

Analisis Pemasaran Bisnis dengan Data Science : Segmentasi Kepribadian Pelanggan berdasarkan Algoritma K-Means Clustering

Mawaddah Harahap^{1,2*}, Yusniar Lubis^{2,3}, Zakarias Situmorang^{2,4}

¹Universitas Prima Indonesia, Fakultas Teknologi dan Ilmu Komputer, Teknik Informatika

²Universitas Prima Indonesia, Fakultas Ekonomi, Doktor Manajemen

³Universitas Medan Area, Fakultas Pertanian, Agribisnis

⁴Universitas Katolik Santo Thomas, Fakultas Ilmu Komputer, Sistem Informasi

Received: 15 Jan 2022

Accepted: 23 Jan 2022

Published: 25 Jan 2022



*Mawaddah Harahap

Keywords: , Analisis Kepribadian Pelanggan, Data Science, CRIS-DM, Pemasaran, Segmentasi

DSI: Jurnal Data Science Indonesia is licensed under a Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0).

Abstrak : Dalam makalah ini kami menyajikan analisis kepribadian pelanggan dalam membantu bisnis untuk memodifikasi produknya berdasarkan target pelanggan dari berbagai jenis segmen pelanggan sehingga menemukan pelanggan yang potensial, membuat pemasaran agar lebih efektif, melihat tren dalam perilaku pembelian pelanggan dan membuat penawaran produk yang relevan kepada pelanggan. Kerangka kerja Data Science (ilmu data) dengan metodologi CRIS-DM diterapkan untuk memberikan pemahaman bisnis, pemahaman data, analisis data dan pemodelan. Pada tahapan pemodelan diusulkan *Principal component analysis* (PCA) untuk pengurangan dimensi fitur, kemudian algoritma K-Means untuk segmentasi pelanggan dengan menggunakan metode elbow dan silhouette yang menghasilkan nilai $k=4$ yang paling optimal. Terakhir, hasil 4 cluster di analisis berdasarkan proposi, belanja, pendidikan dan tingkat keberhasilan kampanye yang disajikan secara visualisasi.

PENDAHULUAN

Segmentasi pelanggan adalah praktik membagi pelanggan menjadi kelompok-kelompok yang mencerminkan kesamaan satu sama lain [1]. Pelanggan akan dibagi menjadi beberapa segmen untuk mengoptimalkan signifikansi setiap pelanggan individu dalam bisnis, untuk memodifikasi produk sesuai dengan kebutuhan dan perilaku yang berbeda dari setiap pelanggan, dan juga untuk membantu bisnis memenuhi berbagai perhatian dari berbagai jenis pelanggan [2]. Semua pelanggan tidak diciptakan sama. Memahami perbedaan dalam potensi teknis, perspektif ekonomi, dan sikap terhadap konservasi sumber daya di seluruh jenis bisnis, wilayah geografis, dan demografi adalah kunci untuk penyampaian program yang efektif [3], [4].

Analisis Sisi Permintaan bekerja dengan klien untuk melihat melampaui pelanggan rata-rata dan menargetkan solusi dan pengiriman pesan kepada pelanggan untuk pemasaran [5]. Memberi pelanggan apa yang mereka inginkan dan butuhkan adalah cara yang pasti untuk meningkatkan penjualan dan kepuasan pelanggan [6]. Hal ini dapat memudahkan dalam mendorong pembelian karena pelanggan tidak perlu terlalu memikirkan apakah mereka ingin membeli atau tidak. Salah satu cara untuk menciptakan minat langsung ini dan siklus penjualan yang lebih pendek adalah dengan menggunakan segmentasi pelanggan [7].

Segmentasi pelanggan memainkan peran penting untuk memungkinkan bisnis membuat pemasaran agar lebih efektif, melihat tren dalam perilaku pembelian pelanggan dan membuat penawaran produk yang relevan kepada pelanggan [8]. Melalui segmentasi pelanggan, perusahaan dapat menghindari penggunaan bahasa dan pesan yang tidak jelas. Penerapan segmentasi pelanggan telah memberikan peran penting dalam berbagai bisnis, seperti koperasi [9], rumah sakit [10], supermarket [11], agen asuransi [12], e-commerce [13] dan lainnya.

Dalam penelitian ini, kami menyajikan analisis segmentasi pelanggan berdasarkan kerangka kerja data

science dengan mengadopsi kerangka CRIS-DM dalam menangani dan memanfaatkan kumpulan data untuk melihat pola dan menciptakan wawasan tentang kampanye pemasaran, sehingga dapat membantu pelaku bisnis dalam menentukan strategi pemasaran yang optimal dan meningkatkan penjualan.

TINJAUAN LITERATUR

Di pasar bisnis saat ini, segmentasi digunakan di sebagian besar proses analisis pasar untuk membantu mengembangkan pedoman bisnis. Dalam lingkungan industri, segmen yang berbeda berkumpul untuk bersaing dalam penyediaan produk dan layanan kepada konsumen. Analisis pemasaran adalah studi tentang data yang dikumpulkan melalui kampanye pemasaran untuk membedakan pola antara hal-hal seperti bagaimana kampanye berkontribusi pada konversi, perilaku konsumen, preferensi regional, preferensi kreatif, dan lainnya [12]. Tujuan analitik pemasaran sebagai praktik adalah menggunakan pola dan temuan ini untuk mengoptimalkan kampanye di masa mendatang berdasarkan apa yang berhasil [14]. Analisis pemasaran merupakan praktik mengukur, mengelola, dan menganalisis kinerja pemasaran untuk memaksimalkan efektivitasnya dan mengoptimalkan Return on Investment (ROI) [15].

Memahami analitik pemasaran memungkinkan pemasar menjadi lebih efisien dalam pekerjaan mereka dan meminimalkan uang pemasaran web yang terbuang sia-sia. Analitik pemasaran dapat menawarkan wawasan mendalam tentang preferensi dan tren pelanggan [16]. Terlepas dari manfaat yang menarik, sebagian besar organisasi gagal mewujudkan janji analisis pemasaran karena tantangan untuk dapat mengukur ROI. Analisis ini memungkinkan pemasar mencapai ROI yang lebih tinggi pada investasi pemasaran dengan memahami apa yang berhasil dalam mendorong konversi, kesadaran merek, atau keduanya. Analytics juga memastikan bahwa konsumen melihat lebih banyak iklan yang ditargetkan dan dipersonalisasi yang sesuai dengan kebutuhan dan minat khusus mereka, daripada komunikasi massal yang cenderung mengganggu. Data pemasaran dapat dianalisis menggunakan berbagai metode dan model tergantung pada KPI yang diukur [15]. Dalam pemasaran modern, analitik harus lebih akurat dari sebelumnya. Sebab, konsumen semakin menuntut dalam memilih jenis media yang ingin mereka ajak berinteraksi. Oleh karena itu, jika ingin menarik perhatian calon pelanggan, manajer perlu mengandalkan analitik untuk membuat iklan yang lebih efektif. Iklan ini akan ditargetkan dengan jelas, berdasarkan minat dan kebutuhan spesifik masing-masing individu, dan ditampilkan pada waktu yang tepat dan pada saluran yang tepat [17]. Dan ketika perusahaan memiliki data yang cukup, kapasitas yang cukup untuk memprosesnya, dan personalisasi kepada pengguna akan menjadi daya tarik yang bermanfaat untuk menghadirkan sumber pelanggan yang potensial (Segmentasi Pelanggan) [15].

Segmentasi pelanggan mengacu pada pembagian sasaran pasar menjadi kelompok yang dapat dikelola dan layak sesuai dengan karakteristik bersama untuk mengembangkan strategi bisnis yang efektif dan tepat. Secara sederhana, pembagian segmentasi ini adalah proses membagi basis pelanggan yang ada menjadi kelompok yang dapat dikelola dan layak berdasarkan karakteristik umum seperti usia, jenis kelamin, loyalitas, frekuensi pembelian, dll. Untuk menargetkan dan mengembangkan strategi pemasaran untuk setiap kelompok sesuai dengan itu. Karakteristik [18]. Tujuan segmentasi pelanggan termasuk membagi pelanggan target menjadi kelompok-kelompok yang lebih kecil yang mencerminkan kesamaan di antara pelanggan di setiap kelompok untuk:

- a) Mengembangkan hubungan yang lebih baik dengan memahami kebutuhan setiap segmen pelanggan
- b) Mengidentifikasi pelanggan yang berharga
- c) Mengidentifikasi peluang cross-selling dan up-selling
- d) Meningkatkan profitabilitas dengan mengembangkan strategi pemasaran yang lebih efektif untuk setiap segmen

BAHAN DAN METODE

Dalam membangun model analisis segmentasi pelanggan untuk menemukan pelanggan yang potensial, kami mengusulkan Siklus hidup proyek *data science* dengan metodologi CRIS-DM yang terdiri dari beberapa fase yaitu:

1. Analisis Bisnis

Analisis kepribadian pelanggan membantu bisnis untuk memodifikasi produknya berdasarkan target pelanggan dari berbagai jenis segmen pelanggan. Misalnya mengeluarkan uang untuk memasarkan produk baru ke setiap pelanggan di database perusahaan, perusahaan dapat menganalisis segmen pelanggan mana yang paling mungkin membeli produk dan kemudian memasarkan produk hanya pada segmen tertentu. Data kepribadian pelanggan dianalisis berdasarkan kebiasaan belanja, pendapatan, pendidikan, dan ukuran keluarga mereka. Analisis Kepribadian Pelanggan adalah analisis terperinci tentang pelanggan ideal perusahaan. Hal ini dapat membantu pelaku bisnis untuk lebih memahami pelanggannya dan memudahkan mereka untuk memodifikasi produk sesuai dengan kebutuhan, perilaku, dan perhatian khusus dari berbagai jenis pelanggan.

2. Pemahaman Data

Dataset ini didasarkan pada kumpulan data kepribadian 2.240 pelanggan dan 29 fitur (kolom). Kumpulan data ini dikelompokkan menjadi empat bagian yang dijadikan satu, yaitu data pelanggan, data produk, data promosi dan data tempat. Rincian data dilihat pada tabel 1 berikut ini:

Tabel 1 Rincian dataset

<p>Data Pelanggan</p> <ul style="list-style-type: none"> • ID: Pengidentifikasi unik pelanggan. • Year_Birth: Tahun lahir pelanggan. • Education: Tingkat pendidikan pelanggan. • Marital_Status: Status pernikahan pelanggan. • Income: Pendapatan rumah tangga tahunan pelanggan. • Kidhome: Jumlah anak dalam rumah tangga pelanggan. • Teenhome: Jumlah remaja dalam rumah tangga pelanggan. • Dt_Customer: Tanggal pendaftaran pelanggan dengan perusahaan. • Recency: Jumlah hari sejak pembelian terakhir pelanggan. • Complain: 1 jika pelanggan mengeluh dalam 2 tahun terakhir, 0 sebaliknya. 	<p>Data Produk</p> <ul style="list-style-type: none"> • MntWines: Jumlah yang dihabiskan untuk anggur dalam 2 tahun terakhir. • MntFruits: Jumlah yang dihabiskan untuk buah-buahan dalam 2 tahun terakhir. • MntMeatProducts: Jumlah yang dihabiskan untuk daging dalam 2 tahun terakhir. • MntFishProducts: Jumlah yang dihabiskan untuk membeli ikan dalam 2 tahun terakhir. • MntSweetProducts: Jumlah yang dihabiskan untuk permen dalam 2 tahun terakhir. • MntGoldProds: Jumlah yang dihabiskan untuk emas dalam 2 tahun terakhir.
<p>Data Promosi</p> <ul style="list-style-type: none"> • NumDealsPurchases: Jumlah pembelian yang dilakukan dengan diskon. • AcceptedCmp1: 1 jika pelanggan menerima penawaran di kampanye pertama, 0 sebaliknya. • AcceptedCmp2: 1 jika pelanggan menerima tawaran di kampanye ke-2, 0 sebaliknya. • AcceptedCmp3: 1 jika pelanggan menerima tawaran di kampanye ke-3, 0 sebaliknya. • AcceptedCmp4: 1 jika pelanggan menerima tawaran dalam kampanye ke-4, 0 sebaliknya. • AcceptedCmp5: 1 jika pelanggan menerima tawaran dalam kampanye ke-5, 0 sebaliknya. 	<p>Data Tempat</p> <ul style="list-style-type: none"> • NumWebPurchases: Jumlah pembelian yang dilakukan melalui situs web perusahaan. • NumCatalogPurchases: Jumlah pembelian yang dilakukan menggunakan katalog. • NumStorePurchases: Jumlah pembelian yang dilakukan langsung di toko. • NumWebVisitsMonth: Jumlah kunjungan ke situs web perusahaan dalam sebulan terakhir.

- Response: 1 jika pelanggan menerima tawaran dalam kampanye terakhir, 0 sebaliknya.

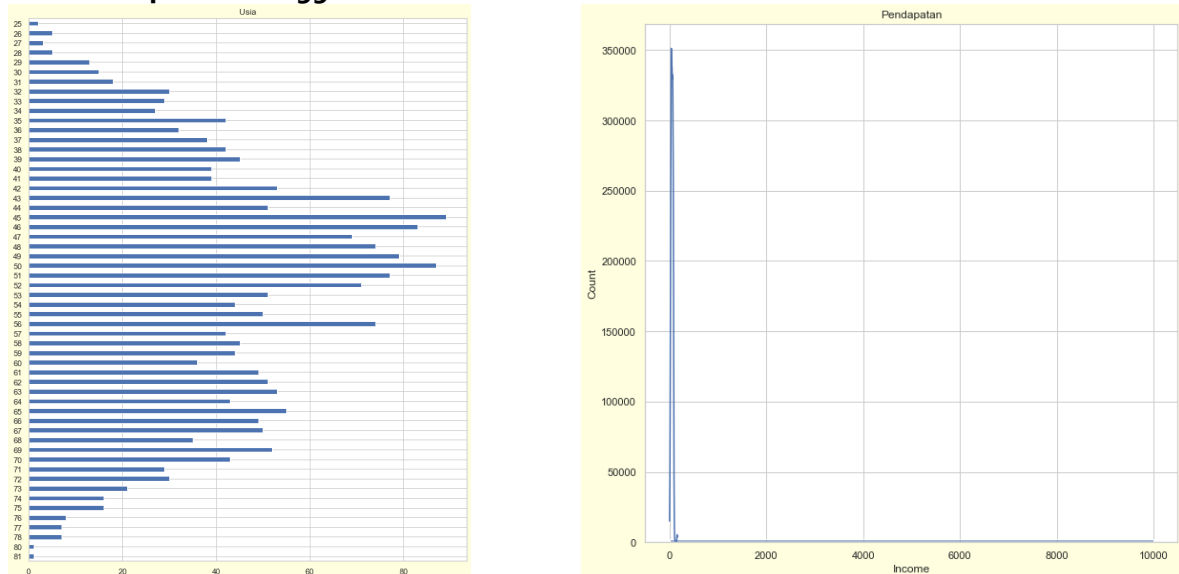
3. Persiapan Data

Persiapan Data, mencakup semua kegiatan dalam membangun kumpulan data akhir dari data mentah awal untuk mempersiapkan data untuk diproses lebih lanjut. Pada tahapan ini beberapa proses yang dilakukan adalah membersihkan data yang hilang, data duplcat, data null (kosong), membuang fitur yang tidak di butuhkan untuk kebutuhan analisis data selanjutnya.

4. Explorasi Data Analysis

Pada bagian ini akan disajikan analisis data dalam bentuk visualisasi agar lebih mudah memahami kumpulan data, penentuan fitur analysis, korelasi data dan lainnya. Dalam tahapan ini beberapa fitur tambahan digunakan untuk memudahkan proses visialuasi dan analisis data, fitur usia yang di ambil dari fitur tanggal lahir pelanggan, fitur Dt_customer diubah menjadi tahun untuk mendapatkan tahun pelanggan, menjumlahkan total pengeluaran dan total kampanye yang diterima untuk setiap pelanggan. Beberapa analisis data yang dilakukan adalah sebagai berikut:

a) Usia dan Pendapatan Pelanggan

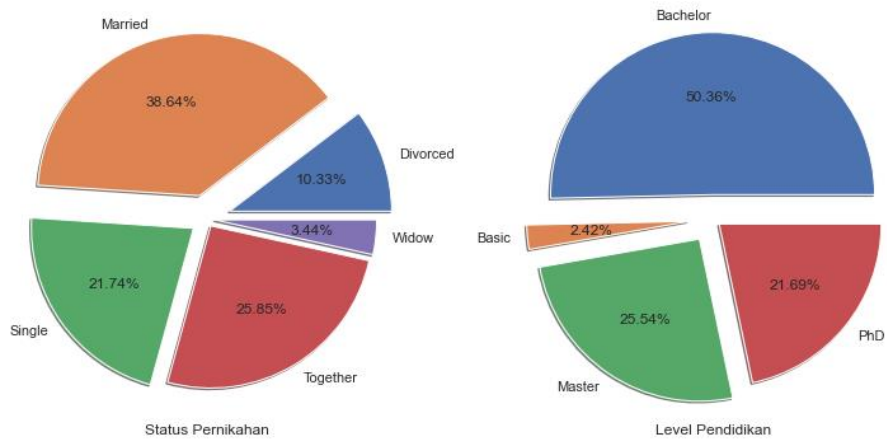


Gambar 1 Analisis Usia dan Pendapatan Pelanggan

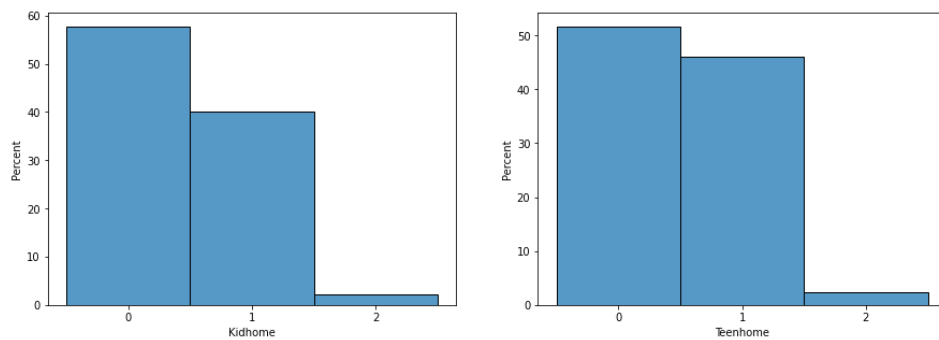
Pada gambar 1 terlihat bahwa sebagian besar pelanggan berusia antara 43–56 tahun dan pelanggan memiliki kisaran pendapatan 30.000–800.000.

b) Status pernikahan dan pendidikan pelanggan

Analisis status pernikahan digunakan untuk memahami bagaimana situasi kehidupan pasangan pelanggan serta kejelasan lebih lanjut tentang rumah tangga mereka. Sedangkan pendidikan pelanggan berguna untuk memahami kondisi pendidikan pelanggan. Pada gambar 2 memberikan gambaran sekilas tentang distribusi pelanggan, sebagian besar pelanggan (64%) menjalin hubungan (Menikah) dan 97% pelanggan setidaknya memiliki gelar sarjana. Selanjutnya, dapat dilihat pada gambar 3 mengenai status kejelasan rumah tangga pelanggan dimana sebagian besar pelanggan memiliki 1 anak atau 0 anak di rumah, sangat sedikit yang memiliki 2 anak, dan tidak ada pelanggan yang memiliki anak di atas 2 tahun.

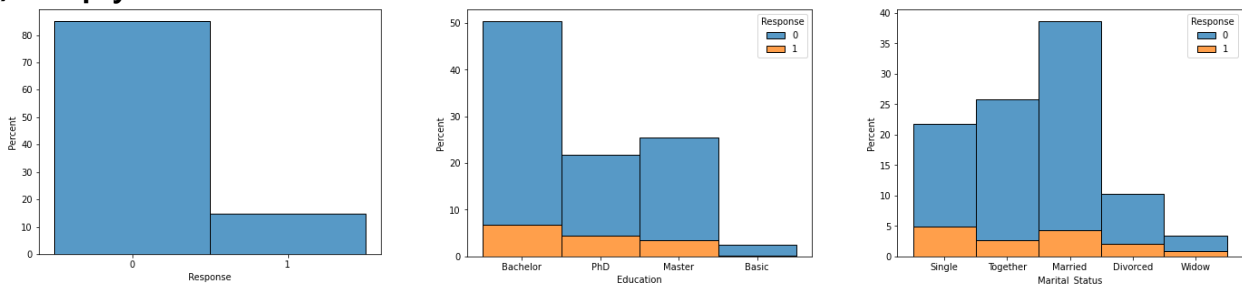


Gambar 2 Analisis Pernikahan dan Pendidikan



Gambar 3 Analisis Kondisi Rumah Tangga Pelanggan

c) Kampanye



Gambar 4 Analisis Kampanye

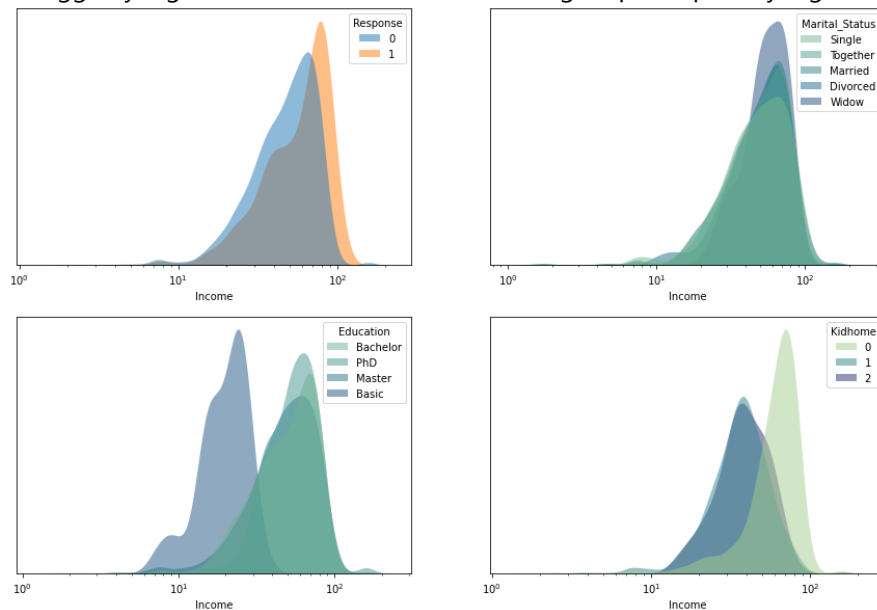
Pada gambar 4 menggambarkan bagaimana respon pelanggan terhadap kampanye yang menghasilkan data tidak seimbang, lebih dari 80% pelanggan mengatakan tidak pada kampanye terakhir. Jadi harus berhati-hati dan menggunakan indikator yang lebih komprehensif dan akurat (seperti F-1) untuk mengevaluasi model klasifikasi. Selanjutnya, kami menemukan bahwa tingkat penerimaan kampanye pada kelompok pendidikan tinggi (Magister dan PhD) lebih tinggi dari pada kelompok pendidikan rendah. Dan para lajang cenderung mengatakan ya untuk kampanye ini. Terakhir, pelanggan tanpa anak dan remaja lebih cenderung menerima tawaran dalam kampanye ini.

d) Pendapatan Pelanggan

Pada bagian ini sangat penting memahami bagaimana distribusi data pendapatan pelanggan terhadap kampanye terakhir, dari gambar 5 dapat dilihat bahwa:

- 1) Kelompok berpenghasilan tinggi memiliki kemungkinan yang lebih besar untuk menerima tawaran dalam kampanye, karena distribusi pendapatan orang-orang yang mengatakan 'ya' dan 'tidak' memiliki sedikit perbedaan.
- 2) Tidak ada perbedaan pendapatan yang jelas antara orang-orang dengan status perkawinan yang berbeda.

- 3) Pelanggan hanya dengan pendidikan dasar memiliki pendapatan yang jauh lebih rendah, sedangkan sarjana, master, dan PhD tidak memiliki perbedaan yang jelas dalam tingkat pendapatan.
- 4) Tampaknya pelanggan yang tidak memiliki anak memiliki tingkat pendapatan yang lebih tinggi.



Gambar 5 Analisis Kampanye

5. Pemodelan

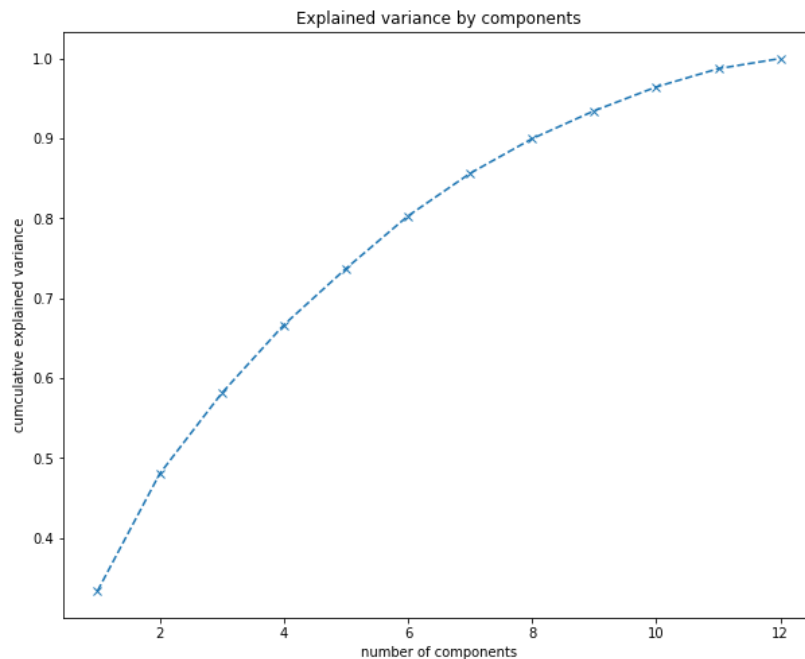
Pada tahapan ini, kami mengusulkan algoritma K-Means untuk pengelompokan pelanggan. Pengelompokan K-Means adalah algoritma pembelajaran mesin yang efisien untuk memecahkan masalah pengelompokan data. Algoritma ini merupakan salah satu algoritma tanpa pengawasan yang cukup cocok untuk memecahkan masalah segmentasi pelanggan. Sebelum menerapkan algoritma ini, kami menggunakan pendekatan *Principal component analysis* (PCA) untuk pengurangan dimensi fitur. Hal ini dilakukan karena ada banyak faktor yang menjadi dasar klasifikasi akhir. Faktor-faktor ini pada dasarnya adalah atribut atau fitur. Semakin tinggi jumlah fitur, semakin sulit untuk mendapatkan model yang akurat. Pengurangan dimensi adalah proses mengurangi jumlah variabel acak yang dipertimbangkan, dengan memperoleh satu set variabel utama dan algoritma PCA adalah teknik untuk mengurangi dimensi kumpulan data, meningkatkan interpretasi tetapi pada saat yang sama meminimalkan kehilangan informasi.

6. Evaluasi

Karena ini adalah pengelompokan tanpa pengawasan. Kami tidak memiliki fitur yang ditandai untuk mengevaluasi atau menilai model yang digunakan. Tujuan dari bagian ini adalah untuk mempelajari pola-pola dalam kluster yang terbentuk dan menentukan sifat dari pola kluster tersebut. Untuk itu, bagian ini, kami akan melihat data berdasarkan cluster melalui analisis data eksplorasi dan penarikan kesimpulan.

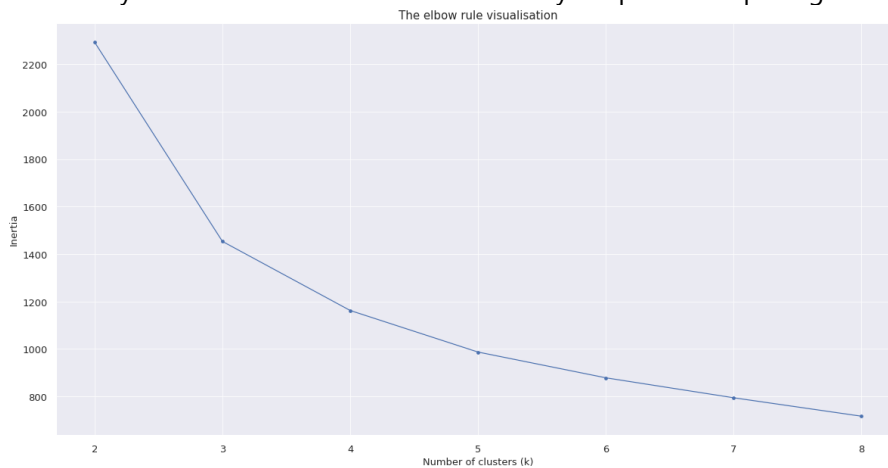
HASIL PENELITIAN

Pada bagian ini, kami uraikan hasil pengelompokan dengan menggunakan algoritma K-Means. Algoritma ini mampu menemukan pola dalam kumpulan data dari data yang tidak berlabel dengan mengelompokkan titik data berdasarkan atribut serupa dalam kumpulan data. Jadi, pertama kami melakukan standarisasi data dengan menggunakan encode untuk menormalisasikan data dengan range 0-1, kemudian menggunakan PCA untuk menemukan variabel penting dari sekumpulan data dengan mengambil jumlah komponen yang menyumbang 80%. Hasil penerapan PCA dapat dilihat pada gambar 6.

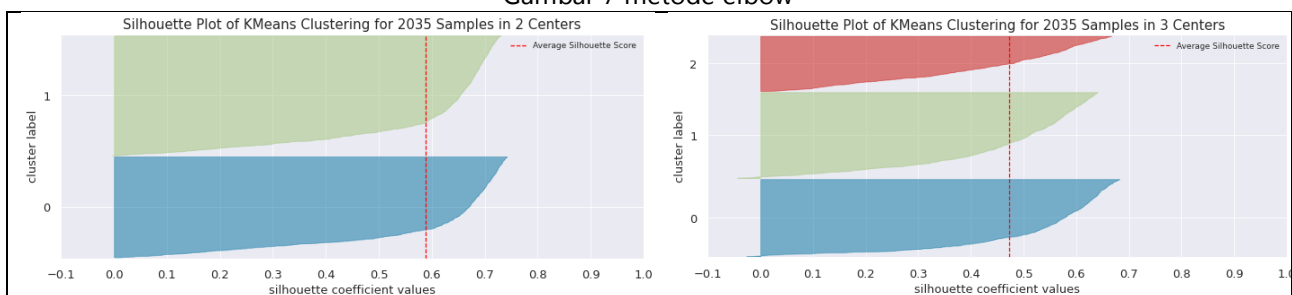


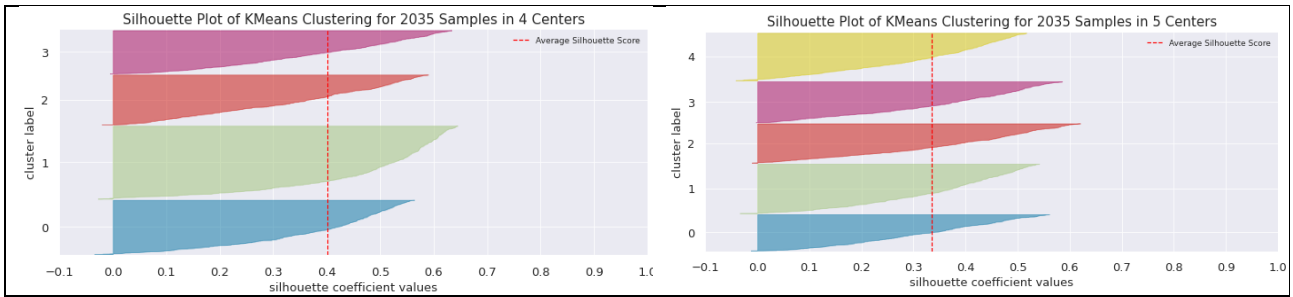
Gambar 6 kumulatif varians variabel dengan PCA

Pada gambar 6 menunjukkan bahwa komponen pertama yaitu balok biru mendapat penjelasan varian tertinggi sebesar 0,2%. Kemudian komponen berikutnya yang menjelaskan persentase varian tertinggi berikutnya. Total 6 komponen untuk menangkap semua varians dalam data dapat mempertahankan sekitar 80% dari total varian data. Dari hasil ini dapat disimpulkan terdapat 6 komponen variabel penting yang akan diterapkan pada algoritma K-Means. Untuk menentukan pusat cluster yang optimal pada algoritma K-Means, kami menggunakan dua model yaitu elbow dan silhouette dan hasilnya dapat dilihat pada gambar 7 dan gambar 8.



Gambar 7 metode elbow

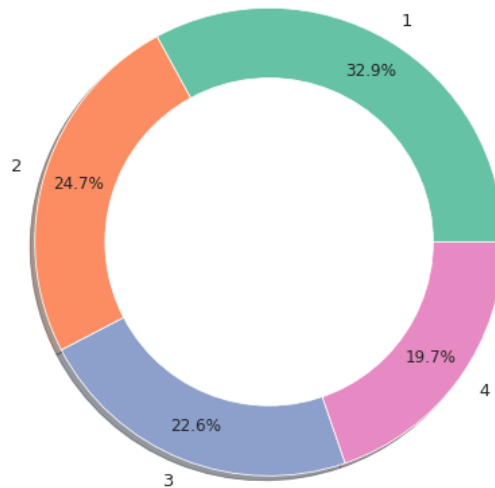




Gambar 8 metode silhouette

Berdasarkan visualisasi kedua metode yang digunakan untuk menentukan nilai k yang optimal pada algoritma K-Means, metode elbow menghasilkan nilai k=3 lebih optimal, sedangkan metode silhouette menghasilkan nilai k=4 dan k=5 perlu di pertimbangkan. Namun, Dilihat dari visualisasi koefisien metode silhouette, 4 cluster memberikan nilai silhouette yang relatif tinggi untuk setiap cluster sehingga kami putuskan menggunakan k=4. Selanjutnya, hasil analisis pengelompokan pelanggan yang dibagi menjadi beberapa hasil analisis, yaitu:

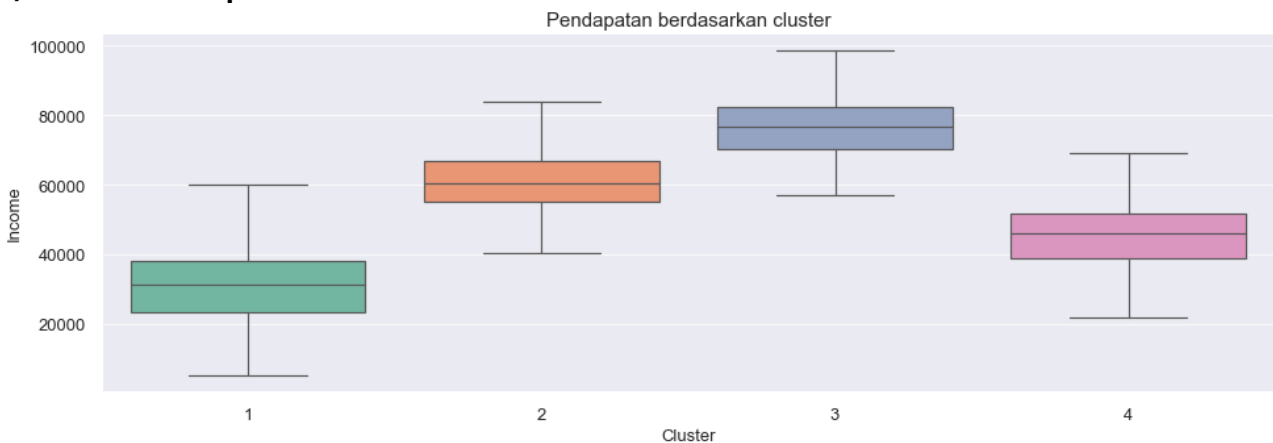
a) Analisis proposi pelanggan



Gambar 9 Analisis Cluster Proporsi Pelanggan

Pada gambar 9 merupakan ilustrasi hasil pengelompokan pelanggan dimana ditemukan bahwa Cluster 1 adalah cluster terbesar (32.9%), sekitar 1/3 dari seluruh pelanggan. Cluster 2, 3 dan 4 berukuran hampir sama.

b) Analisis Pendapatan

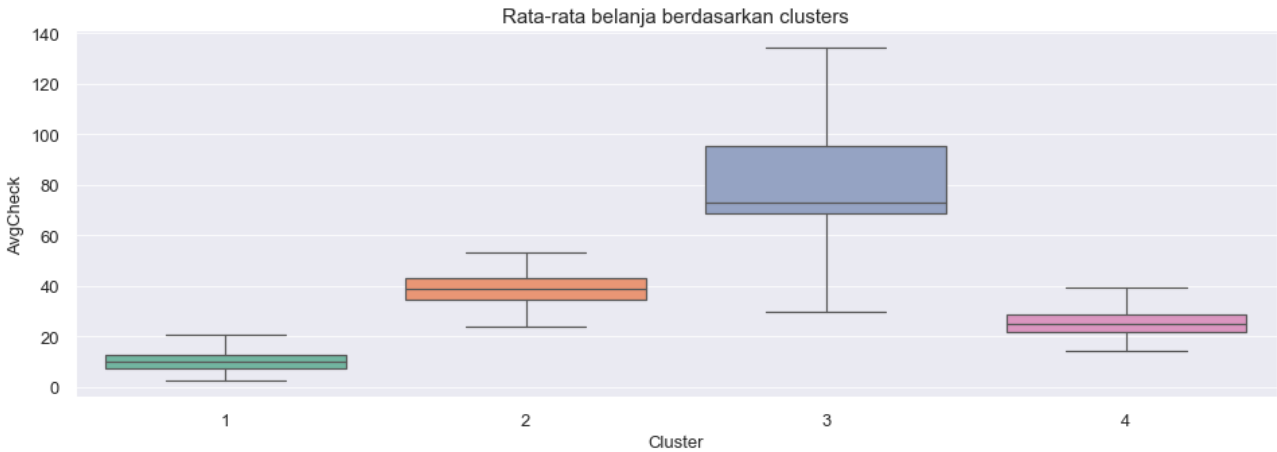


Gambar 9 Analisis Pendapatan pada cluster berbeda

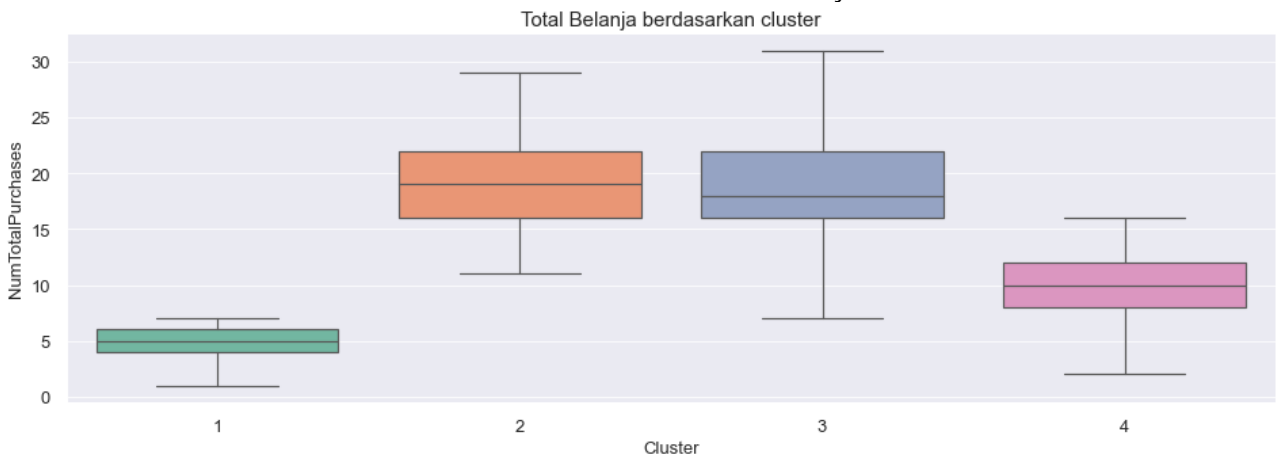
Pada gambar 9 merupakan distribusi pendapatan transaksi dari pelanggan berdasarkan cluster berbeda. Dari hasil ini dapat dilihat bahwa cluster 1 merupakan pelanggan yang berpenghasilan rendah, cluster 2 pelanggan

yang berpenghasilan tinggi, sedangkan di cluster 3 pendapatan sangat tinggi, dan cluster 4 merupakan berpenghasilan menengah.

c) Analisis Belanja Pelanggan



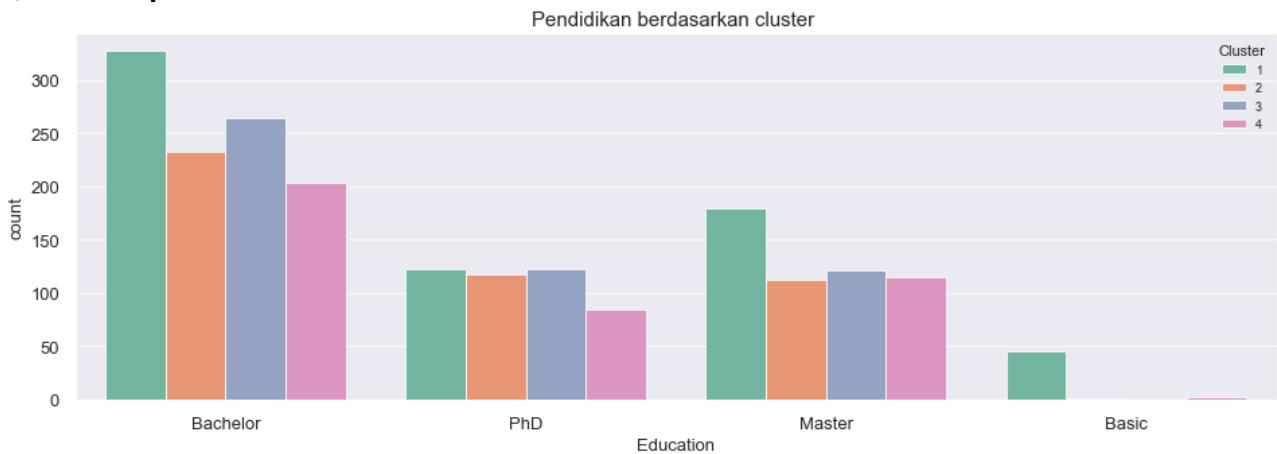
Gambar 10 Analisis Rata-rata belanja



Gambar 11 Analisis Total Belanja cluster berbeda

Pada gambar 10 dan 11 merupakan visualisasi transaksi belanja pelanggan, hasil ini memberikan informasi bahwa pelanggan cluster 2 dan 3 merupakan pembeli paling aktif dan sering, cluster 4 memiliki frekuensi sedang, dan cluster 1 memiliki frekuensi pembelian yang rendah.

d) Analisis pendidikan



Gambar 12 Analisis Pendidikan pada cluster berbeda

e) Analisis Kampanye yang Diterima



Gambar 12 Analisis Kampanye

Dari hasil visualisasi gambar 12 memberikan informasi mengenai tingkat keberhasilan kampanye kepada pelanggan, dimana minat terbesar dalam kampanye 1 menunjukkan Cluster 2 dan 3. Penerimaan kampanye 2 relatif rendah di cluster mana pun, kemudian minat terbesar dalam kampanye 3 menunjukkan Cluster 1. Kampanye 4 relatif berhasil di semua cluster selain Cluster 1 dan minat terbesar dalam kampanye 5 menunjukkan Cluster 3 dan Kampanye terakhir berhasil di semua cluster.

Diskusi

Penelitian ini mengusulkan pendekatan *Principal component analysis* (PCA) untuk pengurangan dimensi fitur yang membuktikan pengaruh yang signifikan terhadap komputasi dan hasil analisis sebagaimana kami uraikan di bagian hasil dan pembahasan, namun kami mempertimbangkan penerapan analisis RFM, sebagian peneliti menggunakan pendekatan ini [19]–[22]. Pertimbangannya ini karena fokus pembahasan lebih pada kepribadian pelanggan yang memiliki segmen berbeda serta tujuan utamanya adalah menemukan strategi pemasaran yang tepat dengan menawarkan jenis produk yang tepat di waktu yang tepat dan sesuai keuangan pelanggan,

Hasil empat cluster dari algoritma K-Means dinamakan cluster 1 pelanggan perunggu, cluster 4 perak, cluster 2 emas dan cluster 3 adalah platinum. Pada cluster platinum merupakan pelanggan yang

penghasilan yang sangat tinggi dengan belanja rata-rata tinggi, sering melakukan pembelian dan kebanyakan adalah pelanggan yang belum menikah. Cluster emas adalah kelompok pelanggan yang hampir sama dengan cluster platinum, namun 80% pelanggan dikelompok ini adalah orang tua. Cluster perak adalah pelanggan berpenghasilan menengah dengan transaksi pembelian memiliki frekuensi menengah dan kebanyakan orang tua. Terakhir cluster perunggu merupakan kelompok pelanggan yang berpenghasilan rendah, sebagian sarjana dengan jumlah pembelian paling rendah, namun paling banyak mengeluh.

KESIMPULAN

Berdasarkan hasil analisis pengelompokan pelanggan berdasarkan algoritma K-Means yang menghasilkan 4 cluster berdasarkan analisis dua metode yaitu elbow dan silhouette. Pada cluster emas merupakan pelanggan dalam kelompok yang memiliki pendapatan rata-rata, membeli jumlah rata-rata produk. Kebanyakan dari mereka hanya terlibat dalam satu kampanye, tetapi mereka melakukan lebih banyak pembelian saat ada diskon. Mereka cenderung lebih sering berbelanja online daripada dua kelompok lainnya. Cluster perunggu merupakan kelompok pelanggan yang memiliki kemampuan pembelian terendah dan membelanjakan sangat sedikit untuk perusahaan. Hampir semuanya tidak ikut kampanye. Namun, mereka suka mengunjungi beranda toko online. Berlawanan dengan cluster perunggu, cluster platinum memiliki rekor pendapatan dan pembelian tertinggi. Pelanggan di cluster ini biasanya memiliki ukuran keluarga kecil, menerima lebih banyak kampanye dan senang melakukan pembelian di toko atau katalog. Mereka tidak banyak mengunjungi situs web toko tetapi mereka masih melakukan pembelian web dalam jumlah sedang. Juga, diskon tidak menarik bagi mereka.

Akan tetapi ukuran kelompok pada cluster perunggu adalah yang terbesar karena menyumbang hampir setengah dari total pelanggan. Ukuran tiga kelompok lainnya serupa. Perusahaan dapat memanfaatkan basis pelanggan saat ini dan menawarkan lebih banyak dan meluncurkan kampanye untuk merangsang keinginan belanja kelompok pelanggan cluster emas dan platinum. Lebih banyak promosi di toko online dapat dilakukan untuk menyerang kelompok pembeli perunggu karena mereka menghabiskan banyak waktu untuk menjelajahi halaman web tetapi melakukan sedikit pembelian dan banyak mengeluh

Supplementary Materials (optional)

Sumber dataset tersedia di <https://www.kaggle.com/rhythmcam/customer-personality-classification-prediction/data>

Kontribusi Penulis

Semua Penulis memiliki kontribusi yang sama dalam makalah ini Semua penulis telah membaca dan menyetujui versi manuskrip yang diterbitkan.

Konflik kepentingan

Para penulis menyatakan tidak ada konflik kepentingan.

REFERENCES

- [1] K. Maheswari, P. Packia, and A. Priya, "Predicting Customer Behavior in Online Shopping Using SVM Classifier," *IEEE Int. Conf. Intell. Tech. Control. SIGNAL Process.*, 2017.
- [2] A. M. Husein, F. K. Waruwu, Y. M. T. B. Bara, M. Donpril, and M. Harahap, "Clustering Algorithm For Determining Marketing Targets Based Customer Purchase Patterns And Behaviors," *Sink. J. dan Penelit. Tek. Inform.*, vol. 6, no. 1, pp. 137–143, 2021, [Online]. Available: <https://www.polgan.ac.id/jurnal/index.php/sinkron/article/view/11191>.
- [3] M. Gull and A. Pervaiz, "Customer behavior analysis towards online shopping using data mining," *5th*

- Int. Multi-Topic ICT Conf. Technol. Futur. Gener. IMTIC 2018 - Proc.*, pp. 1–5, 2018, doi: 10.1109/IMTIC.2018.8467262.
- [4] A. A. Alalwan, "Investigating the impact of social media advertising features on customer purchase intention," *Int. J. Inf. Manage.*, vol. 42, no. June, pp. 65–77, 2018, doi: 10.1016/j.ijinfomgt.2018.06.001.
- [5] M. Harahap, Y. Lubis, and Z. Situmorang, "Data Science bidang Pemasaran: Analisis Prilaku Pelanggan," *Data Sci. Indones.*, vol. 1, no. 1, pp. 21–32, 2021, doi: 10.47709/dsi.v1i1.1194.
- [6] A. M. Husein and M. Harahap, "Pendekatan Data Science untuk Menemukan Churn Pelanggan pada Sector Perbankan dengan Machine Learning," *Data Sci. Indones.*, vol. 1, no. 1, pp. 8–13, 2021, doi: 10.47709/dsi.v1i1.1169.
- [7] A. M. Husein, F. R. Lubis, and M. K. Harahap, "Analisis Prediktif untuk Keputusan Bisnis: Peramalan Penjualan," *Data Sci. Indones.*, vol. 1, no. 1, pp. 32–40, 2021, doi: 10.47709/dsi.v1i1.1196.
- [8] Z. P. Fan, Y. J. Che, and Z. Y. Chen, "Product sales forecasting using online reviews and historical sales data: A method combining the Bass model and sentiment analysis," *J. Bus. Res.*, vol. 74, pp. 90–100, 2017, doi: 10.1016/j.jbusres.2017.01.010.
- [9] O. Tremma, A. Kontogeorgos, P. Karipidis, and F. Chatzitheodoridis, "Mapping the market segments for the consumers of Greek cooperative food products," *Sustain.*, vol. 13, no. 7, p. 3825, Apr. 2021, doi: 10.3390/su13073825.
- [10] S. I. Vuik, E. K. Mayer, and A. Darzi, "Patient segmentation analysis offers significant benefits for integrated care and support," *Health Aff.*, vol. 35, no. 5, pp. 769–775, Aug. 2016, doi: 10.1377/hlthaff.2015.1311.
- [11] R. H. Lin, W. W. Chuang, C. L. Chuang, and W. S. Chang, "Applied big data analysis to build customer product recommendation model," *Sustain.*, vol. 13, no. 9, p. 4985, May 2021, doi: 10.3390/su13094985.
- [12] A. Massaro, A. Panarese, M. Gargaro, C. Vitale, and A. M. Galiano, "Implementation of a Decision Support System and Business Intelligence Algorithms for the Automated Management of Insurance Agents Activities," *Int. J. Artif. Intell. Appl.*, vol. 12, no. 03, pp. 01–13, 2021, doi: 10.5121/ijaia.2021.12301.
- [13] D. Kamthania, A. Pahwa, and S. S. Madhavan, "Market segmentation analysis and visualization using K-mode clustering algorithm for E-commerce business," *J. Comput. Inf. Technol.*, vol. 26, no. 1, pp. 57–68, 2018, doi: 10.20532/cit.2018.1003863.
- [14] Marketing Evolution, "What is Marketing Analytics? Tips and Tools | Marketing Evolution," *marketingevolution.com*, 2021. <https://www.marketingevolution.com/marketing-essentials/marketing-analytics> (accessed Dec. 20, 2021).
- [15] Nisa, "Marketing Analytics: Pengertian, Tujuan, dan Manfaatnya," *inmarketing.id*, 2021. <https://inmarketing.id/marketing-analytics-adalah.html> (accessed Dec. 20, 2021).
- [16] G. Szymański, "Marketing activities of local food producers in e-commerce," *Sustain.*, vol. 13, no. 16, p. 9406, Aug. 2021, doi: 10.3390/su13169406.
- [17] Stitch, "Marketing analytics: definition and uses | Stitch resource," *stitchdata.com*, 2021. <https://www.stitchdata.com/resources/marketing-analytics/> (accessed Dec. 20, 2021).
- [18] Stephan Serrano, "[Guide]: RFM Analysis w/ segmentation examples + how-tos," *barilliance.com*, 2021. <https://www.barilliance.com/rfm-analysis/> (accessed Dec. 20, 2021).
- [19] 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.
- [20] J. Wu *et al.*, "User Value Identification Based on Improved RFM Model and K -Means++ Algorithm for Complex Data Analysis," *Wirel. Commun. Mob. Comput.*, vol. 2021, pp. 1–8, 2021, doi: 10.1155/2021/9982484.
- [21] M. A. Rahim, M. Mushafiq, S. Khan, and Z. A. Arain, "RFM-based repurchase behavior for customer classification and segmentation," *J. Retail. Consum. Serv.*, vol. 61, no. April, p. 102566, 2021, doi: 10.1016/j.jretconser.2021.102566.
- [22] Y. Lee, "Customer Classification and Market Basket Analysis Using K-Means Clustering and Association Rules: Evidence from Distribution Big Data of Korean Retailing Company," *Knowl. Manag. Res.*, vol. 19,

