****

**APLIKASI DATA MINING SEGMENTASI PELANGGAN MENGGUNAKAN METODE K-MEANS CLUSTERING**



**Disusun Oleh:**

**KELOMPOK 11**

**PROGRAM STUDI INFORMATIKA**

**FAKULTAS TEKNIK**

**UNIVERSITAS MUHAMMADIYAH SORONG**

**TAHUN 2024**

# LEMBAR PERSETUJUAN

**APLIKASI DATA MINING SEGMENTASI PELANGGAN MENGGUNAKAN METODE K-MEANS CLUSTERING**

**Diajukan Sebagai Salah Satu Syarat**

**Untuk Memperoleh Nilai UTS dan UAS**

**Mata Kuliah Algoritma dan Pemrograman 2**

**Pada Prodi Informatika Fakultas Teknik**

**Universitas Muhammadiyah Sorong**

**Disusun Oleh:**

**KELOMPOK 11**

****

|  |  |
| --- | --- |
| **Menyetujui dan Mengetahui**  **Dosen Mata Kuliah**  **Fajar R. B Putra, S.Kom., M.Kom.**  **NIDN. 1428099501** | **Sorong, 25 September 2024**  **Menyetujui**  **Ketua Kelompok 11**  **Deliana Andini Sinaga**  **NIM. 202355202131** |

# KATA PENGANTAR

Puji syukur penulis panjatkan kepada Tuhan Yang Maha Esa atas berkat, rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan Tugas Besar dengan judul “APLIKASI DATA MINING SEGMENTASI PELANGGAN MENGGUNAKAN METODE K-MEANS CLUSTERING”.Adapun Tugas Besar ini disusun sebagai salah satu syarat untuk memperoleh nilai UTS dan UAS Mata Kuliah Algortima dan Pemorgraman 2, Program Studi Teknik Informatika, Fakultas Teknik, UNAMIN.Tentunya tidak lupa yang kami hormati kepada:

1. Bapak Dr. H. Muhammad Ali, M.M., M.H. Selaku Rektor Universitas Muhammadiyah Sorong
2. Bapak Ir. Hendrik Pristianto, ST., M.T., IPM. selaku Dekan Fakultas Teknik
3. Bapak Ir. Rendra Soekarta, S.Kom., M.T., IPP. selaku Kaprodi Teknik Informatika
4. Teman-teman dan juga sahabat-sahabatku.

Penulis menyadari bahwa penyusunan Tugas Besar ini masih banyak terdapat kekurangan, maka dari itu kelompok mengharapkan kritikan dan saran yang bersifat membangun.

Sorong, 28 Mei 2024

KELOMPOK 11

# DAFTAR ISI

[LEMBAR PERSETUJUAN ii](#_Toc170952223)

[KATA PENGANTAR iii](#_Toc170952224)

[DAFTAR ISI iv](#_Toc170952225)

[DAFTAR TABEL v](#_Toc170952226)ii

[DAFTAR GAMBAR viii](#_Toc170952227)

[BAB I](#_Toc170952228) [PENDAHULUAN 1](#_Toc170952229)

[1.1 Latar Belakang 1](#_Toc170952230)

[1.2 Rumusan Masalah 3](#_Toc170952231)

[1.3 Tujuan 3](#_Toc170952232)

[1.4 Batasan Masalah 4](#_Toc170952233)

[BAB II](#_Toc170952234) [LANDASAN TEORI 5](#_Toc170952235)

[2.1 State Of The Art 5](#_Toc170952236)

[2.2 Studi Literatur 6](#_Toc170952237)

[2.3 Literatur Terkait 24](#_Toc170952240)

[2.3.1 Pengertian Segmentasi Pelanggan 24](#_Toc170952241)

[2.3.2 Data Mining 25](#_Toc170952242)

[2.3.3 Pengertian Metode K-Means Clustreing 25](#_Toc170952243)

[2.3.4 Pengertian Flowchart 27](#_Toc170952244)

[2.3.5 Pengertian Anroid Studio 29](#_Toc170952245)

[2.3.6 Python 30](#_Toc170952246)

[2.3.7 Kaggle 30](#_Toc170952247)

[2.3.8 Dataset 30](#_Toc170952248)

[2.3.9 Metode Pengembangan Sistem 31](#_Toc170952249)

[2.3.10 Usability Testing 32](#_Toc170952250)

[2.3.11 White Box 32](#_Toc170952251)

BAB III [ANALISA DATA DAN PEMBAHASAN 34](#_Toc170952253)

[3.1 Hasil dan Pembahasan 34](#_Toc170952254)

[3.1.1. Flowchart 34](#_Toc170952255)

[3.1.2 Dataset 36](#_Toc170952256)

[3.1.3 Hasil Klasifikasi 41](#_Toc170952257)

[3.1.4 Studi Kasus 46](#_Toc170952258)

[3.2 Implementasi Interface 50](#_Toc170952259)

[3.2.1 Home Page 50](#_Toc170952260)

[3.2.2 Penginputan Data 51](#_Toc170952261)

[3.2.3 About Page 52](#_Toc170952262)

[3.3 Pengujian 53](#_Toc170952263)

3.4 Usability Testing……………………………………………...…..……….54

[BAB IV](#_Toc170952264) [PENUTUP 55](#_Toc170952265)

[4.1 Kesimpulan 55](#_Toc170952266)

[4.2 Saran. 56](#_Toc170952267)

[DAFTAR PUSTAKA 57](#_Toc170952268)

[Lampiran 1 Evaluasi Pengerjaan Tugas Besar 60](#_Toc170952269)

[Lampiran 2 Dokumentasi 61](#_Toc170952270)

[Lampiran 3 Form Pengisian Tugas Besar 62](#_Toc170952271)

**DAFTAR TABEL**

[Tabel 2. 1 Perbandingan Penelitian Terkait Dan Peneliti 22](#_Toc170981252)

[Tabel 2. 2 Simbol – simbol Flowchart dan Nama Fungsi Terminator 28](#_Toc170981253)

[Tabel 3. 1 Data customers 36](#_Toc170981484)

[Tabel 3. 2 Pengujian Pada User 51](#_Toc170981485)

**DAFTAR GAMBAR**

[Gambar 2.1 State Of the Art 5](#_Toc172066117)

[Gambar 3. 1 Flowchart 34](#_Toc172066125)

[Gambar 3. 2 Klasifikasi Gender 42](#_Toc172066126)

[Gambar 3. 3 Diagram Batang Nilai Rata – Rata Per Cluster 43](#_Toc172066127)

[Gambar 3. 4 Diagram Lingkaran presentase pelanggan 45](#_Toc172066128)

[Gambar 3. 5 Heatmap Korelasi Antar Fitur 48](#_Toc172066129)

[Gambar 3. 6 Home Page 50](#_Toc172066130)

[Gambar 3. 7 Penginputan Data 51](#_Toc172066131)

[Gambar 3. 8 About Page 52](#_Toc172066132)

# 

# BAB I

# PENDAHULUAN

## Latar Belakang

Pelanggan merupakan salah satu aset yang berharga bagi sebuah perusahaan maupun sebuah usaha kecil – kecilan. Dalam persaingan dunia bisnis pada saat ini,kita dituntut untuk selalu mengembangkan bisnis agar bertahan dalam persaingan, khususnya dalam persaingan penjualan menuntut para pengusaha untuk menemukan suatu pola yang dapat meningkatkan penjualan dan pemasaran di perusahaan, salah satunya adalah dengan pemanfaatan data penjualan. Dalam sebuah usaha kita perlu mengetahui tentang informasi mendalam tentang profil dan juga karakteristik pelanggan.

Menurut (Febriani & Putri, 2020) segmentasi pelanggan dipakai untuk mengevaluasi nilai dari setiap pelanggan sehingga perusahaan dapat mengidentifikasi pelanggan yang memberikan keuntungan yang besar dan pelanggan yang tidak. Menurut (Fitriyani & Hendriyani, 2021), dengan mengelola dan memanfaatkan database pelanggan dengan tepat, dapat membantu manajemen untuk mendapatkan informasi seperti : mengetahui berbagai kebutuhan pelanggan, memberikan penawaran khusus, memudahkan dalam membuat program loyalitas, dan membantu dalam segmentasi pelanggan. Informasi yang diperoleh dari pengolahan data penjualan, dapat dijadikan sebagai dasar oleh manajemen untuk mengembangkan usaha.

Untuk itu dibutuhkan proses identifikasi tentang informasi dari pelanggan dan juga dibutuhkan seberpa tingkat ketertarikan pelanggan tentang perusahaan. Proses indentifikasi seberapa tingkat profit pelanggan kepada perusahaan dilakukan dengan menggunakan seluruh data yang dimiliki oleh perusahaan. Seperti teknik data mining, data mining adalah sebuah proses menggali dan mengolah informasi penting dari kumpulan data. Proses penggalian informasi melibatkan teknik seperti memanfaatkan kecerdasan buatan, statistik, matematika, machine learning, dan lainnya. Teknik yang dipakai memiliki tujuan untuk mengenali dan mengekstrak informasi yang bermanfaat dari suatu database yang besar. Saat ini data mining telah menjangkau dan juga dimanfaatkan pada banyak bidang keilmuan lainnya. (Sudarsono dkk, 2021).

Dalam melakukan clustering, maka diperlukan algoritma yang mampu melakukan pengelompokan data objek yang memiliki tingkat kesamaan antar kelas objek. Dari pengertian tersebut maka dapat ditarik kesimpulan jika model clustering cocok untuk mengelompokkan pelanggan berdasarkan tingkat kesamaannya menjadi beberapa cluster atau segmen. Algoritma yang sering diterapkan untuk melakukan clustering salah satunya adalah K-Means.

Karena seiring meningkatnya persaingan bisnis yang semakin ketat, banyak perusahaan hampir di semua bidang yang kemudian tidak lagi memfokuskan aktivitas pemasaran mereka hanya pada pencarian pembeli baru, namun sudah berubah menjadi pada usaha untuk mempertahankan dan meningkatkan kesetiaan pelanggan lama. Kita dapat mempertahankan pelanggan dengan cara melakukan pemetaan pelanggan yang dimana kita mengetahui kemauan maupun keinginan seluruh pelanggan misalnya dalam hal penentuan promo, pemberian hadiah, maupun pelayanan. Permasalahan yang terjadi adalah kita tidak mungkin mengidentifikasi tentang kesamaan antar pelanggan dan juga mengelompokkan antara kesukaan masing – masing pelanggan dengan mengidentifikasi satu per satu.

Dan permasalahan yang terjadi adalah bagaimana cara mengimplementasikan metode K-means clustering dalam aplikasi data mining segmentasi pelanggan. Oleh sebab itu pada penelitian kali ini kami akan merancang dan membangun system mesin learning agar dapat menyelesaikan masalah pada segmentasi pelanggan. Dan juga dengan adanya system ini maka dapat membantu para perusahaan untuk mengelompokkan pelanggan antar kesukaan dan juga kesamaan yang dimiliki sehingga dapat mempertahankan pelanggan. Penulis juga berharap Keberhasilan identifikasi pelanggan ini diharapkan bisa mempertahankan pelanggan dan mencegah pelanggan berpindah pembelian kepada perusahaan lain (pesaing).

## Rumusan Masalah

Berdasarkan Latar Belakang di atas maka dapat dirumuskan beberapa masalah, Rumusan masalah dalam penelitian ini adalah:

1. Bagaimana membangun system untuk segmentasi konsumen, menggunakan k-means clustering ?
2. Bagaimana cara mengimplementasikan metode K-means clustering dalam

aplikasi data mining segmentasi pelanggan ?

## Tujuan

Berdasarkan rumusan masalah yang ditentukan di atas maka tujuan dari penelitian ini adalah:

1. Merancang dan membangun aplikasi sistem mesin learning agar dapat menyelesain masalah pada segmentasi pelanggan.
2. Mengimplementasikan metode K-means clustering dalam aplikasi data mining segmentasi pelanggan.

## Batasan Masalah

Batasan masalah dalam penelitian ini adalah:

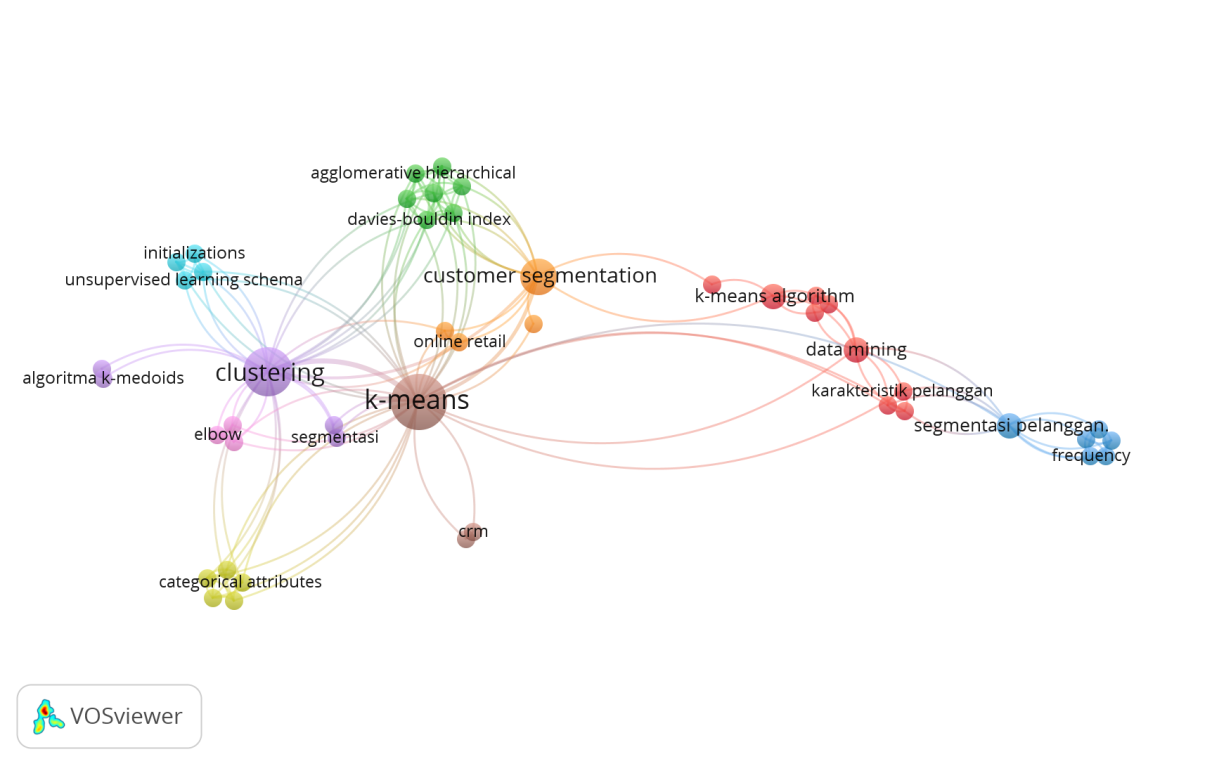
1. Data yang digunakan berdasarkan data studi kasus yang di dapat melalui website kaggel .
2. Aplikasi android hanya support pada handphone versi android 10 ke atas.
3. Metode yang digunakan hanya menggunakan *K-means clustering.*
4. Bahasa Pemrograman, Pembahasan data mining menggunakan bahasa pemrograman tertentu, misalnya Python dengan framework TensorFlow atau PyTorch.
5. Perbandingan Kinerja, Perbandingan kinerja model K-means Clustering dengan metode-metode lain dalam konteks data mining segmentasi tidak termasuk dalam lingkup laporan ini.

# BAB II

# LANDASAN TEORI

## State Of The Art

*State of the art* diambil dari beberapa penelitian yang telah dilakukan sebelumnya sebagai panduan serta menjadi acuan perbandingan dalam penelitian yang akan dilakukan. *State of the art* dalam penelitian ini dapat dilihat pada



*Sumber: VOSviewe*

Gambar 2.1 State Of the Art

Penjelasan terkait State of the Art di atas judul skripsi penelitian ini menggunakan 20 teori yang dimana masing – masing 10 jurnal nasional dan 10 jurnal internasional, teori memiliki keterkaitan dengan judul yang diangkat oleh penulis.

## Studi Literatur

Studi literatur adalah teknik pengumpulan data atau cara untuk menyelesaikan persoalan dengan menelusuri sumber-sumber tulisan yang pernah dibuat sebelumnya. Dengan kata lain, istilah studi literatur ini juga sangat familiar dengan sebutan studi pustaka. Dalam hal ini penulis mengutiip beberapa jurnal yang dijadikan acuan sebagai sumber untuk membuat sebuah aplikasi bla bla bla yang telah dibuat. Berikut beberapa jurnal yang berkaitan dengan penelitian yang akan dilakukan:

* + - 1. Jurnal Nasional: **“Segmentasi Pelanggan Menggunakan K-means Clustering Studi Kasus Pelanggan UHT Milk Greenfield”,**

Penelitian yang dilakukan oleh (Normah, Rifai, & Sari, 2020) Menentukan pola penjualan dengan melihat kecenderungan pembelian konsumen jika dianalisa dan diolah dengan baik maka dapat membantu mengetahui produk mana yang terlaris dan kurang laku terjual, sehingga kontrol stok persediaan dapat dilakukan, dan dapat dijadikan sebagai masukan bagi perusahaan dalam pengembangkan strategi pemasaran. Penelitian ini bertujuan untuk melakukan segmentasi pelanggan menggunakan metode K-Means Clustering dalam kasus pelanggan susu UHT Greenfield. Segmentasi pelanggan penting untuk memahami preferensi, kebutuhan, dan karakteristik pelanggan yang berbeda, sehingga perusahaan dapat mengarahkan upaya pemasaran dengan lebih efektif. Metode K-Means Clustering digunakan untuk mengelompokkan pelanggan berdasarkan atribut tertentu, seperti preferensi rasa, alamat pengiriman, dan depot penjualan.

1. Jurnal Nasional: **“Penerapan Data Mining Untuk Menentukan Segmentasi Pelanggan Dengan Menggunakan Algoritma K-Means dan Model RFM Pada E-Commerce”,**

Penelitian yang dilakukan (A. Wibowo and A. R. Handoko, 2020) membahas tentang perkembangan jual beli yang secara perlahan dari tahun ke tahun kegiatan jual beli yang dilakukan oleh masyarakat secara konvensional tersebut sudah beralih dengan menggunakan peran dari teknologi dan internet. Penggunaan internet bagi masyarakat khususnya di Indonesia sangat meningkat pesat beberapa tahun belakangan ini dan juga meningkat pada setiap tahunnya. Pada penelitian ini selain algoritma k-means, proses segmentasi juga dilakukan dengan melakukan analisis RFM. Model analisis RFM (Recency, Frequency dan Monetary) dapat melakukan identifikasi terhadap karakteristik yang dimiliki oleh para pelanggan dan dimana Model RFM tersebut dapat dikombinasikan dengan algoritma untuk mendapatkan nilai akurasi yang tinggi terhadap hasil clusterisasi (E. Ernawati, S. S. K. Baharin, 2021). Untuk mendapatkan informasi – informasi penting yang dapat dipergunakan bagi para pelaku usaha biasa dipergunakan dengan sebuah cara yang bernama data mining. Maka dari itu, penelitian ini bertujuan untuk menerapakan proses data mining pada E-Commerce untuk mengetahui segmentasi pelanggan.

1. Jurnal Nasional:“S**egmentasi Pelanggan Pada PT. TAB Menggunakan Algoritma K-Means Clustering”,**

Penlitian yang dilakukan (Solichin, A., & Khairunnisa, 2020) menjelaskan Tentang penelitian ini klasterisasi pelanggan yang akan dibentuk dengan menggunakan algoritma K-Means, dengan data yang diolah berasal dari hasil laporan transaksi penjualan periode 2019 sampai dengan 2022. Setelah melalui proses preprocessing, dataset yang telah disiapkan akan dinormalisasikan agar memiliki rentang data yang tidak terlalu jauh. Tujuan dari penelitian ini adalah untuk membentuk kelompok pelanggan, dengan metode K-Means dihasilkan tiga buah klaster dengan populasi 34,96% kelas Bronze, 34,96% kelas Silver dan kelas Gold 28,01%.

1. Jurnal Nasional: **“Algoritma K-Means Clustering Untuk Segmentasi Pelanggan (Studi Kasus : Fashion Viral Solo)”,**

Penelitian yang dilakukan oleh (Febriani & Putri, 2020) menjelaskan tentang segmentasi pelanggan yang dipakai untuk mengevaluasi nilai dari setiap pelanggan sehingga perusahaan dapat mengidentifikasi pelanggan yang memberikan keuntungan yang besar dan pelanggan yang tidak. Menurut (Fitriyani & Hendriyani, 2021), dengan mengelola dan memanfaatkan database pelanggan dengan tepat, dapat membantu manajemen untuk mendapatkan informasi seperti : mengetahui berbagai kebutuhan pelanggan, memberikan penawaran khusus, memudahkan dalam membuat program loyalitas, dan membantu dalam segmentasi pelanggan. Pada penelitian ini, peneliti menggunakan algoritma K-Means untuk melakukan clustering dan penerapan metode Recency, Frequency dan Monetary. K-Means adalah algoritma sederhana, mudah diimplementasikan, tidak lambat, mudah disesuaikan, dan sering digunakan dalam proses data mining khususnya clustering. Nilai atribut Recency menunjukkan waktu terakhir transaksi, Frequency menunjukkan jumlah transaksi, dan Monetary menunjukkan total nominal transaksi. Dengan menggunakan K-Means dan metode Recency, Frequency dan Monetary, peneliti dapat melakukan segmentasi pelanggan Fashion Viral Solo. Dari penelitian ini, diperoleh hasil 2 cluster pelanggan. Cluster 1 dengan jumlah anggota terbanyak, yakni 343 pelanggan dan cluster 2 dengan jumlah anggota 8 pelanggan.

1. Jurnal Nasional: **“Iplementasi Algoritma K-Medoids Clustering Untuk Menentukan Segmentasi Pelanggan”,**

Penelitian yang dilakukan oleh (Mousavi, Boroujeni & Aryanmehr. 2020) menjelaskan tentang Algoritma K-Medoids menggunakan medoid (perwakilan) data sebagai ganti dari mean untuk pusat cluster. Medoid adalah objek data paling sentral di antara titik-titik cluster. Oleh karena itu, K objek dipilih secara acak sebagai medoid untuk mewakili cluster, dan semua objek data ditugaskan ke cluster dengan medoid terdekat. Tujuan penelitian ini adalah untuk melakukan segmentasi pelanggan pada XYZ Store dengan menggunakan teknik data mining clustering yaitu dengan menggunakan algoritma K-Medoids yang menggunakan perwakilan data sebagai pusat cluster. Atribut yang digunakan dalam penelitian ini menggunakan analisis RFM yang terdiri dari tiga atribut yaitu: recency, frequency dan monetary. Penentuan jumlah cluster optimal menggunakan metode Silhouette coefficient dan evaluasi hasil cluster menggunakan metode Davies-Bouldin Index (DBI).

1. Jurnal Nasional: **“Optimalisasi Strategi Pemasaran dengan Segmentasi Pelanggan Menggunakan Penerapan K-Means Clustering pada Transaksi Online Retail”,**

Penelitian yang dilakukan oleh (Carudin, 2023) menjelaskan tentang 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 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. 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.

1. Jurnal Nasional: **“Penerapan Data Mining Menggunakan Metode K-Means Untuk Penentuan Reward Pelanggan”,**

Penelitian yang dilakukan oleh (Ermanto, 2020) menjelaskan tentang sistem yang akan diterapkan dapat memilih pelanggan mana saja yang loyal dan pantas mendapatkan reward sesuai dari kuantitas pembelian dan total harga yang dibeli pada periode tertentu. Penyubur Tani kesulitan dalam menghitung satu persatu pelanggan mana yang frekuensi pembelian dan total harga belanjanya yang paling masuk kategori sangat loyal dan loyal. Oleh karena itu diperlukan sebuah aplikasi untuk mengelompokkan pelanggan sehingga dalam menentukan strategi dalam membangun loyalitas tepat sasaran. Dengan menggunakan konsep CRM dan metode K-Means maka hasil yang diperoleh dari pengolahan data mampu mengelompokkan pelanggan yang harus diprioritaskan dan dapat menentukan pelanggan mana yang pantas mendapatkan reward. Dari 100 data pelanggan, metode K-Means Clustering berhasil mengelompokkan kriteria sangat loyal sebesar 8%, loyal 34% dan potensial sebesar 58%.

1. Jurnal Nasional: **“Segmentasi Pelanggan Berdasarkan Analisis Recency, Frequency, Monetary Menggunakan Algoritma K-Means”,**

Penelitian yang dilakukan oleh (Sri Wahyuni & Fahrullah, 2023) menjelaskan tentang Model RFM merupakan model untuk menentukan segmentasi konsumen berdasarkan kapan transaksi terakhir dilakukan, frequency yaitu tingkat keseringan pelanggan transaksi dan monetary ialah besarnya nilai transaksi yang dilakukan CV. Toedjoe Sinar Group merupakan salah satu percetakan di Samarinda dan bergerak di bidang digital printing dengan beragam produk, tidak hanya fokus pada strategi yang mengutamakan produk, tetapi juga melakukan strategi yang mengutamakan pelanggan. Hal ini diperlukan untuk mengetahui perilaku pelanggan sehingga akan membantu dalam penerapan strategi pemasaran yang tepat untuk meningkatkan pendapatan perusahaan. 8020 data transaksi pelanggan dari Bulan Januari hingga Maret tahun 2022 digunakan dalam penelitian ini. Pemodelan data yang digunakan dalam pengelompokan pelanggan adalah RFM (Recency, Frequency, Monetary) dan teknik data mining dengan menggunakan metode clustering menggunakan algoritma k-means yang terdiri dari business understanding, data understanding, data preparation, modeling, evaluation, dan deployment.

1. Jurnal Nasional: “Penerapan Algoritma K-Means Dalam Segmentasi Pelanggan Pada Toko Sembako Menggunakan Rapidminer”,

Penelitian yang dilakukan oleh (Wahyu Sudrajat, dkk., 2022) menjelaskan tentang bagaimana pengelompokan pelanggan sehingga memudahkan untuk membuat sebuah kebijakan pemasaran. Metode yang digunakan adalah algoritma k-means dengan memanfaatkan rapidminer. Hasil yang di dapatkan bahwa terdapat tiga kluster dengan sebaran data sebanyak 67,3 % (99 pelanggan) yang termasuk kategori pelanggan potensial, untuk pelanggan cukup potensial sebanyak 23,1% (34 pelanggan) serta 9,5% (14 pelanggan) termasuk kategori kurang. Sehingga toko dapat menerapkan kebijakan yang mempertahankan loyalitas pelanggan potensial namun perlu juga membuat strategi kepada kelompok pelanggan yang cukup dan kurang potensial.

1. Jurnal Nasional: “K-means Clustering With Comparison Of Elbow and Silhoutte Methods For Medicines Clustering Based On User Reviews”,

Penelitian yang dilakukan oleh (Safitri Juanita & Raynaldi Dwi Cahyono, 2024) menjelaskan tentang kontribusi dari penelitian ini mengusulkan model terbaik untuk mengklaster ulasan pengguna terhadap obat-obatan menggunakan K-Means dengan membandingkan 2 (dua) teknik untuk menentukan jumlah klaster yang optimal, yaitu Silhouette dan Elbow. Penelitian ini bertujuan untuk merekomendasikan metode clustering K-Means yang terbaik untuk memproses data ulasan dan rating yang luas, dan hasil clustering membantu para ahli medis, pencari informasi obat-obatan, atau bisnis farmasi menentukan pangsa pasar obat-obatan. Hasil penelitian menunjukkan bahwa model K-Means memiliki performaterbaik ketika melakukan clustering menggunakan metode Silhouette dengan nilai DBI sebesar 0.261 dan menghasilkan 2 cluster. Sementara itu, model Elbow memiliki nilai performa terbaik sebesar 0.460 dan menghasilkan 3 cluster. Penelitian ini juga menemukanbahwa hasil clustering dengan kedua metode tersebut menghasilkan tiga kelompok klaster obat berdasarkan ulasan yaitu moderat, tidak populer dan terkenal.

1. Jurnal Internasional: **“An Implementation of K-Means Clustering for Efficient Image Segmentation of Natural Background Images”.**

Penelitian yang dilakukan oleh (Aswin Kumer S V & Dr E Mohan, 2020) menjelaskan tentang segmentasi citra adalah proses menemukan dan mengelompokkan piksel-piksel yang berkorelasi pada citra tertentu. Ada berbagai jenis yang tersedia untuk menemukan piksel yang berkorelasi dalam gambar. Dalam tulisan ini, K-Means Clustering digunakan untuk segmentasi untuk mengamati objek gambar yang berbeda dalam suatu gambar. Mula-mula sampel masukan diubah menjadi citra skala abu-abu, setelah itu citra tersebut diproses dengan clustering K-Means untuk menghasilkan keluaran citra tersegmentasi. K-Means didasarkan pada pengelompokan piksel serupa dan penjatahan piksel tengah. Dengan mengulangi proses yang sama beberapa kali, maka keluaran gambar tersegmentasi akan memiliki diskriminasi objek yang sangat baik. Dari beberapa algoritma segmentasi citra, clustering K-Means akan memberikan hasil yang baik. Diskriminasi objek murni didasarkan pada korelasi piksel yang ada pada gambar. Setelah diproses, dilakukan juga pembentukan ulang gambar untuk visualisasi gambar tersegmentasi yang lebih baik.

1. Jurnal internasional: **“The k-means Algorithm: A Comprehensive Survey and Performance Evaluation”,**

Penelitian yang dilakukan oleh (Ahmed & Raihan Seraj and Syed Mohammed Shamsul Islam, 2020) menjelaskan tentang algoritma k-means clustering dianggap sebagai salah satu algoritma data mining yang paling kuat dan populer di komunitas riset. Namun, terlepas dari popularitasnya, algoritme ini memiliki keterbatasan tertentu, termasuk masalah yang terkait dengan inisialisasi acak dari centroid yang menyebabkan konvergensi yang tidak terduga. Selain itu, algoritme pengelompokan seperti itu memerlukan penentuan jumlah kluster terlebih dahulu, yang bertanggung jawab atas berbagai bentuk kluster dan efek outlier. Masalah mendasar dari algoritma k-means adalah ketidakmampuannya menangani berbagai tipe data. Makalah ini memberikan gambaran terstruktur dan sinoptik dari penelitian yang dilakukan pada algoritma k-means untuk mengatasi kekurangan tersebut. Varian algoritma k-means termasuk perkembangan terkininya dibahas, di mana efektivitasnya diselidiki berdasarkan analisis eksperimental dari berbagai kumpulan data.

1. Jurnal internasional: **“K-Means-Based Nature-Inspired Metaheuristic Algorithms for Automatic Data Clustering Problems: Recent Advances and Future Directions”,**

Penelitian yang dilakukan oleh (Abiodun M. Ikotun, 2021) menjelaksan tentang penelitian ini bertujuan untuk mengidentifikasi, mengambil, merangkum, dan menganalisis penelitian yang diusulkan baru-baru ini terkait dengan perbaikan Algoritma pengelompokan K-means dengan teknik optimasi yang terinspirasi dari alam. Pendekatan pencarian untuk pemilihan artikel diadopsi, yang mengarah pada identifikasi dan pemilihan 147 studi terkait dari jalur akademik dan database bereputasi berbeda. Terlebih lagi, analisis mengungkapkan bahwa meskipun demikian. Algoritma K-means telah diteliti dengan baik dalam literatur, keunggulannya dibandingkan beberapa literatur lainnya algoritma pengelompokan canggih dalam hal kecepatan, aksesibilitas, kesederhanaan penggunaan, dan penerapan untuk memecahkan masalah pengelompokan dengan kumpulan data yang tidak berlabel dan dapat dipisahkan secara nonlinier telah diamati dengan jelas dalam penelitian ini. Penelitian saat ini juga mengevaluasi dan membahas beberapa hal kelemahan yang diketahui dari algoritma pengelompokan K-means, yang mana perbaikannya sudah ada metode dikonseptualisasikan.

1. Jurnal internasional: **“K-Means Clustering Algorithm and Its Simulation Based on Distributed Computing Platform”.**

Penelitian yang dilakukan oleh (Chunqiong Wu dkk., 2021) menjelaskan tentang, saat ini, pertumbuhan data yang eksplosif dan kondisi penyimpanan massal telah membawa banyak masalah seperti komputasi kompleksitas dan daya komputasi yang tidak memadai untuk mengelompokkan penelitian. e platform komputasi terdistribusi melalui beban penyeimbangan secara dinamis mengonfigurasi sejumlah besar sumber daya komputasi virtual, yang secara efektif menerobos kemacetan konsumsi waktu dan energi, dan mewujudkan keunggulan uniknya dalam penambangan data besar-besaran. adalah makalah mempelajari parallel berarti secara luas. adalah artikel pertama menginisialisasi pengambilan sampel secara acak dan kedua memparalelkan proses perhitungan jarak itu memberikan kemandirian antar objek data untuk melakukan analisis cluster secara paralel. Setelah pemrosesan parallel MapReduce, kami menggunakan banyak node untuk menghitung jarak, yang mempercepat efisiensi algoritma. Terakhir, pengelompokan objek data diparalelkan. Hasil menunjukkan bahwa metode kami dapat memberikan layanan secara efisien dan stabil serta memiliki konvergensi yang baik.

1. Jurnal internasional: **“The Application of Data Mining for Predicting Academic Performance Using K-means Clustering and Naïve Bayes Classification”,**

Penelitian yang dilakukan oleh (Zainab Mohammed Ali, dkk, 2020) menjelaskan tentang Data Mining adalah proses analisis multidisiplin yang berkonsentrasi untuk mengekstraksi dan menemukan pengetahuan yang berguna dari data dan informasi. Pengklasifikasi Naïve Bayes mungkin merupakan pengklasifikasi probabilistik yang paling banyak diterapkan pendekatan yang dapat digunakan untuk eksplorasi data. Makalah ini menggunakan pengklasifikasi Naïve Bayes untuk pendidikan proses penambangan data untuk membantu meningkatkan perbedaan kualitas sistem pengajaran di pendidikan tinggi. Ini dengan menggali data evaluasi siswa terkait kinerja instruktur untuk mempelajari atribut utama yang mungkin mempengaruhi kinerja pendidikan di berbagai kursus. Oleh karena itu, makalah ini menggunakan algoritma k-means clustering, yang digunakan untuk menentukan pusat cluster yang ideal, sehingga dapat menjadi centroid cluster. Selanjutnya, Naïve Bayes algoritma proses klasifikasi diterapkan pada data evaluasi akademik untuk menghasilkan aturan yang dipelajari dan dievaluasi untuk memprediksi kinerja pendidikan. Sistem yang diusulkan membantu mengidentifikasi anak putus sekolah dan memberikan nasihat atau konseling yang tepat bagi manajemen pendidikan dalam menyelenggarakan pendidikan yang berilmu keputusan untuk mempertimbangkan dan merestrukturisasi kurikulum pendidikan. Juga, untuk meningkatkan pengalaman akademis instruktur yang pada akhirnya akan meningkatkan kualitas lingkungan pendidikan suatu lembaga pendidikan.

1. Jurnal internasional: **“Incorporating K-means, Hierarchical Clustering and PCA in Customer Segmentation”,**

Penelitian yang dilakukan oleh (Azad Abdulhafedh, 2021) menjelaskan tentang Makalah ini membahas penggunaan algoritma clustering dalam segmentasi pelanggan untuk mendefinisikan strategi pemasaran perusahaan kartu kredit. Segmentasi pelanggan membagi pelanggan menjadi beberapa kelompok berdasarkan karakteristik umum, yang berguna bagi bank, dunia usaha, dan perusahaan untuk meningkatkan produk atau layanannya peluang. Analisis ini mengeksplorasi penerapan K-means, pengelompokan Hierarki, dan Prinsipal Component Analysis (PCA) dalam mengidentifikasi segmen pelanggan suatu perusahaan berdasarkan kartu kreditnya sejarah transaksi. Kumpulan data yang digunakan dalam proyek ini merangkum perilaku penggunaan 8950 pemegang kartu kredit aktif dalam 6 bulan terakhir, dan tujuan kami adalah melakukan segmentasi pelanggan dengan cara yang paling akurat menggunakan clustering teknik. Proyek ini menggunakan dua pendekatan untuk segmentasi pelanggan: pertama, dengan mempertimbangkan seluruh variabel dalam algoritma pengelompokan menggunakan Hierarchical clustering dan K-means. Kedua, dengan menerapkan dimensi pengurangan melalui Principal Component Analysis (PCA) ke dataset, kemudian mengidentifikasi jumlah optimal cluster, dan mengulangi analisis clustering dengan jumlah cluster yang diperbarui. Hasil menunjukkan bahwa PCA bisa secara efektif digunakan dalam proses pengelompokan sebagai alat pemeriksaan untuk K-means dan pengelompokan Hierarki.

1. Jurnal internasional: **“K-Means Clustering Approach for Intelligent Customer Segmentation Using Customer Purchase Behavior Data”,**

Penelitian yang dilakukan oleh (Kayalvily Tabianan, Shubashini Velu & Vinayakumar Ravi, 2022) menjelaskan tentang Sistem e-commerce menjadi lebih populer dan diterapkan di hampir semua bidang bisnis. Sistem e-commerce adalah platform untuk memasarkan dan mempromosikan produk kepada pelanggan melalui online. Segmentasi pelanggan dikenal sebagai proses membagi pelanggan menjadi kelompok-kelompok yang memiliki karakteristik serupa. Tujuan dari segmentasi pelanggan adalah untuk menentukan cara menghadapi pelanggan di setiap kategori guna meningkatkan keuntungan setiap pelanggan terhadap bisnis. Ada beberapa jenis faktor segmentasi pelanggan yaitu psikografis demografis, perilaku, dan geografis. Dalam penelitian ini, faktor perilaku pelanggan menjadi fokus. Oleh karena itu pengguna akan dianalisis menggunakan algoritma clustering dalam menentukan perilaku pembelian sistem E-commerce. Tujuan dari pengelompokan adalah untuk mengoptimalkan kesamaan eksperimental dalam cluster dan untuk memaksimalkan ketidaksamaan antar cluster. Dalam penelitian ini terdapat hubungan antara tiga cluster: jenis acara, produk, dan kategori. Dalam penelitian ini, pendekatan yang diusulkan menganalisis kelompok-kelompok yang memiliki kriteria serupa untuk membantu vendor mengidentifikasi dan fokus pada segmen dengan keuntungan tinggi ke segmen yang paling tidak menguntungkan. Untuk mengolah data yang dikumpulkan dan mengelompokkan pelanggan, digunakan algoritma pembelajaran yang disebut K-Means clustering. Clustering K-Means diimplementasikan untuk menyelesaikan masalah clustering.

1. Jurnal internasional: **“Retracted: An Empirical Study on Customer Segmentation by Purchase Behaviors Using a RFM Model and K-Means Algorithm”,**

Penelitian yang dilakukan oleh (J. Wu, L. Shi, W. Lin dkk., 2020) menjelaskan tentang bahwa dalam berdasarkan penelitian kami dengan menangani masalah dunia nyata dalam suatu perusahaan. Model RFM (kekinian, frekuensi, dan moneter) dan algoritma pengelompokan K-means digunakan untuk melakukan segmentasi pelanggan dan analisis nilai dengan menggunakan data penjualan online. Pelanggan diklasifikasikan menjadi empat kelompok berdasarkan perilaku pembelian mereka. Atas dasar ini, berbagai strategi CRM (manajemen hubungan pelanggan) dikedepankan untuk mendapatkan tingkat kepuasan pelanggan yang tinggi. Efektivitas metode yang kami usulkan dalam makalah ini didukung oleh hasil perbaikan beberapa indeks kinerja utama seperti pertumbuhan pelanggan aktif, total volume pembelian, dan jumlah total konsumsi.

1. Jurnal internasional: **“Unsupervised K-Means Clustering Algorithm”,**

Penelitian yang dilakukan oleh (Kristina P. Sinaga & Miin-Shenyang, 2020) menjelaskan tentang algoritma k-means umumnya merupakan metode clustering yang paling dikenal dan digunakan. Algoritma k-means bukanlah metode pengelompokan tanpa pengawasan. Dalam makalah ini, kami membangun pembelajaran tanpa pengawasan skema untuk algoritma k-means agar bebas inisialisasi tanpa pemilihan parameter dan juga bisa secara bersamaan menemukan jumlah cluster yang optimal. Artinya, kami mengusulkan k-means baru tanpa pengawasan (Uk-means) algoritma clustering dengan secara otomatis menemukan jumlah cluster yang optimal tanpa memberikan apapun inisialisasi dan pemilihan parameter. Kompleksitas komputasi dari pengelompokan Uk-means yang diusulkan algoritma juga dianalisis. Perbandingan antara U-k-means yang diusulkan dan metode lain yang ada adalah dibuat. Hasil eksperimen dan perbandingan sebenarnya menunjukkan aspek-aspek baik dari usulan U-k berarti algoritma pengelompokan.

1. Jurnal internasional: **“Performances of K-Means Clustering Algorithm with Different Distance Metrics”,**

Penelitian yang dilakukan oleh (Taher M. Ghazal dkk., 2021) menjelaskan tentang clustering adalah proses pengelompokan data berdasarkan kemiripannya properti. Salah satu algoritma pengelompokan yang paling populer adalah K-means, di mana jarak diukur antara setiap titik kumpulan data dan pusat massa cluster untuk menemukan objek data serupa dan menugaskannya ke cluster terdekat. Dalam penelitian ini, algoritma clustering K-means dievaluasi dengan tiga metrik matematika yang berbeda dalam hal waktu eksekusi dengan berbeda kumpulan data dan jumlah cluster yang berbeda. Hasilnya menunjukkan bahwa penerapannya tasi metrik ukuran jarak Manhattan mencapai hasil terbaik di sebagian besar Negara kasus. Hasil ini juga menunjukkan bahwa metrik jarak dapat mempengaruhi eksekusi waktu dan jumlah cluster yang dibuat oleh algoritma K-means.

Tabel 2. 1 Perbandingan Penelitian Terkait Dan Peneliti

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| NO | PERBANDINGAN | **PENELITIAN** | | | | | | | | | | | | | | | | | | | | |
| **PP** | **P1** | **P2** | **P3** | **P4** | **P5** | **P6** | **P7** | **P8** | **P9** | **P10** | **P11** | **P12** | **P13** | **P14** | **P15** | **P16** | **P17** | **P18** | **P19** | **P20** |
| Kelompok 11 | (Normah, Rifai, & Sari, 2020) | (A. Wibowo and A. R. Handoko, 2020) | (Solichin, A., & Khairunnisa, 2020) | (Febriani & Putri 2020) | (Mousavi, Boroujeni 2020) | (Carudin, 2023) | (Ermanto, 2020) | (Sri Fahrullah, 2023) | (Wahyu Sudrajat, dkk., 2022) | (Safitri Juanita & Raynaldi Dwi Cahyono, 2024) | ( Dr E Mohan, 2020) | (Ahmed & Raihan Seraj and Syed Mohammed Shamsul Islam, 2020) | (Abiodun M. Ikotun, 2021) | (Chunqiong Wu dkk, 2021) | (Zainab Mohammed Ali, dkk., 2020) | (Azad Abdulhafedh, 2021) | (Shubashini Velu & Vinayakumar Ravi, 2022), | (J. Wu, L. Shi, W. Lin dkk, 2020) | (Kristina P. Sinaga & Miin-Shenyang, 2020) | (Taher M. Ghazal dkk., 2021) |
| FITUR | | | | | | | | | | | | | | | | | | | | | | |
| 1 | Menggunakan Data mining dalam segmentasi pelanggan | √ | √ |  |  |  | √ |  |  | √ |  |  |  |  | √ |  | √ | √ | √ |  |  | √ |
| 2 | Menggunakan Model RFM dalam segmentasi pelnggan |  |  | √ |  | √ | √ |  |  | √ |  | √ |  |  |  |  |  |  |  | √ |  |  |
| 3 | Memberi petunjuk untuk melakukan segmentasi pelanggan | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ |
| 4 | Menggunakan Alogoritma K-Means Clustering untuk melakukan segmentasi pelanggan | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ |
| 5. | Melakukan analisis cluster | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ |
| METODE | | | | | | | | | | | | | | | | | | | | | | |
| 1. | Algoritma k-means clustering | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ |
| **TOOLS** | | | | | | | | | | | | | | | | | | | | | | |
| 1. | Android Studio | √ |  |  |  | √ |  |  |  |  | √ |  |  |  |  |  | √ |  |  |  |  |  |
| 2. | Python | √ |  |  |  |  |  | √ |  | √ |  | √ |  |  |  |  |  |  |  |  |  |  |
| 3. | Java | √ |  | √ |  |  |  |  |  |  |  |  |  | √ |  |  |  |  | √ |  |  |  |

**Keterangan :**

1. PP : Peneliti Penulis.
2. P1 – P20 : Penelitian Jurnal Terkait.

Tata cara pengisian tabel penelitian pada Penelitian Terkait Dan Peneliti silahkan anda cek pada masing-masing referensi apakah ada kesamaan atau tida dari peneliti p1-p20 jika ada maka anda centang jika tidak maka tidak perlu.

## Literatur Terkait

### Pengertian Segmentasi Pelanggan

Segmentasi pelanggan adalah suatu cara untuk mengelompokkan pelanggan ke dalam beberapa kelompok (cluster) dan setiap kelompok memiliki beberapa anggota dengan karakteristik yang sama. Usaha untuk mengklasifkasikan objek-objek dengan suatu kesamaan ke dalam satu grup tersebut juga dapat dikatakan sebagai clustering. Analisis clusterakan membangun suatu clusteryang baik ketika setiap anggota dari clustermemiliki derajat kesamaan yang tinggi. Segmentasi pelanggan juga mempresentasikan elemen kunci dalam identifikasi pelanggan dalam customer relationship management.

Membagi pelanggan berdasarkan karakteristik dan ciri yang sama. Pembagian tersebut berdasarkan pada minat atau permintaan pelanggan dengan ciri yang serupa. Dengan melakukan segmentasi pelanggan, dapat membantu perusahaan untuk memasarkan produknya secara pendekatan individual agar sesuai dengan kebutuhannya. Menurut (Febriani & Putri, 2020) segmentasi pelanggan dipakai untuk mengevaluasi nilai dari setiap pelanggan sehingga perusahaan dapat mengidentifikasi pelanggan yang memberikan keuntungan yang besar dan pelanggan yang tidak. Menurut (Fitriyani & Hemdriyani, 2021) dengan mengelola dan memanfaatkan database pelanggan dengan tepat, dapat membantu manajemen untuk mendapatkan informasi seperti : mengetahui berbagai kebutuhan pelanggan, memberikan penawaran khusus, memudahkan dalam membuat program loyalitas, dan membantu dalam segmentasi pelanggan. Informasi yang diperoleh dari pengolahan data penjualan, dapat dijadikan sebagai dasar oleh manajemen untuk mengembangkan usaha.

### Data Mining

Data mining adalah sebuah proses menggali dan mengolah informasi penting dari kumpulan data. Proses penggalian informasi melibatkan teknik seperti memanfaatkan kecerdasan buatan, statistik, matematika, machine learning, dan lainnya. Teknik yang dipakai memiliki tujuan untuk mengenali dan mengekstrak informasi yang bermanfaat dari suatu database yang besar. Saat ini data mining telah menjangkau dan juga dimanfaatkan pada banyak bidang keilmuan lainnya.. Salah satu tugas umum dari data mining adalah melakukan clustering. Dalam melakukan clustering, maka diperlukan algoritma yang mampu melakukan pengelompokan data objek yang memiliki tingkat kesamaan antar kelas objek. Dari pengertian tersebut maka dapat ditarik kesimpulan jika model clustering cocok untuk mengelompokkan pelanggan berdasarkan tingkat kesamaannya menjadi beberapa cluster atau segmen. (Sudarsono dkk, 2021).

### Pengertian Metode K-Means Clustreing

K-Means adalah metode non-hierarkis untuk mengelompokkan data. Tujuannya adalah untuk membagi data yang ada menjadi satu atau lebih cluster atau kelompok sehingga data dengan karakteristik yang mirip berada dalam cluster yang sama dan data dengan karakteristik yang berbeda berada dalam kelompok yang berbeda (Z Nabila, A. Rahman Isnain, 2021).

Metode K-Means membagi data dengan mengelompokkan objek berdasarkan nilai yang diberikan untuk memisahkan kelompok-kelompok tersebut. Algoritma K-Means adalah metode yang relatif mudah untuk mengatur sejumlah besar objek dengan atribut tertentu ke dalam kelompok (cluster) K. Jumlah k cluster ditentukan sebelumnya dalam algoritma K-Means. Proses menempatkan data atau objek ke dalam kelas atau cluster berdasarkan seberapa mirip atributnya dikenal sebagai clustering. Salah satu metode untuk data mining adalah clustering.

Pengelompokan yang baik menghasilkan objek-objek dengan kesamaan rendah dengan yang ada di cluster lain tetapi kesamaan tinggi dengan yang ada di grup atau cluster yang sama. Metode penambangan data tanpa pengawasan adalah pengelompokan. Dalam proses pengelompokan data, dua jenis pengelompokan data sering digunakan: pengelompokan data non-hierarkis (non-hierarkis) dan pengelompokan data hierarkis (hierarkis).

Algoritma k-means clustering adalah metode yang digunakan untuk mengelompokkan data ke dalam k cluster berdasarkan kesamaan fitur. Berikut adalah rumus dan langkah-langkah dasar dalam algoritma k-means clustering:

1. Inisialisasi: Tentukan jumlah cluster, 𝑘, dan pilih secara acak 𝑘titik sebagai centroid awal (pusat cluster).
2. Assignment Step (Penugasan): Setiap titik data akan ditetapkan ke cluster ​ yang memiliki centroid terdekat. Ini dilakukan dengan menghitung jarak Euclidean antara titik data dan centroid dan menetapkan titik data ke centroid dengan jarak terpendek. Rumus untuk menghitung jarak Euclidean adalah:

di mana adalah nilai fitur ke-m dari titik data dan ​ adalah nilai fitur ke-m dari centroid ​.

1. Update Step (Pembaruan): Hitung ulang posisi centroid berdasarkan rata-rata dari semua titik data dalam cluster tersebut. Rumus untuk menghitung centroid baru dari cluster adalah:

di mana adalah jumlah titik data dalam cluster

1. Konvergensi: Ulangi langkah 2 dan 3 sampai centroid tidak lagi berubah secara signifikan atau jumlah iterasi maksimum tercapai.

### Pengertian Flowchart

*Flowchart* adalah representasi dalam bentuk diagram dari alur pemecahan sebuah masalah atau algoritma. *Flowchart* merupakan representasi dari algoritma yang statefull, karena setiap langkah diwakili oleh satu bentuk diagram. *Flowchart* merupakan salah satu bentuk diagram yang paling sederhana dalam menuliskan algoritma, oleh karena itu tidak semua algoritma dapat dituliskan dalam bentuk *flowchart.* Bagi seorang programer, algoritma digunakan untuk membuat modul-modul program. Salah satu metode untuk menyajikan algoritma secara tepat adalah dengan menggunakan flowchart (diagram alir). Flowchart adalah gambaran dalam bentuk diagram alir dari algoritma-algoritma dalam suatu program, yang menyatakan arah alur program tersebut.

Flowchart adalah penggambara secara grafik dari langkah-langkah dan urut-urutan prosedur dari suatu program. Flowcchart sistem merupakan suatu urutan proses dalam system dengan menunjukkan alat dari media input, output serta jenis media yang digunakan untuk penyimpanan dalam proses pengolahan data sedangkan flowchart program merupakan suatu bagan dengan simbol-simbol tertentu yang menggambarkan suatu urutan dari proses secara detail dan berhubungan antara suatu proses (instruksi) dengan proses lainnya dalam suatu program .

Tabel 2. Simbol – simbol Flowchart dan Nama Fungsi Terminator

|  |  |  |  |
| --- | --- | --- | --- |
| No | Simbol | Nama | Fungsi |
| 1. |  | Terminator | Permulaan/akhir program |
| 2. |  | Garis Alir (Flow Line) | Arah aliran pogram |
| 3. |  | Preparation | Proses inisialisasi/pemberian harga awal |
| 4. |  | Proses | Proses perhitungan/proses pengolahan data |
| 5. |  | Input/Output | Proses input/output data, parameter, dan infomasi |
| 6. |  | Predeined Process (Sub Program) | Permulaan sub program/proses menjalankan sub program |
| 7. |  | Decision | Perbandingan pernyataan, penyeleksian data yang memberikan pilihan untuk langkah selanjutnya |
| 8. |  | *On Page Connector* | Penghubung bagian-bagian *flowchart* yang berada pada satu halaman |
| 9. |  | *Off Page Connector* | Penghubung bagian-bagian *flowchart* yang berada pada halaman berbeda |

*Sumber:* (Zalukhu Agustinus, dkk 2023)

### Pengertian Anroid Studio

Android Studio merupakan sebuah Integrated Development Environment (IDE) khusus untuk membangun aplikasi yang berjalan pada platform android. Android studio ini berbasis pada IntelliJ IDEA, sebuah IDE untuk Bahasa pemrograman Java. Bahasa pemrograman utama yang digunakan adalah Java, sedangkan untuk membuat tampilan atau layout, digunakan bahasa XML. Android studio juga terintegrasi dengan Android Software Development Kit (SDK) untuk deploy ke perangkat android. Android Studio juga merupakan pengembangan dari eclipse, dikembangkan menjadi lebih kompleks dan professional yang telah tersedia didalamnya Android Studio IDE, Android SDK tools (Sondang Sibueaa, Mohammad Ikhsan Saputrob & Widodod, 2022)**.**

### Python

Python adalah bahasa pemrograman dinamis, tingkat tinggi, dimana merupakan bahasa pemrograman interpreter yaitu bahasa yang mengkonversi source code menjadi machine code secara langsung ketika program dijalankan. Bahasa ini juga mendukung pendekatan pemrograman Berorientasi Objek untuk pengembangan aplikasi dan mudah dipelajari serta menyediakan banyak struktur data tingkat tinggi. Python adalah bahasa skrip yang mudah dipelajari namun kuat dan serbaguna, yang membuatnya menarik untuk Pengembangan Aplikasi. Sintaks dan pengetikan Python sangat dinamis dengan sifat interpretasinya menjadikannya bahasa yang ideal untuk skrip dan pengembangan aplikasi yang cepat. Python mendukung banyak pola pemrograman, termasuk gaya pemrograman berorientasi objek, imperatif, dan fungsional serta prosedural. (Suharto, 2023)

### Kaggle

Kaggle, anak perusahaan dari Google LLC, adalah komunitas online untuk para data scientist dan praktisi dari machine learning. Kaggle memungkinkan pengguna untuk menemukan dan menerbitkan dataset, menjelajahi dan membangun model dalam lingkungan data-science berbasis web, bekerja dengan para data scientist dan insinyur machine learning, dan mengikuti kompetisi untuk memecahkan tantangan data science. Salah satu fitur unggulan yang ada pada Kaggle ini adalah terdapatnya banyak public dataset. (Hadi dkk., 2020).

### Dataset

Dataset merupakan sebuah kumpulan data yang berasal dari informasiinformasi pada masa lalu dan siap untuk dikelola menjadi sebuah informasi baru yang biasanya digunakan untuk klasifikasi, prediksi pada machine learning dan deteksi objek. Dataset dapat berupa data apapun yang valid, bisa berupa file excel, bisa berupa gambar, bisa berupa file xml. Jenis dari dataset sendiri ada 2, yaitu ada private dan public dataset. Pada penelitian ini, dataset yang dipakai adalah kumpulan data tentang aplikasi pada Apple App Store, Dataset ini diambil pada website kagle. Dataset ini memiliki 11 ribu aplikasi pada App Store dan 16 kolom atribut. Sebelum data di pakai, data sudah dirapihkan dengan membuang missing value. Setelah itu data dinormalisasikan dengan metode MaxMinScaling. Atribut yang di ambil untuk model K-Means ada sebanyak 3 atribut yaitu : Size Byte, Price, dan User Rating. Melihat dari ukuran aplikasi, harga, dan rating pengguna.(Hadi dkk., 2020).

### Metode Pengembangan Sistem

Menerapkan segmentasi pelanggan menggunakan algoritma k-means clustering melibatkan beberapa langkah penting untuk memastikan hasil yang akurat dan berguna. Pertama, data pelanggan dikumpulkan dari berbagai sumber seperti database internal, survei, dan aktivitas media sosial. Data ini kemudian diproses untuk meningkatkan kualitasnya melalui pembersihan, normalisasi, dan feature engineering. Setelah itu, algoritma machine learning seperti K-means clustering digunakan untuk mengelompokkan pelanggan berdasarkan pola perilaku dan preferensi.

Model dilatih dan dievaluasi menggunakan metrik seperti Silhouette Score untuk menilai kinerjanya. Model yang sudah valid kemudian diintegrasikan ke dalam sistem bisnis perusahaan, seperti CRM, untuk memungkinkan segmentasi otomatis dan real-time. Pemantauan dan pembaruan rutin dilakukan untuk memastikan model tetap relevan dengan data pelanggan yang terus berkembang, memungkinkan perusahaan untuk merancang kampanye pemasaran yang lebih efektif dan personalisasi penawaran produk, yang pada akhirnya meningkatkan kepuasan pelanggan dan penjualan.(Kayalvily Tabianan, 2022)

### Usability Testing

Usability testing dari segmentasi pelanggan menggunakan k-means clustering adalah proses evaluasi untuk memastikan bahwa segmen pelanggan yang dihasilkan dari algoritma k-means clustering bermanfaat dan dapat diimplementasikan secara efektif dalam konteks bisnis atau analisis. Segmentasi pelanggan menggunakan machine learning memungkinkan UMKM untuk mengelompokkan pelanggan berdasarkan karakteristik dan perilaku mereka secara lebih akurat dan efektif. Dengan mengumpulkan data yang relevan seperti informasi demografis dan riwayat pembelian, serta menerapkan algoritma seperti K-Means Clustering atau Gaussian Mixture Models, Pemilik usaha dapat mengidentifikasi pola dan tren yang penting. Hasil segmentasi ini kemudian digunakan untuk mengembangkan strategi pemasaran yang disesuaikan dengan kebutuhan masing-masing segmen, misalnya menawarkan program loyalitas khusus kepada pelanggan yang sering melakukan pembelian besar atau mengirimkan penawaran diskon kepada pelanggan yang sensitif terhadap harga. Dengan demikian, pemilik usaha dapat meningkatkan kepuasan pelanggan, retensi, dan penjualan secara keseluruhan. (John dkk., 2023).

### White Box

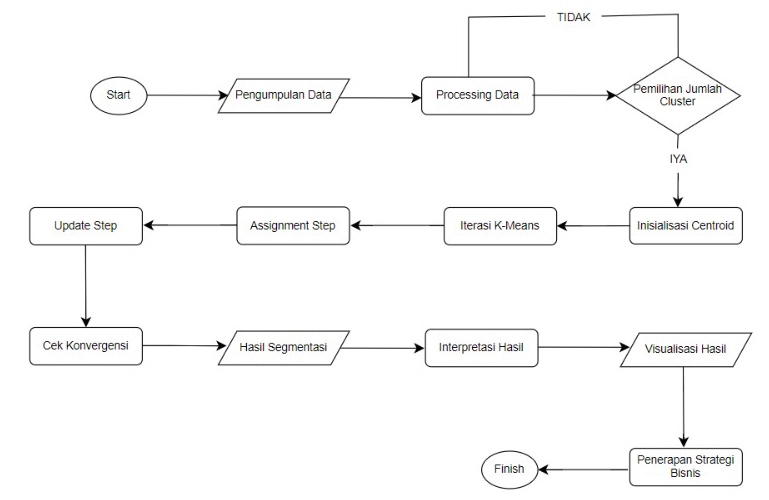
White box, disebut dengan pengujian kotak kaca atau pengujian terstruktur dimana pengujian dikembangkan berdasarkan kode –kode sistem. Pengujian white box memiliki pengetahuan tentang kode dengan menggunakan parameter yang sesuai. White box testing disebut sebagai pengujian structural. Yang mana perangkat lunak yang diuji merupakan hal transparan kepada penguji. Dalam pengujian, uji dirancang dari perspektif pengembang dikarenakan struktur internal dikenal dengan menguji segala bagian kode yang mampu untuk diuji dengan tujuan untuk menentukan kesalahan logis dari kode sumber perangkat lunak. Tipologi acak kesalahan bisa ditemukan dan mampu memunculkan asumsi yang salah dalam sebuah program. Penguji perlu melihat sumber kode dan mampu menentukan bagian mana kode yangtidak berfungsi dengan baik dan benar. Pengujian white box dikenal dengan nama pengujian clear box, glass box, atau open box. Pengujian white box testing dapat dimulai walaupun GUI masih dalam masa pengembangan. Model white box seperti pohon keputusan (decision tree), regresi linier, dan regresi logistik memungkinkan pengguna untuk melihat secara langsung aturan atau hubungan yang digunakan dalam proses prediksi. Dengan model white box, kita dapat mengidentifikasi fitur mana yang paling berpengaruh dan bagaimana mereka mempengaruhi output, sehingga memberikan kepercayaan lebih pada hasil yang dihasilkan oleh model tersebut.(Praniffa dkk., 2023)

# BAB III

# ANALISA DATA DAN PEMBAHASAN

## 3.1 Hasil dan Pembahasan

### 3.1.1. Flowchart



Gambar 3. 1 Flowchart

Flowchart tersebut menggambarkan proses klastering menggunakan algoritma K-Means dalam segmentasi pelanggan. Proses dimulai dari tahap "START" di mana langkah pertama yang dilakukan adialah mengumpulkan data yang digunakan. Kemudian, data yang sudah dikumpulkan kemudian diproses agar siap digunakan dalam algoritma K-Means. Proses ini mungkin termasuk pembersihan data, normalisasi, dan transformasi. Setelah data di proses, lalu kita **pemilihan jumlah cluster (K),** pada langkah ini, kita menentukan jumlah cluster (K) yang akan digunakan dalam algoritma K-Means. Pemilihan ini bisa berdasarkan pengalaman, metode elbow, atau metode lainnya. Selanjutnya, kita melakukan **inisialisasi centroid** yaitu kita menentukan posisi awal centroid untuk setiap cluster, yang di mana centroid adalah titik pusat dari cluster.

Setelah menentukan posisi awal centeroid, kita melakukan iterasi K-Means: Proses iterasi ini yang terdiri dari dua langkah utama yaitu Assignment Step dan Update Step, yang dilakukan berulang-ulang sampai konvergensi tercapai: assignment Step: Menghitung jarak setiap data ke centroid dan mengelompokkan data berdasarkan jarak terdekat ke centroid. Sedangkan Update Step: Mengupdate posisi centroid berdasarkan rata-rata posisi data yang berada dalam cluster tersebut. Kemudian kita melakukan **cek konvergensi** yaitu memeriksa apakah centroid sudah tidak berubah atau perubahan centroid sudah sangat kecil sehingga dianggap telah konvergen. Jika belum konvergen, proses kembali ke Iterasi K-Means. Setelah konvergensi tercapai, hasil segmentasi atau pengelompokan data sudah bisa diperoleh.

Kemudian setelah melakukan cek konvergensi, kita melakukan **interpretasi hasil** langkah ini melibatkan analisis hasil segmentasi untuk memahami karakteristik setiap cluster. Setelah interpretasi hasil, kita melakukan **visualisasi hasil** yaitu hasil segmentasi divisualisasikan agar lebih mudah dipahami dan dianalisis. Kemudian, **penerapan strategi bisnis** dengan berdasarkan hasil interpretasi dan visualisasi, strategi bisnis dapat diterapkan untuk masing-masing cluster. Proses berakhir dengan tahap "FINISH", menandakan bahwa semua langkah telah selesai dilaksanakan dengan baik dan sistem siap untuk digunakan kembali atau untuk dipersiapkan ulang jika diperlukan.

### 3.1.2 Dataset

Tabel 3. 1 Data customers

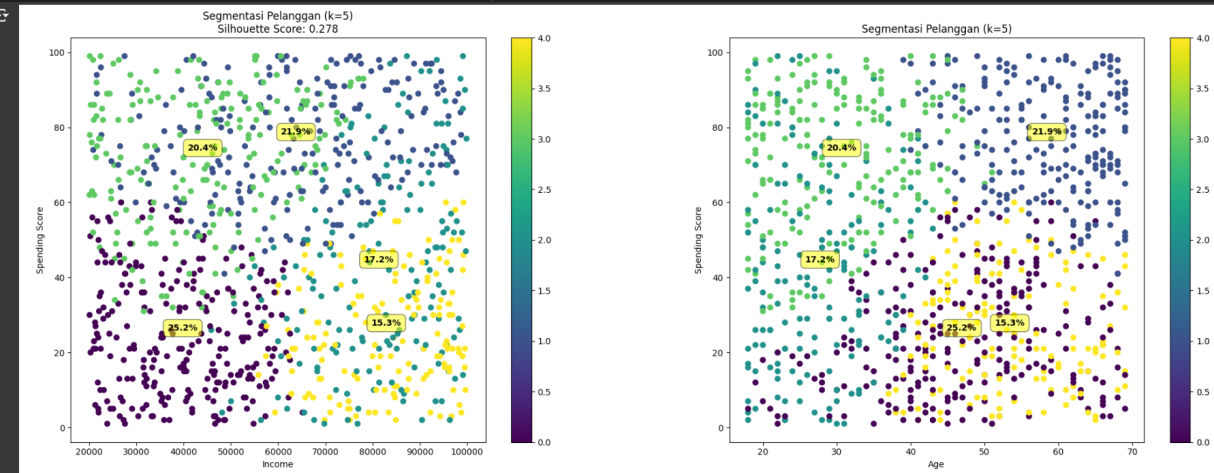
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **CustomerID** | **Genre** | **Age** | **Annual Income (k$)** | **Spending Score (1-100)** |
| 1 | Male | 19 | 15 | 39 |
| 2 | Male | 21 | 15 | 81 |
| 3 | Female | 20 | 16 | 6 |
| 4 | Female | 23 | 16 | 77 |
| 5 | Female | 31 | 17 | 40 |
| 6 | Female | 22 | 17 | 76 |
| 7 | Female | 35 | 18 | 6 |
| 8 | Female | 23 | 18 | 94 |
| 9 | Male | 64 | 19 | 3 |
| 10 | Female | 30 | 19 | 72 |
| 11 | Male | 67 | 19 | 14 |
| 12 | Female | 35 | 19 | 99 |
| 13 | Female | 58 | 20 | 15 |
| 14 | Female | 24 | 20 | 77 |
| 15 | Male | 37 | 20 | 13 |
| 16 | Male | 22 | 20 | 79 |
| 17 | Female | 35 | 21 | 35 |
| 18 | Male | 20 | 21 | 66 |
| 19 | Male | 52 | 23 | 29 |
| 20 | Female | 35 | 23 | 98 |
| 21 | Male | 35 | 24 | 35 |
| 22 | Male | 25 | 24 | 73 |
| 23 | Female | 46 | 25 | 5 |
| 24 | Male | 31 | 25 | 73 |
| 25 | Female | 54 | 28 | 14 |
| 26 | Male | 29 | 28 | 82 |
| 27 | Female | 45 | 28 | 32 |
| 28 | Male | 35 | 28 | 61 |
| 29 | Female | 40 | 29 | 31 |
| 30 | Female | 23 | 29 | 87 |
| 31 | Male | 60 | 30 | 4 |
| 32 | Female | 21 | 30 | 73 |
| 33 | Male | 53 | 33 | 4 |
| 34 | Male | 18 | 33 | 92 |
| 35 | Female | 49 | 33 | 14 |
| 36 | Female | 21 | 33 | 81 |
| 37 | Female | 42 | 34 | 17 |
| 38 | Female | 30 | 34 | 73 |
| 39 | Female | 36 | 37 | 26 |
| 40 | Female | 20 | 37 | 75 |
| 41 | Female | 65 | 38 | 35 |
| 42 | Male | 24 | 38 | 92 |
| 43 | Male | 48 | 39 | 36 |
| 44 | Female | 31 | 39 | 61 |
| 45 | Female | 49 | 39 | 28 |
| 46 | Female | 24 | 39 | 65 |
| 47 | Female | 50 | 40 | 55 |
| 48 | Female | 27 | 40 | 47 |
| 49 | Female | 29 | 40 | 42 |
| 50 | Female | 31 | 40 | 42 |
| 51 | Female | 49 | 42 | 52 |
| 52 | Male | 33 | 42 | 60 |
| 53 | Female | 31 | 43 | 54 |
| 54 | Male | 59 | 43 | 60 |
| 55 | Female | 50 | 43 | 45 |
| 56 | Male | 47 | 43 | 41 |
| 57 | Female | 51 | 44 | 50 |
| 58 | Male | 69 | 44 | 46 |
| 59 | Female | 27 | 46 | 51 |
| 60 | Male | 53 | 46 | 46 |
| 61 | Male | 70 | 46 | 56 |
| 62 | Male | 19 | 46 | 55 |
| 63 | Female | 67 | 47 | 52 |
| 64 | Female | 54 | 47 | 59 |
| 65 | Male | 63 | 48 | 51 |
| 66 | Male | 18 | 48 | 59 |
| 67 | Female | 43 | 48 | 50 |
| 68 | Female | 68 | 48 | 48 |
| 69 | Male | 19 | 48 | 59 |
| 70 | Female | 32 | 48 | 47 |
| 71 | Male | 70 | 49 | 55 |
| 72 | Female | 47 | 49 | 42 |
| 73 | Female | 60 | 50 | 49 |
| 74 | Female | 60 | 50 | 56 |
| 75 | Male | 59 | 54 | 47 |
| 76 | Male | 26 | 54 | 54 |
| 77 | Female | 45 | 54 | 53 |
| 78 | Male | 40 | 54 | 48 |
| 79 | Female | 23 | 54 | 52 |
| 80 | Female | 49 | 54 | 42 |
| 81 | Male | 57 | 54 | 51 |
| 82 | Male | 38 | 54 | 55 |
| 83 | Male | 67 | 54 | 41 |
| 84 | Female | 46 | 54 | 44 |
| 85 | Female | 21 | 54 | 57 |
| 86 | Male | 48 | 54 | 46 |
| 87 | Female | 55 | 57 | 58 |
| 88 | Female | 22 | 57 | 55 |
| 89 | Female | 34 | 58 | 60 |
| 90 | Female | 50 | 58 | 46 |
| 91 | Female | 68 | 59 | 55 |
| 92 | Male | 18 | 59 | 41 |
| 93 | Male | 48 | 60 | 49 |
| 94 | Female | 40 | 60 | 40 |
| 95 | Female | 32 | 60 | 42 |
| 96 | Male | 24 | 60 | 52 |
| 97 | Female | 47 | 60 | 47 |
| 98 | Female | 27 | 60 | 50 |
| 99 | Male | 48 | 61 | 42 |
| 100 | Male | 20 | 61 | 49 |
| 101 | Female | 23 | 62 | 41 |
| 102 | Female | 49 | 62 | 48 |
| 103 | Male | 67 | 62 | 59 |
| 104 | Male | 26 | 62 | 55 |
| 105 | Male | 49 | 62 | 56 |
| 106 | Female | 21 | 62 | 42 |
| 107 | Female | 66 | 63 | 50 |
| 108 | Male | 54 | 63 | 46 |
| 109 | Male | 68 | 63 | 43 |
| 110 | Male | 66 | 63 | 48 |
| 111 | Male | 65 | 63 | 52 |
| 112 | Female | 19 | 63 | 54 |
| 113 | Female | 38 | 64 | 42 |
| 114 | Male | 19 | 64 | 46 |
| 115 | Female | 18 | 65 | 48 |
| 116 | Female | 19 | 65 | 50 |
| 117 | Female | 63 | 65 | 43 |
| 118 | Female | 49 | 65 | 59 |
| 119 | Female | 51 | 67 | 43 |
| 120 | Female | 50 | 67 | 57 |
| 121 | Male | 27 | 67 | 56 |
| 122 | Female | 38 | 67 | 40 |
| 123 | Female | 40 | 69 | 58 |
| 124 | Male | 39 | 69 | 91 |
| 125 | Female | 23 | 70 | 29 |
| 126 | Female | 31 | 70 | 77 |
| 127 | Male | 43 | 71 | 35 |
| 128 | Male | 40 | 71 | 95 |
| 129 | Male | 59 | 71 | 11 |
| 130 | Male | 38 | 71 | 75 |
| 131 | Male | 47 | 71 | 9 |
| 132 | Male | 39 | 71 | 75 |
| 133 | Female | 25 | 72 | 34 |
| 134 | Female | 31 | 72 | 71 |
| 135 | Male | 20 | 73 | 5 |
| 136 | Female | 29 | 73 | 88 |
| 137 | Female | 44 | 73 | 7 |
| 138 | Male | 32 | 73 | 73 |
| 139 | Male | 19 | 74 | 10 |
| 140 | Female | 35 | 74 | 72 |
| 141 | Female | 57 | 75 | 5 |
| 142 | Male | 32 | 75 | 93 |
| 143 | Female | 28 | 76 | 40 |
| 144 | Female | 32 | 76 | 87 |
| 145 | Male | 25 | 77 | 12 |
| 146 | Male | 28 | 77 | 97 |
| 147 | Male | 48 | 77 | 36 |
| 148 | Female | 32 | 77 | 74 |
| 149 | Female | 34 | 78 | 22 |
| 150 | Male | 34 | 78 | 90 |
| 151 | Male | 43 | 78 | 17 |
| 152 | Male | 39 | 78 | 88 |
| 153 | Female | 44 | 78 | 20 |
| 154 | Female | 38 | 78 | 76 |
| 155 | Female | 47 | 78 | 16 |
| 156 | Female | 27 | 78 | 89 |
| 157 | Male | 37 | 78 | 1 |
| 158 | Female | 30 | 78 | 78 |
| 159 | Male | 34 | 78 | 1 |
| 160 | Female | 30 | 78 | 73 |
| 161 | Female | 56 | 79 | 35 |
| 162 | Female | 29 | 79 | 83 |
| 163 | Male | 19 | 81 | 5 |
| 164 | Female | 31 | 81 | 93 |
| 165 | Male | 50 | 85 | 26 |
| 166 | Female | 36 | 85 | 75 |
| 167 | Male | 42 | 86 | 20 |
| 168 | Female | 33 | 86 | 95 |
| 169 | Female | 36 | 87 | 27 |
| 170 | Male | 32 | 87 | 63 |
| 171 | Male | 40 | 87 | 13 |
| 172 | Male | 28 | 87 | 75 |
| 173 | Male | 36 | 87 | 10 |
| 174 | Male | 36 | 87 | 92 |
| 175 | Female | 52 | 88 | 13 |
| 176 | Female | 30 | 88 | 86 |
| 177 | Male | 58 | 88 | 15 |
| 178 | Male | 27 | 88 | 69 |
| 179 | Male | 59 | 93 | 14 |
| 180 | Male | 35 | 93 | 90 |
| 181 | Female | 37 | 97 | 32 |
| 182 | Female | 32 | 97 | 86 |
| 183 | Male | 46 | 98 | 15 |
| 184 | Female | 29 | 98 | 88 |
| 185 | Female | 41 | 99 | 39 |
| 186 | Male | 30 | 99 | 97 |
| 187 | Female | 54 | 101 | 24 |
| 188 | Male | 28 | 101 | 68 |
| 189 | Female | 41 | 103 | 17 |
| 190 | Female | 36 | 103 | 85 |
| 191 | Female | 34 | 103 | 23 |
| 192 | Female | 32 | 103 | 69 |
| 193 | Male | 33 | 113 | 8 |
| 194 | Female | 38 | 113 | 91 |
| 195 | Female | 47 | 120 | 16 |
| 196 | Female | 35 | 120 | 79 |
| 197 | Female | 45 | 126 | 28 |
| 198 | Male | 32 | 126 | 74 |
| 199 | Male | 32 | 137 | 18 |
| 200 | Male | 30 | 137 | 83 |

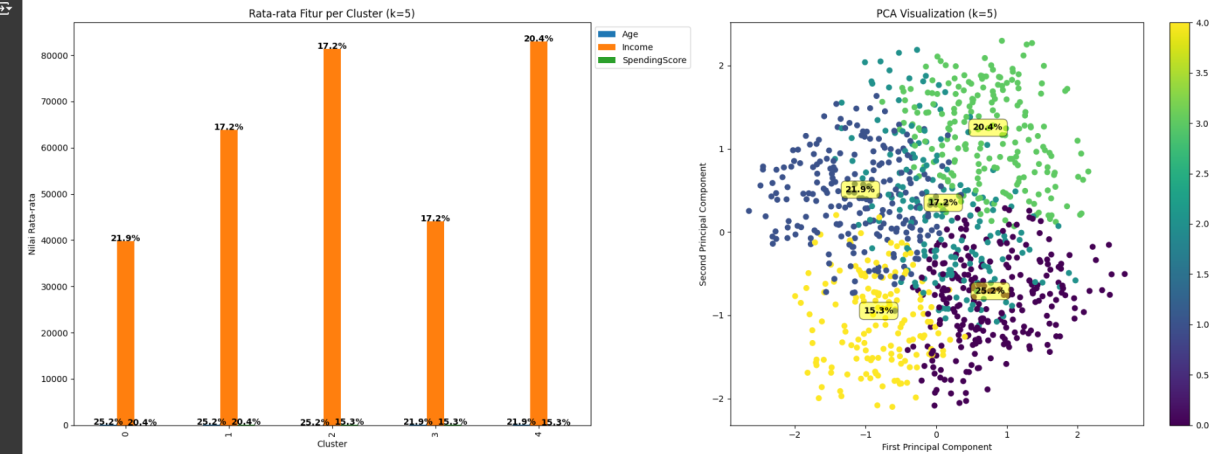
Jadi pada tabel data pelanggan diatas menjelaskan mengandung kolom-kolom berikut: **CustomerID**: Identifikasi unik untuk setiap pelanggan. **Genre**: Jenis kelamin pelanggan (Pria atau Wanita). **Age**: Usia pelanggan. **Annual Income (k$)**: Pendapatan tahunan pelanggan dalam ribuan dolar. **Spending Score (1-100)**: Skor yang diberikan kepada pelanggan berdasarkan perilaku belanja dan kebiasaan pembelian mereka, dengan rentang dari 1 hingga 100. Data ini tampaknya digunakan untuk beberapa bentuk segmentasi pelanggan atau analisis clustering, di mana pelanggan yang berbeda dikelompokkan berdasarkan karakteristik serupa seperti usia, pendapatan, dan skor belanja mereka.

Tujuan dilakukannya pengumpulan data tersebut untuk mengelompokkan pelanggan berdasarkan usia, pendapatan, dan skor belanja untuk mengidentifikasi segmen-segmen yang berbeda. Misalnya, pelanggan dengan pendapatan tinggi dan belanja tinggi, pendapatan rendah dan belanja rendah, dll. Seperti merancang strategi pemasaran yang disesuaikan dengan segmen pelanggan yang berbeda, menawarkan penawaran dan promosi yang dipersonalisasi berdasarkan profil pelanggan, menganalisis pola belanja untuk memahami perilaku pelanggan, mengidentifikasi tren dan memprediksi penjualan di masa depan, meningkatkan pengalaman pelanggan dengan memahami demografi dan kebiasaan belanja mereka, meningkatkan retensi pelanggan melalui keterlibatan yang ditargetkan, menggunakan wawasan dari dataset untuk membuat keputusan bisnis yang lebih baik, dan menyesuaikan harga, penawaran produk, dan penempatan toko berdasarkan data pelanggan.

### 3.1.3 Hasil Klasifikasi

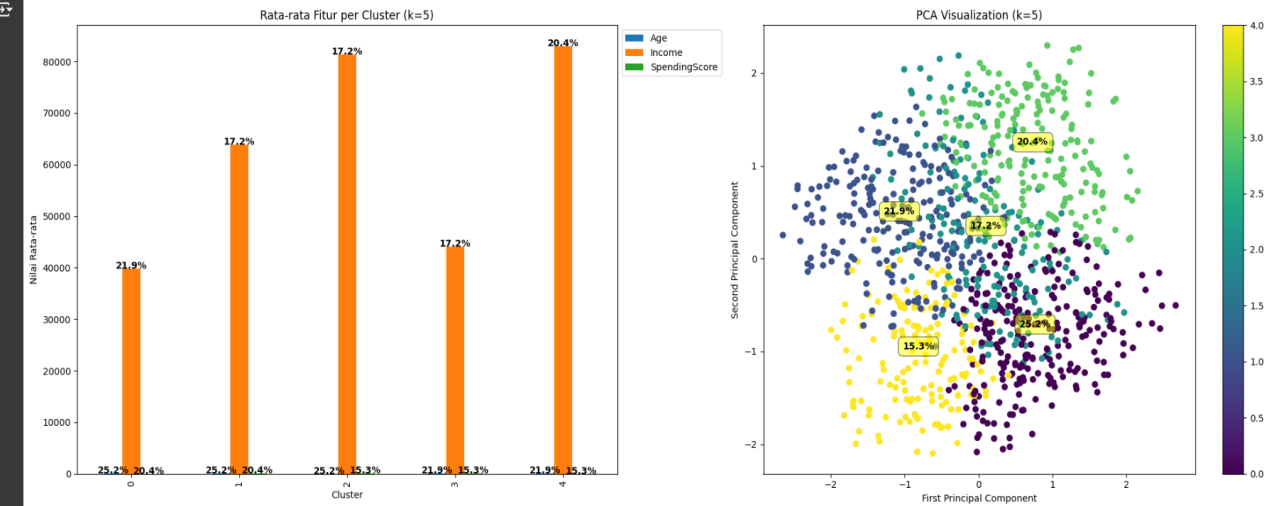
Pada bagian ini hasil klasifikasi di dapat dari kaggle, kemudian setelah mendapatkan data set kemudian di masukkan ke dalam google colab yang di mana google (atau “Colaboratory”) adalah lingkungan notebook berbasis awan yang mirip dengan Jupyter. Colab juga memungkinkan kita untuk menulis dan menjalankan kode Python langsung di browser tanpa perlu konfigurasi tambahan*.*

****

****

Gambar 3. 2 Klasifikasi Gender

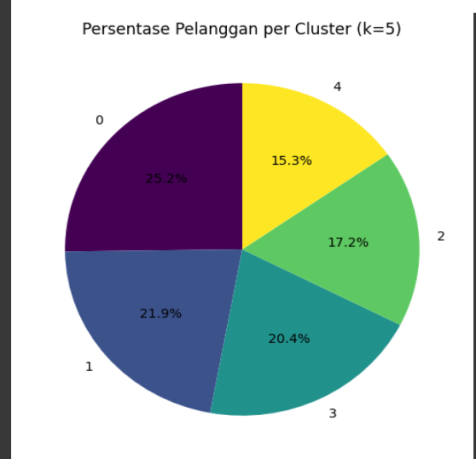
Gambar-gambar tersebut menampilkan hasil segmentasi pelanggan menggunakan teknik clustering dengan k=5 (5 cluster). Mari saya jelaskan secara rinci plot kiri, Income vs Spending Score :Sumbu X: Income (pendapatan) dari 20.000 hingga 100.000, sumbu Y: Spending Score (skor pengeluaran) dari 0 hingga 100, titik-titik berwarna menunjukkan cluster yang berbeda, persentase untuk setiap cluster ditampilkan: 20.4%, 21.9%, 17.2%, 15.3%, 25.2%. Plot kanan: Age vs Spending Score: sumbu X: Age (usia) dari 20 hingga 70, sumbu Y: Spending Score (skor pengeluaran) dari 0 hingga 100, distribusi cluster serupa dengan plot kiri. Kedua plot ini menunjukkan Silhouette Score 0.278, yang mengindikasikan kualitas clustering. Plot bawah: Ini menampilkan visualisasi PCA (Principal Component Analysis) dari data yang sama: sumbu X: First Principal Component, sumbu Y: Second Principal Component, titik-titik berwarna menunjukkan 5 cluster yang sama, persentase cluster ditampilkan: 20.4%, 21.9%, 17.2%, 15.3%, 25.2%. Analisis: cluster berwarna ungu (25.2%): Cenderung memiliki pendapatan dan skor pengeluaran rendah, usia bervariasi, cluster berwarna kuning (15.3%): Pendapatan tinggi tapi skor pengeluaran rendah, cenderung berusia lebih tua. cluster berwarna biru (21.9%): Pendapatan tinggi dan skor pengeluaran tinggi, usia bervariasi, cluster berwarna hijau (20.4%): pendapatan menengah ke rendah tapi skor pengeluaran tinggi, cenderung berusia lebih muda, cluster berwarna merah/oranye (17.2%): Karakteristik menengah untuk semua variabel. Visualisasi PCA menunjukkan pemisahan cluster yang cukup jelas, mengindikasikan bahwa segmentasi ini berhasil menangkap perbedaan karakteristik antar kelompok pelanggan.



Gambar 3. 3 Diagram Batang Nilai Rata – Rata Per Cluster

Gambar tersebut menampilkan grafik batang yang berjudul "Rata-rata Fitur per Cluster (k=5)". Grafik ini menunjukkan rata-rata nilai fitur untuk setiap cluster dalam analisis segmentasi pelanggan dengan 5 cluster. Mari saya jelaskan secara rinci: sumbu X: Menunjukkan 5 cluster yang berbeda, diberi label dari 0 hingga 4. Sumbu Y: Menunjukkan nilai rata-rata fitur, dengan skala dari 0 hingga 80000. Ini kemungkinan merepresentasikan pendapatan rata-rata atau metrik keuangan lainnya. Batang: Setiap cluster memiliki satu batang oranye yang menunjukkan nilai rata-rata fiturnya. Persentase: Di atas setiap batang, terdapat persentase yang menunjukkan proporsi pelanggan dalam cluster tersebut. Detail per cluster: Cluster 0: 25.2% pelanggan, nilai rata-rata sekitar 40000, cluster 1: 21.9% pelanggan, nilai rata-rata sekitar 60000, cluster 2: 17.2% pelanggan, nilai rata-rata tertinggi, sekitar 80000, cluster 3: 15.3% pelanggan, nilai rata-rata terendah, di bawah 20000, cluster 4: 20.4% pelanggan, nilai rata-rata tertinggi kedua, mendekati 80000. Interpretasi: cluster 2 dan 4 memiliki nilai rata-rata tertinggi, mungkin merepresentasikan pelanggan dengan pendapatan atau pengeluaran tinggi, cluster 3 memiliki nilai terendah, mungkin pelanggan dengan pendapatan atau pengeluaran rendah. cluster 0 dan 1 berada di tengah-tengah, mungkin pelanggan dengan nilai menengah, distribusi pelanggan: cukup merata di antara cluster, dengan rentang dari 15.3% hingga 25.2%, menunjukkan segmentasi yang relatif seimbang.

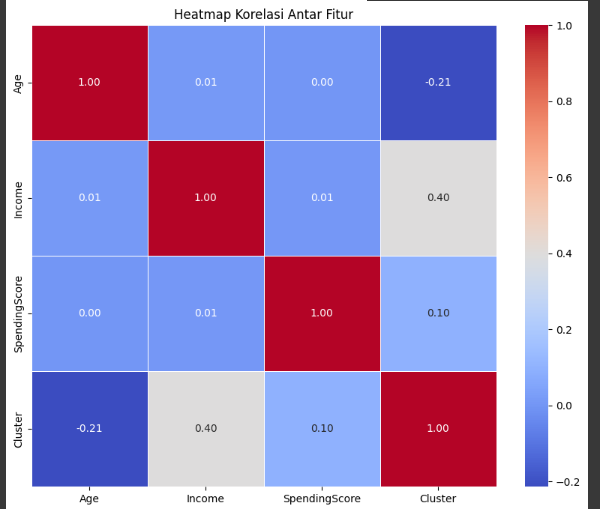
Grafik ini memberikan gambaran cepat tentang karakteristik masing-masing cluster dalam hal nilai rata-rata fitur tertentu, membantu dalam memahami perbedaan antar segmen pelanggan yang telah diidentifikasi melalui analisis clustering.



Gambar 3. 4 Diagram Lingkaran presentase pelanggan

Gambar tersebut menampilkan diagram lingkaran (pie chart) yang berjudul "Persentase Pelanggan per Cluster (k=5)". Diagram ini menggambarkan distribusi pelanggan ke dalam 5 cluster yang berbeda. Mari saya jelaskan secara rinci: Cluster 0 (Ungu): Persentase: 25.2% dan kelompok terbesar dari semua cluster. Cluster 1 (Biru Tua): Persentase: 21.9% dan kelompok terbesar kedua. Cluster 2 (Hijau): Persentase: 17.2% dan kelompok menengah. Cluster 3 (Biru Muda): Persentase: 20.4% dan kelompok terbesar ketiga. Cluster 4 (Kuning): Persentase: 15.3% dan kelompok terkecil dari semua cluster. Analisis: Distribusi pelanggan cukup merata di antara kelima cluster, dengan perbedaan sekitar 10% antara cluster terbesar dan terkecil Tidak ada cluster yang sangat dominan, menunjukkan segmentasi yang relatif seimbang. Cluster 0 dan 1 bersama-sama mewakili hampir setengah dari total pelanggan (47.1%). Tiga cluster lainnya (2, 3, dan 4) memiliki persentase yang lebih kecil tetapi masih signifikan.

Kesimpulan: Diagram ini memberikan gambaran visual yang jelas tentang bagaimana pelanggan terdistribusi di antara lima segmen yang berbeda. Informasi ini berguna untuk memahami komposisi basis pelanggan dan dapat membantu dalam pengambilan keputusan terkait strategi pemasaran atau layanan pelanggan yang ditargetkan untuk masing-masing segmen.



Gambar 3. 5 Heatmap Korelasi Antar Fitur

Gambar tersebut menampilkan sebuah heatmap yang berjudul "Heatmap Korelasi Antar Fitur". Heatmap ini menunjukkan korelasi antara berbagai fitur yang digunakan dalam analisis segmentasi pelanggan. Mari saya jelaskan secara rinci: Fitur-fitur yang ditampilkan: Age (Usia), Income (Pendapatan), SpendingScore (Skor Pengeluaran), Cluster (Kelompok hasil segmentasi). Skala warna: Merah tua menunjukkan korelasi positif kuat (mendekati 1). Biru tua menunjukkan korelasi negatif (mendekati -1). Warna putih atau abu-abu muda menunjukkan korelasi lemah atau tidak ada (mendekati 0). Interpretasi korelasi: a. Age (Usia): Berkorelasi sempurna dengan dirinya sendiri (1.00). Korelasi sangat lemah dengan Income (0.01) dan SpendingScore (0.00). Korelasi negatif lemah dengan Cluster (-0.21). Income (Pendapatan): Berkorelasi sempurna dengan dirinya sendiri (1.00). Korelasi sangat lemah dengan Age (0.01) dan SpendingScore (0.01). Korelasi positif moderat dengan Cluster (0.40). SpendingScore (Skor Pengeluaran): Berkorelasi sempurna dengan dirinya sendiri (1.00), Korelasi sangat lemah dengan Age (0.00) dan Income (0.01), Korelasi positif lemah dengan Cluster (0.10). Cluster: Berkorelasi sempurna dengan dirinya sendiri (1.00), Korelasi negatif lemah dengan Age (-0.21), Korelasi positif moderat dengan Income (0.40), Korelasi positif lemah dengan SpendingScore (0.10).

Kesimpulan: Tidak ada korelasi kuat antar fitur selain dengan diri sendiri, Income memiliki korelasi terkuat dengan Cluster, menunjukkan bahwa pendapatan mungkin menjadi faktor penting dalam pengelompokan, Age memiliki korelasi negatif lemah dengan Cluster, mengindikasikan bahwa cluster cenderung sedikit berbeda berdasarkan usia, SpendingScore memiliki korelasi paling lemah dengan fitur lainnya. Heatmap ini memberikan gambaran visual yang jelas tentang hubungan antar fitur dalam dataset, membantu dalam memahami faktor-faktor yang memengaruhi segmentasi pelanggan dan potensi interaksi antar variabel.

### 3.1.4 Studi Kasus

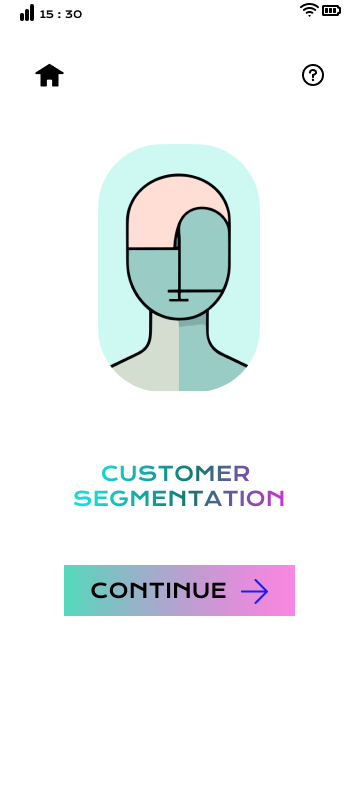
Perusahaan XYZ adalah sebuah perusahaan e-commerce yang menjual berbagai produk mulai dari pakaian, elektronik, hingga barang-barang rumah tangga. Seiring dengan pertumbuhan bisnis, XYZ ingin lebih memahami pelanggannya agar dapat menawarkan produk dan promosi yang lebih relevan, serta meningkatkan retensi pelanggan. Dengan tujuan untuk Tujuan mengidentifikasi segmen-segmen pelanggan yang berbeda berdasarkan perilaku pembelian dan karakteristik demografis, mengembangkan strategi pemasaran yang ditargetkan untuk setiap segmen, dan meningkatkan kepuasan dan retensi pelanggan melalui penawaran yang lebih personal. Untuk melakukan segmentasi pelanggan, XYZ menggunakan pendekatan berbasis data yang melibatkan analisis data transaksi pelanggan dan informasi demografis. **Pengumpulan Data:** Data transaksi: jumlah pembelian, frekuensi pembelian, nilai total pembelian, kategori produk yang dibeli dan data demografis: usia, jenis kelamin, lokasi geografis. **Praproses Data:** Pembersihan data untuk mengatasi duplikasi dan data yang hilang dam normalisasi data untuk memastikan skala yang konsisten. **Analisis Segmentasi:** Menggunakan algoritma K-means clustering untuk mengelompokkan pelanggan berdasarkan perilaku pembelian dan karakteristik demografisdan menentukan jumlah klaster yang optimal menggunakan metode k-means clustering. **Profiling Segmen:** Mengidentifikasi karakteristik utama dari setiap segmen dan menamakan segmen berdasarkan karakteristik yang menonjol. Hasil Setelah melakukan segmentasi, XYZ berhasil mengidentifikasi empat segmen utama pelanggan: **Segmen A: Pembeli Setia** Ciri-ciri: Frekuensi pembelian tinggi, nilai pembelian besar, sering membeli produk-produk premium, demografi: Usia 30-45 tahun, pendapatan menengah ke atas, strategi: menawarkan program loyalitas eksklusif dan diskon khusus untuk produk premium. **Segmen B: Pemburu Diskon** ciri-ciri: Pembelian terjadi saat ada promosi atau diskon besar, nilai pembelian lebih rendah, demografi: Usia 20-35 tahun, pendapatan menengah, strategi: Mengirimkan informasi tentang promosi dan diskon melalui email dan aplikasi mobile. **Segmen C: Pembeli Kasual** Ciri-ciri: Frekuensi pembelian rendah, nilai pembelian sedang, membeli berbagai kategori produk, demografi: Usia 25-40 tahun, pendapatan menengah, strategi: Mengirimkan rekomendasi produk yang dipersonalisasi dan penawaran bundling. **Segmen D: Pembeli Satu Kali** Ciri-ciri: Hanya membeli sekali atau dua kali, nilai pembelian variatif, demografi: Usia bervariasi, pendapatan bervariasi, strategi: Menawarkan insentif untuk pembelian ulang, seperti voucher diskon atau gratis ongkos kirim.

Setelah enam bulan implementasi, XYZ melakukan evaluasi terhadap hasil strategi segmentasi. Mereka menemukan peningkatan signifikan dalam retensi pelanggan, kepuasan pelanggan yang lebih tinggi, dan peningkatan rata-rata nilai pembelian per pelanggan. Segmentasi pelanggan telah membantu perusahaan e-commerce XYZ untuk lebih memahami pelanggannya dan mengembangkan strategi pemasaran yang lebih efektif. Dengan pendekatan yang lebih terfokus, XYZ berhasil meningkatkan loyalitas pelanggan dan pendapatan perusahaan. Segmentasi pelanggan adalah alat yang sangat efektif bagi perusahaan untuk meningkatkan efisiensi pemasaran dan pengalaman pelanggan. Studi kasus ini menunjukkan bagaimana analisis data yang tepat dapat memberikan wawasan berharga dan menghasilkan hasil bisnis yang positif.

## 3.2 Implementasi Interface

Dibawah ini merupakan tampilan implementasi *interface* menu yang ada pada aplikasi segmentasi customer yang dapat di lihat pada gambar di bawah ini :

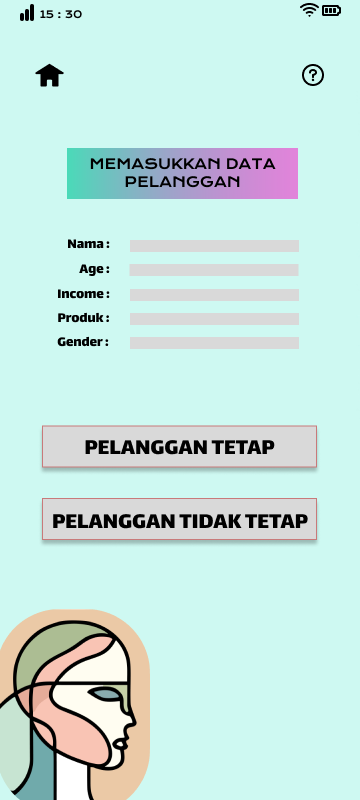
### 3.2.1 Home Page



Gambar 3. 6 Home Page

Pada bagian atas layar, terdapat representasi grafis berbentuk kepala manusia yang terbagi menjadi empat segmen oleh dua garis yang saling berpotongan, membentuk desain seperti bidikan di atas wajah. Di bawah grafik ini, terdapat teks yang berbunyi “CUSTOMER SEGMENTATION” dalam huruf kapital tebal. Di bawah teks tersebut, ada tombol bertuliskan “CONTINUE” dengan panah mengarah ke kanan, menandakan kemungkinan interaktif dan akan membawa pengguna ke layar atau fungsi lain dalam aplikasi jika ditekanSegmentasi pelanggan adalah strategi pemasaran yang melibatkan pembagian pelanggan ke dalam kelompok berdasarkan karakteristik bersama, sehingga bisnis dapat memasarkan kepada setiap kelompok dengan efektif dan sesuai..

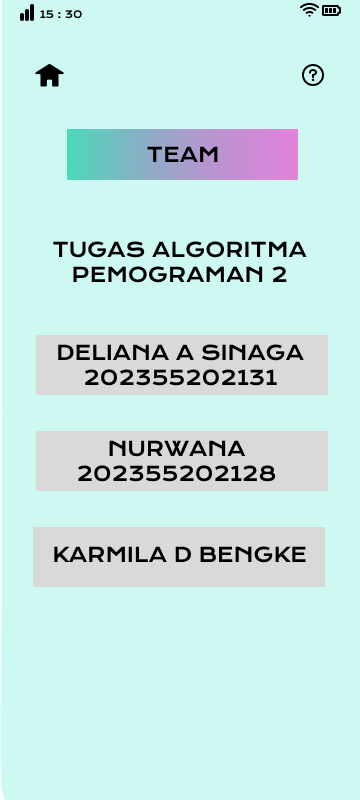
### 3.2.2 Penginputan Data

****Gambar 3. 7 Penginputan Data

Di bagian atas, terdapat judul dengan teks “ DATA PELANGGAN”. Di samping kanan judul ini, ada ikon yang dimana Ikon ini menyerupai profil pengguna atau orang secara sederhana. Dan di bagian pojok kiri ada terdapat ikon yang menyerupai bentuk rumah yang bertujuan untuk kembali ke halama utama, kemudian di atas pojok kiri terdapat ikon tanda Tanya yang bertujuan untuk menampilkan about page dari aplikasi tersebut.

Terdapat beberapa kolom input data: “NAMA”: Kolom untuk memasukkan nama pelanggan. “INCOME”, yang dimana ini bertujuan memasukkan seberapa besar income atau pendapatan yang didapat oleh pelanggan setiap bulannya. “AGE”: Kolom untuk memasukkan usia pelanggan. “GENDER”: Kolom untuk memasukkan jenis kelamin pelanggan. “SKOR”: Kolom kosong, untuk memasukkan skor pelanggan, pada bagian skor ini adalah kita mengisi seberapa besar keseringan oleh penjual dalam melakukan pembelian di dalam usaha kita dengan range 1-100. Di bagian bawah, terdapat teks instruksional yang berbunyi “INPUT DATA”. Ini mengarahkan pengguna untuk memasukkan data sesuai dengan format yang telah ditentukan dan jika sudah maka silahkan menekan tombol input data tersebut. Kemudian di bagian pojok kanan bawah terdapat ikon yang di bagian atas berbentuk + untuk menambahkan data pelanggan yang baru dan juga ikon dengan garis tiga bertujuan untuk melihat semua data pelanggan yang telah kita tambahkan.

### 3.2.3 About Page



Gambar 3. 8 About Page

Jadi pada bagian About Page ini berisi tentang user dapat melihat atau *credit* dari aplikasi yang dimana disini tertera nama dan NIM dari pembuat project tugas besar laporan ini, pada menu dan halaman ini tidak ada bagian atau *interface* yang dapat interaksi langsung Bersama user dan user hanya cukup melihat saja.

## 3.3 Pengujian

Tabel 3. 2 Pengujian Pada User

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **ID** | **Deskripsi Pengujian** | **Test Case** | **Hasil yang Diharapkan** | **Hasil Pengujian** | **Kesimpulan** |
| T01 | Menekan tombol Continue | *Mulai* | Menampilkan halaman untuk memasukkan data pelanggan | Sesuai Harapan | [] *Valid*  [] *Invalid* |
| T02 | Menekan tombol tanda tanya (About) | About | Menampilkan halaman pembuat aplikasi | Sesuai harapan | [] *Valid*  [] *Invalid* |
| T03 | Menekan pada bagian nama | Data Input | Memasukkan nama pelanggan | Sesuai Harapan | [] *Valid*  [] *Invalid* |
| T04 | Menekan pada bagian umur | Data Input | Memasukkan umur pelanggan | Sesuai Harapan | [] *Valid*  [] *Invalid* |
| T05 | Menekan pada bagian gender | Data input | Memasukkan jenis kelamin pelanggan | Sesuai Harapan | [] *Valid*  [] *Invalid* |
| T06 | Menekan pada bagian income | Data input | Memasukkan pendapatan pelanggan | Sesuai Harapan | [] *Valid*  [] *Invalid* |
| T07 | Menekan pada bagian skor | Data Input | Memasukkan skor pelanggan | Sesuai Harapan | [] *Valid*  [] *Invalid* |
| T07 | Melihat hasil inputan data | Output klasifikasi | Menampilkan hasil segementasi berdasarkan data yang telah dimasukkan | Sesuai Harapan | [] *Valid*  [] *Invalid* |

**3.4 Usability Testing**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| No | Tugas | Tingkat Kesulitan (1-5) | Waktu  Penyelesaian | Tingkat Keberhasilan  (1-5) | Komentar |
| 1 | Mengakses halaman utama melalui tombol Home | 1 | 1 detik | 5 | Sangat responsif |
| 2 | Menemukan dan memahami informasi singkat tentang aplikasi di halaman utama | 1 | 1 detik | 5 | Sangat respoinsif |
| 3 | Mengakses fitur input data untuk memasukkan informasi pelanggan | 2 | 7 detik | 5 | Sangat responsif |
| 4 | Memasukkan data pelanggan | 2 | 15 detik | 5 | Sangat responsif |
| 5 | Memahami hasil segementasi dari pelanggan, seperti pelanggan tetap atau pelanggan tidak tetap | 1 | 1 detik | 5 | Sangat responsif |
| 6 | Navigasi kembali ke halaman utama setelah segmentasi | 1 | 1 detik | 5 | Sangat responsif |

# BAB IV

# PENUTUP

### 4.1 Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan dapat disimpulkan beberapa hal sebagai berikut bahwa segmentasi pelanggan menggunakan K-means clustering adalah teknik yang populer dalam analisis data untuk mengelompokkan pelanggan berdasarkan karakteristik tertentu. K-means clustering adalah algoritma pembelajaran tanpa pengawasan yang digunakan untuk mengelompokkan data ke dalam sejumlah k kelompok berdasarkan kesamaan karakteristik. Dalam konteks segmentasi pelanggan, tujuan utamanya adalah mengidentifikasi kelompok-kelompok pelanggan yang memiliki karakteristik serupa sehingga perusahaan dapat memahami kebutuhan dan preferensi setiap kelompok dengan lebih baik. Dengan mengelompokkan pelanggan, perusahaan dapat membuat strategi pemasaran yang lebih tepat sasaran untuk setiap segmen, meningkatkan efektivitas kampanye. K-means clustering adalah alat yang sangat berguna dalam segmentasi pelanggan, memungkinkan perusahaan untuk memahami dan melayani pelanggan mereka dengan lebih efektif. Namun, seperti semua alat analisis, penggunaannya memerlukan pemahaman yang baik tentang data dan metodologi yang digunakan untuk menghasilkan hasil yang bermakna dan dapat ditindaklanjuti.

### 4.2 Saran

Beberapa saran yang dapat diberikan untuk pengembangan lebih lanjut aplikasi segmentasi pelanggan menggunakan k-means clustering : Pastikan data dinormalisasi sehingga setiap fitur memiliki skala yang sama, menghindari dominasi fitur dengan skala yang lebih besar. Hilangkan data yang tidak relevan atau duplikasi, serta tangani data yang hilang (missing data) dengan cara yang tepat, seperti imputasi atau penghapusan. Pilih fitur yang benar-benar relevan dengan tujuan segmentasi. Misalnya, jika ingin mengelompokkan pelanggan berdasarkan perilaku belanja, pilih fitur seperti frekuensi pembelian, jumlah pengeluaran, dan jenis produk yang dibeli. Lakukan clustering secara berkala karena perilaku dan preferensi pelanggan dapat berubah seiring waktu. Perbarui model dengan data terbaru untuk menjaga relevansi segmentasi. Terus eksperimen dengan fitur dan metode clustering yang berbeda untuk menemukan pendekatan yang paling efektif.

Dengan mengikuti saran-saran ini, Anda dapat meningkatkan akurasi dan efektivitas segmentasi pelanggan menggunakan K-means clustering, yang pada akhirnya dapat membantu dalam merancang strategi bisnis yang lebih baik dan meningkatkan kepuasan serta loyalitas pelanggan.

# DAFTAR PUSTAKA

A. Wibowo and A. R. Handoko. (2020). Penerapan Data Mining Untuk Menentukan Segmentasi Pelanggan Dengan Menggunakan Algoritma K-Means dan Model RFM Pada E-Commerce. *JURIKOM (Jurnal Riset Komputer)*, *7*(3), 866. https://doi.org/10.25126/jtiik.2020702925.

Abiodun M. Ikotun, M. S. A. & A. E. E. (2021). K-Means-Based Nature-Inspired Metaheuristic Algorithms for Automatic Data Clustering Problems: Recent Advances and Future Directions. *Applied Sciences*, *11*(23). https://doi.org/https://doi.org/10.3390/app112311246

Ahmed, M., & Raihan Seraj and Syed Mohammed Shamsul Islam. (2020). The k-means Algorithm: A Comprehensive Survey and Performance Evaluation. *MDPI*, *9*(8). https://doi.org/https://doi.org/10.3390/electronics9081295

Aswin Kumer S V & Dr E Mohan. (2020). *An Implementation of K-Means Clustering for Efficient Image Segmentation of Natural Background Images*.

Azad Abdulhafedh. (2021). Incorporating K-means, Hierarchical Clustering and PCA in Customer Segmentation. *Journal of City and Development*, *3*(1). https://doi.org/10.12691/jcd-3-1-3

Carudin. (2023). Optimalisasi Strategi Pemasaran dengan Segmentasi Pelanggan Menggunakan Penerapan K-Means Clustering pada Transaksi Online Retail. *Jurnal Teknologi Dan Informasi (JATI)*, *13*(2), 124. https://doi.org/https://doi.org/10.34010/jati.v13i2.10090

Chunqiong Wu dkk. (2021). k-Means Clustering Algorithm and Its Simulation Based on Distributed Computing Platform. *Hindawi*, 1. https://doi.org/https://doi.org/10.1155/2021/9446653

E. Ernawati, S. S. K. Baharin, and F. K. (2021). A review of data mining methods in RFM-based customer segmentation. *Journal of Physics. Conf. Ser.*, *1869*(1). https://doi.org/10.1088/1742-6596/1869/1/012085.

Ermanto. (2020). PENERAPAN DATA MINING MENGGUNAKAN METODE K-MEANS UNTUK PENENTUAN REWARD PELANGGAN. *METHOMIKA: Jurnal Manajemen Informatika & Komputerisasi Akuntansi*, *7*(2), 202. https://doi.org/https://doi.org/10.46880/jmika.Vol7No2.pp201-207

Febriani & Putri. (2020). ALGORITMA K-MEANS CLUSTERING UNTUK SEGMENTASI PELANGGAN (STUDI KASUS : FASHION VIRAL SOLO). *Infotech Journal*. https://doi.org/https://doi.org/10.31949/infotech.v9i1.5214

Febriani, A., & Putri, S. A. (2020). Segmentasi Konsumen Berdasarkan Model Recency, Frequency, Monetary dengan Metode K-Means. *JIEMS (Journal of Industrial Engineering and Management Systems)*, *13*(2), 125. https://doi.org/10.30813/jiems.v13i2.2274

Fitriyani & Hemdriyani. (2021). algoritma k-means clustering untuk segmentasi pelanggan (studi kasus : fashion viral solo). *Infotech Journal*, *9*(1), 125. https://doi.org/https://doi.org/10.33050/icit.v7i2.1645

Fitriyani & Hendriyani. (2021). algoritma k-means clustering untuk segmentasi pelanggan (studi kasus : fashion viral solo). *Infotech Journal*, *9*(1), 125. https://doi.org/https://doi.org/10.33050/icit.v7i2.1645

Hadi, S. W., Julianto, M. F., Rahmatullah, S., & Gata, W. (2020). Analisa Cluster Aplikasi Pada App Store Dengan Menggunakan Metode K-Means. *Bianglala Informatika*, *8*(2), 86–90. https://doi.org/10.31294/bi.v8i2.8191

J. Wu, L. Shi, W. Lin dkk. (2020). Retracted: An Empirical Study on Customer Segmentation by Purchase Behaviors Using a RFM Model and K-Means Algorithm. *Hindawi*. https://doi.org/https://doi.org/10.1155/2020/8884227

John, J. M., Shobayo, O., & Ogunleye, B. (2023). An Exploration of Clustering Algorithms for Customer Segmentation in the UK Retail Market. *Analytics*, *2*(4), 809–823. https://doi.org/10.3390/analytics2040042

Kayalvily Tabianan, S. V. & V. R. (2022). K-Means Clustering Approach for Intelligent Customer Segmentation Using Customer Purchase Behavior Data. *Sustainability*, *14*(12). https://doi.org/https://doi.org/10.3390/su14127243

Kristina P. Sinaga & Miin-Shenyang. (2020). Unsupervised K-Means Clustering Algorithm. *IEEE Accsess*. https://doi.org/10.1109/ACCESS.2020.2988796

Mousavi, B. & A. (2020). Iplementasi Algoritma K-Medoids Clustering Untuk Menentukan Segmentasi Pelanggan. *Jurnal Nuansa Informatika*, *17*(1), 197. https://doi.org/https://doi.org/10.25134/nuansa

Normah, Rifai, & S. (2020). Segmentasi Pelanggan Menggunakan K-means Clustering Studi Kasus Pelanggan UHT Milk Greenfield. *Cerdika : Jurnal Ilmiah Indonesia*, *3*(7), 631. https://doi.org/https://doi.org/10.59141/cerdika.v3i7.639

Praniffa, A. C., Syahri, A., Sandes, F., Fariha, U., Giansyah, Q. A., & Hamzah, M. L. (2023). Pengujian Black Box Dan White Box Sistem Informasi Parkir Berbasis Web Black Box and White Box Testing of Web-Based Parking Information System. *Jurnal Testing Dan Implementasi Sistem Informasi*, *1*(1), 1–16.

Safitri Juanita & Raynaldi Dwi Cahyono. (2024). K-MEANS CLUSTERING WITH COMPARISON OF ELBOW AND SILHOUETTE METHODS FOR MEDICINES CLUSTERING BASED ON USER REVIEWS. *Jurnal Teknik Informatika (JUTIF)*, *5*(1), 283. https://doi.org/https://doi.org/10.52436/1.jutif.2024.5.1.1349

Solichin, A., & Khairunnisa, K. (2020). SEGMENTASI PELANGGAN PADA PT. TAB MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING. *JIKI (JURNAL ILMU KOMPUTER DAN INFORMATIKA)*, *4*(2), 138.

Sondang Sibueaa, Mohammad Ikhsan Saputrob, A. A., & Widodod, Y. B. (2022). APLIKASI MOBILE COLLECTION BERBASIS ANDROID PADA PT. SUZUKI FINANCE INDONESIA. *JURNAL INFORMATIKA DAN TEKNOLOGI KOMPUTER*. https://journal.amikveteran.ac.id/index.php

Sri Wahyuni & Fahrullah. (2023). SEGMENTASI PELANGGAN BERDASARKAN ANALISIS RECENCY, FREQUENCY, MONETARY MENGGUNAKAN ALGORITMA K-MEANS. *Jurnal Simantec*, *12*(1), 29. https://journal.trunojoyo.ac.id/simantec/article/view/23164/9510

Sudarsono dkk. (2021). algoritma k-means clustering untuk segmentasi pelanggan (studi kasus : fashion viral solo). *Infotech Journal*, *9*(1), 125. https://doi.org/10.30813/jbase.v4i1.2729

Suharto, A. (2023). Fundamental Bahasa Pemrograman Python. *Eureka Media Aksara*, 1–25.

Taher M. Ghazal et dkk. (2021). Performances of K-Means Clustering Algorithm with Different Distance Metrics. *Intelligent Automation & Soft Computing*. https://doi.org/10.32604/iasc.2021.019067

Wahyu Sudrajat, dkk. (2022). Penerapan Algoritma K-Means Dalam Segmentasi Pelanggan Pada Toko Sembako Menggunakan Rapidminer. *Jurnal Ilmiah Binar*, *3*(2), 54. https://doi.org/https://doi.org/10.52303/jb.v3i2.57

Z Nabila, A. Rahman Isnain, Z. A. (2021). PENERAPANDATA MININGPADA PENJUALAN PRODUK PAKAIAN DAMEYRA FASHIONMENGGUNAKAN METODE K-MEANS CLUSTERING. *JATI(Jurnal Mahasiswa Teknik Informatika)*, *7*(1). https://doi.org/https://doi.org/10.36040/jati.v7i1.6336

Zainab Mohammed Ali, dkk. (2020). The Application of Data Mining for Predicting Academic Performance Using K-means Clustering and Naïve Bayes Classification. *International Journal of Psychosocial Rehabilitation*, *24*(03). https://doi.org/10.37200/IJPR/V24I3/PR200962

Zalukhu Agustinus, S. P. & D. D. (2023). PERANGKAT LUNAK APLIKASI PEMBELAJARAN FLOWCHART. *Jurnal Teknologi Informasi Dan Industri*, *4*(1).

# Lampiran 1 Evaluasi Pengerjaan Tugas Besar

Kelompok 11 :

1. Deliana A. Sinaga : BAB 1, BAB 2, BAB 3, BAB 4 (laporan full) desain UI/UX
2. Nurwana : Cover, Flowchart, Cari Jurnal
3. Karmila D. Bengke : Cari jurnal

Mengetahui Dosen Pengganti Mata Kuliah

Mata Kuliah Algoritma Pemrograman 2

FAJAR R. B PUTRA, S.Kom., M.Kom.

# Lampiran 2 Dokumentasi



# Lampiran 3 Form Pengisian Tugas Besar

**FORM PENGISIAN PENGERJAAN TUGAS BESAR**

**Jenis Tugas :** segmentasi pelanggan menggunakan k-means clusterin

|  |  |  |  |
| --- | --- | --- | --- |
| No | Hari/Tanggal | Kegiatan | Paraf |
| 1 | 21/04/2024 | Mencari dataset |  |
| 2 | 24/04/2024 | Membuat BAB 1 |  |
| 3 | 27/04/2024 | Membuat UI/UX |  |
| 4 | 28/04/2024 | Develop Aplikasi |  |
| 5 | 28/04/2024 | Membuat BAB 2 |  |
| 6 | 30/05/2024 | Membuat BAB 3 |  |
| 7 | 03/06/2024 | Membuat BAB 4 |  |