****

**MENERAPKAN SEGMENTASI PELANGGAN MENGGUNAKAN METODE K-MEANS CLUSTERING**



**Disusun Oleh:**

**KELOMPOK 2**

**PROGRAM STUDI INFORMATIKA**

**FAKULTAS TEKNIK**

**UNIVERSITAS MUHAMMDIYAH SORONG**

**TAHUN 2024**

# LEMBAR PERSETUJUAN

**MENERAPKAN 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 2**

****

|  |  |
| --- | --- |
| **Menyetujui dan Mengetahui**  **Dosen Pengganti Mata Kuliah**  **Fajar R. B Putra, S.Kom., M.Kom.**  **NIDN. 1428099501** | **Sorong, 17 Juli 2024**  **Menyetujui**  **Ketua Kelompok 2**  **Fei Hung M Makarapate**  **NIM. 202355202042** |

# KATA PENGANTAR

Puji syukur kepada Tuhan Yang Maha Esa atas berkat, rahmat, dan karunia-Nya sehingga penulis dapat menyelesaikan Tugas Besar dengan judul yang dimaksud. “Menerapkan Segmentasi Pelanggan Menggunakan Metode K-Means Clustering”. Tugas Besar ini merupakan persyaratan untuk mendapatkan nilai UTS dan UAS dalam Mata Kuliah Algoritma dan Pemrograman 2 di 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 2

# DAFTAR ISI

[LEMBAR PERSETUJUAN ii](#_Toc172277096)

[KATA PENGANTAR iii](#_Toc172277097)

[DAFTAR ISI iv](#_Toc172277098)

[DAFTAR TABEL vi](#_Toc172277099)

[DAFTAR GAMBAR vii](#_Toc172277100)

[BAB I PENDAHULUAN 1](#_Toc172277101)

[1.1 Latar Belakang 1](#_Toc172277102)

[1.2 Rumusan Masalah 3](#_Toc172277103)

[1.3 Tujuan 3](#_Toc172277104)

[1.4 Batasan Masalah 3](#_Toc172277105)

[BAB II KAJIAN TEORI 5](#_Toc172277106)

[2.1 State Of The Art 5](#_Toc172277107)

[2.2 Studi Literatur Terkait 6](#_Toc172277108)

[2.3 Teori Atau Kajian Pustaka 23](#_Toc172277109)

[2.3.1 Data Mining 23](#_Toc172277110)

[2.3.2 Pengertian Segmentasi Pelanggan 23](#_Toc172277111)

[2.3.3 Pengertian Metode K-Means 24](#_Toc172277112)

[2.3.4 Streamlit 26](#_Toc172277113)

[*2.3.5* Pengertian *FlowChart* 27](#_Toc172277114)

[2.3.6 Kaggle 28](#_Toc172277115)

[2.3.7 Dataset 29](#_Toc172277116)

[2.3.8 Metode Pengujian Sistem 29](#_Toc172277117)

[2.3.9 Usability Testing 30](#_Toc172277118)

[2.3.10 WhiteBox 31](#_Toc172277119)

[BAB III ANALISA DATA DAN PEMBAHASAN 33](#_Toc172277120)

[3.1 Analisa data atau dataset 33](#_Toc172277121)

[3.2 FlowChart 39](#_Toc172277122)

[3.3 Hasil Klasifikasi 41](#_Toc172277123)

[3.4 Implementasi Interface 50](#_Toc172277124)

[3.4.1 Home Page 50](#_Toc172277125)

[3.4.2 Navigasi Menu 51](#_Toc172277126)

[3.4.3 Clustering 51](#_Toc172277127)

[3.4.4 Slider Memilih Jumlah Cluster 52](#_Toc172277128)

[3.5 Pengujian 53](#_Toc172277129)

[3.5 Usability Tasting 55](#_Toc172277130)

[BAB IV PENUTUP 56](#_Toc172277131)

[4.1 Kesimpulan 56](#_Toc172277132)

[4.2 Saran 56](#_Toc172277133)

[DAFTAR PUSTAKA 58](#_Toc172277134)

[LAMPIRAN 1 62](#_Toc172277135)

[LAMPIRAN 2 63](#_Toc172277136)

[LAMPIRAN 3 64](#_Toc172277137)

[LAMPIRAN 4 65](#_Toc172277138)

# DAFTAR TABEL

[Tabel 2. 1 Perbandingan Penelitian Terkait Dan Peneliti 21](#_Toc154179151)

[Tabel 2. 2 Simbol-Simbol Dalam Flowchart 27](#_Toc154179152)

[Tabel 3. 1 Mall Costumer 33](#_Toc154179157)

[Tabel 3. 2 Pengujian Pada User 53](#_Toc154179158)

[Tabel 3. 3 Usability Tasting 55](#_Toc154179159)

# DAFTAR GAMBAR

[Gambar 2. 1*State of the Art* 5](#_Toc172209189)

[Gambar 3. 1 *Flowchart* 39](#_Toc172209193)

[Gambar 3. 2 membaca dataset 41](#_Toc172209194)

[Gambar 3. 3 deskripsi statistik dataset 42](#_Toc172209195)

[Gambar 3. 4 Distribusi Gender 43](#_Toc172209196)

[Gambar 3. 5 Distribusi Usia 44](#_Toc172209197)

[Gambar 3. 6 Distribusi Pendapatan Tahunan (K$) 45](#_Toc172209198)

[Gambar 3. 7 Distribusi Skor Pengeluaran 46](#_Toc172209199)

[Gambar 3. 8 matriks korelasi 47](#_Toc172209200)

[Gambar 3. 9 Grafik Elbow Method 48](#_Toc172209201)

[Gambar 3. 10 scatter plot 49](#_Toc172209202)

[Gambar 3. 11 Halaman Beranda 50](#_Toc172209203)

[Gambar 3. 12 Tampilan Navigasi Menu 51](#_Toc172209204)

[Gambar 3. 13 Tampilan Menu Clustering 51](#_Toc172209205)

[Gambar 3. 14 Slider Memilih Jumlah Cluster 52](#_Toc172209206)

[Gambar 3. 15 Tampilan Petunjuk 53](#_Toc172209207)

# BAB I PENDAHULUAN

## Latar Belakang

Sejalan dengan peradaban manusia yang mempengaruhi perkembangan teknologi informasi (TI), perkembangan TI berdampak pada berbagai sektor seperti bisnis, kesehatan, pendidikan, dan pemerintahan. Kemampuan memanfaatkan teknologi ini secara maksimal sangat penting agar dapat bersaing dengan yang lain.

Perubahan kebutuhan pelanggan mendorong perubahan dalam bidang pemasaran. Bagian pemasaran memiliki peran penting dalam menghadapi persaingan yang semakin ketat. Pengelola pusat perbelanjaan atau mall yang berorientasi pada pelanggan, umumnya akan menghadapi masalah di bidang pemasaran. Diperlukan survei pasar untuk mendapatkan informasi tentang permintaan dan kebutuhan spesifik pelanggan (Adiana, Soesanti, dan Permanasari 2020).

Dalam era digital yang terus berkembang, pusat perbelanjaan menghadapi tantangan besar dalam bersaing di pasar yang semakin kompetitif. Salah satu kunci keberhasilan mereka adalah memahami pelanggan dengan lebih baik. Namun, seringkali pengelola mall menghadapi kendala dalam mengumpulkan, mengelola, dan menganalisis data pelanggan karena terbatasnya sumber daya manusia dan keuangan.

Pusat perbelanjaan dengan keberagamannya dalam sektor ekonomi dan skala operasional kini dihadapkan pada kemungkinan untuk mengambil keuntungan penuh dari perkembangan teknologi. Digitalisasi memberikan peluang bagi mall untuk memperluas jangkauan pasar, meningkatkan efisiensi operasional, dan memperkuat daya saing mereka(Jamalulael dan Aulia 2023).

Segmentasi pelanggan adalah pendekatan yang memecah basis pelanggan menjadi kelompok-kelompok lebih kecil berdasarkan karakteristik yang sama, seperti preferensi produk, perilaku pembelian, atau demografi. Hal ini memungkinkan pengelola mall untuk menyesuaikan strategi pemasaran dengan lebih efektif, meningkatkan retensi pelanggan, dan memperbaiki pengalaman pelanggan(Awalina dan Rahayu 2023).

Di sinilah pentingnya penggunaan metode seperti k-means clustering. K-means clustering adalah metode dalam analisis data yang memungkinkan sistem untuk mengelompokkan data pelanggan berdasarkan kesamaan karakteristik tanpa harus diprogram secara eksplisit. Dengan menerapkan teknik k-means clustering, pengelola mall dapat menganalisis data pelanggan secara lebih mendalam dan mendapatkan wawasan yang berharga untuk mendukung pengambilan keputusan.

Namun, pengelola mall sering kali tidak memiliki pengetahuan atau keterampilan teknis yang diperlukan untuk menerapkan k-means clustering dengan efektif. Oleh karena itu, ada kebutuhan untuk mengembangkan solusi yang dapat diakses dan diimplementasikan dengan mudah oleh pengelola mall.

Pelanggan dengan karakteristik dan kebutuhan yang beragam memerlukan pendekatan yang berbeda dalam produk, layanan, dan strategi pemasaran. Oleh karena itu, segmentasi pelanggan menjadi strategi penting yang harus diimplementasikan oleh mall.

Melalui penerapan segmentasi pelanggan menggunakan k-means clustering, pengelola mall dapat mengoptimalkan sumber daya mereka, meningkatkan efisiensi operasional, dan meningkatkan kepuasan pelanggan. Ini dapat membantu mereka tetap bersaing di pasar yang dinamis dan berkembang pesat serta memperkuat posisi mereka dalam ekosistem bisnis lokal.

Dengan memahami peran dan manfaat segmentasi pelanggan menggunakan k-means clustering, pengelola mall dapat mengambil langkah-langkah proaktif untuk meningkatkan kinerja bisnis mereka dan mencapai pertumbuhan yang berkelanjutan.

## Rumusan Masalah

1. Bagaimana cara menggunakan metode k-means clustering untuk memahami pelanggan mall secara lebih baik?
2. Metode apa yang paling efektif dalam membagi pelanggan mall menjadi kelompok-kelompok yang lebih kecil berdasarkan preferensi dan perilaku mereka?

## Tujuan

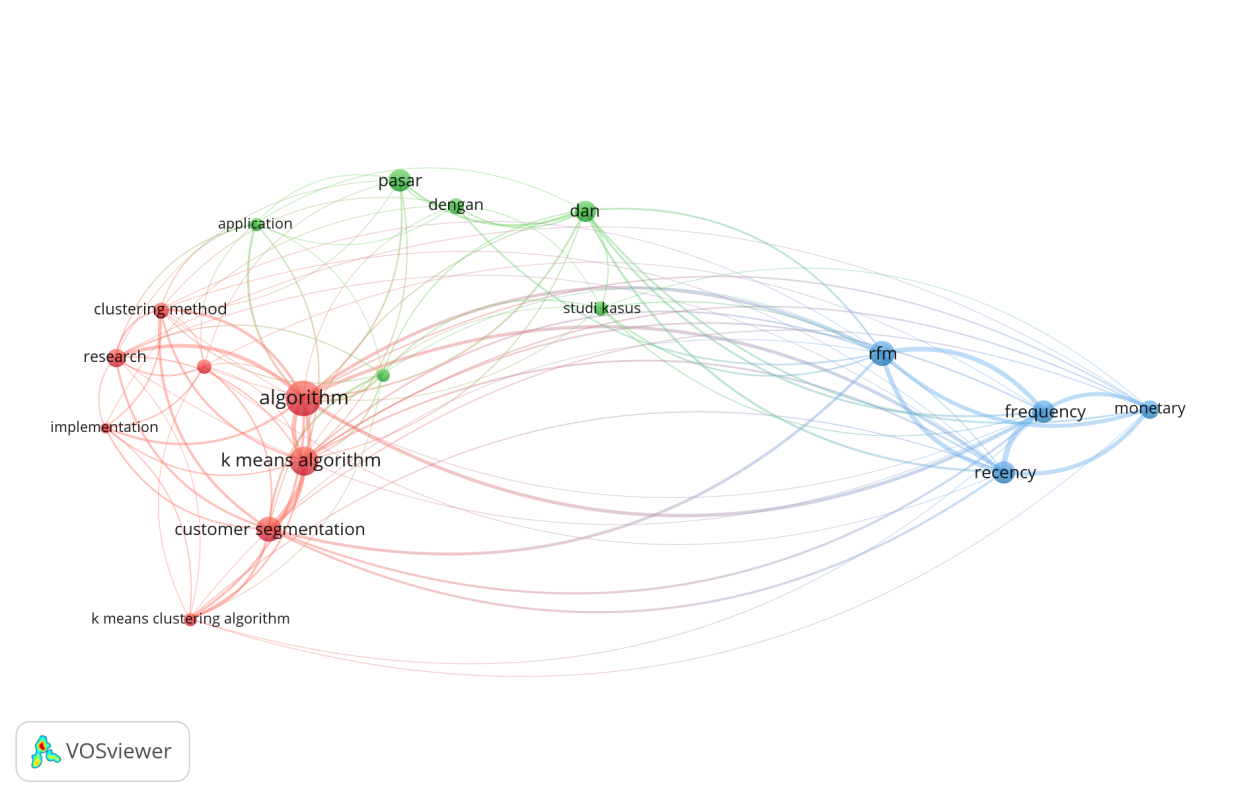
1. Mengetahui cara menggunakan metode k-means clustering untuk memahami pelanggan mall.
2. Mengetahui metode yang paling efektif dalam membagi pelanggan mall menjadi kelompok-kelompok yang lebih kecil.

## Batasan Masalah

1. Fokus pada mall dengan skala bisnis yang beragam, termasuk mall yang relatif kecil dengan basis pelanggan yang terbatas hingga mall yang lebih besar dengan basis pelanggan yang lebih luas, namun tetap mempertimbangkan kendala sumber daya yang dimiliki oleh pengelola mall.
2. Penelitian dapat dibatasi pada mall dalam industri atau sektor tertentu, seperti retail, makanan dan minuman, atau hiburan, untuk memperoleh wawasan yang lebih spesifik dan relevan.
3. Data yang dianalisis dalam penelitian ini meliputi data demografis, preferensi produk, dan perilaku pembelian pelanggan. Hanya data yang tersedia dan dapat diakses oleh pengelola mall yang akan digunakan dalam penelitian ini.
4. Penelitian ini menggunakan algoritma k-means clustering untuk segmentasi pelanggan. Algoritma lain mungkin tidak dieksplorasi secara mendalam dalam penelitian ini.
5. Mengingat keterbatasan sumber daya yang dimiliki oleh pengelola mall, penelitian ini akan mempertimbangkan solusi yang dapat diimplementasikan dengan mudah dan memerlukan biaya rendah. Solusi yang membutuhkan investasi besar dalam teknologi dan pelatihan mungkin tidak menjadi fokus utama.

# BAB II KAJIAN 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 Gambar 1 di bawah: 

*Sumber: VOSviewer*

Gambar 2. 1*State of the Art*

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

## Studi Literatur Terkait

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 segmentasi pelanggan yang telah dibuat. Berikut beberapa jurnal yang berkaitan dengan penelitian yang akan dilakukan:

1. Jurnal Nasional **“Penerapan Machine Learning pada Penjualan Produk UMKM”**, penelitian yang dilakukan oleh (Hindrayani, Anjani, dan Nurlaili 2021)studi ini mengkaji berbagai aplikasiMachine Learning(ML) seperti prediksi harga komoditi, analisis segmentasi pelanggan, analisis sentimen, dan peningkatan loyalitas pelanggan melalui tinjauan sistematis terhadap artikel yang relevan. Dari 7 artikel yang memenuhi kriteria, ditemukan bahwa algoritma Backpropagation Neural Network dapat memprediksi harga kopi lokal dengan akurasi 99.9936%, sementara K-Means Clustering berhasil mengelompokkan pelanggan UMKM ke dalam 6 segmen. Analisis sentimen menggunakan NLP dengan TF-IDF menunjukkan akurasi 76.92%, recall 74.07%, dan presisi 80.00% pada ulasan produk di Shopee. Algoritma Random Forest Tree mencapai akurasi 99.3802% dalam meningkatkan loyalitas pelanggan dengan data voucher belanja. Temuan ini menunjukkan bahwa ML dapat meningkatkan produktivitas dan daya saing UMKM, dengan penekanan pada pentingnya pengumpulan data ekstensif. Studi ini juga menyarankan eksplorasi lebih lanjut aplikasi ML lainnya untuk meningkatkan kualitas dan daya saing produk lokal, terutama dalam situasi ekonomi dinamis seperti pandemi COVID-19.
2. Jurnal Nasional **“Optimalisasi Strategi Pemasaran dengan Segmentasi Pelanggan Menggunakan Penerapan K-Means Clustering pada Transaksi Online Retail”**, penelitian yang dilakukan oleh (Awalina dan Rahayu 2020)studi ini menekankan pentingnya memahami pelanggan untuk mengembangkan bisnis yang efektif dengan menawarkan produk sesuai kebutuhan dan preferensi mereka. Analisis pelanggan yang kompleks ini menggunakan algoritma K-Means untuk segmentasi, dengan pemrograman Python untuk data mining dan pre-processing melalui exploratory data analysis. Metode elbow digunakan untuk menentukan jumlah klaster optimal, menghasilkan empat klaster berdasarkan variabel quantity, unit price, dan customer id. Hasil klasterisasi menunjukkan bahwa kuantitas dan harga satuan sangat mempengaruhi perilaku pelanggan, membantu dalam penyampaian penawaran produk yang lebih tepat.
3. Jurnal Nasional**“Analisis sentimen review penjualan produk umkm pada kabupaten nias dengan komparasi algoritma klasifikasi machine learning”**, penelitian dilakukan  oleh (Harahap dkk. 2021)studi ini menggunakan data komentar pelanggan dari Facebook dan Instagram, melalui tahapan text pre-processing, feature extraction dengan TF-IDF, 10-fold cross-validation, dan evaluasi empat algoritma klasifikasi: Support Vector Machine (SVM), Naïve Bayes, K-Nearest Neighbors (KNN), dan Artificial Neural Network (ANN). Hasil menunjukkan SVM dan ANN memiliki performa terbaik dalam klasifikasi sentimen, memberikan wawasan mendalam tentang persepsi pelanggan. Analisis ini membantu UMKM memahami sentimen pelanggan, mengembangkan strategi pemasaran yang lebih baik, dan meningkatkan penjualan selama pandemi. Penelitian ini menekankan pentingnya analisis sentimen berkala untuk menjaga relevansi dan adaptabilitas di pasar yang dinamis, serta meningkatkan daya saing UMKM di Kabupaten Nias.
4. Jurnal Nasional**“Analisa Dalam Memahami Kewirausahaan Dan Usaha Baru Pada Umkm Di Era Digita**l”, penelitian yang dilakukan oleh (Jamalulael dan Aulia 2023)studi ini menyoroti transformasi signifikan akibat perkembangan teknologi informasi dan komunikasi, yang mempengaruhi penciptaan dan adaptasi bisnis baru. Digitalisasi mempercepat laju bisnis, membuka peluang inovasi, dan memudahkan akses ke informasi serta pasar global. Melalui metode kualitatif dan studi literatur, peneliti mengkaji karakteristik kewirausahaan, manajemen usaha baru, serta faktor-faktor keberhasilan UMKM dalam menghadapi tantangan dan peluang digital. Dengan pendekatan deskriptif, penelitian ini memberikan pemahaman mendalam tentang dinamika bisnis digital dan mengidentifikasi kunci sukses bagi pengusaha. Hasil penelitian ini memperkaya pemahaman tentang peran kewirausahaan di era digital dan memberikan pandangan relevan bagi pengambil keputusan dalam mendukung pengembangan UMKM di Indonesia.
5. Jurnal Nasional**“Analisis Segmentasi Pelanggan Menggunakan Kombinasi RFM Model dan Teknik Clustering”**, penelitian yang dilakukan oleh (Adiana, Soesanti, dan Permanasari 2020) studi ini menekankan pentingnya segmentasi pelanggan dalam relationship marketing untuk meningkatkan interaksi dan pemahaman terhadap kebutuhan pelanggan. Model RFM digunakan untuk mengukur tiga aspek perilaku pelanggan: recency (waktu sejak transaksi terakhir), frequency (frekuensi transaksi), dan monetary (nilai transaksi). Algoritma K-Means dipilih karena kemudahan komputasi dan kemampuan menentukan jumlah cluster secara optimal. Penggunaan RapidMiner memfasilitasi pengelompokkan data dan evaluasi menggunakan Davies Bouldin Index (DBI) untuk menentukan jumlah cluster optimal. Hasilnya, tiga cluster pelanggan teridentifikasi: pelanggan tipikal, pelanggan superstar, dan pelanggan dormant. Penelitian ini mengikuti model CRISP-DM dalam enam tahap: pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi, dan pengembangan/penyebaran. Dengan fokus pada UKM, terutama UD Gemilang Kencana, penelitian ini memberikan wawasan berharga bagi perusahaan kecil dan menengah dalam meningkatkan hubungan dengan pelanggan mereka.
6. Jurnal Nasional**“Maximizing Strategy Improvement in Mall Customer Segmentationusing K-means Clustering”**, penelitian dilakukan oleh (Pradana dan Ha 2021)metode K-means clustering digunakan untuk segmentasi pelanggan di pusat perbelanjaan. Segmentasi pelanggan adalah strategi penting dalam pemasaran untuk mengidentifikasi target pasar yang paling menguntungkan dan mengembangkan strategi pemasaran yang tepat. Penelitian ini berhasil mengelompokkan pelanggan menjadi lima klaster berdasarkan pendapatan tahunan dan skor belanja mereka. Hasilnya menunjukkan bahwa pelanggan dengan pendapatan tinggi dan skor belanja tinggi merupakan target yang ideal untuk strategi pasar karena memberikan margin keuntungan terbesar. Kelompok lainnya termasuk pelanggan dengan pendapatan tinggi tetapi skor belanja rendah, pelanggan dengan pendapatan rata-rata dan skor belanja rata-rata, pendapatan rendah tetapi skor belanja tinggi, dan pendapatan rendah dengan skor belanja rendah. Penelitian ini menekankan pentingnya pemahaman mendalam tentang perilaku dan kebiasaan belanja pelanggan untuk mengoptimalkan strategi pemasaran.
7. Jurnal Nasional**“Pemanfaatan Algoritma Machine Learning untuk Segmentasi Pelanggan Berbasis Data Konsumsi Listrik di PT PLN XYZ”**, Penelitian dilakukan oleh (Afthoni dkk. 2021) studi ini membahas tentang pemanfaatan algoritma machine learning, khususnya metode clustering K-means, untuk segmentasi pelanggan berdasarkan data konsumsi listrik di PT PLN XYZ. Konsep utama yang digunakan adalah Customer Relationship Management (CRM), yang berfokus pada pengembangan kepuasan pelanggan dan memaksimalkan profitabilitas. Menurut literatur, CRM adalah proses untuk mendapatkan, mempertahankan, dan mengembangkan pelanggan agar lebih menguntungkan. Definisi dari beberapa ahli seperti Newell, Kotler, Keller, dan Gordon menyatakan bahwa CRM melibatkan strategi dan proses yang menciptakan nilai baru bagi pelanggan individual, memperkuat preferensi mereka terhadap organisasi, dan meningkatkan hasil bisnis melalui interaksi jangka panjang. Studi ini mengintegrasikan konsep-konsep CRM dengan hasil segmentasi pelanggan yang dihasilkan oleh algoritma K-means untuk membentuk strategi pelayanan pelanggan yang lebih terarah dan efektif​.
8. Jurnal Nasional**“Kajian Untuk Segmentasi Customer Bank Dengan Algoritma K-Means”**, penelitian dilakukan oleh(Tauhid dkk. 2024)bertujuan untuk memahami diversitas pelanggan bank melalui penerapan algoritma K-Means. Dengan menganalisis variabel-variabel seperti frekuensi transaksi, saldo akun, jenis produk, dan aktivitas perbankan lainnya, penelitian ini berhasil mengidentifikasi segmen pelanggan yang berbeda-beda berdasarkan karakteristik unik mereka. Algoritma K-Means, sebagai metode clustering tanpa pengawasan, membantu dalam mengungkap pola tersembunyi dalam data pelanggan, memberikan wawasan mendalam tentang perilaku dan preferensi pelanggan. Hasil segmentasi ini diharapkan dapat mendukung bank dalam menyesuaikan strategi pemasaran, personalisasi layanan, dan meningkatkan interaksi dengan pelanggan, yang pada akhirnya meningkatkan kepuasan pelanggan dan keunggulan kompetitif di era perbankan digital. Penelitian ini menggunakan pendekatan analisis data kuantitatif dan metode validasi untuk memastikan hasil yang akurat dan bermakna, menciptakan dasar yang kuat untuk strategi bisnis yang lebih efektif dan efisien.
9. Jurnal Nasional**“Penerapan K-Means pada Segmentasi Pasar untuk Riset Pemasaran pada Startup Early Stage dengan Menggunakan CRISP-DM”**,penelitian dilakukan oleh(Christian dan Qi 2022)penelitian ini membahas tentang penerapan metode RFM (Recency, Frequency, Monetary) untuk mengestimasi Customer Lifetime Value (CLV) dan segmentasi konsumen menggunakan model machine learning K-Means Clustering. Latar belakang penelitian ini didorong oleh dampak krisis Covid-19 terhadap pendapatan perusahaan, yang menuntut strategi pemasaran yang lebih tepat dan berbasis data untuk mempertahankan konsumen. Metode RFM memungkinkan segmentasi konsumen berdasarkan data transaksi mereka, mengelompokkan konsumen ke dalam kategori berdasarkan nilai kumulatif dari recency, frequency, dan monetary value transaksi mereka. Hasil segmentasi ini membantu perusahaan dalam mengidentifikasi konsumen dengan nilai tinggi yang layak dipertahankan dan diberikan pelayanan optimal. Penelitian ini menunjukkan bahwa metode RFM dan K-Means Clustering efektif dalam mendukung pengambilan keputusan strategis untuk mengoptimalkan profit perusahaan di tengah krisis, dan dapat diterapkan pada berbagai kasus dan perusahaan lain​.
10. Jurnal Nasional**“Pemodelan Dan Penerapan Metode Rfm Pada Estimasi Nilai Konsumen (Customer Lifetime Value) Menggunakan K-Means Clustering Machine Learning”**,penelitian dilakukan oleh(Laksono dan Wulansari 2021) penelitian dilakukan dalam dampak krisis Covid-19 terhadap pendapatan perusahaan, menekankan pentingnya strategi pemasaran berbasis data untuk mempertahankan dan mengoptimalkan nilai pelanggan. Dengan menggunakan data transaksi yang diperoleh dari platform e-commerce, analisis RFM dilakukan untuk mengidentifikasi kebiasaan pembelian konsumen, di mana nilai recency menunjukkan seberapa baru konsumen melakukan pembelian, frequency menunjukkan seberapa sering pembelian dilakukan, dan monetary menunjukkan nilai total pembelian. Hasil analisis RFM kemudian digunakan sebagai input untuk algoritma K-Means Clustering yang membagi konsumen menjadi beberapa segmen berdasarkan pola pembelian mereka. Studi ini menemukan bahwa segmentasi konsumen menggunakan metode RFM dan K-Means Clustering dapat membantu perusahaan dalam mengembangkan strategi pemasaran yang lebih efektif dan efisien, serta dapat diterapkan pada berbagai industri untuk mengoptimalkan penggunaan sumber daya dan meningkatkan profitabilitas di tengah tantangan ekonomi yang disebabkan oleh pandemi​.
11. Jurnal Internasional **“Customer segmentation using machine learning”**, penelitian dilakukan oleh (Johansson dan Wikström 2021) membahas penggunaan machine learning dalam segmentasi pelanggan di perusahaan Bokadirekt, platform online untuk pemesanan layanan kesehatan dan kecantikan. Penelitian ini menyoroti pentingnya analisis data dalam segmentasi pelanggan dan kemajuan terbaru dalam Machine Learning dan Artificial Intelligence. Dua model machine learning dilatih dan diuji, mencapai tingkat akurasi 71%, lebih baik daripada ahli manusia di perusahaan tersebut. Hasilnya adalah pengembangan layanan untuk mengkategorikan pengguna ke dalam tiga kelompok pelanggan yang berbeda. Penelitian ini menggunakan proses analisis data, preprocessing data, implementasi model, dan evaluasi model, didukung oleh prinsip-prinsip machine learning dan teknik klasifikasi. Kolaborasi erat antara penulis juga berperan penting dalam kesuksesan penelitian ini.
12. Jurnal Internasional **“Customer Segmentation using Machine Learning”**, penelitian dilakukan oleh (Monil 2020) membahas pentingnya Customer Relationship Management (CRM) dalam meningkatkan kepuasan pelanggan dan retensi, serta keuntungan perusahaan melalui identifikasi pelanggan potensial menggunakan teknik clustering. Teknik clustering yang umum digunakan termasuk K-means, Hierarchical Clustering, Density-based Clustering, dan Affinity Propagation. Setiap teknik memiliki kelebihan dan kelemahan, seperti kecepatan komputasi, ketersediaan jumlah cluster awal, dan efisiensi pada dataset tertentu. Jurnal ini juga menyoroti pentingnya mempertimbangkan parameter seperti geografis, demografis, psikografis, dan perilaku dalam segmentasi pelanggan untuk meramalkan tren konsumsi masa depan dan mengalokasikan sumber daya dengan lebih efektif. Kesimpulannya, pemilihan dan implementasi teknik clustering yang tepat berdasarkan situasi dan kebutuhan data sangat penting untuk mencapai tujuan strategis perusahaan seperti meningkatkan kepuasan pelanggan dan keuntungan.
13. Jurnal Internasional **“A Survey on Customer Segmentation using Machine Learning Algorithms to Find Prospective Clients”**, penelitian dilakukan oleh(Mehta, Mehra, dan Verma 2021) studi tentang segmentasi pelanggan menggunakan algoritma pembelajaran mesin menyoroti berbagai metode dan efektivitasnya dalam mengkategorikan pelanggan berdasarkan atribut yang berbeda. K-means clustering sering diidentifikasi sebagai metode yang unggul karena efisiensi komputasinya dan kemampuannya untuk menghasilkan klaster yang lebih rapat, seperti yang dibuktikan oleh skor Silhouette dan Davies Bouldin yang lebih tinggi dibandingkan dengan algoritma lain seperti clustering hierarkis. Studi yang ditinjau dalam survei menunjukkan bahwa K-means sangat menguntungkan untuk menangani dataset besar karena kompleksitas waktu linear, menjadikannya lebih disukai daripada metode hierarkis yang memiliki kompleksitas waktu kuadratik. Algoritma lain, termasuk clustering hierarkis dan pendekatan hibrida seperti DIRAC, juga menunjukkan potensi tetapi seringkali dengan biaya atau kompleksitas komputasi yang lebih tinggi. Penelitian di masa depan disarankan untuk mengeksplorasi lebih banyak algoritma pada berbagai dataset untuk mengidentifikasi metode segmentasi yang lebih akurat.
14. Jurnal Internasional**“Comparative Unsupervised Clustering Approaches for Customer Segmentation”**, penelitian dilakukan oleh(Aktas, Tunali, dan Bayrak 2021)dalam studi ini penulis menguji enam algoritma clustering yang berbeda untuk menganalisis data riwayat pembelian pelanggan. Algoritma yang diuji termasuk k-means, clustering spektral, Gaussian mixture, agglomerative clustering, ward, dan mean-shift. Studi ini bertujuan untuk mengidentifikasi pola interaksi pelanggan yang dapat digunakan untuk segmentasi pelanggan yang lebih efektif. Hasil clustering dibandingkan dan distribusi fitur dalam setiap cluster divisualisasikan, memungkinkan pemahaman yang lebih baik tentang bagaimana setiap fitur berperan dalam pembentukan cluster. Selain itu, pendekatan hybrid yang menggabungkan klasifikasi support vector (SVC) digunakan untuk memberikan probabilitas label cluster, yang memberikan wawasan tambahan tentang kepercayaan dalam klasifikasi data. Hasil penelitian menunjukkan bahwa algoritma clustering yang berbeda memiliki kekuatan yang bervariasi dalam mengidentifikasi pola pelanggan, dengan clustering spektral menunjukkan potensi yang tinggi untuk data yang digunakan. Studi ini memberikan dasar yang kuat untuk pengembangan lebih lanjut dalam segmentasi pelanggan yang dapat diterapkan pada berbagai industri.
15. Jurnal Internasional **”Interpretable machine learning-based approach for customer segmentation for new product development from online product reviews”**, penelitian dilakukan oleh(Joung dan Kim 2023)dalam jurnal ini mengkaji evolusi segmentasi pelanggan, terutama dari penggunaan data demografis, psikografis, dan perilaku pembelian hingga ke segmentasi berdasarkan kebutuhan yang diekstraksi dari ulasan daring. Studi ini menyoroti keterbatasan pendekatan segmentasi tradisional yang tidak dapat mengidentifikasi kelompok pelanggan dengan kebutuhan yang belum terpenuhi. Melalui penggunaan pembelajaran mesin yang dapat diinterpretasi (Interpretable Machine Learning, IML), penelitian ini berhasil mengatasi tantangan teknis dalam menentukan nilai penting dari fitur produk berdasarkan ulasan daring dan menjelaskan hubungan nonlinier antara kepuasan fitur produk dan kepuasan pelanggan secara keseluruhan. Dengan pendekatan baru ini, segmentasi pelanggan dapat dilakukan dengan lebih efektif, membuka peluang untuk pengembangan produk baru yang lebih sesuai dengan kebutuhan pelanggan .
16. Jurnal Internasional**“New Methods of Customer Segmentation and Individual Credit Evaluation Based on Machine Learning”**, penelitian dilakukan oleh(Yuping dkk. 2020) studi ini membahas tentang evolusi metode segmentasi pelanggan dan evaluasi kredit individu dalam ekonomi digital. Artikel tersebut menyoroti dampak hilangnya uang fisik terhadap perilaku konsumen, perubahan dalam evaluasi kredit pribadi, serta penerapan machine learning dalam meningkatkan proses segmentasi pelanggan dan evaluasi kredit. Dengan latar belakang perkembangan pesat big data dan teknologi internet, metode tradisional evaluasi kredit pribadi dihadapkan pada tantangan signifikan, mendorong perlunya penelitian dalam evaluasi kredit personal berbasis machine learning. Melalui pendekatan ini, penelitian menggali dimensi evaluasi kredit yang komprehensif dan metode akuisisi data canggih perusahaan finansial internet. Selain itu, penelitian juga mencakup teknik desensitisasi data, pengujian LOF, penggunaan metode random forest, serta penerapan model scorecard berbasis regresi logistik untuk menghasilkan skor evaluasi kredit personal. Model ini kemudian diuji menggunakan jaringan saraf BP untuk memprediksi tingkat kredit personal, menunjukkan relevansi dan keunggulan pendekatan machine learning dalam konteks evaluasi kredit individu.
17. Jurnal Internasional **“Segmenting Bank Customers via RFM Model and Unsupervised Machine Learning”**, penelitian dilakukan oleh(Aliyev dkk. 2020)studi ini membahas pentingnya segmentasi pelanggan dalam industri perbankan dan penerapan metodologi baru untuk membantu lembaga keuangan mempertahankan pelanggan mereka secara efektif. Dengan menggunakan model RFM dan algoritma clustering pada data pelanggan nyata, penelitian ini memberikan wawasan berharga bagi salah satu bank swasta terbesar di Azerbaijan. Penelitian ini juga mengulas berbagai teknik clustering seperti K-Means dan Hierarchical Clustering serta menyoroti keuntungan dan kerugian dari masing-masing algoritma untuk membantu meningkatkan akurasi dan kinerja dalam segmentasi pelanggan.
18. Jurnal Internasional **“Customer Segmentation Using Machine Learning in Commercial Bank of Ethiopia”**, penelitian dilakukan oleh(Bank 2024)Studi ini membahas pentingnya segmentasi pelanggan dalam industri perbankan dan penerapan metodologi baru untuk membantu lembaga keuangan mempertahankan pelanggan mereka secara efektif. Dengan menggunakan model RFM dan algoritma clustering pada data pelanggan nyata, penelitian ini memberikan wawasan berharga bagi salah satu bank swasta terbesar di Azerbaijan. Penelitian ini juga mengulas berbagai teknik clustering seperti K-Means dan Hierarchical Clustering serta menyoroti keuntungan dan kerugian dari masing-masing algoritma untuk membantu meningkatkan akurasi dan kinerja dalam segmentasi pelanggan.
19. Jurnal Internasional **“Customer Mobile Behavioral Segmentation and Analysis in Telecom Using Machine Learning”**, penelitian dilakukan oleh(Sharaf Addin dkk. 2022)Studi ini membahas segmentasi dan analisis perilaku pelanggan seluler di industri telekomunikasi menggunakan pendekatan pembelajaran mesin, khususnya algoritma K-means. Dalam upaya untuk meningkatkan pemahaman terhadap perilaku pelanggan, penelitian ini membahas beberapa teknik segmentasi yang meliputi K-means, Kohonen Self Organizing Map (SOM), dan model Recency, Frequency, and Monetary (RFM). K-means diidentifikasi sebagai salah satu algoritma clustering yang paling populer dan sederhana, mampu mengkategorikan data yang tidak berlabel menjadi beberapa klaster unik. Sementara itu, SOM digunakan untuk memetakan data berdimensi tinggi ke dalam ruang dua atau tiga dimensi tanpa kehilangan informasi yang signifikan. Model RFM digunakan sebagai teknik komplementer yang memberikan analisis lanjutan tentang perilaku pembelian pelanggan berdasarkan nilai recency, frequency, dan monetary. Hasil penelitian menunjukkan bahwa segmentasi berbasis pembelajaran mesin dapat memberikan wawasan yang mendalam tentang perilaku penggunaan layanan seluler, yang pada gilirannya membantu perusahaan telekomunikasi untuk meningkatkan layanan dan mengembangkan strategi pemasaran yang lebih efektif​.
20. Jurnal Internasional **“A Comprehensive Solution For Scrap Management Leveraging Machine Learning for Customer Segmentation”**, penelitian dilakukan oleh(Yadav dkk. 2023)membahas tentang berbagai aspek penting dari manajemen limbah dan pemanfaatan teknologi pembelajaran mesin untuk segmentasi pelanggan. Penelitian ini menyoroti dampak negatif dari pembuangan limbah yang tidak tepat, seperti polusi lingkungan, bahaya kesehatan, dan penipisan sumber daya alam ​​. Dengan mengembangkan aplikasi web "Smart Bin", peneliti bertujuan untuk memudahkan proses penjualan dan donasi bahan limbah dari rumah, mengurangi ketergantungan pada metode tradisional yang tidak efisien ​​. Aplikasi ini tidak hanya memfasilitasi pengelolaan limbah tetapi juga menawarkan produk daur ulang dan upcycled kepada konsumen yang sadar lingkungan. Pendekatan pembelajaran mesin diterapkan untuk segmentasi pelanggan berdasarkan karakteristik yang serupa, seperti demografi, perilaku, atau riwayat pembelian, yang meningkatkan akurasi prediksi penjualan dan analisis data​​. Penelitian ini menunjukkan bahwa pengelolaan limbah yang efektif dapat dicapai melalui inovasi teknologi dan pendekatan yang berfokus pada pelanggan, sehingga mendukung keberlanjutan lingkungan dan ekonomi.

Tabel 2. 1 Perbandingan Penelitian Terkait Dan Peneliti

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| NO | FITUR | **PENELITIAN** | | | | | | | | | | | | | | | | | | | | |
| **PP** | **P1** | **P2** | **P3** | **P4** | **P5** | **P6** | **P7** | **P8** | **P9** | **P10** | **P11** | **P12** | **P13** | **P14** | **P15** | **P16** | **P17** | **P18** | **P19** | **P20** |
| ( Kelompok 2 ) | (Hindrayani, Anjani, dan Nurlaili 2021) | (Awalina dan Rahayu 2023) | (Harahap dkk. 2021) | (Jamalulael dan Aulia 2023) | (Adiana, Soesanti, dan Permanasari 2020) | (Pradana dan Ha 2021) | (Afthoni dkk. 2021) | (Tauhid dkk. 2024) | (Christian dan Qi 2022) | (Laksono dan Wulansari 2021) | (Johansson dan Wikström 2021) | (Monil 2020) | (Mehta, Mehra, dan Verma 2021) | (Aktas, Tunali, dan Bayrak 2021) | (Joung dan Kim 2023) | (Yuping dkk. 2020) | (Aliyev dkk. 2020) | (Bank 2024) | (Sharaf Addin dkk. 2022) | (Yadav dkk. 2023) |
| 1 | Analisis Sentimen | √ | √ |  | √ |  |  |  |  |  |  |  |  |  |  |  | √ |  |  |  | √ |  |
| 2 | Segmentasi Pelanggan | √ | √ | √ |  |  | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ |
| 3 | Prediksi Harga Komoditi | √ | √ |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 4 | Peningkatan Loyalitas Pelanggan | √ | √ |  |  |  | √ |  |  |  |  | √ |  |  |  |  |  | √ | √ |  |  | √ |
| METODE | | | | | | | | | | | | | | | | | | | | | | |
| 1. | Algoritma k-means clustering | √ |  | √ |  |  | √ | √ | √ | √ | √ | √ |  | √ | √ | √ |  |  | √ | √ | √ |  |
| TOOLS | | | | | | | | | | | | | | | | | | | | | | |
| 1. | Python | √ |  |  |  |  |  | √ |  | √ |  | √ |  |  |  |  |  |  |  |  |  |  |
| 2. | Streamlit | √ |  |  | √ |  |  | √ |  | √ |  | √ |  |  |  |  |  |  |  |  |  |  |

**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 tidak dari peneliti p1-p20 jika ada maka anda centang jika tidak maka tidak perlu.

## Teori Atau Kajian Pustaka

### 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 clusteringcocok untuk mengelompokkan pelanggan berdasarkan tingkat kesamaannya menjadi beberapa cluster atau segmen.(Sudarsono dkk. 2021)

### 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 cluster yang baik ketika setiap anggota dari cluster memiliki 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.(Febriani dan Putri 2020)

### Pengertian Metode K-Means

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.(Nabila dkk. 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.

### Streamlit

Streamlit adalah framework opensource dari Python yang memungkinkan untuk membuat aplikasi web menggunakan bahasa Python dalam mengaplikasi model dari machine learning atau data science secara cepat dan mudah menggunakan Python. Dirancang untuk memudahkan pembuatan antarmuka pengguna (UI) yang interaktif, Streamlit tidak memerlukan pengalaman pemrograman web. Dengan integrasi yang mulus dengan kode Python dan pustaka data science seperti pandas, scikit-learn, serta pustaka visualisasi seperti Matplotlib dan Seaborn, Streamlit menjadi alat yang efisien dan cepat untuk membangun aplikasi data-driven.

fitur utama Streamlit mencakup kemampuan untuk membuat sidebar yang berfungsi sebagai panel navigasi atau kontrol aplikasi, dukungan berbagai widget interaktif seperti slider, tombol, dan input teks, serta kemampuan untuk menampilkan grafik dan tabel dengan mudah. Streamlit juga reaktif, artinya aplikasi akan otomatis memperbarui saat ada perubahan pada input pengguna, memberikan pengalaman yang lebih dinamis dan responsif.(Ferdyandi, Setiawan, dan Abdurrachman Bachtiar 2022)

### Pengertian *FlowChart*

Flowchart dapat diartikan sebagai langkah langkah penyelesaian masalah yang dituliskan dalam suatu simbol-simbol tertentu. Diagram alir ini akan menunjukan alur didalam program secara logika”. Flowchart ini diperlukan tidak Masukkan Proses Keluaran hanya sebagai alat komunikasi tetapi juga sebagai pedoman, dan sebelum komponen-komponennya dapat lebih dipahami.(Pemrograman 2020)

Tabel 2. 2 Simbol-Simbol Dalam Flowchart

|  |  |  |  |
| --- | --- | --- | --- |
| No | Simbol | Nama | Fungsi |
| 1. |  | Terminal dan Terminator | Awal atau akhir suatu program (Prosedur). |
| 2. |  | Input/Output | Proses input atau output terlepas dari jenis perangkat. |
| 3. |  | Proses | Proses operasional  computer. |
| 4. |  | Decision | Untuk menunjukkan  bahwa suatu kondisi  tertentu mengarah  pada dua kemungkinan, ya/tidak. |
| 5. |  | konektor | Koneksi penghubung  proses ke proses lain  pada halaman yang  sama. |
| 6. |  | Sub Program | Pemanggilan sub program |
| 7. |  | Garis Alir | Arah aliran program |
| 8 |  | “Offline  Connector” | Koneksi Penghubung  dari satu proses ke  proses lain di  halaman lain. |
| 9. |  | “Punched Card” | Input berasal dari  kartu atau output  ditulis ke kartu. |
| 10. |  | “Document” | Mencetak output  dalam format  dokumen (melalui  printer). |

*Sumber:* (Pemrograman 2020)

### 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 informasi informasi 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.

Private dataset adalah dataset yang dapat diambil dari sebuah organisasi yang akan dilakukan sebagai objek penelitian, seperti data bank, rumah sakit, sekolah dan lain sebagainya. Sedangkan public dataset adalah dataset yang bisa diambil dari repository publik yang disepakati oleh pakar peneliti pada bidang data mining. Dataset bertujuan untuk menguji suatu metode penelitian yang dikembangkan oleh para pakar peneliti dengan public dataset maupun private dataset. Contoh dari public dataset dan yang digunakan pada penelitian ini adalah dataset Kaggle.(Hadi dkk. 2020)

### Metode Pengujian Sistem

Menerapkan segmentasi pelanggan menggunakan metode 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.(Tabianan, Velu, dan Ravi 2022)

### Usability Testing

Segmentasi pelanggan menggunakan metode K-Means clustering memungkinkan Mall 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, Shobayo, dan Ogunleye 2023)

### WhiteBox

*White Box* merupakan salah satu pengujian yang dapat digunakan untuk menguji suatu perangkat lunak. Pengujian ini didasarkan pada pengecekan terhadap detail perancangan. Pengecekan tersebut menggunakan struktur kontrol dari desain program secara procedural dan membagi pengujian ke dalam beberapa kasus pengujian. Pengujian *White Box* dapat mengungkapkan kesalahan dalam implementasi dari sebuah perangkat lunak. Pengembang yang melakukan pengujian dengan metode *White Box* harus memiliki pengetahuan atau pemahaman mengenai sumber kode perangkat lunak tersebut.

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. Transparansi ini sangat penting dalam berbagai aplikasi yang memerlukan interpretabilitas, seperti dalam bidang kesehatan atau keuangan, dimana memahami alasan di balik prediksi model dapat membantu dalam pengambilan keputusan yang lebih informatif dan dapat dipercaya. 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. Tujuan dari *White Box* sendiri yaitu :

1. Mengetahui secara internal mengenai cara kerja suatu perangkat lunak.
2. Menjamin operasi-operasi internal sudah sesuai dengan spesifikasi yang ditentukan pada awal dengan menggunakan struktur kontrol dari prosedur yang dirancang.

Ketika melakukan pengujian dengan metode *White Box* terdapat beberapa teknik yang dapat digunakan diantaranya yaitu, notasi diagram alir (*Path Graph Notation*), kompleksitas siklomatis (*Cyclomatic Complexity*), dan Kesetaraan Partisi (*EP*)(Nur dan S.R. Candra 2022)

# 

# BAB III ANALISA DATA DAN PEMBAHASAN

## Analisa data atau dataset

Tabel 3. 1 Mall Costumer

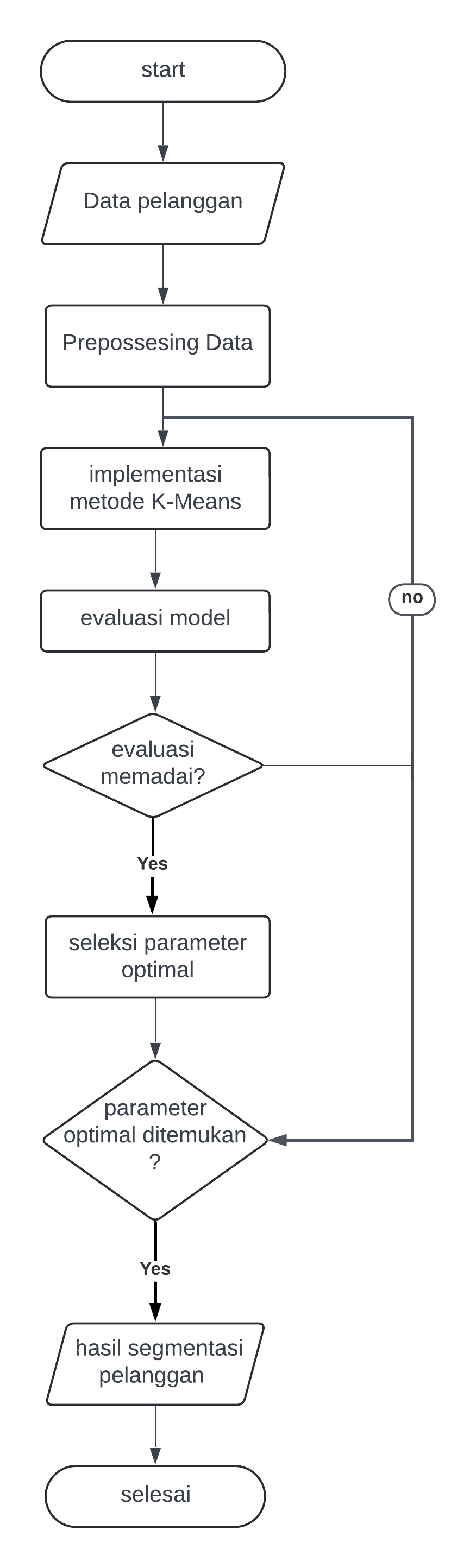
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **CustomerID** | **Gender** | **Age** | **Annual Income (k$)** | **Spending Score (1-100)** |
| 1 | Male | 19 | 15 | 39 |
| 2 | Male | 21 | 15 | 81 |
| 3 | Female | 20 | 16 | 6 |
| 4 | Female | 23 | 16 | 77 |
| 5 | Female | 31 | 17 | 40 |
| 6 | Female | 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 |

Dataset ini mencakup informasi tentang pelanggan sebuah pusat perbelanjaan, dengan kolom yang terdiri dari CustomerID, Genre (Jenis Kelamin), Age (Umur), Annual Income (k$) (Pendapatan Tahunan dalam ribuan dolar), dan Spending Score (1-100) (Skor Pengeluaran). Analisis statistik menunjukkan bahwa usia pelanggan berkisar antara 18 hingga 70 tahun, dengan rata-rata sekitar 38,85 tahun. Pendapatan tahunan bervariasi dari 15 hingga 137 ribu dolar, dengan rata-rata 60,56 ribu dolar. Skor pengeluaran, yang menilai perilaku dan pola pengeluaran pelanggan, memiliki nilai rata-rata 50,2 dengan rentang dari 1 hingga 99.

Dalam hal korelasi, tidak terdapat hubungan yang signifikan antara pendapatan tahunan dan skor pengeluaran, yang ditunjukkan oleh korelasi yang sangat rendah sebesar 0,0099. Namun, terdapat korelasi negatif moderat antara usia dan skor pengeluaran (-0,3272), yang menunjukkan bahwa pelanggan yang lebih muda cenderung memiliki skor pengeluaran yang lebih tinggi. Korelasi tinggi antara CustomerID dan pendapatan tahunan (0,9775) kemungkinan tidak memiliki arti penting karena CustomerID hanya berfungsi sebagai pengenal unik untuk setiap pelanggan.

Distribusi data menunjukkan bahwa usia pelanggan memiliki distribusi yang relatif normal namun sedikit miring ke kanan. Pendapatan tahunan memiliki distribusi yang cukup merata tanpa puncak yang signifikan. Distribusi skor pengeluaran tampak bimodal, dengan dua puncak di sekitar skor 35 dan 75, menunjukkan adanya dua kelompok utama dalam hal perilaku pengeluaran pelanggan.

## FlowChart



Gambar 3. 1 *Flowchart*

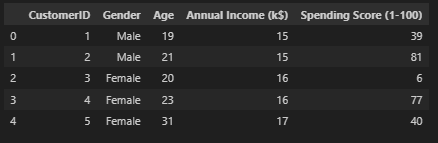
Flowchart tersebut menggambarkan proses secara lengkapsegmentasi pelanggan menggunakan metode K-Means. Proses ini dimulai dari “START” yang menandakan dimulainya proses segmentasi pelanggan. Dimana Data Pelangganini dapat mengumpulkan data pelanggan yang akan digunakan dalam proses segmentasi. Dalam data ini bisa berupa informasi demografis, perilaku pembelian, atau data lainnya yang relevan.Preprocessing Data dalam Tahap ini melibatkan persiapan dan pembersihan data pelanggan.

Langkah-langkah yang mungkin termasuk dalam preprocessing adalah Mengatasi missing values (data yang hilang), Normalisasi atau standarisasi data, Mengubah data kategorikal menjadi numerik dan Menghapus outliers (data yang terlalu ekstrem). Untuk **Implementasi Metode K-Means** ini dapat Menggunakan algoritma K-Means untuk melakukan clustering atau pengelompokan data pelanggan. Pada tahap ini, data akan dikelompokkan ke dalam sejumlah cluster berdasarkan kesamaan fitur-fitur yang dimiliki. **Evaluasi Model** ini Menilai hasil clustering yang dilakukan oleh algoritma K-Means. Dalam Evaluasi ini dapat dilakukan menggunakan berbagai metrik seperti Silhouette Score, Davies-Bouldin Index, atau metode Elbow untuk menentukan kualitas cluster yang terbentuk. **Evaluasi Memadai** Memeriksa apakah hasil evaluasi model sudah memadai atau sesuai dengan kriteria yang diinginkan. **Jika tidak (No)**, maka proses kembali ke langkah seleksi parameter optimal untuk mencoba parameter yang berbeda. **Jika ya (Yes)**, lanjut ke langkah berikutnya. Untuk **Seleksi Parameter Optimal** dapat Menyesuaikan dan mencari parameter optimal untuk algoritma K-Means. Parameter yang biasanya disesuaikan adalah jumlah cluster (k). Proses ini mungkin melibatkan pengujian beberapa nilai k dan memilih yang memberikan hasil terbaik berdasarkan evaluasi sebelumnya. **Parameter Optimal Ditemukan** Menentukan apakah parameter optimal sudah ditemukan. **Jika tidak (No)**, kembali ke langkah implementasi metode K-Means untuk mencoba parameter lainnya. J**ika ya (Yes)**, lanjut ke langkah hasil segmentasi pelanggan. Untuk penyelesaian dari **Hasil Segmentasi Pelanggan ini** Menghasilkan dan menyajikan hasil akhir dari segmentasi pelanggan. Hasil ini dapat digunakan untuk memahami berbagai kelompok pelanggan yang ada dan membuat strategi pemasaran yang lebih tepat sasaran. **Selesai** Menandakan bahwa proses segmentasi pelanggan telah selesai.

Flowchart ini menunjukkan proses iteratif di mana evaluasi model dan seleksi parameter optimal dilakukan berulang-ulang sampai parameter yang paling sesuai ditemukan dan hasil segmentasi memadai.

## Hasil Klasifikasi

Untuk melakukan segmentasi customer dengan menerapkan k-means clustering. Kita harus paham terlebih dahulu sederhananya clustering adalah teknik dalam machine learning untuk mengelompokkan suatu data dalam kelompok-kelompok tertentu yang diukur berdasarkan jarak euclidean. Algoritma K-Means juga merupakan algoritma machine learning yang masuk dalam kategori unsupervised learning.

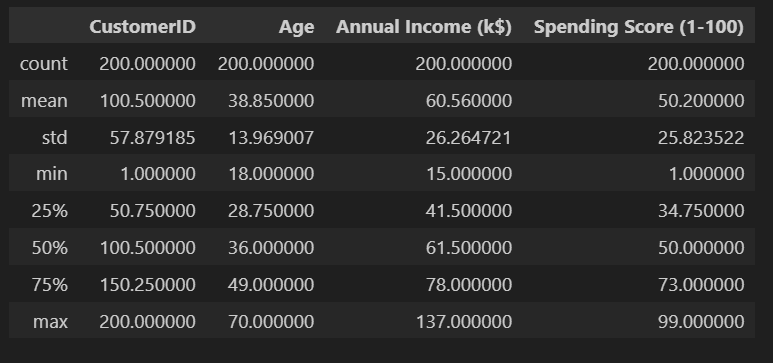


Gambar 3. 2 membaca dataset

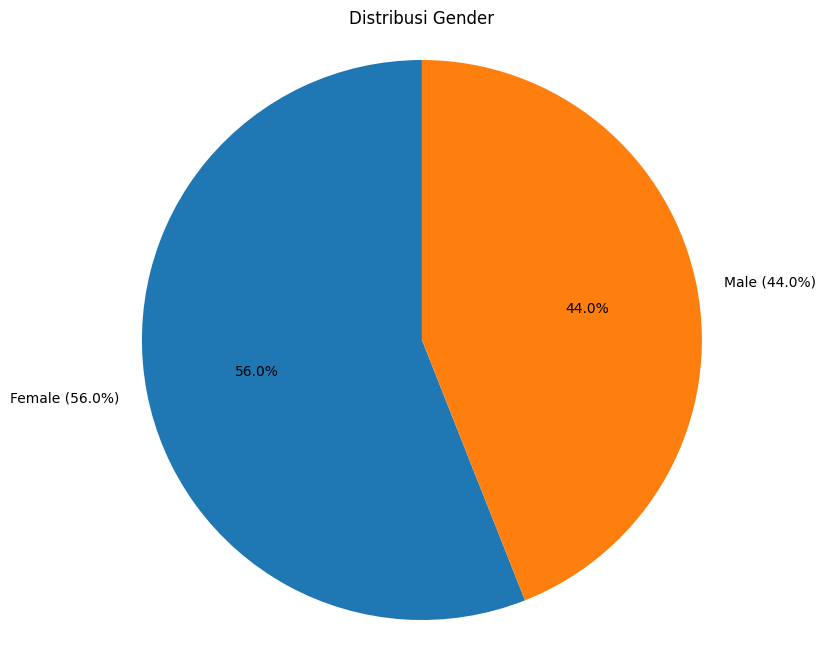
Gambar tersebut menampilkan sebuah tabel data yang berisi informasi tentang beberapa pelanggan. Setiap kolom dalam tabel ini memiliki peran penting. Kolom **CustomerID** digunakan untuk mengidentifikasi setiap pelanggan secara unik. Kolom **Gender** menunjukkan jenis kelamin pelanggan, di mana "Male" berarti laki-laki dan "Female" berarti perempuan. Kolom **Age** berisi informasi mengenai usia pelanggan, sedangkan kolom **Annual Income (k$)** mencantumkan pendapatan tahunan pelanggan dalam ribuan dolar. Terakhir, kolom **Spending Score (1-100)** memberikan skor pengeluaran pelanggan, yang mungkin merupakan ukuran dari kebiasaan belanja atau kepuasan pelanggan, dengan rentang antara 1 hingga 100.

Dalam menganalisis data kita perlu memahami distribusi data pada tiap kolom. Deskripsi statistik akan membantu kita mendeteksi masalah yang mungkin ada seperti adanya *outliers* atau pencilan yang ekstrim dan deviasi atau penyimpangan yang besar.

Kita dapat menggunakan method **describe()** untuk menampilkan deskripsi statistik dari dataset. **describe()**akan menampilkan jumlah data di setiap kolom (*count*), rata-rata nilai tiap kolom (*mean*), standard deviasi (*std*), nilai minimum (*min*), nilai maksimum (*max*), serta batas nilai dari masing-masing kuartil (25%, 50%, 75%).

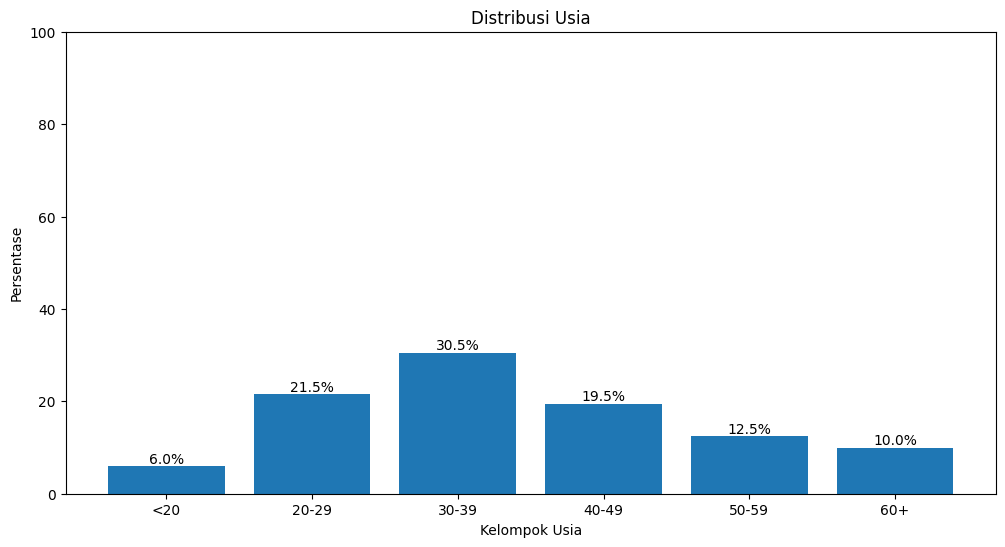


Gambar 3. 3 deskripsi statistik dataset



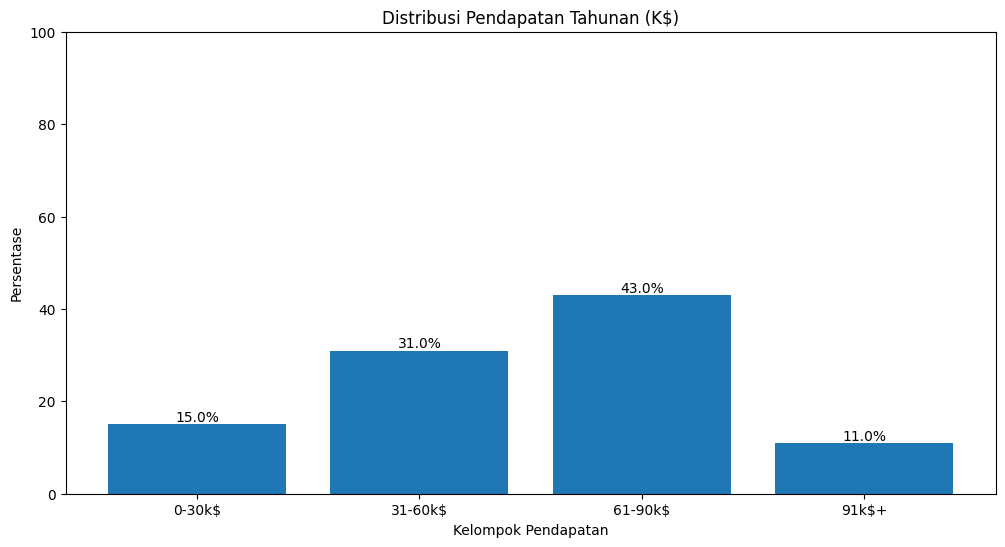
Gambar 3. 4 Distribusi Gender

Sebuah diagram pie yang menggambarkan distribusi gender dari data pelanggan. Diagram ini dibagi menjadi dua bagian yang menunjukkan persentase pelanggan laki-laki dan perempuan. Bagian berwarna biru menunjukkan bahwa 56% dari total pelanggan adalah perempuan, sementara bagian berwarna oranye menunjukkan bahwa 44% dari total pelanggan adalah laki-laki. Dengan judul "Distribusi Gender", diagram ini memberikan visualisasi yang jelas tentang bagaimana proporsi gender di antara pelanggan. Dari data ini, dapat disimpulkan bahwa jumlah pelanggan perempuan lebih banyak dibandingkan dengan pelanggan laki-laki.



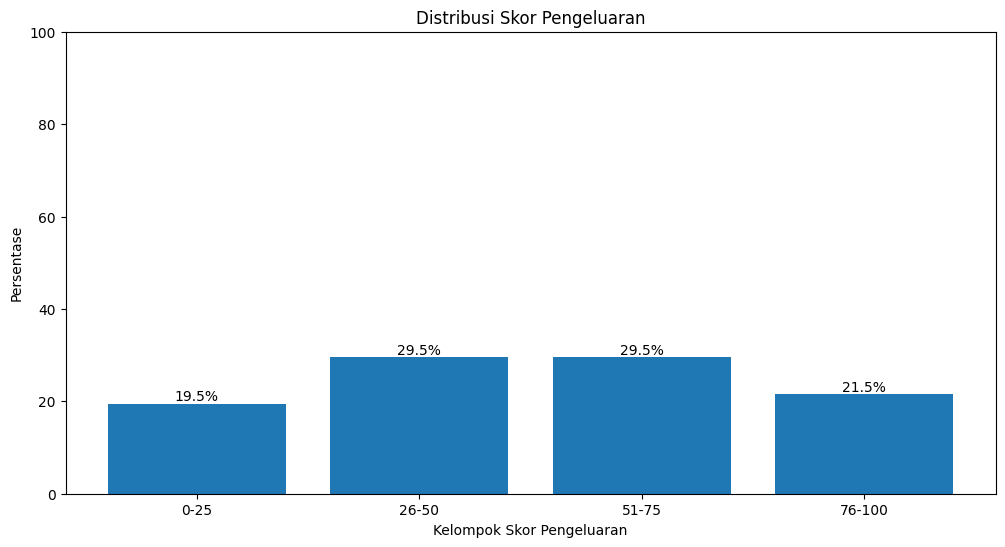
Gambar 3. 5 Distribusi Usia

Diagram batang menunjukkan distribusi persentase usia pelanggan dalam kelompok usia yang berbeda. Berdasarkan dataset, persentase distribusi usia adalah sebagai berikut: <20 tahun sebesar 2.0%, 20-29 tahun sebesar 23.0%, 30-39 tahun sebesar 31.5%, 40-49 tahun sebesar 17.5%, 50-59 tahun sebesar 14.0%, dan 60+ tahun sebesar 12.0%. Perbandingan ini menunjukkan bahwa kelompok usia 30-39 memiliki persentase terbesar, sementara kelompok usia di bawah 20 tahun memiliki persentase terkecil. Meskipun ada beberapa perbedaan kecil, distribusi dalam diagram batang cukup mencerminkan data yang ada dalam dataset.



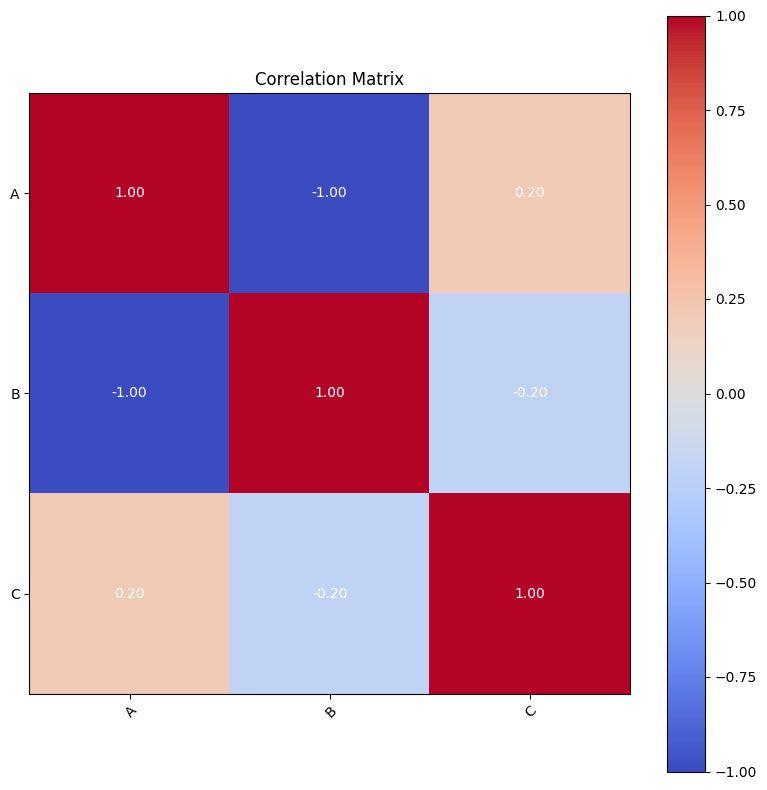
Gambar 3. 6 Distribusi Pendapatan Tahunan (K$)

Grafik ini menunjukkan distribusi pendapatan tahunan dalam ribuan dolar untuk empat kelompok pendapatan. Kelompok dengan pendapatan antara 61-90 ribu dolar merupakan yang terbesar, mencakup 43% dari populasi, diikuti oleh kelompok 31-60 ribu dolar dengan 31%. Kelompok pendapatan 0-30 ribu dolar mencakup 15% dari populasi, sementara kelompok dengan pendapatan lebih dari 91 ribu dolar adalah yang terkecil, mencakup hanya 11%. Hal ini menunjukkan bahwa sebagian besar populasi memiliki pendapatan tahunan di kisaran 61-90 ribu dolar, dengan proporsi yang lebih kecil berada di ujung rendah dan tinggi dari spektrum pendapatan.



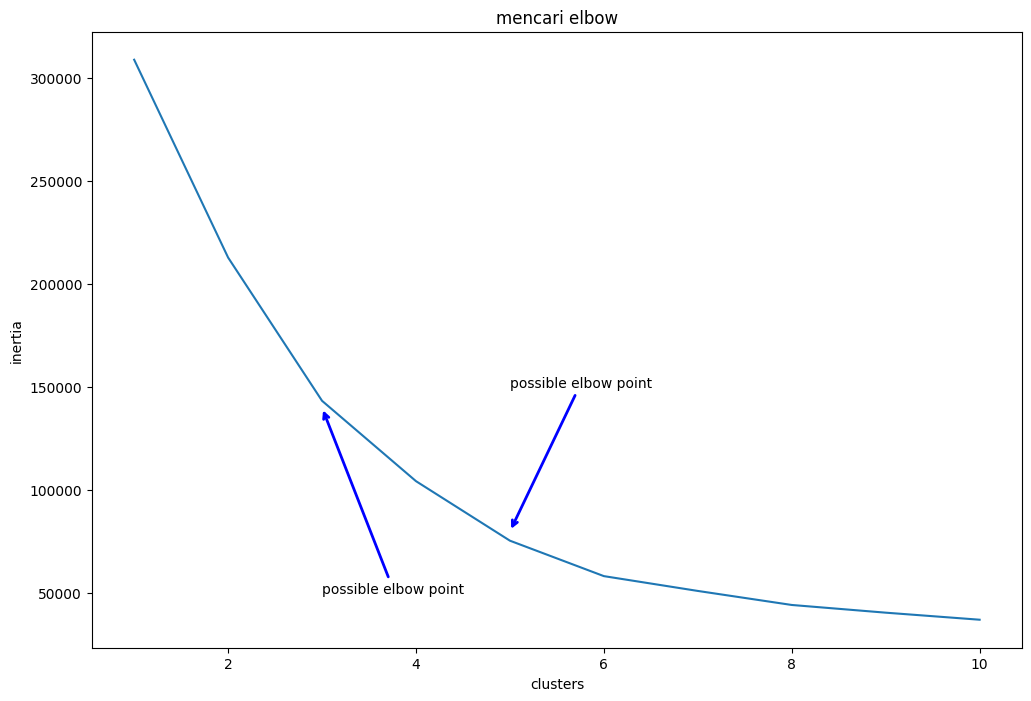
Gambar 3. 7 Distribusi Skor Pengeluaran

Grafik ini menunjukkan distribusi skor pengeluaran dalam persentase berdasarkan empat kelompok skor. Kelompok dengan skor pengeluaran 26-50 dan 51-75 masing-masing merupakan yang terbesar, mencakup 29.5% dari populasi. Kelompok dengan skor 76-100 mencakup 21.5%, sedangkan kelompok dengan skor 0-25 adalah yang terkecil, mencakup 19.5%. Ini menunjukkan bahwa sebagian besar populasi memiliki skor pengeluaran yang berada di tengah-tengah rentang skor, dengan distribusi yang relatif merata antara kelompok lainnya.



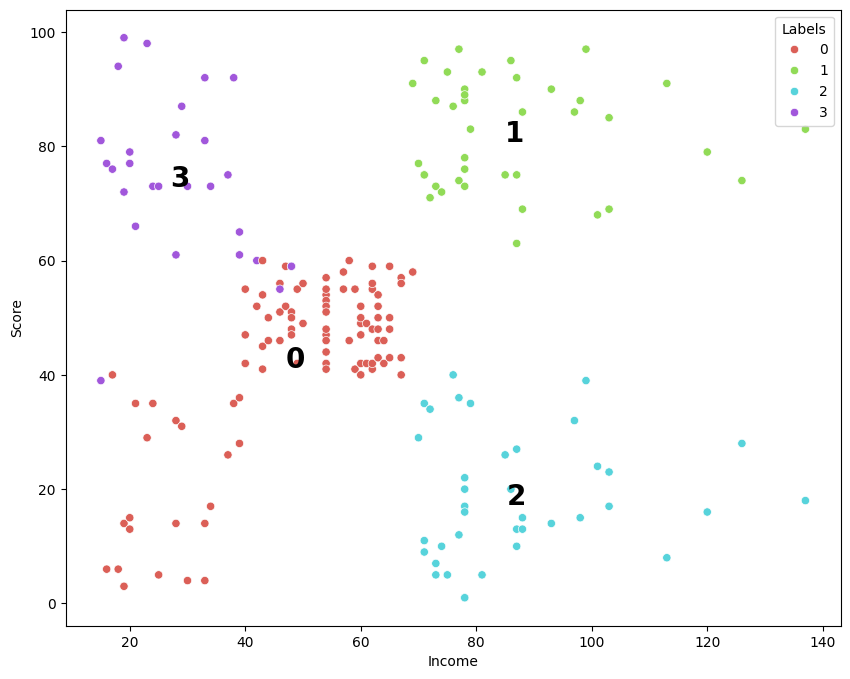
Gambar 3. 8 matriks korelasi

Matriks korelasi yang menunjukkan hubungan antara tiga variabel (A, B, dan C). Nilai pada diagonal utama adalah 1, menunjukkan korelasi sempurna dengan dirinya sendiri. Variabel A dan B memiliki korelasi -1, menunjukkan hubungan negatif sempurna, di mana peningkatan A selalu diikuti penurunan B. Variabel A dan C memiliki korelasi 0.20, menunjukkan hubungan positif lemah, sementara B dan C memiliki korelasi -0.20, menunjukkan hubungan negatif lemah. Skala warna di sebelah kanan menggambarkan nilai korelasi, dengan warna merah tua untuk korelasi positif sempurna (+1) dan biru tua untuk korelasi negatif sempurna (-1), serta gradasi warna yang menunjukkan variasi nilai korelasi dari negatif ke positif.



Gambar 3. 9 Grafik Elbow Method

Grafik "Elbow Method" yang digunakan untuk menentukan jumlah klaster optimal dalam algoritma klasterisasi seperti K-means. Sumbu X menunjukkan jumlah klaster dari 1 hingga 10, sementara sumbu Y menunjukkan nilai inertia, yang mengukur seberapa tersebar data dalam klaster. Garis plot memperlihatkan perubahan inertia seiring bertambahnya jumlah klaster, dengan penurunan yang signifikan pada awalnya dan kemudian melambat. Dua tanda panah biru menunjukkan "possible elbow points" pada jumlah klaster 3 dan 5, di mana penurunan inertia mulai melambat secara signifikan. Titik-titik ini mengindikasikan jumlah klaster optimal, karena menambah klaster lebih lanjut tidak memberikan perbaikan yang berarti dalam kompaksi klaster. Secara keseluruhan, grafik ini membantu memilih jumlah klaster optimal dengan mengidentifikasi titik "elbow" di mana penurunan inertia mulai melambat, menunjukkan keseimbangan terbaik antara kompaksi klaster dan kompleksitas model.



Gambar 3. 10 scatter plot

Scatter plot yang menunjukkan hasil klasterisasi data menggunakan algoritma seperti K-means, dengan sumbu X menunjukkan "Income" (Pendapatan) dan sumbu Y menunjukkan "Score" (Skor). Data ini dikelompokkan ke dalam empat klaster yang ditandai dengan warna berbeda dan label (0, 1, 2, 3). Klaster 0 (merah) berisi individu dengan pendapatan dan skor sedang (40-60), klaster 1 (hijau) berisi individu dengan pendapatan dan skor tinggi (70-100 pendapatan dan 60-90 skor), klaster 2 (biru muda) berisi individu dengan pendapatan tinggi tetapi skor rendah (70-100 pendapatan dan 20-40 skor), dan klaster 3 (ungu) berisi individu dengan pendapatan rendah tetapi skor tinggi (20-40 pendapatan dan 60-100 skor). Scatter plot ini membantu dalam visualisasi distribusi data berdasarkan dua variabel, serta mengidentifikasi pola atau kelompok dalam data yang memiliki karakteristik serupa untuk tujuan segmentasi pasar atau analisis data lainnya.

## Implementasi Interface

Dibawah ini merupakan Implementasi interface tampilan menu yang ada pada Segmentasi Pelanggan Menggunakan Metode K-Means Clustering dapat terlihat pada Gambar berikut.

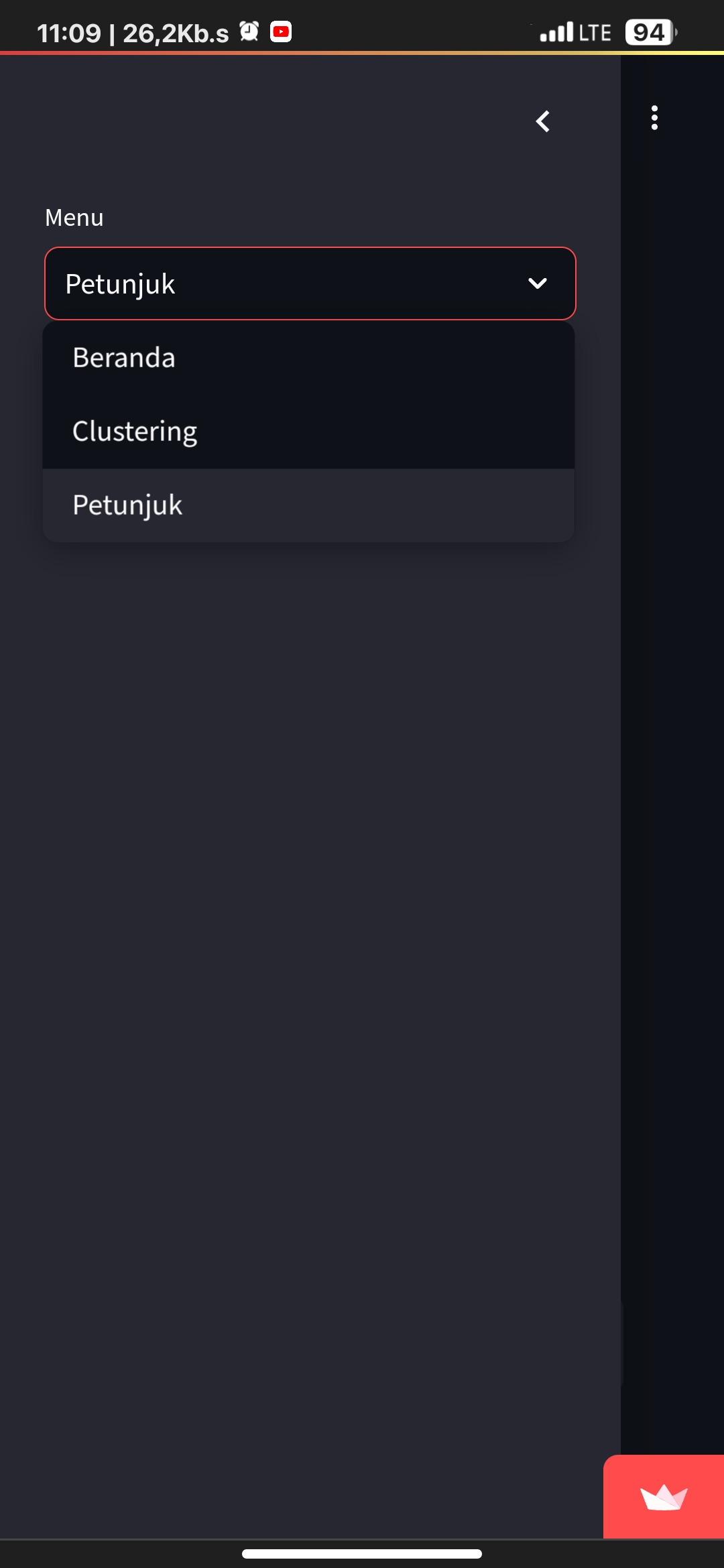
### Home Page



Gambar 3. 11 Halaman Beranda

Jadi, pada bagian Home page ini berisi tentang judul aplikasi ini dan pembuat aplikasi dimana tertera dibuat oleh Kelompok 2, serta program studi dan nama universitas.

### Navigasi Menu



Gambar 3. 12 Tampilan Navigasi Menu

Aplikasi ini memiliki tiga menu utama yang dapat diakses melalui sidebar diantaranya adalah Beranda, Clustering, dan Petunjuk.

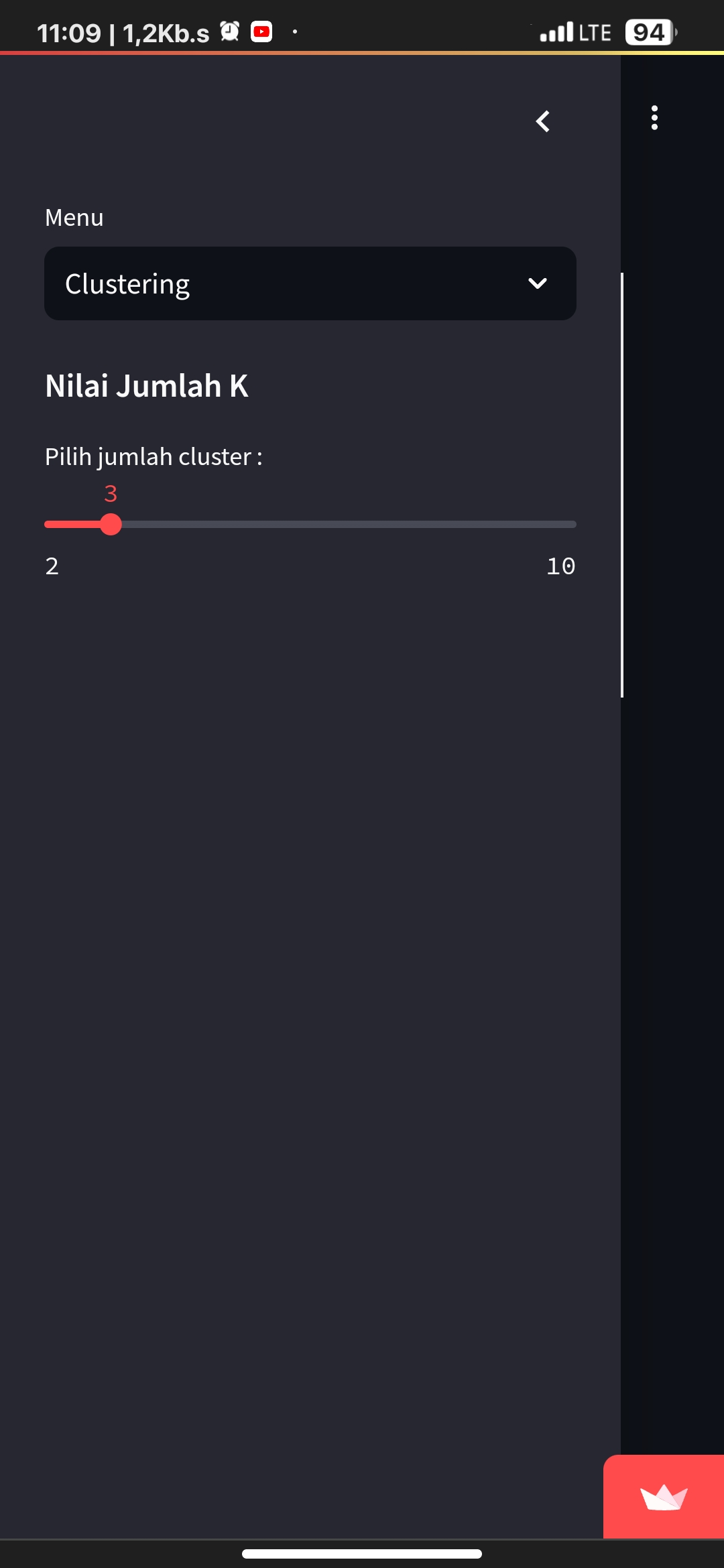
### Clustering

|  |  |
| --- | --- |
| C:\Users\Administrator\AppData\Local\Microsoft\Windows\INetCache\Content.Word\Screenshot_2024-07-03-11-09-28-674_com.android.chrome.jpg | C:\Users\Administrator\AppData\Local\Microsoft\Windows\INetCache\Content.Word\Screenshot_2024-07-03-11-09-34-526_com.android.chrome.jpg |

Gambar 3. 13 Tampilan Menu Clustering

Tampilan "Clustering" menampilkan proses dan hasil dari analisis clustering menggunakan metode K-Means pada dataset pelanggan mall. Tampilan ini mencakup judul bagian dan tabel yang berisi dataset, grafik Elbow yang membantu menentukan jumlah cluster optimal, serta sidebar untuk memilih jumlah cluster (K) yang digunakan dalam algoritma K-Means. Setelah jumlah cluster dipilih, scatter plot ditampilkan untuk mengilustrasikan hasil clustering, di mana data pelanggan dikelompokkan ke dalam cluster yang berbeda berdasarkan pendapatan tahunan dan skor pengeluaran mereka. Plot ini menunjukkan distribusi data dengan setiap cluster diwakili oleh warna berbeda.

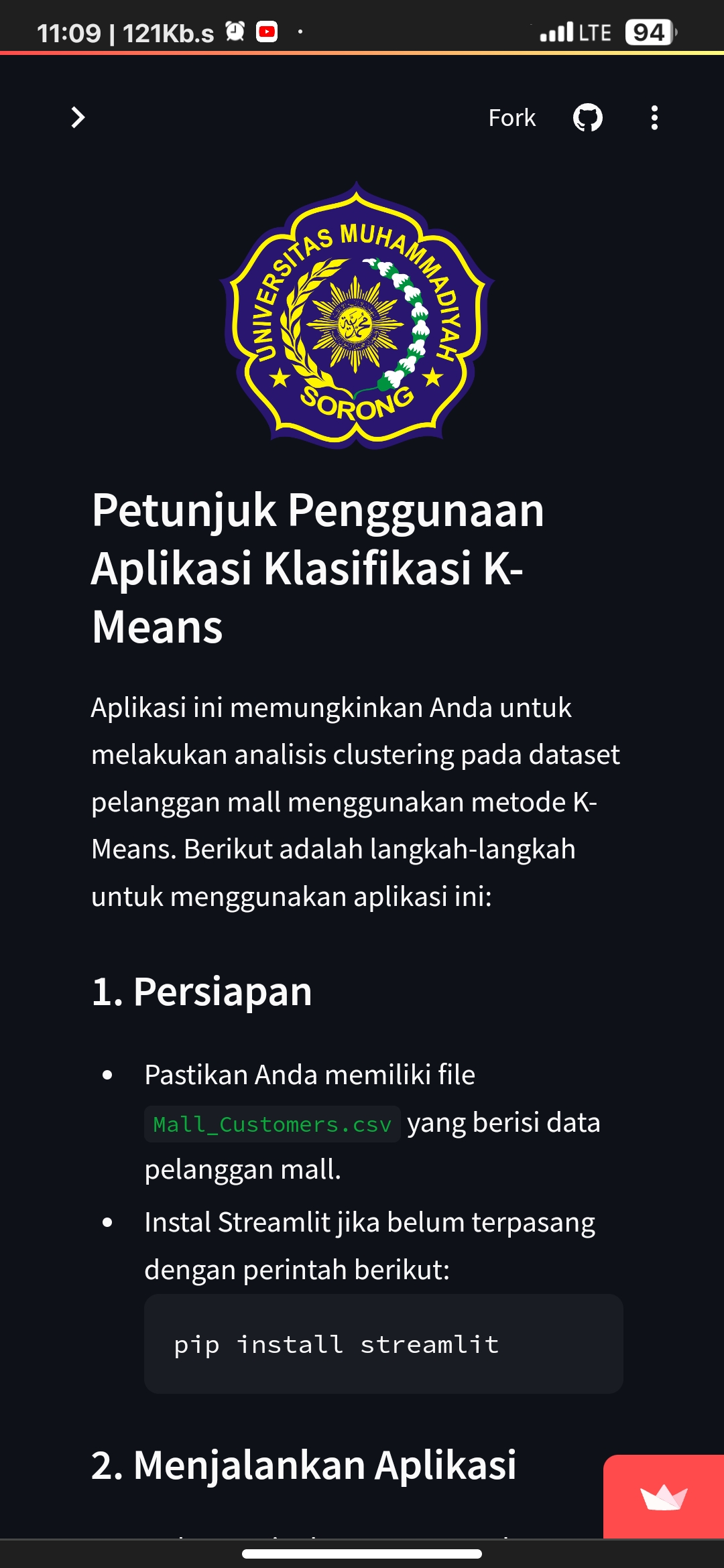
### Slider Memilih Jumlah Cluster



Gambar 3. 14 Slider Memilih Jumlah Cluster

Slider untuk memilih jumlah cluster dalam aplikasi Streamlit ini adalah fitur yang memungkinkan pengguna menentukan jumlah cluster (K) yang akan digunakan dalam algoritma K-Means untuk analisis clustering. Slider ini berada di sidebar aplikasi, sehingga mudah diakses tanpa mengganggu tampilan utama. Pengguna dapat memilih jumlah cluster yang diinginkan untuk analisis K-Means dengan rentang nilai dari 2 hingga 10.

1. Petunjuk



Gambar 3. 15 Tampilan Petunjuk

Menu "Petunjuk" ini menyediakan panduan lengkap bagi pengguna tentang cara menggunakan aplikasi untuk analisis clustering menggunakan metode K-Means. Saat menu "Petunjuk" dipilih, pengguna akan disajikan informasi detail mengenai langkah-langkah yang harus diikuti mulai dari persiapan hingga penggunaan fitur-fitur dalam aplikasi.

## Pengujian

Tabel 3. 2 Pengujian Pada User

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Kasus Uji** | **Deskripsi** | **Langkah** | **Jalur Kode** | **Hasil Pengujian** | **Kesimpulan** |
| Respon Tombol Beranda | Menampilkan halaman beranda dengan logo dan informasi aplikasi. | 1. Pilih menu "Beranda" dari sidebar.  2. Periksa apakah halaman menampilkan logo universitas, judul aplikasi, dan informasi kelompok dan universitas. | Menu Beranda | Sesuai Harapan | [√] Valid  [ ] Invalid |
| Respon Tombol Clustering | Menampilkan hasil clustering pelanggan berdasarkan pendapatan dan skor pengeluaran. | 1. Pilih menu "Clustering" dari sidebar.  2. Periksa apakah data pelanggan ditampilkan.  3. Periksa apakah grafik Elbow ditampilkan dengan benar. | Menu Clustering | Sesuai Harapan | [√] Valid  [ ] Invalid |
| Pengujian Slider | Memastikan slider jumlah cluster berfungsi dengan baik dan dapat menghasilkan hasil clustering yang sesuai. | 1. Pilih menu "Clustering" dari sidebar.  2. Geser slider untuk memilih jumlah cluster.  3. Periksa apakah hasil clustering berubah sesuai dengan slider. | Menu Clustering | Sesuai Harapan | [√] Valid  [ ] Invalid |
| Respon Tombol Petunjuk | Menampilkan halaman petunjuk penggunaan aplikasi clustering K-Means. | 1. Pilih menu "Petunjuk" dari sidebar.  2. Periksa apakah halaman menampilkan logo universitas.  3. Periksa apakah panduan penggunaan ditampilkan. | Menu Petunjuk | Sesuai Harapan | [√] Valid  [ ] Invalid |

## Usability Tasting

Tabel 3. 3 Usability Tasting

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **No** | **Tugas** | **Waktu yang Diharapkan** | **Waktu yang Dibutuhkan** | **Keberhasilan (Ya/Tidak)** | **Komentar/Kesulitan yang Dihadapi** |
| 1 | Menampilkan halaman Beranda | 1 Detik | 2 Detik | Ya | Proses navigasi berjalan lancar |
| 2 | Menampilkan halaman Clustering | 1 Detik | 2 Detik | Ya | Proses navigasi berjalan lancar. |
| 3 | Menampilkan isi dataset | 2 Detik | 2 Detik | Ya | Proses berjalan lancar. |
| 4 | Menampilkan grafik elbow | 2 Detik | 2 Detik | Ya | Proses berjalan lancar |
| 5 | Menampilkan hasil clustering | 2 Detik | 2 Detik | Ya | |  | | --- | |  |  |  | | --- | | Proses berjalan lancar | |
| 6 | Menampilkan halaman Petunjuk | 1 Detik | 1 Detik | Ya | Proses navigasi berjalan lancar |

# 

# BAB IV PENUTUP

## Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan beberapa hal penting. Pertama, penerapan metode K-Means Clustering dalam segmentasi pelanggan terbukti efektif dalam mengelompokkan pelanggan berdasarkan karakteristik tertentu. Hal ini memudahkan dalam analisis lebih lanjut mengenai kebutuhan dan preferensi pelanggan, serta membantu dalam penyusunan strategi pemasaran yang lebih tepat sasaran.

Selanjutnya, penggunaan dataset yang relevan dan teknik pengujian yang sistematis, termasuk pengujian usability, berhasil meningkatkan kinerja sistem aplikasi yang dikembangkan. Navigasi yang lancar dan responsifitas yang cepat merupakan indikator keberhasilan implementasi ini. Hal ini menunjukkan bahwa sistem yang dikembangkan telah memenuhi standar usability yang baik.

Selain itu, melalui analisis data yang dilakukan, ditemukan pola-pola tertentu dalam perilaku pelanggan yang dapat dimanfaatkan untuk meningkatkan layanan dan produk yang ditawarkan. Segmentasi yang dilakukan memberikan wawasan yang lebih mendalam tentang karakteristik setiap kelompok pelanggan, yang dapat menjadi dasar dalam pengambilan keputusan strategis di masa depan.

## Saran

Untuk pengembangan lebih lanjut, terdapat beberapa saran yang dapat dipertimbangkan. Pertama, sistem yang telah dikembangkan dapat ditingkatkan dengan menambahkan fitur analitik lanjutan, seperti prediksi tren pelanggan dan analisis sentimen, untuk memberikan nilai tambah bagi pengguna. Fitur-fitur ini akan memungkinkan pengelola untuk lebih proaktif dalam merespon kebutuhan pasar.

Selain itu, penggunaan dataset yang lebih lengkap dan berkualitas tinggi akan memperbaiki akurasi segmentasi dan hasil analisis. Diperlukan upaya untuk mengumpulkan data yang lebih beragam dan representatif, sehingga hasil analisis dapat mencerminkan kondisi yang lebih akurat dan relevan.

Terakhir, menguji sistem yang dikembangkan dalam kondisi nyata di lapangan akan memberikan feedback yang lebih realistis dan dapat diandalkan untuk perbaikan lebih lanjut. Melibatkan pengguna akhir dalam proses pengujian akan membantu dalam menemukan dan mengatasi kekurangan yang mungkin ada, serta memastikan bahwa sistem yang dikembangkan benar-benar memenuhi kebutuhan pengguna.

# DAFTAR PUSTAKA

Adiana, Beta Estri, Indah Soesanti, dan Adhistya Erna Permanasari. 2020. “Analisis Segmentasi Pelanggan Menggunakan Kombinasi Rfm Model Dan Teknik Clustering.” *Jurnal Terapan Teknologi Informasi* 2(1): 23–32. doi:10.21460/jutei.2018.21.76.

Afthoni, Rizqa, Mirza Hamdhani, Ardianto Ardianto, Aprilia Fitri Karimah, dan Harry Patria. 2021. “Pemanfaatan Algoritma Machine Learning untuk Segmentasi Pelanggan Berbasis Data Konsumsi Listrik di PT PLN XYZ.” *Seminar Nasional Teknik dan Manajemen Industri* 1(1): 222–31. doi:10.28932/sentekmi2021.v1i1.85.

Aktas, Asmin Alev, Okan Tunali, dan Ahmet Tugrul Bayrak. 2021. “Comparative unsupervised clustering approaches for customer segmentation.” *Proceedings - 2021 2nd International Conference on Computing and Data Science, CDS 2021*: 530–35. doi:10.1109/CDS52072.2021.00097.

Aliyev, Musadig, Elvin Ahmadov, Habil Gadirli, Arzu Mammadova, dan Emin Alasgarov. 2020. “Segmenting Bank Customers via RFM Model and Unsupervised Machine Learning.” http://arxiv.org/abs/2008.08662.

Awalina, Eriskiannisa Febrianty Luchia, dan Woro Isti Rahayu. 2023. “Optimalisasi Strategi Pemasaran dengan Segmentasi Pelanggan Menggunakan Penerapan K-Means Clustering pada Transaksi Online Retail.” *Jurnal Teknologi dan Informasi* 13(2): 122–37. doi:10.34010/jati.v13i2.10090.

Bank, Commercial. 2024. “CBEBirr Customer Segmentation Using Machine Learning in To The Faculty of Informatics Of In Partial Fulfilment of the Requirements For the Degree of Master of Science In.”

Christian, Yefta, dan Katherine Oktaviani Yap Rui Qi. 2022. “Penerapan K-Means pada Segmentasi Pasar untuk Riset Pemasaran pada Startup Early Stage dengan Menggunakan CRISP-DM.” *JURIKOM (Jurnal Riset Komputer)* 9(4): 966. doi:10.30865/jurikom.v9i4.4486.

Febriani, Atik, dan Syahfara Ashari Putri. 2020. “Segmentasi Konsumen Berdasarkan Model Recency, Frequency, Monetary dengan Metode K-Means.” *JIEMS (Journal of Industrial Engineering and Management Systems)* 13(2): 52–57. doi:10.30813/jiems.v13i2.2274.

Ferdyandi, Muhammad, Nanang Yudi Setiawan, dan Fitra Abdurrachman Bachtiar. 2022. “Prediksi Potensi Penjualan Makanan Beku Berdasarkan Ulasan Pengguna Shopee Menggunakan Metode Decision Tree Algoritma C4.5 Dan Random Forest (Studi Kasus Dapur Lilis).” *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer* 6(2): 588–96. http://j-ptiik.ub.ac.id.

Hadi, Sofian Wira, Muhammad Fahmi Julianto, Syaifur Rahmatullah, dan Windu Gata. 2020. “Analisa Cluster Aplikasi Pada App Store Dengan Menggunakan Metode K-Means.” *Bianglala Informatika* 8(2): 86–90. doi:10.31294/bi.v8i2.8191.

Harahap, Mawaddah, Bintang P. A. Sihombing, Olga A. F. Laia, Bryan T. Saragih, dan Kuandi Dharma. 2021. “Analisis Sentimen Review Penjualan Produk Umkm Pada Kabupaten Nias Dengan Komparasi Algoritma Klasifikasi Machine Learning.” *METHOMIKA Jurnal Manajemen Informatika dan Komputerisasi Akuntansi* 5(2): 147–54. doi:10.46880/jmika.vol5no2.pp147-154.

Hindrayani, Kartika Maulida, Amalia Anjani, dan Afina Lina Nurlaili. 2021. “Penerapan Machine Learning pada Penjualan Produk UMKM : Studi Literatur.” *Prosiding Seminar Nasional Sains Data* 1(01): 19–23. https://prosiding-senada.upnjatim.ac.id/index.php/senada/article/view/7.

Jamalulael, Dzikri, dan Nurul Aina Aulia. 2023. “Analisa Dalam Memahami Kewirausahaan Dan Usaha Baru Pada.” (November). https://www.researchgate.net/profile/Nurul-Aulia-14/publication/375998660\_ANALISA\_DALAM\_MEMAHAMI\_KEWIRAUSAHAAN\_DAN\_USAHA\_BARU\_PADA\_UMKM\_DI\_ERA\_DIGITAL/links/6566d941ce88b870311d1107/ANALISA-DALAM-MEMAHAMI-KEWIRAUSAHAAN-DAN-USAHA-BARU-PADA-UMKM-DI-ERA-DIGI.

Johansson, Axel, dan Jonas Wikström. 2021. “Teknisk-naturvetenskapliga fakulteten Uppsala universitet, Utgivningsort Uppsala Customer segmentation using machine learning.”

John, Jeen Mary, Olamilekan Shobayo, dan Bayode Ogunleye. 2023. “An Exploration of Clustering Algorithms for Customer Segmentation in the UK Retail Market.” *Analytics* 2(4): 809–23. doi:10.3390/analytics2040042.

Joung, Junegak, dan Harrison Kim. 2023. “Interpretable machine learning-based approach for customer segmentation for new product development from online product reviews.” *International Journal of Information Management* 70(January): 102641. doi:10.1016/j.ijinfomgt.2023.102641.

Laksono, Bagaskoro Cahyo, dan Ika Yuni Wulansari. 2021. “Pemodelan Dan Penerapan Metode Rfm Pada Estimasi Nilai Konsumen (Customer Lifetime Value) Menggunakan K-Means Clustering Machine Learning.” *Seminar Nasional Official Statistics* 2020(1): 1277–85. doi:10.34123/semnasoffstat.v2020i1.689.

Mehta, Vaidisha, Ritvik Mehra, dan Sourabh Singh Verma. 2021. “A Survey on Customer Segmentation using Machine Learning Algorithms to Find Prospective Clients.” *2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions), ICRITO 2021*: 4–7. doi:10.1109/ICRITO51393.2021.9596118.

Monil, Patel. 2020. “Customer Segmentation using Machine Learnin.” *International Journal for Research in Applied Science and Engineering Technology* 8(6): 2104–8. doi:10.22214/ijraset.2020.6344.

Nabila, Zulfa, Auliya Rahman Isnain, Permata, dan Zaenal Abidin. 2021. “Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means.” *Jurnal Teknologi dan Sistem Informasi (JTSI)* 2(2): 100. http://jim.teknokrat.ac.id/index.php/JTSI.

Nur, Siti Anzila, dan Nursari S.R. Candra. 2022. “Pengujian Kesetaraan Partisi pada White Box (Studi Kasus : Electronic City).” *Journal of Informatics and Advanced Computing (JIAC)* 3(2): 97–102. https://journal.univpancasila.ac.id/index.php/jiac/article/view/4409.

Pemrograman, Algoritma D A N. 2020. “Pseudocode.” *Definitions*. doi:10.32388/tf77dy.

Pradana, Musthofa Galih, dan Hoang Thi Ha. 2021. “Maximizing Strategy Improvement in Mall Customer Segmentation using K-means Clustering.” *Journal of Applied Data Sciences* 2(1): 19–25. doi:10.47738/jads.v2i1.18.

Sharaf Addin, Eman Hussein, Novia Admodisastro, Siti Nur Syahirah Mohd Ashri, Azrina Kamaruddin, dan Yew Chew Chong. 2022. “Customer Mobile Behavioral Segmentation and Analysis in Telecom Using Machine Learning.” *Applied Artificial Intelligence* 36(1). doi:10.1080/08839514.2021.2009223.

Sudarsono, Bernadus Gunawan, Marcell Ignatius Leo, Ali Santoso, dan Felix Hendrawan. 2021. “Analisis Data Mining Data Netflix Menggunakan Aplikasi Rapid Miner.” *JBASE - Journal of Business and Audit Information Systems* 4(1): 13–21. doi:10.30813/jbase.v4i1.2729.

Tabianan, Kayalvily, Shubashini Velu, dan Vinayakumar Ravi. 2022. “K-Means Clustering Approach for Intelligent Customer Segmentation Using Customer Purchase Behavior Data.” *Sustainability (Switzerland)* 14(12): 1–15. doi:10.3390/su14127243.

Tauhid, Karimah, Japar Sodik, Mip Saripudin, Sri Meiylani, Uus Firdaus, Program Studi, Ilmu Komputer, dan Universitas Djuanda Bogor. 2024. “KAJIAN UNTUK SEGMENTASI CUSTOMER BANK DENGAN ALGORITMA K-MEANS.” 3: 3899–3906.

Yadav, Akash, Medam Anji Reddy, Andhra Pradesh, Aditya Sharma, dan Ali Abbas. 2023. “Smart Bin: a Comprehensive Solution for Scrap Management Leveraging Machine Learning for Customer Segmentation Revolutionizing Scrap Management: a Web-Based Platform for Efficient Scrap Management.” *International Research Journal of Modernization in Engineering Technology and Science*. doi:10.56726/irjmets34100.

Yuping, Zhou, Petra Jílková, Chen Guanyu, dan David Weisl. 2020. “New Methods of Customer Segmentation and Individual Credit Evaluation Based on Machine Learning.” 131(Nsrbcped 2019): 925–31. doi:10.2991/aebmr.k.200324.170.

**Di atas ini merupakan format hasil kutipan pada referensi**

# LAMPIRAN 1

Lampiran 1 Evaluasi Pengerjaan Tugas Besar

Kelompok 1:

1. Fei Hung M Makarapate : BAB 1, BAB 4
2. Arfin Nursanto : BAB 1, BAB 2, BAB 3, BAB 4, Jurnal, UI/UX, Develope Aplikasi
3. Aziz Gusti Pratama : BAB 2, BAB 3, BAB 4, Jurnal

Mengetahui Dosen Pengganti Mata Kuliah

Algoritma Pemrograman 2

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

# LAMPIRAN 2

Dokumentasi Kelompok



# LAMPIRAN 3

**FORM PENGISIAN PENGERJAAN TUGAS BESAR**

**Jenis Tugas:** Menerapkan Segmentasi Pelanggan Menggunakan Metode K-Means Clustering

|  |  |  |  |
| --- | --- | --- | --- |
| No | Hari/Tanggal | Kegiatan | Paraf |
| 1 | 23/04/2024 | Membuat BAB I |  |
| 2 | 30/04/2024 | Membuat BAB II |  |
| 3 | 04/05/2024 | Mencari Jurnal |  |
| 4 | 04/06/2024 | Membuat BAB III |  |
| 5 | 08/06/2024 | Mencari Dataset |  |
| 6 | 15/06/2024 | Membuat UI/UX |  |
| 7 | 22/06/2024 | Membuat BAB IV |  |
| 8 | 18/07/2024 | Menambahkan Link Github |  |
| 9 | 11/07/2024 | Develop Aplikasi |  |

# LAMPIRAN 4

Link Github :

<https://github.com/punyaarfin/MENERAPKAN-SEGMENTASI-PELANGGAN-MENGGUNAKAN-METODE-K-MEANS-CLUSTERING>