Jurnal Teknologi dan Informasi (JATI)
Volume 13 Nomor 2 Edisi September 2023
P-ISSN 2088-2270, E-ISSN 2655-6839
DOI 10.34010/jati.v13i2

Optimalisasi Strategi Pemasaran dengan Segmentasi Pelanggan Menggunakan Penerapan K-Means Clustering pada Transaksi Online Retail

Optimizing Marketing Strategies with Customer Segmentation Using K-Means Clustering on Online Retail Transactions

Eriskiannisa Febrianty Luchia Awalina[1], **Woro Isti Rahayu**[2]
Program Studi D4 Teknik Informatika, Universitas Logistik dan Bisnis Internasional, Indonesia[1]
Program Studi Data Science, Universitas Logistik dan Bisnis Internasional, Indonesia[2]
eriskiannisaluch20@gmail.com*[1], woroisti@ulbi.ac.id[2]

Abstrak

Pemahaman yang baik mengenai pelanggan sangat penting untuk menjalankan bisnis bagi suatu perusahaan. Mengenali dan memahami setiap pelanggan dapat membantu menciptakan komunikasi dalam menyampaikan penawaran produk dengan menyesuaikan kebutuhan dan memberikan layanan yang disesuaikan setiap pelanggan. Namun, dalam mengidentifikasi setiap kebutuhan pelanggan tidak mudah, karena faktanya menganalisis pelanggan adalah area yang sangat luas. Hal ini dapat mencakup berbagai karakteristik dan perilaku pelanggan yang berbeda. Oleh karena itu, diperlukan segmentasi pelanggan untuk mengelompokkan pelanggan berdasarkan perilaku dan karakteristik. Untuk melakukan segmentasi pelanggan berdasarkan data, banyak model dan algoritma telah digunakan, dan dalam penelitian ini, metode clustering menggunakan algoritma K-means menjadi salah satu pilihan yang efektif. Metode ini telah menjadi tren dan banyak digunakan dimana hal tersebut dibuktikan dengan banyak nya jurnal terkait dari rentang tahun 2018 - 2022. Penelitian ini menggunakan pemrograman python untuk proses data mining dan pre-processing yang dilakukan pada data melalui exploratory data analysis untuk memahami informasi dari data yang digunakan sebelum melakukan klasterisasi. Dalam penerapan metode K-means, digunakan metode elbow untuk menentukan jumlah klaster yang optimal. Hasil dari metode elbow menunjukkan bahwa penggunaan 4 klaster adalah pilihan yang tepat dalam kasus ini. Selanjutnya, pemodelan K-means dengan 4 klaster dilakukan menggunakan variabel quantity, unit price, dan customer id, dan menghasilkan 4 klaster yang berbeda dengan karakteristik yang spesifik pada masing-masingnya. dapat diamati bahwa kuantitas dan harga satuan berperan penting dalam mempengaruhi perilaku pelanggan.
Kata kunci: Clustering; K-Means; Online Retail; Python; Segmentasi Pelanggan.

Abstract

A good understanding of customers is very important to run a business for a company. Recognising and understanding each customer can help create communication in delivering product offerings by tailoring needs and providing tailored services to each customer. However, identifying each customer's needs is not easy, due to the fact that analysing customers is a very broad area. It can include a variety of different customer characteristics and behaviours. Therefore, customer segmentation is required to group customers based on behaviour and characteristics. To perform customer segmentation based on data, many models and algorithms have been used, and in this research, the clustering method using the K-means algorithm is one of the effective choices. This method has become a trend and is widely used which is proven by many related journals from the range of 2018 - 2022. This research uses python programming for the data mining process and pre-processing is done on the data through exploratory data analysis to understand the information from the data used before clustering. In applying the K-means method, the elbow method is used to determine the optimal number of clusters. The results of the elbow method show that the use of 4 clusters is the right choice in this case. Furthermore, K-means modelling with 4 clusters was performed using the variables quantity, unit price, and customer id, and resulted in 4 different clusters with specific characteristics in each. It can be observed that quantity and unit price play an important role in influencing customer behaviour.
Keywords: Clustering; Customer Segmentation; K-Means; Online Retail; Python.

Naskah diterima 19 Juni 2023; direvisi 1 Agustus 2023; dipublikasi 1 September 2023.

1. Pendahuluan

Perkembangan teknologi informasi semakin meningkat tentunya sejalan dengan kondisi peradaban manusia, termasuk dalam bidang bisnis. Dalam bisnis, persaingan menuntut perusahaan untuk memaksimalkan kemampuan dengan sebaik-baiknya agar dapat bertahan dan bersaing di antara perusahaan lainnya. Perusahaan harus mampu memahami dan memasukkan karakteristik pelanggan yang menjadi salah satu hal penting untuk dipertimbangkan[1]. Salah satu bisnis yang sedang berkembang adalah belanja online melalui e-commerce yang memperlihatkan bagaimana perilaku dari pembelian pelanggan yang berubah secara dinamis. Seiring

Jurnal Teknologi dan Informasi (JATI)
Volume 13 Nomor 2 Edisi September 2023
P-ISSN 2088-2270, E-ISSN 2655-6839
DOI 10.34010/jati.v13i2

dengan semakin ketatnya persaingan dalam industri bisnis tersebut, perusahaan dituntut untuk mengalihkan fokusnya tidak hanya pada produk, namun juga berfokus pada pelanggan[2]. Dalam mempertahankan dan memperluas bisnis, perusahaan perlu menyadari bahwa mempertahankan pelanggan yang sudah ada sama pentingnya dengan mencari pelanggan baru. Layanan terhadap pelanggan juga merupakan salah satu hal yang penting untuk memberikan Customer Experience yang baik. Penyebab Customer Experience yang buruk seperti perusahaan yang tidak mampu memberikan solusi atas permasalahan yang dihadapi pelanggan atau mengalami kesulitan dalam mengidentifikasi pelanggan akan menyebabkan resiko kehilangan pelanggan dan penurunan jumlah transaksi[3]. Selain itu, banyaknya kompetitor menyebabkan harga produk yang ditawarkan bersaing dengan perusahaan lain, sehingga pelanggan perlu memilih harga yang lebih cocok untuk dibeli. Oleh karena itu, dibutuhkan segmentasi pelanggan untuk membagi pelanggan ke dalam beberapa kelompok dan mengidentifikasi setiap kebutuhan pelanggan yang berbeda.

Segmentasi merupakan proses membagi pasar menjadi segmen yang lebih kecil dan segmentasi pelanggan mengacu pada bagaimana membagi sasaran pasar menjadi kelompok yang sesuai dengan karakteristik untuk mengembangkan strategi bisnis yang tepat. Pembagian segmentasi merupakan proses untuk membagi pelanggan menjadi beberapa kelompok atau segmen yang dapat dikelola berdasarkan dengan karakteristik seperti loyalitas, demografis, kebiasaan membeli atau frekuensi pembelian untuk mengembangkan strategi dalam pemasaran ke setiap kelompok tersebut. Segmentasi pelanggan merupakan pendekatan untuk memahami kebutuhan dan perilaku pelanggan. Tujuan dari segmentasi pelanggan ini adalah untuk lebih mengetahui dan memahami target pasar serta promosi apa yang paling cocok untuk diberikan kepada setiap segmen pelanggan[4]. Namun, untuk dapat melakukan analisis dan identifikasi pelanggan dengan segmentasi tidak mudah, karena area dari segmentasi yang sangat luas dan data transaksi yang semakin hari semakin banyak jumlahnya serta pengelolaan analisis data dalam skala yang besar menjadi tantangan yang kompleks dalam melakukan segmentasi. Dalam hal ini, diperlukan adanya sebuah metode yang tepat dan dapat membantu proses segmentasi pelanggan agar lebih mudah. Penelitian ini menggunakan metode penelitian studi literatur yang dilakukan guna mendapatkan metode apa yang paling efisien dilakukan untuk permasalahan segmentasi. Pada peneliti sebelumnya yang diacu dari beberapa jurnal terpilih dengan rentang waktu 5 tahun terakhir yaitu tahun 2018 hingga tahun 2022 menggunakan algoritma clustering untuk mengatasi permasalahan segmentasi.

Penelitian sebelumnya berjudul "Pemanfaatan Data Transaksi Untuk Dasar Membangun Strategi Berdasarkan Karakteristik Pelanggan Dengan Algoritma K-Means Clustering Dan Model RFM" yang dilakukan oleh peneliti Carudin [5]. Penelitian ini fokus pada proses transaksi pelanggan di sebuah perusahaan retail. Perusahaan mengalami penurunan jumlah transaksi dari tahun 2017 hingga 2019, yaitu 2092 transaksi dari 1040 pelanggan pada tahun 2017, 1754 transaksi dari 922 pelanggan pada tahun 2018, dan 486 transaksi dari 250 pelanggan pada tahun 2019. penelitian ini menggunakan model RFM (Recency, Frequency, Monetary) dan menerapkan algoritma K-means untuk mengelompokkan data berdasarkan karakteristik pelanggan. Terdapat 6 klaster yang digunakan mengacu pada karakteristik pelanggan. Klaster yang dihasilkan adalah: Dormant Customer (318 pelanggan), Everyday Shopper (316 pelanggan), Occasional Customer (315 pelanggan), Typical Customer (316 pelanggan), Golden Customer (319 pelanggan), dan Super Star (314 pelanggan). Hasil uji kinerja klaster menunjukkan bahwa pengelompokkan dengan 6 klaster menghasilkan nilai Davies Bouldin sebesar 0.500, yang menandakan bahwa klaster ini memiliki kinerja terbaik dan optimal. Kemudian pada penelitian sebelumnya berjudul "Penerapan Data Mining Untuk Menentukan Segmentasi Pelanggan Dengan Menggunakan Algoritma K-Means dan Model RFM Pada E-Commerce" yang dilakukan oleh peneliti S. Sharyanto dan D. Lestari [6]. Penelitian ini mengatasi masalah dimana beberapa pelaku usaha menghadapi kesulitan dalam mengenali karakteristik dan kebutuhan konsumen serta kesulitan dalam melakukan segmentasi. Oleh karena itu, penelitian ini menggunakan metode data mining dengan algoritma K-means untuk klasterisasi dan analisis RFM. Data yang digunakan adalah dataset online transaksi pada e-commerce yang terdiri dari 8 atribut dan 532 record. Melalui uji klasterisasi dengan menggunakan nilai cluster K dari 2 hingga 5, ditemukan bahwa terdapat 4 segment yang paling optimal pada cluster K=4, dengan nilai Davies Bouldin Index (DBI) sebesar 0,6788. Selanjutnya pada Penelitian sebelumnya berjudul "Analisis Segmentasi Pelanggan Menggunakan Kombinasi RFM Model dan Teknik Clustering" yang dilakukan oleh peneliti B. E. Adiana, I. Soesanti, dan A. E. Permanasari [1]. Penelitian ini membahas tentang perusahaan UD Gemilang yang menghadapi kesulitan dalam mengidentifikasi minat dan selera konsumen terhadap produknya. Dalam penelitian ini, dilakukan segmentasi pelanggan dengan menggunakan teknik k-means dan analisis data transaksi menggunakan RFM Model. Hasil penelitian menunjukkan terdapat 3 klaster dengan karakteristik masing-masing. Klaster pertama terdiri dari 30 pelanggan yang termasuk dalam kategori "typical customers". Klaster kedua terdiri dari 8 pelanggan yang termasuk dalam kategori "superstar". Sedangkan klaster ketiga terdiri dari 89 pelanggan yang termasuk dalam kategori "dormant customer". Sehingga untuk melihat perbandingan antara hasil peneliti sebelumnya, maka disajikan dalam bentuk tabel 1.

Jurnal Teknologi dan Informasi (JATI)
Volume 13 Nomor 2 Edisi September 2023
P-ISSN 2088-2270, E-ISSN 2655-6839
DOI 10.34010/jati.v13i2

Tabel 1. Hasil perbandingan penelitian

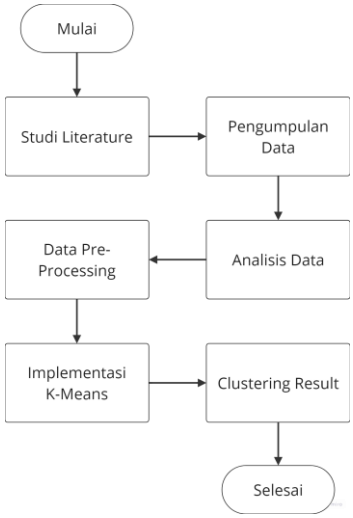
Judul Jurnal dan Peneliti	Tahun Terbit	Metode	Objek Penelitian	Perbandingan yang dijadikan alasan tinjauan penelitian
Pemanfaatan Data Transaksi Untuk Dasar Membangun Strategi Berdasarkan Karakteristik Pelanggan Dengan Algoritma K-Means Clustering Dan Model RFM Peneliti: Carudin [5]	2021	RFM Model dan K-Means	Pelanggan perusahaan retail	Penelitian ini menggunakan 6 cluster yang telah ditentukan secara acak, namun jumlah cluster tersebut bisa dioptimalkan dengan menggunakan metode elbow. Sehingga hasil penelitian ini digunakan sebagai patokan bagaimana menentukan cluster yang optimal dengan penentuan nilai k menggunakan metode lainnya seperti elbow.
Penerapan Data Mining Untuk Menentukan Segmentasi Pelanggan Dengan Menggunakan Algoritma K-Means dan Model RFM Pada E-Commerce Peneliti: S. Sharyanto dan D. Lestari [6]	2022	RFM Model dan K-Means	Data transaksi e-commerce	Penelitian ini juga menggunakan pengujian cluster k=2 sampai k=5, kemudian nilai optimal ditentukan dengan penggunaan metode davies bouldin index. Sehingga hasil penelitian ini digunakan sebagai patokan bagaimana menentukan cluster yang optimal dengan penentuan nilai k menggunakan metode lain seperti elbow. Pada penelitian juga dilakukan proses data mining yang menjadi sumber untuk dilakukannya kembali proses tersebut oleh peneliti sebelum implementasi pemodelan menggunakan k-means.
Analisis Segmentasi Pelanggan Menggunakan Kombinasi RFM Model dan Teknik Clustering Peneliti: B. E. Adiana, I. Soesanti, dan A. E. Permanasari [1]	2018	RFM Model dan K-Means	Pelanggan UD Gemilang Kencana	Penelitian ini menggunakan kombinasi model RFM dan K-Means serta penggunaan metode DBI dan silhouette untuk uji nilai cluster. Sehingga hasil dari penelitian ini digunakan untuk melihat bagaimana jika metode yang digunakan hanya menggunakan K-Means tanpa RFM. Teori pada penelitian ini juga digunakan kembali sebagai acuan, berupa kepercayaan dan loyalitas konsumen merupakan hal yang penting untuk mendapatkan kepercayaan dari konsumen.

Sehingga berdasarkan dari penelitian sebelumnya, penelitian ini bertujuan untuk menginvestigasi bagaimana perkembangan topik penelitian sebelumnya yang berkaitan dengan segmentasi pelanggan dan metode yang banyak digunakan untuk mengatasi permasalahan segmentasi dalam penelitian sebelumnya. Segmentasi pelanggan adalah proses pengelompokkan pelanggan berdasarkan karakteristik yang serupa, sehingga memungkinkan perusahaan untuk lebih memahami kebutuhan dan preferensi pelanggan serta meningkatkan efektivitas strategi pemasaran. Selain itu, pre-processing dan eksplorasi data juga menjadi langkah penting sebelum melakukan pemodelan klasterisasi dan bagaimana analisis deskriptif terhadap data penjualan produk dan pelanggan [5],[6]. Melalui eksplorasi data, peneliti dapat memahami pola-pola, tren, dan karakteristik penting dalam data sebelum memulai proses pemodelan klasterisasi. Selain itu, tujuan dari pre-processing adalah agar data lebih terorganisir dan mudah digunakan saat melakukan proses analisis serta klasifikasi pada data[8]. Metode yang digunakan dalam penelitian ini adalah teknik klasterisasi dengan menggunakan algoritma K-Means. K-Means adalah salah satu algoritma klasterisasi yang populer dan efisien untuk mengelompokkan data ke dalam beberapa klaster berdasarkan kesamaan karakteristik. Peneliti juga akan menggunakan metode elbow untuk menentukan nilai klaster yang optimal. Penelitian ini juga bertujuan untuk menemukan jumlah segmen pelanggan yang tepat untuk dikelompokkan berdasarkan karakteristiknya masing-masing. Dengan mengidentifikasi segmen pelanggan yang berbeda, perusahaan dapat mengarahkan upaya pemasaran dan layanan mereka sesuai dengan preferensi dan kebutuhan masing-masing segmen, sehingga meningkatkan kepuasan pelanggan dan efisiensi operasional.Dengan melakukan penelitian ini, diharapkan akan didapatkan wawasan yang lebih mendalam tentang perkembangan topik penelitian segmentasi pelanggan dan metode yang paling efektif dalam mengatasi permasalahan segmentasi. Selain itu, hasil eksplorasi data, analisis deskriptif, dan pemodelan klasterisasi akan memberikan wawasan yang berharga bagi perusahaan dalam meningkatkan strategi pemasaran dan kepuasan pelanggan, serta meningkatkan daya saing di pasar yang semakin kompetitif. Penelitian ini menggunakan data transaksi online retail yang merupakan kumpulan data transnasional berisi data transaksi online. Data transaksi tersebut dimanfaatkan dan dilakukan pre-processing untuk menemukan informasi penting di dalam data kemudian dikelompokkan dengan menggunakan pendekatan algoritma k-means dan metode elbow untuk menentukan nilai kluster yang optimal, selain itu pada penelitian ini akan digunakan bahasa pemrograman python untuk analisis data dan penerapan metode [9].

Jurnal Teknologi dan Informasi (JATI)
Volume 13 Nomor 2 Edisi September 2023
P-ISSN 2088-2270, E-ISSN 2655-6839
DOI 10.34010/jati.v13i2

2. Metode Penelitian

Pada bagian ini akan dijelaskan mengenai alur dari metode penelitian yang dilakukan. Berikut yang terlihat pada gambar 1 merupakan diagram dari alur metode pada penelitian.



Gambar 1. Tahapan Alur Metode Penelitian

Berikut merupakan tahapan yang dilakukan pada penelitian berdasarkan gambar dari diagram alur metode penelitian:

1. Pengumpulan Data
   Pada tahap data collecting atau pengumpulan data, akan digunakan data yang berkaitan dengan penelitian. Dataset yang digunakan adalah data transaksi Online Retail yang merupakan data transnasional dengan total data 54910 dengan jumlah kolom 8[10].
2. Studi Literatur
   Tahap selanjutnya dalam penelitian yaitu studi literatur, dimana pada tahap ini akan dilakukan pencarian serta pengkajian jurnal yang membahas tentang metode apa saja yang digunakan dari penelitian sebelumnya yang berkaitan dengan topik penelitian, kemudian hasil dari studi literatur ini akan menentukan metode mana yang paling efektif untuk diterapkan pada permasalahan segmentasi pelanggan.
3. Analisis Data
   Pada tahap kedua ini, akan dilakukan data understanding atau penyelidikan awal untuk mengenali data yang digunakan seperti atribut dan variabel apa saja yang ada pada data. Analisis data juga merupakan salah satu tahap dalam pengolahan data dengan tujuan untuk mengambil informasi penting di dalam data.
4. Data Pre-Processing
   Pre-processing merupakan proses awal dalam pengolahan data pada exploratory data analysis sehingga akan menghasilkan data dengan format yang sesuai dan siap untuk digunakan pada tahap selanjutnya. Tujuan dari pre-processing adalah agar data lebih mudah digunakan saat melakukan klasifikasi. Berikut merupakan tahap yang dilakukan saat pre-processing[7]:
   - Handle Missing Value: menemukan data kosong, kemudian drop variable yang memiliki missing value atau mengganti missing value tersebut dengan average, median atau modus dari data.
   - Handle Outlier: merupakan value yang memiliki nilai ekstrem. Cara menangani data yang memiliki outlier adalah dengan menghitung inter kuartil range.
   - Data Transformation: mengconvert data categorical dan data numeric dengan one hot encode atau feature scalling.
5. Implementasi K-Means Clustering
   Tahap selanjutnya yaitu mengimplementasikan model machine learning menggunakan model K-Means Clustering yang merupakan salah satu algoritma unsupervised learning. Dalam tahapan ini akan dievaluasi hasil klasterisasi dari penerapan dan pembagian cluster ke segmen pelanggan. Alur segmentasi pelanggan dimulai dari tahap menyiapkan data kemudian seleksi data yang akan di cluster dan menentukan nilai cluster dengan penerapan metode elbow.
6. Clustering Result
   Dari proses pembagian cluster maka dapat di identifikasi hasil segmentasi pelanggan menggunakan K-Means. Hasil clustering berupa kelompok pelanggan yang memiliki kontribusi dalam penjualan yang terbagi menjadi beberapa segmen tertentu sesuai dengan perilaku pembelian.

Jurnal Teknologi dan Informasi (JATI)
Volume 13 Nomor 2 Edisi September 2023
P-ISSN 2088-2270, E-ISSN 2655-6839
DOI 10.34010/jati.v13i2

3. Hasil dan Pembahasan

Bagian ini berisi hasil penelitian berdasarkan metode penelitian yang sudah ditentukan sebelumnya. Berikut merupakan hasil dan pembahasan dari penelitian yang dilakukan.

3.1 Studi Literatur

Studi literatur pada penelitian ini dilakukan dengan search proses yang merupakan tahap pencarian yang dilakukan untuk mendapatkan sumber yang sesuai dengan objek penelitian yang dicari. Proses penelusuran jurnal dalam penelitian ini dapat bersumber dari database Google Scholar, IEEE Explore atau Semantic Scholar dengan syarat paper jurnal dapat di download dan memiliki open access. Pencarian dilakukan dengan bantuan tools "Publish or Perish" untuk melakukan pencarian literatur dengan memasukkan keyword atau kata kunci yang sesuai dengan topik seperti "Segmentasi Pelanggan"+"K-means" dan mencari berbagai literatur yang membahas metode apa yang paling sering digunakan untuk menyelesaikan masalah segmentasi pelanggan agar mendapat hasil yang optimal. Hasil dari studi literatur yang telah dilakukan pada penelitian ini telah dipilih sebanyak 23 paper jurnal sesuai dengan kriteria paper jurnal yang diterbitkan dengan rentang 5 tahun terakhir yaitu tahun 2018 – 2022 dan memiliki bahasan terkait topik "Segmentasi Pelanggan" dan "metode k-means". Informasi tersebut selanjutnya dikelompokkan berdasarkan tabel 2.

Tabel 2. Hasil jurnal terpilih sesuai dengan topik

Cites	Penulis	Variable Jurnal Penelitian	Sumber Jurnal Penelitian	Jumlah	Tahun
[11]	N Ahsina, F Fatimah	K-Means Algorithm, Metode elbow, Pelanggan Kredit Bank	Jurnal Ilmiah Teknologi Informasi Terapan	1	2022
[1]	BE Adiana, I Soesanti	Metode CRISP-DM, RFM model, K-Means algorithm, usaha kecil dan menengah (UKM)	Jurnal Terapan Teknologi Informasi	1	2018
[12]	NH Harani, C Prianto, FA Nugraha	K-Means Algorithm, Metode Elbow, Customer Profiling	Jurnal Manajemen Informatika (JAMIKA)	1	2020
[13]	S Sharyanto, D Lestari	Data Mining, Model RFM, K-Means Algorithm, E-Commerce Data	JURIKOM (Jurnal Riset Komputer)	2	2022
[14]	MA Satriawan, R Andreswari	Data Pelanggan Pengguna Telkomsel, RFM Model, K-Means Algorithm	e-Proceeding of Engineering	2	2021
[15]	NW Wardani, GR Dantes…	Prediksi customer churn, algoritma decision tree C4.5, data perusahaan retail, RFM Model	Jurnal RESISTOR (Rekayasa Sistem Komputer)	1	2018
[16]	K Anam, D Sudrajat, DA Kurnia	K-Means Algorithm	Jurnal ICT: Information Communication & Technology	1	2022
[17]	S Monalisa	Customer Purchase Behavior, K-Means Algorithm, RFM	Query: Journal of Information Systems	1	2018
[18]	S Setiawan, H Amani	K-means Algorithm, Model Rfm, Klinik Kecantikan Seoul Secret	SISTEMASI: Jurnal Sistem Informasi	1	2021
[7]	R Siagian, P Sirait, A Halim	K-Means Algorithm, K-Medoids Algorithm, E-commerce Data Transactions	Seminar Nasional Teknik dan Manajemen Industri	1	2022
[19]	R Afthoni, M Hamdhani, A Ardianto…	Machine Learning, K-Means Algorithm, Data Konsumsi Listrik di PT PLN XYZ	SNIA (Seminar Nasional Informatika dan Aplikasinya)	1	2021
[8]	Y Christian, KOYR Qi	K-Means Algorithm, Startup Early Stage, CRISP-DM	Scientific Journal of Informatics	1	2022
[3]	A Fauzi	K-Means Algorithm, Data Transaksi Superstore	EXPLORE Jurnal	1	2019
[20]	A Alamsyah, PE Prasetyo, S Sunyoto	RFM Model, K-Means Algorithm, Elbow Method	Mathematical Problems in Engineering	1	2022
[21]	T Juhari, A Juarna	RFM Model, K-Means Algorithm, Online Bookstore	Proceedings - 2018 IEEE 15th International Conference on e-Business Engineering, ICEBE 2018	1	2022
[2]	J. Wu	RFM Model, K -Means Algorithm	Journal of King Saud University - Computer and Information Sciences	1	2020
[22]	M. Tavakoli	User Behavior Analysis, RFM Model, Data Mining	IEEE Access	1	2018
[23]	P. Anitha	RFM model, customer purchase behavior, K-Means Algorithm	SRPH (Scientific Research Publishing House)	1	2022
[24]	R Taghi Livari, N	RFM Model, K-Means Algorithm, Food Distribution Industry	Jurnal Repositor	1	2021

Cites	Penulis	Variable Jurnal Penelitian	Sumber Jurnal Penelitian	Jumlah	Tahun
	Zarrin Ghalam				
[25]	ABH Kiat	K-Means Algorithm, Metode Elbow, RFM Model	Jurnal Sistem Informasi (Journal of Information Systems)	1	2020
[26]	Turkmen B	machine learning, k-means clustering, hierarchical clustering, DBSCAN clustering	The European Journal of Social and Behavioural Sciences	1	2022
[27]	Devarapalli D, Sowjanya Virajitha A, Sai G	K-Means, RFM, DBSCAN	International Journal of Mechanical Engineering	1	2022
[28]	Suharti P, Suryandari A	K-Means, pelanggan Alfagift, metode elbow, data mining	Sebatik	1	2022

Berdasarkan pada tabel 2, hasil analisis dalam berbagai studi literatur pustaka yang dilakukan, dapat disimpulkan bahwa algoritma k-means tetap menjadi salah satu pilihan utama dalam analisis data dan masih banyak digunakan oleh peneliti sebelumnya dibuktikan dari penelitian sebelumnya yang menggunakan metode k-means dalam rentang wkatu tahun 2018- 2022 terutama digunakan untuk melakukan segmentasi data, suatu proses yang membagi data ke dalam kelompok – kelompok yang lebih terdefinisi. Selain itu, beberapa penelitian juga telah mengadopsi metode elbow untuk mencari nilai klaster yang optimal. Metode elbow memungkinkan peneliti untuk menemukan jumlah klaster yang paling sesuai untuk data yang dianalisis.

Hasil analisis dari berbagai studi literatur terlihat bahwa algoritma k-means memberikan hasil yang konsisten dan dapat diandalkan dalam berbagai aplikasi, termasuk dalam masalah segmentasi. Metode elbow juga membantu peneliti dalam menentukan parameter penting seperti jumlah kluster yang optimal untuk memaksimalkan efektivitas analisis data. Sehingga, penggunaan k-means dan metode elbow telah terbukti memberikan kontribusi yang signifikan dalam bidang analisis data dan pengelompokan data secara keseluruhan.

3.2 Pengumpulan Data

Pada penelitian ini digunakan dataset transaksi online retail dengan jumlah 54910 dengan 8 kolom. Pada tabel 3 merupakan keterangan dan deskripsi dari variable yang akan digunakan pada penelitian.

Tabel 3. Deskripsi dataset

Kategori	Keterangan
Invoice No	Berisi 6 digit angka yang secara unique di assigned untuk masing - masing transaksi atau biasanya disebut dengan transaction ID.
Stock Code	Produk atau item code dengan 5 digit angka yang secara unique di assigned ke masing - masing produk atau biasa disebut dengan product ID.
Description	Detail atau nama produk.
Quantity	Quantity atau kuantitas barang yang terjual dari setiap produk per transaksi.
Invoice Date	Informasi terkait waktu kapan transaksi di generate.
Unit Price	Harga produk per unit.
Customer ID	5 digit angka identitas untuk masing - masing customer.
Country	Informasi mengenai negara mana yang melakukan transaksi.

3.3 Analisis Data dan Pre-processing

Selanjutnya, setelah data berhasil dikumpulkan, akan dilakukan tahap analisis data dan pre-processing. Tahap pre-processing bertujuan untuk membersihkan data dari potensi masalah seperti data yang hilang, outlier atau data yang tidak relevan, sehingga data yang digunakandalam proses pemodelan klasterisasi menjadi lebih terstruktur.

3.3.1 Import library

Dalam penelitian ini, akan dilakukan pre-processing data dan penerapan algoritma k-means pada data transaksi online retail. Proses pre-processing dan penerapan k-means akan menggunakan berbagai library python yang relevan untuk analisis data. Selain itu, untuk membantu dalam implementasi kode program, penelitian ini menggunakan Google Colab sebagai platform untuk menjalankan kode dan analisis data. Berikut merupakan daftar library python yang diperlukan untuk penelitian ini, sebagaimana terlihat pada gambar 2.

Jurnal Teknologi dan Informasi (JATI)
Volume 13 Nomor 2 Edisi September 2023
P-ISSN 2088-2270, E-ISSN 2655-6839
DOI 10.34010/jati.v13i2

```
# Import modules
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

from sklearn.cluster import KMeans
from sklearn.preprocessing import MinMaxScaler
from sklearn.preprocessing import StandardScaler
from sklearn import metrics
from mpl_toolkits.mplot3d import Axes3D

import warnings
warnings.filterwarnings("ignore")
pd.set_option('display.float_format', lambda x: '%.2f' % x)
```

Gambar 2. Import Library

Pada library python yang digunakan memiliki masing – masing fungsi yang digunakan dalam membantu proses analisis data dan tahap pre-processing. Berikut merupakan fungsi dari tiap – tiap library python yang digunakan:

- Pandas: untuk membaca file data dengan ekstensi csv atau excel.
- Numpy: untuk proses komputasi numeric data.
- Matplotlib: untuk plotting angka dan membuat diagram.
- Seaborn: untuk menentukan hubungan 2 variabel dan menganalisis perbedaan distribusi univariate dan bivariate.
- Sklearn: sklearn atau scikit-learn merupakan modul Bahasa pemrograman python untuk membantu dalam melakukan processing data, training data dan kebutuhan yang berkaitan dengan machine learning dan data science.
- Import k-means: untuk menerapkan algoritma pengelompokkan dan clustering menggunakan k-means.
- Import MinMaxScaler: untuk normalisasi data dan bekerja untk scalling data menyesuaikan data dalam rentang tertentu.
- Import StandardScaler: library ini digunakan sebagai metode pre-processing dalam melakukan standarisasi fitur.
- Import Metrics: library ini digunakan untuk evaluasi algoritma machine learning.

3.3.2 Dataset

Setelah tahap pre-processing data, dilanjutkan dengan eksplorasi data untuk memperoleh pemahaman yang lebih mendalam mengenai dataset yang digunakan. Pada gambar 3 terlihat isi dari data transaksi online retail yang menjadi fokus penelitian ini. Untuk membaca data tersebut, digunakan fungsi read_csv yang merupakan bagian dari library pandas. Fungsi ini berfungsi untuk membaca file data dalam format csv dan mengonversikannya menjadi struktur data DataFrame yang lebih mudah dikelola dalam analisis data[29]. Selanjutnya, digunakan pula fungsi data.head() untuk melihat beberapa baris data teratas dari dataset tersebut. Fungsi head() ini memungkinkan kita untuk dengan cepat meninjau beberapa baris pertama data tanpa perlu melihat seluruh dataset secara keseluruhan.

Melalui tahap eksplorasi data ini, penelitian dapat mengidentifikasi pola-pola awal, tren, dan karakteristik penting dalam data transaksi online retail sebelum melangkah ke tahap selanjutnya dalam analisis dan pemodelan klasterisasi. Informasi yang diperoleh dari eksplorasi data ini membantu peneliti dalam mempersiapkan langkah-langkah selanjutnya dengan lebih baik dan memastikan data yang digunakan dalam penelitian ini telah terbaca dan dipahami dengan benar.

```
data = pd.read_csv("online_retail.csv",encoding='unicode escape')
data
```

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	2010-12-01 08:26:00	2.55	17850.00	United Kingdom
1	536365	71053	WHITE METAL LANTERN	6	2010-12-01 08:26:00	3.39	17850.00	United Kingdom
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	2010-12-01 08:26:00	2.75	17850.00	United Kingdom
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	2010-12-01 08:26:00	3.39	17850.00	United Kingdom
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	2010-12-01 08:26:00	3.39	17850.00	United Kingdom
...	...	...	...	...	...	...	...	...
541904	581587	22613	PACK OF 20 SPACEBOY NAPKINS	12	2011-12-09 12:50:00	0.85	12680.00	France
541905	581587	22899	CHILDREN'S APRON DOLLY GIRL	6	2011-12-09 12:50:00	2.10	12680.00	France
541906	581587	23254	CHILDRENS CUTLERY DOLLY GIRL	4	2011-12-09 12:50:00	4.15	12680.00	France
541907	581587	23255	CHILDRENS CUTLERY CIRCUS PARADE	4	2011-12-09 12:50:00	4.15	12680.00	France
541908	581587	22138	BAKING SET 9 PIECE RETROSPOT	3	2011-12-09 12:50:00	4.95	12680.00	France

Gambar 3. Cek isi data

Pada tahap eksplorasi data selanjutnya, dilakukan pemeriksaan terhadap bentuk data untuk mengetahui berapa banyak jumlah baris dan kolom yang terdapat dalam dataset. Proses ini dilakukan dengan menggunakan fungsi shape pda data yang telah dibaca sebagai DataFrame. Output dari data.shape akan menampilkan informasi tentang jumlah baris dan kolom dari data transaksi online retail yang terlihat pada gambar 4. Dengan

Jurnal Teknologi dan Informasi (JATI)
Volume 13 Nomor 2 Edisi September 2023
P-ISSN 2088-2270, E-ISSN 2655-6839
DOI 10.34010/jati.v13i2

mengevaluasi bentuk data melalui fungsi shape, penelitian dapat memastikan bahwa dataset telah diimpor dengan benar dan sesuai. Informasi mengenai jumlah baris dan kolom juga membantu peneliti dalam memahami seberapa besar ukuran data yang sedang dihadapi.

```
[7]  # Rows and Column
     print("Rows: {}, Columns: {}".format(data.shape[0], data.shape[1]))

     Rows: 536641, Columns: 8
```

Gambar 4. Cek baris dan kolom data

Selain Langkah pemeriksaan bentuk data, tahap eksplorasi data juga melibatkan pengecekan terhadap tipe data yang dimiliki oleh setiap kolom dalam dataset. Tujuan dari pengecekan ini adalah untuk memastikan bahwa tipe data yang ada telah sesuai dengan karakteristik dan representasi yang seharusnya. Jika terdapat tipe data yang tidak sesuai, maka diperlukan transformasi data untuk mengubahnya menjadi tipe data yang tepat sesuai kebutuhan analisis. Pada gambar 5 merupakan tipe data dari masing – masing kolom sebelum dilakukan transformasi data.

```
# Cek data types
data.dtypes

InvoiceNo       object
StockCode       object
Description     object
Quantity         int64
InvoiceDate     object
UnitPrice      float64
CustomerID     float64
Country         object
dtype: object
```

Gambar 5. Cek Tipe Data

Langkah eksplorasi data selanjutnya adalah dengan melakukan pengecekan terhadap deskriptif data dari dataset. Dalam analisis data, deskriptif data sangat penting untuk memberikan gambaran ringkas tentang distribusi dan karakteristik dari masing-masing variabel. Untuk melakukan ini, digunakan fungsi describe dari library pandas. Fungsi ini akan menampilkan summary statistics seperti jumlah data (count), rata-rata (mean), standar deviasi (std), nilai minimum (min), dan nilai maksimum (max) dari setiap variabel. Gambar 6 menampilkan deskripsi data dari dataset yang digunakan, yang mencakup statistik ringkas dari setiap variabel yang diamati. Deskripsi data ini membantu peneliti dalam memahami karakteristik data secara keseluruhan dan dapat menjadi titik awal untuk menjalankan analisis yang lebih mendalam, termasuk penanganan outlier atau pengambilan keputusan tentang pre-processing lanjutan yang diperlukan.

```
data.describe()
```

	Quantity	UnitPrice	CustomerID	Sales
count	380580.00	380580.00	380580.00	380580.00
mean	12.82	3.13	15293.53	22.05
std	127.73	22.43	1712.59	158.23
min	1.00	0.00	12346.00	0.00
25%	2.00	1.25	13969.00	4.95
50%	6.00	1.95	15159.00	11.90
75%	12.00	3.75	16793.00	19.80
max	74215.00	8142.75	18287.00	77183.60

Gambar 6. Deskriptif data

3.3.3 Duplikat Data

Untuk mengidentifikasi duplikat data dalam dataset, digunakan fungsi duplicated().sum() dari library pandas. Fungsi ini akan menghitung jumlah data yang duplikat dalam dataset. Duplikat data merujuk pada data yang memiliki kondisi di mana dua atau lebih baris dalam dataset memiliki kesamaan data dan nilai yang sama. Gambar 7 menunjukkan langkah - langkah yang dilakukan untuk mengecek dan menghapus data yang merupakan duplikat. Pengecekan dilakukan dengan fungsi duplicated().sum(), untuk menghitung jumlah data yang duplikat. Jika hasilnya lebih dari 0, berarti ada duplikasi data dalam dataset. Setelah identifikasi dilakukan, langkah selanjutnya adalah menghapus data yang duplikat. Deduplikasi dilakukan menggunakan fungsi drop_duplicates() yang akan menghilangkan baris-baris yang memiliki kesamaan data dan nilai yang sama.

Jurnal Teknologi dan Informasi (JATI)
Volume 13 Nomor 2 Edisi September 2023
P-ISSN 2088-2270, E-ISSN 2655-6839
DOI 10.34010/jati.v13i2

```
# Check duplicate
data.duplicated().sum()
```

```
5268
```

```
# Drop duplicate
data = data.drop_duplicates()
data.duplicated().sum()
```

```
0
```

Gambar 7. Cek duplikat data

3.3.4 Penanganan Missing Value

Langkah eksplorasi data selanjutnya adalah dengan memeriksa apakah terdapat data yang hilang pada dataset. Data yang hilang atau missing value merupakan nilai yang tidak ada atau tidak terdefinisi dalam data. Pada gambar 8 terlihat bahwa beberapa kolom dalam dataset memiliki data yang hilang. Adanya data hilang akan mempengaruhi analisis data dan menyebabkan ketidakakuratan dalam hasil yang diperoleh. Untuk menangani missing value tersebut, maka dapat dilakukan dengan cara menghapus kolom atau memasukkan nilai ke dalam missing value. Fungsi isnull().sum() digunakan untuk mengecek nilai null dalam data dan menjumlahkan bilangan numerik yang ada pada dataframe berdasarkan variabel.

```
# Cek missing value
data.isnull().sum()
```

```
InvoiceNo            0
StockCode            0
Description       1454
Quantity             0
InvoiceDate          0
UnitPrice            0
CustomerID      135037
Country              0
dtype: int64
```

Gambar 7. Penanganan Data yang Hilang

Data yang memiliki missing value dapat diisi dengan nilai ‘NaN’ saat melakukan analisis statistik, nilai tersebut tidak akan mempengaruhi hasilnya, atau dapat menggunakan fungsi replace() untuk mengganti nilai kosong dengan salah satu nilai yang ada pada data. Pada penelitian ini missing value ditangani dengan cara menghapus beberapa row yang memiliki missing value seperti yang terlihat pada gambar 8.

```
# Remove rows that have missing value
data = data.dropna()
```

Gambar 8. Menghapus Data yang Hilang

3.3.5 Transformasi Data dan Menambahkan Kolom Baru

Pada tahap data transformation, dilakukan manipulasi data dengan mengubah format sehingga data siap untuk digunakan dalam analisis dan pemodelan. Dalam penelitian ini, data transformation dilakukan pada beberapa variabel untuk menyesuaikan tipe data yang digunakan. Misalnya, pada kolom "invoice date" yang sebelumnya memiliki tipe data "object" akan diubah menjadi tipe data "datetime", sehingga memungkinkan untuk melakukan analisis berbasis tanggal dan waktu dengan lebih efisien. Selain itu, pada kolom "customer id" yang awalnya berbentuk "float64" akan diubah menjadi tipe data "object", sehingga memperlakukan "customer id" sebagai data kategori atau label yang lebih tepat. Salah satu fungsi utama dari data transformation adalah mengubah data dari format categorical (non-numerical) menjadi format numerik. Hal ini berguna ketika data akan digunakan untuk analisis yang memerlukan operasi matematika atau pemodelan statistik.

```
# Convert data types for some column
data["InvoiceDate"] = pd.to_datetime(data["InvoiceDate"])
data["InvoiceDate_date"] = data["InvoiceDate"].dt.date
data["CustomerID"] = data["CustomerID"].astype("str")

data.dtypes
```

```
InvoiceNo                  object
StockCode                  object
Description                object
Quantity                    int64
InvoiceDate        datetime64[ns]
UnitPrice                 float64
CustomerID                 object
Country                    object
InvoiceDate_date           object
dtype: object
```

Gambar 9. Cek Transformasi Data

Selanjutnya menambahkan kolom penjualan baru dengan rumus "Kuantitas * harga satuan" merupakan langkah dalam tahap preprocessing data untuk menghitung nilai penjualan untuk setiap transaksi atau item dalam dataset. Dengan menambahkan kolom penjualan baru, kita dapat dengan mudah melihat nilai penjualan untuk setiap transaksi atau item dalam dataset, yang akan sangat berguna untuk analisis lebih lanjut dan pemodelan.

Jurnal Teknologi dan Informasi (JATI)
Volume 13 Nomor 2 Edisi September 2023
P-ISSN 2088-2270, E-ISSN 2655-6839
DOI 10.34010/jati.v13i2

```
# Add new column sales with formula = Quantity * unit price
data["Sales"] = data["Quantity"] * data["UnitPrice"]

data.head()
```

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country	InvoiceDate_date	Sales
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	2010-12-01 08:26:00	2.55	17850.0	United Kingdom	2010-12-01	15.30
1	536365	71053	WHITE METAL LANTERN	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom	2010-12-01	20.34
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	2010-12-01 08:26:00	2.75	17850.0	United Kingdom	2010-12-01	22.00
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom	2010-12-01	20.34
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom	2010-12-01	20.34

Gambar 10. Tambah kolom sales

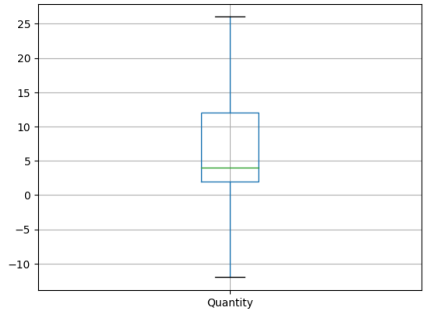
3.3.6 Deteksi dan Penanganan Outlier

Outlier merupakan titik data yang memiliki nilai yang signifikan berbeda dengan data lainnya. Pada penelitian ini, untuk mengidentifikasi outlier, digunakan visualisasi boxplot. Boxplot adalah metode grafis yang memungkinkan kita untuk dengan cepat melihat adanya outlier dalam dataset. Selain itu, pada penelitian ini, outlier juga diidentifikasi menggunakan perhitungan quantile dengan metode interquantile range (IQR). Metode IQR menghitung jarak antara kuartil atas dan kuartil bawah dalam data. Proses penanganan outlier dimulai dengan menghitung nilai IQR pada setiap variabel dalam dataset. Setelah nilai IQR diperoleh, batas atas (upper bound) dan batas bawah (lower bound) ditentukan dengan menghitung nilai kuartil atas (Q3) ditambah dengan 1.5 kali nilai IQR dan nilai kuartil bawah (Q1) dikurangi dengan 1.5 kali nilai IQR. Kemudian, nilai-nilai yang berada di luar batas atas dan batas bawah akan dianggap sebagai outlier dan akan dihapus dari dataset. Pada Gambar 10, ditunjukkan cara menangani outlier pada penelitian ini.

```
# Remove outliers
def remove_outlier(df_in, col_name):
  q1 = df_in[col_name].quantile(0.25)
  q3 = df_in[col_name].quantile(0.75)
  iqr = q3-q1 #Interqurtile range
  fence_low = q1-1.5*iqr
  fence_high = q3+1.5*iqr
  df_out = df_in.loc[(df_in[col_name] > fence_low) & (df_in[col_name] < fence_high)]
  return df_out
```

Gambar 11. Penanganan Outlier

Kemudian setelah dilakukan penghapusan data outlier, maka dapat dilihat pada boxplot yang akan menunjukkan apakah masih ada outlier pada data. Pada gambar 11 merupakan diagram boxplot setelah dilakukan penghapusan outlier menggunakan interquantile range.



Gambar 12. Diagram Boxplot

3.3.7 Korelasi Data

Korelasi merupakan teknik yang digunakan untuk mencari hubungan antar dua variabel numerik pada data atau variabilitas gabungan dari dua variabel dan korelasi memiliki nilai antar -1 dan 1. Berikut merupakan aturan nilai korelasi:

- Jika korelasi bernilai 1 atau mendekati 1 maka variabel tersebut memiliki hubungan perfect positif correlation.
- Jika korelasi bernilai 0 maka variabel tersebut tidak berhubungan sama sekali.
- Jika korelasi bernilai -1 maka kedua variabel memiliki hubungan perfect negative correlation.

Analisis korelasi ini akan fokus pada variabel-unit price (harga unit), quantity (jumlah produk), dan sales (penjualan) pada dataset. Variabel ini dipilih karena analisis korelasi hanya dapat dilakukan pada data yang bersifat numerik. Pada penelitian ini, dilakukan analisis korelasi antara unit price, quantity, dan sales untuk melihat sejauh mana harga produk, jumlah produk yang terjual, dan total penjualan saling berhubungan. Hal ini penting karena dapat membantu dalam pemahaman tentang bagaimana perubahan dalam harga atau jumlah produk mempengaruhi penjualan secara keseluruhan.

Jurnal Teknologi dan Informasi (JATI)
Volume 13 Nomor 2 Edisi September 2023
P-ISSN 2088-2270, E-ISSN 2655-6839
DOI 10.34010/jati.v13i2

Gambar 13. Diagram Korelasi

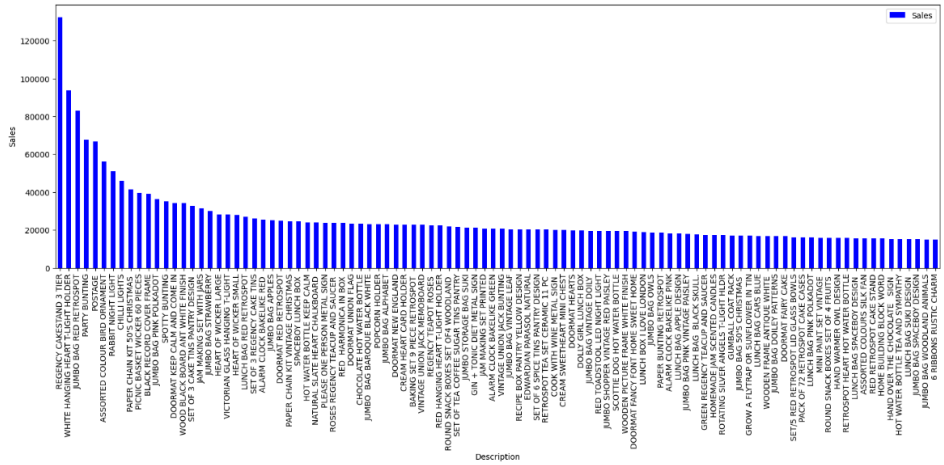
Pada gambar 13, dapat dilihat hubungan antara variabel quantity, unit price, dan sales. Dalam analisis ini, nilai korelasi antara sales dengan quantity terlihat sebesar 0,92. Nilai korelasi ini menunjukkan bahwa antara sales dan quantity memiliki hubungan yang sangat erat dan positif, karena mendekati 1. Artinya, semakin tinggi nilai quantity (jumlah produk yang terjual), maka semakin tinggi juga nilai sales (total penjualan). Sementara itu, antara sales dengan unit price terlihat nilai korelasi sebesar -0,13. Nilai korelasi yang mendekati 0 menunjukkan bahwa tidak ada hubungan yang signifikan antara sales dan unit price (harga produk). Artinya, perubahan harga produk tidak berpengaruh secara kuat terhadap total penjualan.

3.4 Deskriptif Analisis

Setelah dilakukan ekplorasi data, informasi terkait dengan data yang diteliti dapat dihasilkan. Deskriptif analisis memberikan gambaran yang lebih mendalam tentang berbagai tren penjualan, pembelian, dan tren pelanggan selama periode waktu tertentu. Analisis deskriptif juga dapat memberikan gambaran tentang sebaran data, seperti nilai penjualan maksimum dan minimum, rata-rata penjualan, dan variasi data. Informasi ini membantu dalam memahami karakteristik data dan memperoleh wawasan tentang variabilitas dalam penjualan, pembelian, atau tren pelanggan.

3.4.1 Kinerja Produk

Dalam konteks ini, dimiliki tiga produk dengan penjualan tertinggi yaitu "REGENCY CAKESTAND 3 TIER," "WHITE HANGING HEART T-LIGHT HOLDER," dan "JUMBO BAG RED RETROSPOT". Berdasarkan data tersebut, dapat disimpulkan bahwa "REGENCY CAKESTAND 3 TIER" adalah produk dengan kinerja terbaik dari ketiga produk yang diamati, karena memiliki nilai penjualan paling tinggi sebesar 132567.70. Sementara itu, "WHITE HANGING HEART T-LIGHT HOLDER" memiliki penjualan yang lebih rendah dibandingkan dengan "REGENCY CAKESTAND 3 TIER," tetapi masih cukup baik karena berada di peringkat kedua dalam penjualan dengan total penjualan sebesar 93767.80. Sedangkan "JUMBO BAG RED RETROSPOT" memiliki penjualan lebih rendah dibandingkan dengan kedua produk lainnya, namun tetap merupakan produk yang memiliki penjualan yang signifikan dengan total penjualan sebesar 83056.52. Pada gambar 14 terlihat grafik produk dengan total penjualan tertinggi hingga produk dengan total penjual terendah.

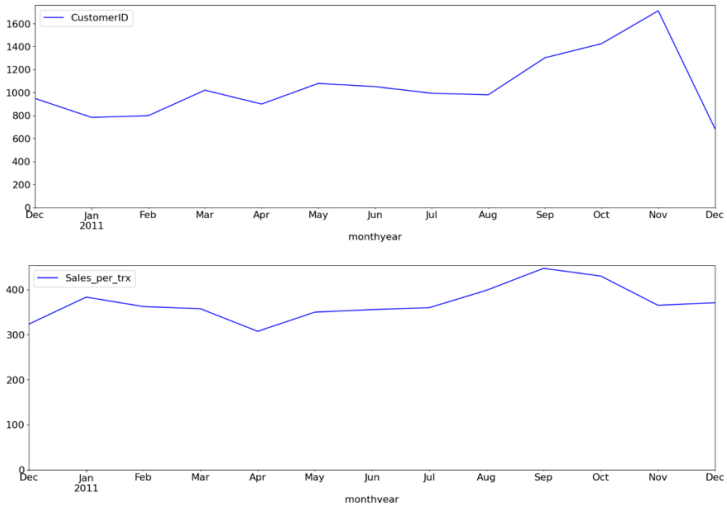


Gambar 14. Diagram Kinerja Produk

Jurnal Teknologi dan Informasi (JATI)
Volume 13 Nomor 2 Edisi September 2023
P-ISSN 2088-2270, E-ISSN 2655-6839
DOI 10.34010/jati.v13i2

3.4.2 Tren Penjualan dan Pelanggan

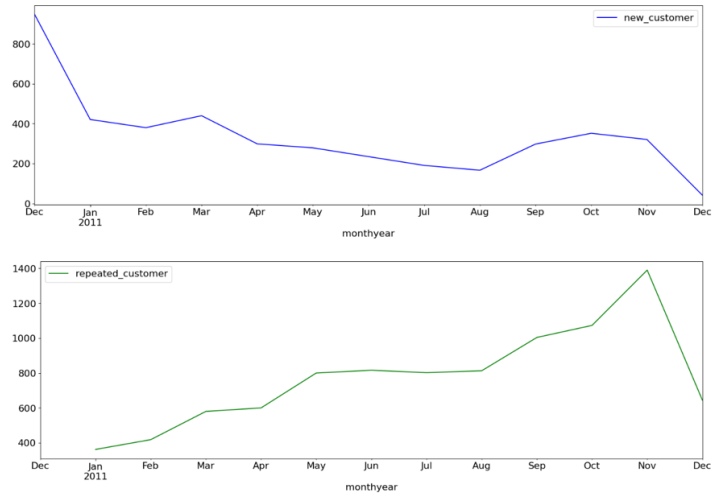
Tren penjualan dan pelanggan digunakan untuk memahami perkembangan bisnis dari waktu ke waktu, khususnya dalam hal penjualan dan jumlah pelanggan. Analisis tren ini memungkinkan untuk melihat bagaimana kinerja bisnis berubah dari bulan ke bulan atau dari tahun ke tahun, sehingga membantu perusahaan dalam mengidentifikasi pola dan mengambil keputusan strategis. Pada gambar 15 merupakan tren pelanggan dari setiap bulannya dan rata-rata penjualan per transaksi selama periode tertentu yang Ini menunjukkan besarnya nilai transaksi rata-rata yang dihasilkan dari setiap pembelian pelanggan.



Gambar 15. Diagram Tren Penjualan dan Pelanggan

3.4.3 Visualisasi Proporsi Pelanggan Baru dan Pelanggan Lama

Proporsi antara pelanggan baru (new customers) dan pelanggan tetap (repeated customers) dapat diidentifikasi dan dipahami melalui analisis data yang mencatat informasi tentang pelanggan yang melakukan pembelian dalam periode waktu tertentu. Dalam hal ini, perhatian tertuju pada bagaimana jumlah pelanggan baru dan pelanggan tetap berubah seiring berjalannya waktu. Pada gambar 16 grafik yang menggambarkan proporsi antara pelanggan baru dan pelanggan tetap selama periode waktu tertentu, kita dapat melihat pola perubahan yang terjadi dari bulan ke bulan. Repeated Customer mengalami kenaikan dari bulan September hingga November, terlihat adanya kenaikan dalam jumlah pelanggan tetap (repeated customers). Ini menunjukkan bahwa selama periode tersebut, lebih banyak pelanggan yang melakukan pembelian secara berulang daripada sebelumnya. New Customer mengalami kenaikan di Bulan Maret meskipun tidak mencapai angka yang tinggi, terlihat bahwa jumlah pelanggan baru (new customers) mengalami kenaikan di bulan Maret.

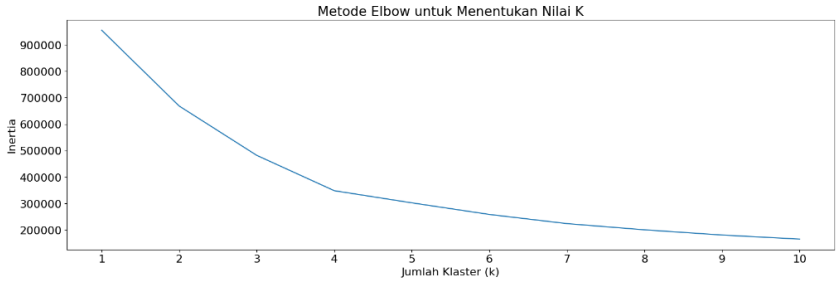


Gambar 16. Diagram proporsi pelanggan baru dan pelanggan lama

3.5 Implementasi K-Means dan Metode Elbow

Jurnal Teknologi dan Informasi (JATI)
Volume 13 Nomor 2 Edisi September 2023
P-ISSN 2088-2270, E-ISSN 2655-6839
DOI 10.34010/jati.v13i2

Setelah tahap pre-processing dan analisis data, selanjutnya tahap implementasi menggunakan metode k-means. Sebelum masuk ke tahap implementasi, pada metode k-means perlu diketahui nilai k atau jumlah kluster. Penentuan nilai k akan berpengaruh pada baik atau tidaknya klaster tersebut. Dalam algoritma K-Means, jumlah kluster atau nilai k akan diinisialisasi secara acak untuk memulai proses klasterisasi pada data. Proses ini melibatkan pemilihan titik pusat atau centroid untuk setiap kluster, dan iterasi akan terus berlangsung hingga tidak ada perubahan posisi centroid yang signifikan. Untuk menentukan jumlah kluster yang optimal dari data, digunakan metode elbow (siku). Metode elbow adalah salah satu pendekatan yang umum digunakan dalam analisis K-Means untuk menemukan jumlah kluster yang paling tepat. Metode ini berdasarkan pada pemikiran bahwa penambahan kluster akan mengurangi Within Cluster Sum of Square (WCSS), yaitu jumlah kuadrat jarak antara setiap titik data dengan pusat kluster [30]. Semakin banyak kluster yang ditambahkan, semakin kecil WCSS yang diperoleh. WCSS akan terus menurun ketika nilai k meningkat karena setiap titik data akan lebih dekat dengan centroid dalam kluster yang lebih kecil [31].



Gambar 17. Grafik Metode Elbow

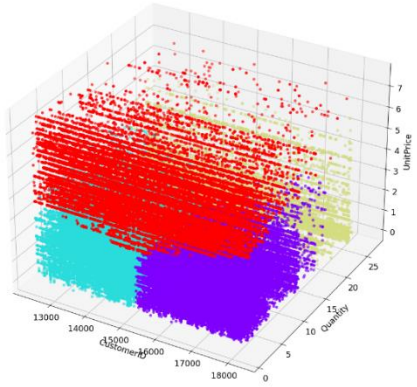
Berikut pada tabel 4 merupakan nilai wcss dari kluster 1 sampai 10.

Tabel 4. Nilai WCSS setiap kluster

Clusters	WCSS
1	955337.9999999835
2	668540.7921399678
3	482055.1282752048
4	348351.5301644874
5	302478.7943320485
6	258523.2754754402
7	223864.6874139582
8	200377.0897622632
9	180127.83713655904
10	165306.06142894237

3.5.1 Hasil Clustering

Pada penelitian ini, nilai k yang telah ditentukan adalah 4, sehingga metode K-Means akan diterapkan dengan membentuk 4 kluster berbeda. hasil kluster yang telah terbentuk akan ditunjukkan melalui plotting data. Pada gambar 16, terlihat hasil klasterisasi dari metode K-Means yang telah dilakukan dengan 4 kluster. Plotting tersebut akan memvisualisasikan titik-titik data yang telah dikelompokkan ke dalam masing-masing kluster dengan warna atau tanda yang berbeda untuk masing-masing kluster.



Gambar 16. Plotting hasil cluster

Pengelompokan data transaksi berdasarkan 3 variabel yaitu quantity (kuantitas produk), unit price (harga satuan), dan customerid (identitas pelanggan). Dari hasil clustering yang terbentuk, kita dapat menyimpulkan karakteristik dari masing-masing cluster sebagai berikut:

1. Cluster 0: Kluster ini memiliki rata – rata kuantitas produk adalah 7.20, rata – rata harga satuan adalah 1.70, dan rata – rata Customer ID adalah 13905.00. Kluster ini terdiri dari pelanggan dengan jumlah barang yang dibeli rendah dan harga satuan barang yang tinggi. Kode pelanggan di kluster ini juga memiliki nilai yang tinggi, menandakan potensi loyalitas pelanggan atau kemungkinan pelanggan dengan tingkat pembelian yang lebih sering. Rekomendasi yang dapat dilakukan adalah dengan memberikan program diskon atau berikan penawaran ekslusif seperti produk – produk baru kepada pelanggan.
2. Cluster 1: Kluster ini memiliki rata – rata kuantitas produk adalah 5.30, rata – rata harga satuan adalah 1.70, dan rata – rata Customer ID adalah 16927.90. Kluster ini terdiri dari pelanggan dengan jumlah barang yang dibeli menengah dan harga satuan barang yang cukup rendah. Kode pelanggan di kluster ini juga memiliki nilai yang tinggi, yang menunjukkan potensi loyalitas. Rekomendasi yang dapat dilakukan adalah penawaran khusus untuk pembelian dalam jumlah tertentu untuk mendorong pelanggan di kluster ini untuk membeli lebih banyak.
3. Cluster 2: Kluster ini memiliki rata – rata kuantitas produk adalah 23.00, rata – rata harga satuan adalah 1.10, dan rata - rata Customer ID adalah 15017.30. Kluster ini terdiri dari pelanggan dengan jumlah barang yang dibeli tinggi dan harga satuan barang yang rendah. Kode pelanggan di kluster ini juga memiliki nilai yang menengah. Rekomendasi yang dapat dilakukan adalah program hadiah atau undian bagi pelanggan di kluster ini dengan setiap pembelian untuk memberikan kejutan dan meningkatkan kepuasan pelanggan.
4. Cluster 3: Kluster ini memiliki rata – rata kuantitas produk adalah 3.60, rata – rata harga satuan adalah 14.80, dan rata – rata Customer ID adalah 15217.90. Kluster ini terdiri dari pelanggan dengan jumlah barang yang dibeli rendah dan harga satuan barang yang tinggi. Kode pelanggan di kluster ini juga memiliki nilai yang menengah. Berikan diskon berkala atau penawaran khusus pada waktu-waktu tertentu untuk mendorong pembelian berulang dari pelanggan di kluster ini.

4. Kesimpulan

Berdasarkan beberapa penelitian, sudah digunakan banyak model dan algoritma yang digunakan untuk mengklasifikasikan dan mengelompokkan pelanggan sesuai dengan data. Dalam penelitian ini, metode dari salah satu algoritma unsupervised learning direkomendasikan untuk digunakan pada permasalahan segmentasi pelanggan. Salah satu model segmentasi yang masih menjadi tren dari tahun 2018 hingga tahun 2022 adalah dengan penggunaan algoritma clustering yaitu K-means. Penelitian ini melakukan proses data mining dan pre-processing pada data dengan exploratory data analysis dan menghasilkan temuan informasi di dalam data mnegenai tren penjualan dan tren pelanggan. Kemudian dilakukan penerapan metode k-means dan metode elbow pada penelitian ini untuk mencari nilai kluster berdasarkan nilai WCSS yang ditunjukkan melalui grafik, sehingga pada penelitian ini dihasilkan 4 kluster. Pengelompokkan K-Means menggunakan 3 variabel data yaitu quantity, unit price dan customer id menghasilkan 4 karakteristik pelanggan. Kluster 0 ini terdiri dari pelanggan dengan jumlah barang yang dibeli rendah dan harga satuan barang yang tinggi dan kode pelanggan yang tinggi. Kluster 1 ini terdiri dari pelanggan dengan jumlah barang yang dibeli menengah dan harga satuan barang yang cukup rendah dan kode pelanggan yang tinggi, Kluster 2 ini terdiri dari pelanggan dengan jumlah barang yang dibeli tinggi dan harga satuan barang yang rendah dan kode pelanggan yang menengah. Kluster 3 ini terdiri dari pelanggan dengan jumlah barang yang dibeli rendah dan harga satuan barang yang tinggi dan kode pelanggan yang menengah. Dari hasil klasterisasi, dapat diamati bahwa kuantitas dan harga satuan berperan penting dalam mempengaruhi perilaku pelanggan. Perbandingan antara metode elbow dengan metode penentuan klaster lainnya dapat menjadi pertimbangan untuk peneliti selanjutnya guna menentukan nilai k yang paling optimal.

Daftar Pustaka

[1] B. E. Adiana, I. Soesanti, and A. E. Permanasari, "Analisis Segmentasi Pelanggan Menggunakan Kombinasi RFM Model dan Teknik Clustering," *JUTEI (Jurnal Terapan Teknologi Informasi)*, vol. 2, no. 1, pp. 23–32, 2018, doi: 10.21460/jutei.2017.21.76.

[2] J. Wu *et al.*, "An Empirical Study on Customer Segmentation by Purchase Behaviors Using a RFM Model and K -Means Algorithm," *Math Probl Eng*, vol. 2020, 2020, doi: 10.1155/2020/8884227.

[3] A. Fauzi Sistem Informasi, F. H. Universitas Buana Perjuangan Karawang Jl Ronggowaluyo, T. Timur, and K. priati, *Data Mining dengan Teknik Clustering Menggunakan Algoritma K-Means pada Data Transaksi Superstore*. 2017. [Online]. Available: http://community.tableau.com.

[4] D. Devarapalli *et al.*, "Analysis of RFM Customer Segmentation Using Clustering Algorithms," *International Journal of Mechanical Engineering*, vol. 7, no. 1, 2022, Accessed: Apr. 11, 2023.

[Online]. Available: https://www.researchgate.net/publication/358285794_Analysis_of_RFM_Customer_Segmentation_Using_Clustering_Algorithms

[5] Carudin, "Pemanfaatan Data Transaksi Untuk Dasar Membangun Strategi Berdasarkan Krekteristik Pelanggan dengan Algoritma K-Means Clustering dan Model RFM," *Jurnal Teknologi Terpadu*, vol. 7, no. 1, pp. 7–14, Jul. 2021, Accessed: Apr. 09, 2023. [Online]. Available: https://journal.nurulfikri.ac.id/index.php/jtt

[6] S. Sharyanto and D. Lestari, "Penerapan Data Mining Untuk Menentukan Segmentasi Pelanggan Dengan Menggunakan Algoritma K-Means dan Model RFM Pada E-Commerce," *JURIKOM (Jurnal Riset Komputer)*, vol. 9, no. 4, p. 866, Aug. 2022, doi: 10.30865/jurikom.v9i4.4525.

[7] R. Siagian, P. Sirait, and A. Halim, "Penerapan Algoritma K-Means dan K-Medoids untuk Segmentasi Pelanggan pada Data Transaksi E-Commerce," May 2022. [Online]. Available: http://sistemasi.ftik.unisi.ac.id

[8] Y. Christian and K. O. Y. R. Qi, "Penerapan K-Means pada Segmentasi Pasar untuk Riset Pemasaran pada Startup Early Stage dengan Menggunakan CRISP-DM," *JURIKOM (Jurnal Riset Komputer)*, vol. 9, no. 4, p. 966, Aug. 2022, doi: 10.30865/jurikom.v9i4.4486.

[9] A. Satriawan, R. Andreswari, and O. N. Pratiwi, "Segmentasi Pelanggan Telkomsel Menggunakan Metode Clustering Dengan RFM Model dan Algoritma K-Means," *e-Proceeding of Engineering*, vol. 8, no. 2, pp. 2876–2883, 2021, Accessed: Jun. 02, 2023. [Online]. Available: https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/14687

[10] T. Juhari, A. Juarna, and U. Gunadarma, "Implementation Rfm Analysis Model For Customer Segmentation Using The K-Means Algorithm Case Study Xyz Online Bookstore," 2022. Accessed: May 02, 2023. [Online]. Available: https://utmmataram.ac.id/ojs/index.php/explore/article/view/548/pdf

[11] N. Huda Ahsina, F. Fatimah, F. Rachmawati, U. Ibn Khaldun Bogor JlKH Sholeh Iskandar Km, and K. Bogor, "Analisis Segmentasi Pelanggan Bank Berdasarkan Pengambilan Kredit dengan Menggunakan Metode K-Means Clustering," 2022.

[12] N. H. Harani, C. Prianto, and F. A. Nugraha, "Segmentasi Pelanggan Produk Digital Service Indihome Menggunakan Algoritma K-Means Berbasis Python," *Jurnal Manajemen Informatika (JAMIKA)*, vol. 10, no. 2, pp. 133–146, 2020, doi: 10.34010/jamika.v10i2.

[13] S. Sharyanto and D. Lestari, "Penerapan Data Mining Untuk Menentukan Segmentasi Pelanggan Dengan Menggunakan Algoritma K-Means dan Model RFM Pada E-Commerce," *JURIKOM (Jurnal Riset Komputer)*, vol. 9, no. 4, p. 866, Aug. 2022, doi: 10.30865/jurikom.v9i4.4525.

[14] A. Satriawan, R. Andreswari, and O. N. Pratiwi, "Segmentasi Pelanggan Telkomsel Menggunakan Metode Clustering Dengan RFM Model dan Algoritma K-Means," Apr. 2021. Accessed: May 02, 2023. [Online]. Available: https://repository.telkomuniversity.ac.id/pustaka/files/168042/jurnal_eproc/segmentasi-pelanggan-telkomsel-menggunakan-metode-clustering-dengan-rfm-model-dan-algoritma-k-means.pdf

[15] N. W. Wardani, G. Dantes, and G. Indrawan, "PREDIKSI CUSTOMER CHURN DENGAN ALGORITMA DECISION TREE C4.5 BERDASARKAN SEGMENTASI PELANGGAN PADA PERUSAHAAN RETAIL," *JURNAL RESISTOR*, vol. 1, no. 1, 2018.

[16] K. Anam, D. Sudrajat, D. A. Kurnia, and N. Masuk, "Analisis Segmentasi Pelanggan Menggunakan Metode K-Means Clustering," *Jurnal ICT : Information Communication & Technology*, vol. 21, pp. 273–278, 2022.

[17] S. Monalisa, "Segmentasi Perilaku Pembelian Pelanggan Berdasarkan Model RFM dengan Metode K-Means," *Jurnal Sistem Informasi*, vol. 2, no. 1, p. 1, 2018, Accessed: Jun. 08, 2023. [Online]. Available: http://jurnal.uinsu.ac.id/index.php/query/article/view/1553

[18] H. Amani and W. Tripiawan, "Perancangan Segmentasi Pelanggan dengan Metode Clustering K-Means dan Model RFM pada Klinik Kecantikan Seoul Secret," *e-Proceeding of Engineering* , vol. 8, no. 2, pp. 2286–2293, 2021, Accessed: Jun. 08, 2023. [Online]. Available: https://repository.telkomuniversity.ac.id/pustaka/files/167904/jurnal_eproc/perancangan-segmentasi-pelanggan-dengan-metode-clustering-k-means-dan-model-rfm-pada-klinik-kecantikan-seoul-secret.pdf

[19] R. Afthoni, M. Hamdhani, A. Fitri Karimah, H. Patria, J. Analitika Bisnis, and F. Magister Manajemen Teknologi, "Pemanfaatan Algoritma Machine Learning untuk Segmentasi Pelanggan Berbasis Data Konsumsi Listrik di PT PLN XYZ," *Seminar Nasional Teknik dan Manajemen Industri dan Call for Paper(SENTEKMI 2021)*, vol. 1, no. 1, pp. 222–231, 2021, Accessed: Jun. 08, 2023. [Online]. Available: https://sentekmi.maranatha.edu/index.php/sentekmi2021/article/view/85

[20] A. Alamsyah *et al.*, "Customer Segmentation Using the Integration of the Recency Frequency Monetary Model and the K-Means Cluster Algorithm," *Scientific Journal of Informatics*, vol. 9, no. 2, pp. 189–196, Nov. 2022, doi: 10.15294/sji.v9i2.39437.

[21] T. Juhari, A. Juarna, and U. Gunadarma, "Implementation Rfm Analysis Model For Customer Segmentation Using The K-Means Algorithm Case Study Xyz Online Bookstore," *EXPLORE*, vol. 12, no. 1, pp. 107–118, 2022, Accessed: Jun. 02, 2023. [Online]. Available: https://www.semanticscholar.org/paper/IMPLEMENTATION-RFM-ANALYSIS-MODEL-FOR-CUSTOMER-THE-Juhari-Juarna/c4c6aaaf0ed60d442fbfc8f4f77abf28f7b1b3ca

[22] M. Tavakoli, M. Molavi, V. Masoumi, M. Mobini, S. Etemad, and R. Rahmani, "Customer Segmentation and Strategy Development Based on User Behavior Analysis, RFM Model and Data Mining Techniques: A Case Study," in *Proceedings - 2018 IEEE 15th International Conference on e-Business Engineering, ICEBE 2018*, Institute of Electrical and Electronics Engineers Inc., Dec. 2018, pp. 119–126. doi: 10.1109/ICEBE.2018.00027.

[23] P. Anitha and M. M. Patil, "RFM model for customer purchase behavior using K-Means algorithm," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 5, pp. 1785–1792, May 2022, doi: 10.1016/j.jksuci.2019.12.011.

[24] R. Livari and N. Ghalam, "Customers Grouping Using Data Mining Techniques in the Food Distribution Industry (A Case Study)," *SRPH Journal of Applied management and Agile Organisation*, Oct. 2021, doi: 10.47176/sjamao.3.1.1.

[25] A. Burhan, H. Kiat, Y. Azhar, and V. Rahmayanti, "Penerapan Metode K-Means Dengan Metode Elbow Untuk Segmentasi Pelanggan Menggunakan Model RFM (Recency, Frequency & Monetary)," *REPOSITOR*, vol. 2, no. 7, pp. 945–952, 2020.

[26] B. Turkmen, "Customer Segmentation With Machine Learning for Online Retail Industry," *The European Journal of Social and Behavioural Sciences*, vol. 31, no. 2, pp. 111–136, Apr. 2022, doi: 10.15405/ejsbs.316.

[27] D. Devarapalli *et al.*, "Analysis of RFM Customer Segmentation Using Clustering Algorithms," *International Journal of Mechanical Engineering*, vol. 7, no. 1, 2022, Accessed: Jun. 02, 2023. [Online]. Available: https://www.researchgate.net/publication/358285794_Analysis_of_RFM_Customer_Segmentation_Using_Clustering_Algorithms

[28] P. H. Suharti, A. S. Suryandari, and R. N. Amalia, "Analisis Segmentasi Pelanggan Menggunakan K-Means Clustering Studi Kasus Aplikasi Alfagift," *Sebatik*, vol. 26, no. 2, pp. 420–427, Dec. 2022, doi: 10.46984/sebatik.v26i2.2134.

[29] Y. Christian and K. O. Y. R. Qi, "Penerapan K-Means pada Segmentasi Pasar untuk Riset Pemasaran pada Startup Early Stage dengan Menggunakan CRISP-DM," *JURIKOM (Jurnal Riset Komputer)*, vol. 9, no. 4, p. 966, Aug. 2022, doi: 10.30865/jurikom.v9i4.4486.

[30] A. D. Savitri, F. Abdurrachman Bachtiar, and N. Y. Setiawan, "Segmentasi Pelanggan Menggunakan Metode K-Means Clustering Berdasarkan Model RFM Pada Klinik Kecantikan (Studi Kasus : Belle Crown Malang)," Sep. 2018. [Online]. Available: http://j-ptiik.ub.ac.id