, *Promising Customer*, *New Customer*, *Need Attention*, *At Risk*, *Potential*, *About To Sleep*, *Old Potensial*, *Hibernating*, dan *Lost*. Dengan kategori segmentasi pelanggan yang dihasilkan RFM dapat mengetahui karakteristik pada setiap pelanggan.

4.2 Saran

Berdasarkan hasil penelitian yang telah dilakukan, terdapat beberapa saran yang dapat diberikan pada penelitian ini sebagai berikut.

1. Untuk peneliti selanjutnya diharapkan dapat meninjau kembali kategori dalam segmentasi pelanggan. Terdapat 4 dari 11 jenis kategori dalam segmentasi RFM yang belum memuat berdasarkan ketiga nilai komponen RFM sehingga perlu dilakukan kesesuaian dan perhitungan kembali agar keseluruhan kategori dalam segmentasi berdasarkan ketiga nilai komponen RFM. Begitu juga untuk *Outlier* yang dimiliki pada data harus diberi perhatian lebih lanjut dengan melakukan penelitian khusus mengenai jumlah *outlier* yang ditemukan sangatlah besar.
2. Untuk perusahaan ritel atau e-commerce diharapkan dapat mempertimbangkan menggunakan metode EDA dan analisis RFM seperti pada penelitian ini untuk mengetahui dan mengidentifikasi karakteristik pelanggan yang telah dikategorikan ke dalam segmentasi pelanggan sehingga dapat memudahkan perusahaan dalam melakukan proses lanjutan ke dalam bidang pemasaran guna membuat strategi pemasaran yang efektif dan berdasarkan tren, membantu pembuatan keputusan dan tata kelola strategis, serta memprediksi perilaku pelanggan di masa mendatang.

5. REFERENSI

- Harani, N. H. (dkk.). (2020). Segmentasi Pelanggan Produk Digital Service Indihome Menggunakan Algoritma K-Means Berbasis Python. *Jurnal Manajemen Informatika (JAMIKA)*, 10(2), 133-146.
- Angela, J. B. dan Irsyad, A. (2023). Implementasi Visualisasi Data Berbasis Web Pada Exploratory Data Analysis Profil Kesehatan Kota Samarinda. *Kreatif Teknologi dan Sistem Informasi (KRETISI)*, 1(1), 9-16.
- Tivano, D., Ruslianto, I., & Prawira, D (2023). SISTEM SEGMENTASI PELANGGAN POTENSIAL MENGGUNAKAN METODE K-MEANS DAN ANALISIS RFM. *Coding Jurnal Komputer dan Aplikasi*, 11(03), 325-337.
- Rizaty, Monavia Ayu. (2023, Januari 24). Transaksi e-Commerce RI Tak Capai Target pada 2022. *DataIndonesia.id* [Online]. Tersedia: <https://dataindonesia.id/ekonomi-digital/detail/transaksi-ecommerce-ri-tak-capai-target-pada-2022>. [2023, Maret 1].
- Mustajab, Ridhwan. (2023, September 4). Pengguna E-Commerce RI Diproyeksi Capai 196,47 Juta pada 2023. *dataindonesia.id* [Online]. Tersedia : <https://dataindonesia.id/ekonomi-digital/detail/pengguna-ecommerce-ri-diproyeksi-capai-19647-juta-pada-2023> [2023, Maret 1].
- Nugraha, Jefri Putri. (dkk.). 2021. Teori Perilaku Konsumen. Pekalongan: Penerbit NEM.
- KATRAGADDA, V. (2022). Dynamic Customer Segmentation: Using Machine Learning to Identify and Address Diverse Customer Needs in Real-Time. *IRE Journals*, 5(10), 278-279.
- Amutha, R. (dkk.). 2023. Customer Segmentation Using Machine Learning Techniques. *Tuijin Jishu / Journal of Propulsion Technology*, 44(3).
- Leni, D. (dkk.). (2023). Evaluasi Sifat Mekanik Baja Paduan Rendah Berdasarkan Komposisi Kimia Dan Suhu Perlakuan Panas Menggunakan Teknik *Exploratory Data Analysis* (EDA). *Dinamika Teknik Mesin*, 13(1), 74-83.
- Da Poian, V., Theiling, B., Clough, L., McKinney, B., Major, J., Chen, J., & Hörst, S. (2023). Exploratory data analysis (EDA) machine learning approaches for ocean world analog mass spectrometry. *Frontiers in Astronomy and Space Sciences*, 10, 1134141.
- Iriawan, Nur dan Permadina Kanah. (dkk.). 2021. Aplikasi *Data Science* dengan Mudah Menggunakan Python. Penerbit Andi.

-
- Alzami, F., Sambasri, F. D., Nabila, M., Megantara, R. A., Akrom, A., Premunendar, R. A., ... & Sulistiyawati, P. (2023). Implementation of RFM method and K-Means algorithm for Customer Segmentation in E-Commerce with Streamlit. *ILKOM Jurnal Ilmiah*, 15(1), 32-44.
- 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), 1785-1792.
- Supangat, S., & Mulyani, Y. (2023). Customer Loyalty Analysis Using Recency, Frequency, Monetary (RFM) and K-means Cluster for Labuan Bajo Souvenirs in Online Store. *Journal of Information Systems and Informatics*, 5(1), 285-299.
- Wicaksana, Putu Aryasuta. (dkk.). (2022). "Literature Review Analisis Perilaku Pelanggan Menggunakan RFM Model." *Majalah Ilmiah Teknologi Elektro*. 21(1), 21-30.