SEGMENTASI NASABAH BANK MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING DAN VISUALISASI DINAMIS DENGAN STREAMLIT

TUGAS AKHIR

Oleh :

Nimas Widyaningrum

NPM 20670070

PROGRAM STUDI S1 INFORMATIKA

FAKULTAS TEKNIK DAN INFORMATIKA

UNIVERSITAS PGRI SEMARANG

2024

SEGMENTASI NASABAH BANK MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING DAN VISUALISASI DINAMIS DENGAN STREAMLIT

TUGAS AKHIR

Diajukan kepada Fakultas Teknik dan Informatika Universitas PGRI Semarang untuk Penyusunan Skripsi

Oleh :

Nimas Widyaningrum

NPM 20670070

PROGRAM STUDI S1 INFORMATIKA

FAKULTAS TEKNIK DAN INFORMATIKA

UNIVERSITAS PGRI SEMARANG

2024

HALAMAN PERSETUJUAN

SKRIPSI

SEGMENTASI NASABAH BANK MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING DAN VISUALISASI DINAMIS DENGAN STREAMLIT

Disusun dan diajukan oleh

NIMAS WIDYANINGRUM

NPM 20670070

Telah disetujui olem pembimbing untuk dilanjutkan untuk di hadapan Dewan Penguji

Semarang, 24 Juli 2024

Pembimbing 1	Pembimbing 2
Mega Novita,S.Si.,M.Si., M.Nat.Sc.,Ph.D	Khoiriya Latifah, S.Kom., M. Kom.
NIDN. 061511880	NIDN. 0623058802

SKRIPSI

SEGMENTASI NASABAH BANK MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING DAN VISUALISASI DINAMIS DENGAN STREAMLIT

Disusun dan diajukan oleh

NIMAS WIDYANINGRUM

NPM 20670070

Telah dipertahankan di depan Dewan Penguji pada tanggal 29/7/2024 dan dinyatakan telah memenuhi syarat

Dewan Penguji

Ketua,

Ibnu Toto Husodo, S.T., M.

NIP/NPP 136901387

Sekretaris,

Bambang Agus H,S.Kom,M,Kom

NIP/NPP 14 8201433

Penguji 1,

Mega Novita,S.Si.,M.Si.,M.Nat.Sc.,Ph.D

NIP/NPP 158801493

Penguji 2,

Khoriyah Latifah,S.Kom, M.Kom

NIP/NPP 147801434

Penguji 3,

Aris Tri Joko Harjanto,S.Kom,M.Kom

NIP/NPP 1482014443

iv

MOTO DAN PERSEMBAHAN

Moto

"Tidak ada yang lebih indah dari pada melihat diri sendiri bahagia dan kembali semangat pada semua hal yang disukai, sembuh dari semua hal yang membuat sakit, bangkit dari semua keterpurukan dan tersenyum tulus kembali."

"Jadilah kuat, untuk semua hal yang membuatmu patah. Karena pada akhirnya, yang harus kita pelajari dari hidup adalah bagaimana menjadi kuat sendiri, tanpa ada bahu yang bisa untuk bersandar."

"Tidak ada proses yang mudah untuk tujuan yang indah, karena kita masih dalam zona berjuang. Takdir milik Allah SWT, tapi doa dan usaha milik kita."

Persembahan :

Saya persembahkan skripsi ini untuk :

1. Kedua orang tua saya yang selalu memberi support dan selalu memberi masukan dalam hidup saya.
2. Saudara-saudara saya
3. Teman-teman saya yang selalu memberi semangat
4. Almamaterku Universitas PGRI Semarang

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan dibawah ini:

Nama : Nimas Widyaningrum

NPM : 20670070

Progdi : Informatika

Fakultas : Teknik dan Informatika

Menyatakan dengan sebenarnya bahwa skripsi yang saya buat ini benar-benar merupakan hasil karya saya sendiri, bukan plagiarism.

Apabila pada kemudian hari skripsi ini terbukti hasil plagiarism, saya bersedia menerima sanksi atas perbuatan tersebut.

Semarang, 24 Juli 2024

Yang membuat pernyataan

Nimas Widyaningrum

NPM 20670070

ABSTRAK

Permaslahan dalam penelitian ini adalah, industri perbankan menghadapi persaingan yang semakin ketat, mendorong bank untuk mengembangkan strategi pemasaran yang efektif guna menarik dan mempertahankan nasabah. Segmentasi pasar merupakan salah satu strategi yang digunakan untuk mengidentifikasi dan mengelompokkan nasabah berdasarkan karakteristik tertentu, dengan tujuan meningkatkan efisiensi pemasaran dan penawaran produk yang tepat sasaran. Penelitian ini bertujuan untuk mengimplementasikan algoritma K-Means Clustering dalam segmentasi nasabah bank, mengembangkan aplikasi web interaktif menggunakan Streamlit untuk visualisasi hasil segmentasi, serta menyediakan alat bantu analisis yang memudahkan bank dalam menginterpretasikan hasil segmentasi dan mengambil keputusan strategis. Dalam penelitian ini, data nasabah bank yang mencakup informasi demografi, jumlah kredit, dan lama kredit dianalisis menggunakan algoritma K-Means Clustering untuk mengelompokkan nasabah ke dalam beberapa klaster berdasarkan kemiripan karakteristik. . Dengan batasan yang diterapkan, yaitu menggunakan data nasabah dari satu bank dan tidak mempertimbangkan faktor eksternal lainnya, penelitian ini memberikan kontribusi penting dalam penerapan teknik data mining dan pengembangan aplikasi visualisasi interaktif untuk meningkatkan strategi pemasaran di industri perbankan. Hasil penelitian menunjukkan bahwa algoritma K-Means Clustering efektif dalam mengelompokkan nasabah bank ke dalam klaster yang berbeda, berdasarkan karakteristik yang relevan. Hasil segmentasi tersebut kemudian divisualisasikan dalam bentuk dashboard interaktif menggunakan Streamlit, sehingga memudahkan pihak bank untuk memahami dan menganalisis hasil segmentasi secara lebih intuitif.

Kata kunci: K-Means Clustering, segmentasi nasabah bank, Streamlit, data mining.

PRAKATA

Puji syukur atas kehadirat Allah SWT yang telah memberikan rahmat dan hidayah-Nya, serta kedua orang tua saya yang selalu ada dan mendukung saya, peneliti dapat menyusun dan menyelesaikan skripsi ini dengan lancar. Skripsi yang berjudul "Segmentasi Nasabah Bank Menggunakan Algoritma K-Means Clustering dan Visualisasi Dinamis dengan Streanlit" ini disusun untuk memenuhi sebagai syarat memperoleh gelar sarjana Teknik.

Penyusunan skripsi ini tidak lepas dari hambatan dan rintanganserta kesulitan-kesulitan. Namun berkat bimbingan, bantuan, nasihat dan dorongan serta saran-saran dari berbagai pihak, khususnya Pembimbing, segala hambatan dan rintangan serta kesulitan tersebut dapat teratasi dengan baik. Oleh karena itu, dalam kesempatan ini dengan tulus hati penulis sampaikan terima kasih kepada:

1. Ibu Dr.Sri Suciati, M.Hum selaku Rektor Universitas PGRI Semarang, yang telah memberi kesempatan kepada penulis untuk menimba ilmu di Universitas PGRI Semarang.
2. Bapak Ibnu Toto Husodo, S.T., M.T. selaku Dekan Fakultas Teknik dan Informatika Universitas PGRI Semarang yang telah memberikan izin penulis untuk melakukan penelitian.
3. Bapak Bambang Agus Herlambang, S.Kom., M.Kom. selaku Ketua Program Studi Informatika Fakultas Teknik dan Informatika Universitas PGRI Semarang.
4. Ibu Mega Novita,S.Si.,M.Si., M.Nat.Sc.,Ph.D. selaku Pembimbing I yang telah mengarahkan penulis dengan penuh ketekunan dan kecermatan.
5. Ibu Khoiriya Latifah, S.Kom., M. Kom. selaku Pembmbing II yang telah membimbing penulis dengan penuh dedikasi yang tinggi.
6. Bapak dan Ibu Dosen Program Studi Informatika yang telah memberikan bekal ilmu kepada penulis selama belajar di Universitas PGRI Semarang.
7. Teruntuk saudara saya yang telah memberikan semangat dan masukan selama kuliah.

8. Teruntuk teman-teman terdekat penulis yang telah memberikan semangat, yang telah membantu penulis dan bertukar pikiran dalam penulisan dan penyusunan tugas akhir ini.

Akhir kata penulis mengucapkan terima kasih kepada semua pihak yang telah membantu dan memberikan dukungan. Penulis berharap semoga skripsi ini bermanfaat bagi pendidik, khususnya dibidang Informatika.

Semarang, 24 Juli 2024

Nimas Widyaningrum

20670070

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

DAFTAR LAMPIRAN

BAB I
PENDAHULUAN

A. Latar Belakang

Pada masa globalisasi saat ini suatu bank harus mampu bersaing serta mengikuti perubahan yang ada, khususnya saat ini lembaga keuangan juga berkembang pesat, faktor terjadinya perubahan yaitu dalam suatu bank yaitu karena terdapat persaingan yang semakin hari semakin banyak dan beragam. Dalam persaingan yang ketat tentu bank tersebut mempunyai cara dalam menghadapinya, seperti strategi yang cocok untuk digunakan bila strategi itu tepat maka menguntungkan bagi bank jika tidak maka akan sebaliknya. Untuk memperoleh keuntungan di bank, suatu bank harus ada sebuah strategi pemasaran yang merupakan langkah awal dalam pengenalan produk kepada nasabah. Hal tersebut bisa maksimal jika didukung dengan perencanaan yang baik, baik secara eksternal dan internal. Segmentasi pasar adalah strategi yang banyak digunakan karena segmentasi pasar berhubungan dengan penentuan posisi, dan penetapan pasar sehingga di persepsikan produk suatu perusahaan unik dan unggul dibandingkan dari yang lainnya [1].

Dalam menarik minat nasabah perusahaan harus dapat menentukan posisi pasar yang sesuai hal tersebut harus sesuai dengan apa yang nasabah inginkan sehingga nasabah tertarik dalam membeli produk dan jasa tersebut. Kejayaan suatu bank tergantung banyak dan tidaknya jumlah nasabah, suatu bank mengalami kenaikan dan penurunan nasabah pasti ada sebab dan akibatnya. Segmentasi pasar dilakukan agar perusahaan mendapatkan hasil yang maksimal, salah satu caranya yaitu perusahaan sendiri mampu melihat kemampuan perusahaan dalam melakukan segmentasi pasar sendiri. Setelah mengerti luas pasar yang ada serta jumlahnya langkah berikutnya yaitu menentukan saran pasar yang sesuai. Segmentasi perlu ada variabel utama dalam melakukan pertimbangan yang diperhatikan. Penetapan harga, menyerahkan serta menyiapkan produk berserat jasa untuk memutuskan sasaran pasar [2].

Segmentasi nasabah merupakan proses aktual untuk mengidentifikasi atau membagi basis nasabah yang luas menjadi sub-kelompok nasabah berdasarkan variabel dari data nasabah. Sebagian besar bank memiliki nasabah yang besar dengan karakteristik berbeda dalam hal usia, pendapatan, nilai, gaya hidup, dan banyak lagi.Segmentasi nasabah dapat diterapkan berdasarkan data nasabah dari internet banking. *Clustering* merupakan teknik *unsupervised* data mining yang dapat digunakan untuk melakukan segmentasi nasabah bank berdasarkan data demografi, jumlah kredit, dan lama kredit, sehingga menghasilkan segmentasi pasar nasabah yang sesuai target pemasaran [3]. Data mining adalan proses yang menggunakan teknik *stitisk*, matematika, kecerdasan buatan dan machine learning untuk mengekstrasikan dan mengidentifikasi informasi terkait dari berbagai database besar [4].

Dalam segmentasi nasabah seringkali menggunakan algoritma K-Means *Clustering* sebagai metode untuk membatu pengelompokan menjadi beberapa klaster sehingga mendapatkan visualisasi data yang signifikan hasilnya. Algoritma K-Means *clustering* merupakan salah satu teknik data mining yang dapat digunakan untuk mengelompokkan berdasarkan kemiripan karakteristik tertentu dimana data-data yang memiliki kemiripan akan berada pada klaster yang sama [5]. K-Means merupakan salah satu metode dalam data mining yang digunakan untuk mempartisi data yang ada kedalam beberapa cluster sehingga data yang memiliki karakteristik yang sama akan dikelompokkan kedalam satu cluster dan data dengan karakteristik yang berbeda akan dikelompokkan kedalam *cluster* lain. Dari data yang dianalisa dengan algoritma K-Means *Clustering* akan menghasil satu tujuan yaitu mendapatkan kelompok data yang akan dipromosikan [6].

Namun, hasil dari segmentasi ini sering kali sulit dipahami dan dianalisis tanpa visualisasi yang memadai. Oleh karena itu, diperlukan alat yang mampu menyajikan hasil segmentasi dalam bentuk yang mudah dipahami dan interaktif. Streamlit adalah salah satu *framework* Python yang memungkinkan pengembangan aplikasi web interaktif dengan mudah dan cepat. Dengan

Streamlit, hasil segmentasi dapat divisualisasikan dalam bentuk dashboard yang dinamis, sehingga memudahkan analisis dan pengambilan keputusan.

Berdasarkan uraian diatas maka dilakukan analisis segnentasi nasabah dengan menggunakan metode *K-Means Clustering*. Maka penulis mengangkat topik penelitian "**SEGMENTASI NASABAH BANK MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING DAN VISUALISASI DINAMIS DENGAN STREAMLIT"** topic penelitian ini menekankan pada analisis segmentasi nasabah berdasarkan data pengelompokan dalam beberapa klaster dan menentukan *customer profiling*.

B. Identifikasi Masalah

Berdasarkan latar belakang yang telah diuraikan, beberapa masalah yang diidentifikasi dalam penelitian ini adalah:

1. Bank memerlukan cara untuk mengenali dan memahami karakteristik serta kebutuhan setiap nasabahnya agar dapat menawarkan produk dan layanan yang sesuai.
2. Implementasi dan mengintegrasi Algoritma K-Means untuk pemahaman mendalam tentang cara kerja dan penentuan jumlah cluster yang optimal dan pemilihan fitur yang relevan dan diintegrasikan ke dalam platform visualisasi interaktif memerlukan keterampilan teknis dalam pemrograman dan pengembangan aplikasi web.
3. Diperlukan metode untuk memvalidasi dan mengevaluasi hasil segmentasi untuk memastikan bahwa *cluster* yang dihasilkan benar-benar mencerminkan segmen nasabah yang berbeda.

C. Rumusan Masalah

Berdasarkan indentifikasi masalah di atas, rumusan masalah dalam penelitian ini adalah:

1. Bagaimana cara mengimplementasikan algoritma K-Means *Clustering* untuk segmentasi nasabah bank?

2. Bagaimana cara mengintegrasikan hasil segmentasi nasabah ke dalam aplikasi web interaktif menggunakan Streamlit?
3. Bagaimana visualisasi hasil segmentasi dapat membantu dalam pengambilan keputusan strategi di bank?

D. Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Mengimplementasikan algoritma K-Means *Clustering* untuk melakukan segmentasi nasabah bank berdasarkan karakteristik tertentu.
2. Mengembangkan aplikasi web interaktif menggunakan Streamlit untuk visualisasi hasil segmentasi nasabah.
3. Menyediakan alat bantu analisis yang memudahkan pihak bank dalam menginterpretasikan hasil segmentasi dan mengambil keputusan strategis.

E. Batasan Masalah

Agar penelitian ini lebih terfokus dan dapat diselesaikan dengan baik, terdapat beberapa batasan yang diterapkan:

1. Data yang digunakan dalam penelitian ini terbatas pada data nasabah dari suatu bank.
2. Segmentasi nasabah hanya dilakukan berdasarkan data yang tersedia seperti data demografi dan transaksi, tanpa mempertimbangkan faktor eksternal lainnya.
3. Aplikasi yang dikembangkan menggunakan Streamlit hanya mencakup visualisasi dasar dan hasil segmentasi tanpa fitur analisis lanjutan.

F. Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Bagi Bank

   Memperoleh pemahaman yang lebih baik mengenai karakteristik dan kebutuhan setiap segmen nasabah, sehingga dapat meningkatkan layanan dan strategi pemasaran. Dengan visualisasi interaktif hasil segmentasi memudahkan pihak manajemen bank dalam menganalisis data

dan mengambil keputusan strategi berdasarkan informasi yang akurat dan mudah dipahami. Dengan memahami kebutuhan dan preferensi masing-masing segmen nasabah, bank dapat menyediakan layanan yang lebih personal dan meningkatkan pengalaman nasabah secara keseluruhan.

2. Bagi Mahasiswa

   Mahasiswa akan mendapatkan penglaman praktis dalam mengimplementasikan algoritma K-Means *Clustering*, serta menggunakan Streamlit untuk pengembangan aplikasi web interaktif. Penelitian ini memungkinkan mahasiswa untuk mendalami konsep-konsep teoritis terkait segmentasi nasabah, mechine learning, dan visualisasi data. Dan mahasiswa akan memperoleh pengalaman dalam melakukan penelitian yang mencakup seluruh proses, mulai identifikasi masalah, pengumpulan data, hingga penyusunan laporan penelitian.

3. Bagi Akademik

   Penelitian ini menambah literatur dan wawasan mengenai penggunaan algoritma K-Means *Clustering* dan visualisasi data dengan Streamlit dalam bidang perbankan. Hasil penelitian ini dapat menjadi referansi bagi penelitian selanjutnya. Penilitian ini dapat menjadi contoh dan inspirasi akademis dan peneliti lainnya untuk mengekplorasi lebih lanjut penggunaan algoritma *clustering* dan alat visualisasi dalam berbagai bidang lainnya. Hasil penelitian ini dapat digunakan sebagai bahan ajar atau studi kasus dalam mata kuliah yang berkaitan dengan data science, machine learning, dan sistem informasi.

BAB II
KAJIAN PUSTAKA/TEORI

A. Tinjauan Pustaka

Sebelumnya telah terdapat beberapa penelitian yang berhubungan dengan penerapan metode K-Means *Clustering* pada Tabel 2.1. Di mana setiap penelitian memiliki kriteria dan pola yang sama bahkan berbeda satu sama lain. Berikut merupakan table penelitian sebelumnya:

Tabel 2. 1 Hasil Penelitian Sebelumnya

No.	**Nama Peneliti dan Tahun**	**Judul**	**Metode**	**Hasil**
1.	Widyawati , Wawan Laksito Yuly Saptomo, Yustina Retno Wahyu Utami (2020)	Penerapan Agglomerative Hierarchical Clustering Untuk Segmentasi Pelanggan	Agglomerative Hierarchical Clustering (AHC)	Berhasil dibuat
2.	Nita Mirantika, Tri Septiar Syamfithriani, Ragel Trisudarmo (2023)	Implementasi Algoritma K-Medoids Clustering Untuk Menentukan Segmentasi Pelanggan	K-Medoids	Berhasil dibuat
3.	Satria Ardi Perdana , Sara Famayla Florentin , dan Agus Santoso (2022)	Analisis Segmentasi Pelanggan Menggunakan K-Means Clustering Studi Kasus Aplikasi Alfagift	K-Means Clustering	Berhasil dibuat
4.	Ira Ariati , Reza Nugraha Norsa ,	Segmentasi Pelanggan Menggunakan K-	K-Means Clustering	Berhasil dibuat

	Lurinjani Akhsan , Jerry Heikal (2023)	Means Clustering Studi Kasus Pelanggan UHT Milk Greenfield		
5.	Gustientiedina , M.Hasmil Adiya , Yenny Desnelita (2019)	Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan Pada RSUD Pekanbaru	K-Means Clustering	Berhasil dibuat
6.	Nita Mirantikka, Annisa Tsamratul'Ain, Futry Diviana Agnia (2021)	Penerapan Algoritma K-Means Clustering Untuk Pengelompokkan Penyebaran COVID-19 Di Provinsi Jawa Barat	K-Means Clustering	Berhasil dibuat
7.	Qorik Indah Mawarni, Eko Setia Budi(2022)	Implementasi Algoritma K-Means Clustering Dalam Penilaian Kedisiplinan Siswa	K-Means Clustering	Berhasil dibuat

Menurut Widyawati, Wawan Laksito Yuly Saptomo, Yustina Retno Wahyu Utami (2020), penelitian yang berjudul Penerapan Agglomerative Hierarchical Clustering Untuk Segmentasi Pelanggan. Hasil dari penerapan Agglomerative Hierarchical Clustering (AHC) diperoleh 7 cluster yang mana cluster 1 terdapat 20 anggota, cluster 2 terdapat 43 anggota, cluster 3 terdapat 75 anggota, cluster 4 terdapat 158 anggota, cluster 5 terdapat 9 anggota, cluster 6 terdapat 2 anggota dan cluster 7 terdapat 1 anggota. Sedangkan pada pengujian Koefisien Silhouette mendapatkan hasil dalam cluster 1 terdapat 20 data tepat berada di cluster 1 dari 20 data, dalam cluster 2 terdapat 43 data yang tepat berada di cluster 2 dari 43 data, dalam cluster 3 terdapat 74 data yang tepat berada di cluster 3 dari 75 data, dalam cluster 4 terdapat 157 data yang tepat berada di

cluster 4 dari 158 data, dalam cluster 5 terdapat 5 data yang tepat berada di cluster 5 dari 9 data, dalam cluster 6 terdapat 2 data yang tepat berada di cluster 6 dari 2 data dan dalam cluster 7 hanya terdapat 1 data [7].

Menurut (2023), penelitian yang berjudul Implementasi Algoritma K-Medoids *Clustering* Untuk Menentukan Segmentasi Pelanggan. Hasil segmentasi pelanggan diperoleh tiga jenis pelanggan yaitu *loyal customer* sebanyak 21 pelanggan, *typical customer* sebanyak 31 pelanggan dan *newcomer* sebanyak 61 pelanggan. *Loyal customer* adalah pelanggan yang mempunyai nilai *recency* tinggi (baru), *frequency* paling banyak dan *monetary* paling banyak. Typical customer adalah pelanggan yang memiliki recency rendah (lama), frequency sedang dan monetary sedang. Newcomer adalah pelanggan yang mempunyai nilai recency paling tinggi (paling baru), frequency dan monetary paling rendah [8].

Menurut Satria Ardi Perdana, Sara Famayla Florentin, dan Agus Santoso (2022), penelitian yang berjudul Analisis Segmentasi Pelanggan Menggunakan K-Means Clustering Studi Kasus Aplikasi Alfagift. Setelah melalui keseluruhan proses data mining, dapat disimpulkan bahwa pengelompokan pelanggan terbentuk sesuai jumlah cluster terbaik yaitu sebanyak 3 cluster atau k=3. Cluster pertama berjumlah 7.219 pelanggan, cluster 2 sebanyak 6.902 pelanggan dan cluster 3 sebanyak 5.371 pelanggan. Dari tabel 9 dapat dilihat bahwa C1, C2. C3 memiliki persamaan karakteristik yaitu Pelanggan dalam kluster ini paling banyak berusia antara 26-35 tahun dan berjenis kelamin perempuan dengan rata-rata frekuensi pembelian sebanyak 1-5 kali dalam 1 bulan. C1 dan C2 memiliki persamaan karakteristik untuk metode pembayaran yang paling banyak dipilih adalah COD (Cash on Delivery) dengan transaksi terbanyak di wilayah Jakarta Barat, sedangkan untuk C2 metode pembayaran yang paling banyak dipilih adalah Gopay dengan transaksi terbanyak di wilayah Jakarta Selatan [5].

Menurut Ira Ariati, Reza Nugraha Norsa, Lurinjani Akhsan, Jerry Heikal (2023), penelitian yang berjudul Segmentasi Pelanggan Menggunakan K-

Means Clustering Studi Kasus Planggan UHT Milk Greenfield. Dari penelitian Analisis Segmentasi Pemasaran Susu Greenfields didapatkan hasil. Dari Cluster yang sudah di dapat yaitu 3 cluster (Menengah, Premium, Massal) direkomendasikan development persona yaitu cluster massal proses pengiriman yang terbanyak ke Kota Jembrana. Walaupun intensitas pembelian sedikit namun pembelian mereka sangat banyak sehingga meningkatkan penjualan. Rekomendation on Top 8ps marketing (Product, Price, Place, Promotion, People, Process, Prysical, Performance) sudah ada, hal ini menjadikan Susu UHT Greenfield dapat mengidentifikasi dan memahami komponenkomponen penting dari pemasaran dalam penjualan Susu UHT. Analisis ini membantu dalam merencanakan dan mengelola strategi pemasaran yang efektif dan tepat sasaran [9].

Menurut Gustientiedina, M.Hasmil Adiya, Yenny Desnelita (2019), penelitian yang berjudul Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan Pada RSUD Pekanbaru. Clusterisasi data obat yang dilakukan dengan algoritma k-means didapatkan hasil cluster nya setelah melakukan iterasi ke-4 yaitu terdapat kelompok obat yang pemakaian sedikit terdapat pada cluster 1 yang memiliki 224 anggota, kelompok obat yang pemakaian sedang terdapat pada cluster 2 yang memiliki 55 anggota, dan kelompok obat yang pemakaian tinggi terdapat pada cluster 3 yang memiliki 16 anggota [10].

Menurut Nita Mirantikka, Annisa Tsamratul'Ain, Futry Diviana Agnia (2021), penelitian yang berjudul Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Penyebaran COVID-19 Di Provinsi Jawa Barat. Berdasarkan hasil penelitian yang dilakukan, dapat kami simpulkan bahwa perhitungan secara manual algoritma K-Means Clustering dengan menggunakan metode CRISP-DM dan perhitungan menggunakan tools aplikasi R diperoleh klaster atau pengelompokan yang sama. Untuk klaster 1 sebagai klaster dengan jumlah penyebaran covid-19 yang paling tinggi diperoleh 2 kota/kabupaten yaitu kota Depok dan Kota Bekasi. Untuk klaster 2 sebagai klaster dengan jumlah penyebaran covid-19 menegah diperoleh 5

kabupaten/ kota yaitu Kota Bandung, Kabupaten bandung, Kabupaten Bekasi, kabupaten Bogor, dan kabupaten Karawang. Sisanya sebanyak 20 kota/ kabupaten masuk dalam klaster 3 dengan jumlah penyebaran covid-19 yang lebih sedikit [11].

Menurut Qorik Indah Mawarni, Eko Setia Budi (2022), penelitian yang berjudul Implementasi Algoritma K-Means Clustering Dalam Penilaian Kedisiplinan Siswa. Dari hasil penelitian dan pembahasan terkait implementasi metode k-means clustering terhadap penilaian kedisiplinan siswa maka dapat di tarik kesimpulan yaitu Penilaian kedisiplinan siswa/i dapat diimplementasikan menggunakan metode k-means clustering. Penelitian ini menerapkan metode Algoritma K-Means clustering dengan menggunakan Microsoft Excel 2013 dan Orange yang melakukan proses data mining. Hasil dalam penelitiann prosedur pemecahan algoritma k-means clustering terhadap kedisiplinan siswa/i dibagi menjadi tiga cluster. Dari 133 sempel siswa terdapat 41 siswa masuk dalam cluster satu (C1), kemudian 33 siswa masuk kedalam cluster kedua (C2), dan 59 siswa masuk kedalam cluster tiga (C3) [12].

B. Landasan Teori

Landasan teori merupakan konsep, prinsip, atau teori-teori yang digunakan sebagai kerangka sebuah penelitian. Landasan teori akan memberikan dasar yang kuat untuk penelitian ini. Oleh karena itu, berikut ini landasan teori yang akan digunakan penulis:

1. Segmentasi Pelanggan

   Segmentasi adalah proses membagi pelanggan menjadi beberapa klaster dengan kategori loyalitas pelanggan untuk membangun strategi pemasaran. Segmentasi pelanggan adalah salah satu langkah awal dalam membuat model bisnis [13].Segmentasi pelanggan juga merupakan suatu aktivitas mengelompokan pelanggan bersasarkan kategori tertentu seperti perilaku, minat, demografi geologi, loyalitas, transaksi atau lainnya. Tujuan pengelompokan ini untuk mempermudah bisnis menyusun strategi pemasaran yang lebih efektif. Tujuan utama dari segmentasi pelanggan

tersebut adalah mengelompokan target pasar agar lebih mudan menjangkaunya [14].

Ada 5 jenis – jenis segmentasi pelanggan berdasarkan karakteristik pelanggan adalah sebagai berikut:

a. Nilai dan Manfaat

Katagori segmentasi pelanggan ini berdasarkan pada nilai atau manfaat dari produk perusahaan yang akan diterima pelanggan. Dalam penentuan pelanggan yang masuk jenis customer segmentation, ada beberapa hal yang perlu diperhatikan seperti berikut:

- Khusus hanya untuk pelanggan lama
- Membutuhkan wawasan analitis dan substasional terkait perilaku pelanggan melalui kontak customer, brand, hingga transaksi. Contoh segmentasi pelanggan dari jenis ini dengan fokus pelanggan untuk meningkatkan nilai, mencari keuntungan, ragu-ragu dan lainnya.

b. Demografis

Segmentasi demografis digunakan untuk mengelompokkan pelanggan berdasarkan jenis kelamin, pendapatan, usia, etnis, pendidikan, pekerjaan dan lainnya.Contoh segmentasi demografis adalah seorang pekerja kantoran berusia 25-30 tahun. Maka pelanggan diluar karakteristik tersebut bukan menjadi bagiannya.

c. Pelanggan Baru

Sebagian bisnis mencurahkan fokusnya untuk mendapatkan pelanggan sebanyak mungkin. Terutama saat baru mengeluarkan produk baru. Maka dari itu, bisnis juga perlu mengelompokkan mereka berdasarkan beberapa poin berikut:

1) Pelanggan yang belun pernah membeli produk sama sekali.
2) Pelanggan yang sudah satu atau dua kali membeli produk

d. Pelanggan setia

Bisnis harus bisa menjaga pelanggan lama agar setia pada brand. Pasalnya pelanggan lama akan senang dilayani dengan baik, sehingga tidak ragu untuk melakukan pembelian berulang.Maka dari itu, segmentasi pelanggan setia juga banyak dibentuk oleh bisnis. Mereka bisa diklasifikasikan menjadi beberapa bagian berdasarkan kunjungan, pembelian hingga alasan pembelian.

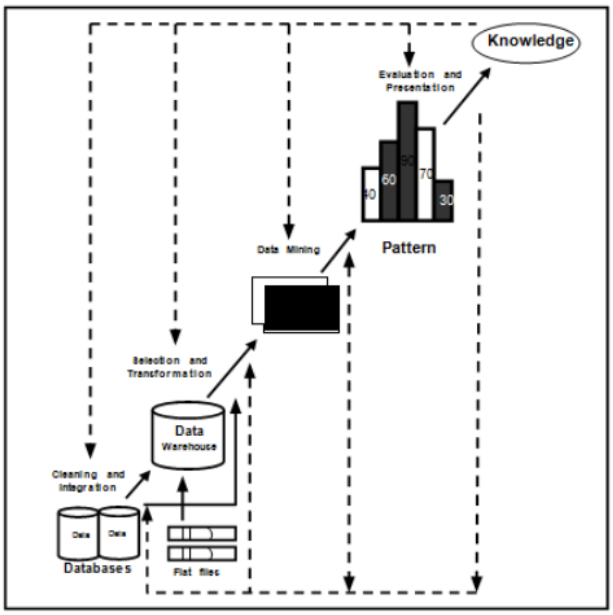
e. Gaya Hidup/ *Lifestyle*

Bisnis juga terkadang mengelompokkan pelanggan berdasarkan gaya hidup mereka seperti minat, kegiatan atau perilaku. Hal ini biasanya bisa bisnis ketahui dari aktivitas pelanggan di dunia maya melalui website, sosial media dan lainnya.

2. Data mining

Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar. Istilah data mining memiliki hakikat sebagai disiplin ilmu yang tujuan utamanya adalah untuk menemukan, menggali, atau menambang pengetahuan dari data atau informasi yang kita miliki. Data mining, sering juga disebut sebagai Knowledge Discovery in Database (KDD). KDD adalah kegiatan yang meliputi pengumpulan, pemakaian data, historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar. Data mining ialah analisa atau pengamatan data dengan jumlah yang besar untuk menemukan hubungan yang belum diketahui sebelumnya, dan dua metode baru guna meringkas data supaya lebih mudah dimengerti serta kegunaannya untuk pemilih data [15].

Pendapat lain mengatakan bahwa data mining terdiri dari 6 tahap yang terpola dalam Gambar 2.1 sebagai berikut [16] :



Gambar 2.1 Tahapan Data Mining

Tahap-tahap *data mining* yaitu :

a. Pembersihan data (*Data Cleaning*)

   Pembersihan data merupakan proses menghilangkan *noise* dan data yang tidak konsisten atau data tidak relevan. Pada umumnya data yang diperoleh, baik dari database suatu perusahaan maupun hasil eksperimen, memiliki isian-isian yang tidak sempurna seperti data yang hilang, data yang tidak valid atau juga hanya sekedar salah ketik. Selain itu, ada juga atribut-atribut data yang tidak relevan dengan hipotesa data mining yang dimiliki. Data-data yang tidak relevan itu juga lebih baik dibuang. Pembersihan data juga akan mempengaruhi performasi dari teknik data mining karena data yang ditangani akan berkurang jumlah dan kompleksitasnya.

b. Integrasi data (*Data Integration*)

   Integrasi data merupakan penggabungan data dari berbagai database ke dalam satu database baru. Tidak jarang data yang diperlukan untuk data mining tidak hanya berasal dari satu database tetapi juga berasal dari beberapa database atau file teks. Integrasi data dilakukan pada atribut-

aribut yang mengidentifikasikan entitas-entitas yang unik seperti atribut nama, jenis produk, nomor pelanggan dan lainnya. Integrasi data perlu dilakukan secara cermat karena kesalahan pada integrasi data bisa menghasilkan hasil yang menyimpang dan bahkan menyesatkan pengambilan aksi nantinya. Sebagai contoh bila integrasi data berdasarkan jenis produk ternyata menggabungkan produk dari kategori yang berbeda maka akan didapatkan korelasi antar produk yang sebenarnya tidak ada.

c. Seleksi data (*Data Selection*)

   Data yang ada pada database sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari database. Sebagai contoh, sebuah kasus yang meneliti faktor kecenderungan orang membeli dalam kasus market basket *analysis*, tidak perlu mengambil nama pelanggan, cukup dengan id pelanggan saja.

d. Transformasi data (*Data Transformation*)

   Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam data mining. Beberapa metode data mining membutuhkan format data yang khusus sebelum bisa diaplikasikan. Sebagai contoh beberapa metode standar seperti analisis asosiasi dan clustering hanya bisa menerima input data kategorikal. Karenanya data berupa angka numerik yang berlanjut perlu dibagibagi menjadi beberapa *interval*. Proses ini sering disebut transformasi data.

e. Proses *Mining*

   Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data.

f. Evaluasi pola (*Pattern Evaluation*)

   Dalam tahap ini hasil dari teknik data mining berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesa yang ada memang tercapai. Bila ternyata hasil yang diperoleh tidak sesuai hipotesa ada beberapa alternatif yang dapat diambil seperti

menjadikannya umpan balik untuk memperbaiki proses data mining, mencoba metode data mining lain yang lebih sesuai, atau menerima hasil ini sebagai suatu hasil yang di luar dugaan yang mungkin bermanfaat.

g. Presentasi Pengetahuan (*Knowledge Presentation*)

Merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna. Tahap terakhir dari proses data mining adalah bagaimana memformulasikan keputusan atau aksi dari hasil analisis yang didapat.

3. Algoritma K-Means

Teknik pengelompokkan paling populer dalam bidang ilmiah salah satunya adalah algoritma K-Means. Algoritma K-Means adalah salah satu algoritma dengan teknik *clustering* berdasarkan pembagian jarak dalam data *mining*. Keuntungan menggunakan algoritma ini mudah dipahami, diterapkan, serta memiliki efek pengelompokan yang baik sehingga KMeans banyak digunakan dalam bidang penelitian . Namun, algoritma ini juga memiliki kekurangan, yakni memiliki ketergantungan yang kuat pada pemilihan pusat cluster awal [17].

K-Means adalah salah satu metode dalam data mining yang dapat mengelompokan data atau *clustering* sebuah data kedalam bentuk satu *cluster* atau lebih *cluster* sehingga data yang memiliki kaarakteristik yang sama dikelompokkan ke dalam satu *cluster* yang sama dan data dengan karakteristik yang berbeda dikelompokkan ke dalam kelompok berbeda yang lainnya. Proses menggunakan Algoritma K-Means untuk pengelompokan beberapa klaster melakukan tahapan sebagai berikut [18] :

1) Menentukan nilai k sebagai jumlah cluster yang ingin di bentuk.
2) Menentukan nilai acak atau random untuk pusat *cluster* awal *centroid* sebanyak k, untuk menghitung jarak setiap data input terhadap masing-masing *centroid* dengan menggunakan rumus jarak *Eulidean Distance* yaitu :

$$d(xi, \mu j) = \sqrt{\sum(xi - \mu j)^2} \quad (1)$$

Dimana :

xi = data kriteria

µj = *centroid* pada *cluster* ke-js

3) Mengelompokan setiap data berdasarkan kedekatannya dengan *centroid* atau mencari jarak terkecil.
4) Memperbaharui nilai *centroid* baru, nilai *centroid* baru diperoleh dari rata-rata *cluster* yang bersngkutan dengan menggunakan rumus yaitu :

$$\mu j(t+1) = \frac{1}{Nsj} \sum_{jesj} xj \quad (2)$$

Keterangan:

$\mu j(t+1)$ = *centroid* baru pada iterasi (t+1)

Nsj = data pada *cluster* Sj

5) Apabila data setiap *cluster* belum berhenti, lakukan perulangan dari langkah 2 hingga 5, sampai anggota tiap *cluster* tidak ada yang berubah.

4. PCA (*Principal Component Analysis*)

PCA merupakan kombinasi linear dari variabel awal yang secara geometris kombinasi linear ini merupakan sistem koordinat baru yang diperoleh dari rotasi sistem semula. Metoda PCA sangat berguna digunakan jika data yang ada memiliki jumlah variabel yang besar dan memiliki korelasi antar variabelnya. Perhitungan dari *principal component analysis* didasarkan pada perhitungan nilai eigen dan vektor eigen yang menyatakan penyebaran data dari suatu dataset [19]. Dengan menggunakan PCA, variabel yang tadinya sebanyak n variabel akan diseleksi menjadi k variabel baru yang disebut *principal component*, dengan jumlah k lebih sedikit dari n. Dengan hanya menggunakan k *principal component* akan menghasilkan nilai yang sama dengan menggunakan n variabel [20]. Variabel hasil dari

seleksi disebut *principal component*. PCA digunakan untuk menjelaskan struktur matriks varians-kovarians dari suatu set variabel melalui kombinasi linier dari variabel-variabel tersebut. Secara umum *principal component* (PC) dapat berguna untuk seleksi fitur dan interpretasi variabel-variabel [21].

Prinsip kerja metode ini adalah dengan mengekstaksi atribut sehingga menyisakan atribut yang bertujuan untuk memperoleh hasil lebih optimal. Metode ini terdiri dari 4 tahapan yaitu [22] :

a. Mencari sejumlah data yang berdimensi m x n, dimana m adalah jumlah sampel data sedangkan n adalah jumlah atribut

$$X^*_{i.j} = X_{i.j} - \overline{X} \quad (3)$$

Keterangan:

$X^*_{i.j}$ = Elemen Matrik X*

$X_{i.j}$ = Elemen Matrik X

$\overline{X}$ = Nilai rata-rata marik X

b. Mencari nilai kovarian (C_x) dari sejumlah data dengan persamaan 2 :

$$C_x = \frac{1}{m-1} . X^{*T}_{i.j} . X^*_{i.j} \quad (4)$$

c. Menghitung nilai eigen (λ) dengan persamaan 3, dimana I merupakan matrik identitas dan v merupakan vector eigen:

$$|C_x - \lambda I| = 0 \; dan \; (C_x - \lambda I) . v = 0 \quad (5)$$

d. Menghitung persentase kontribusi komulatif variansi (V), dimana d adalah jumlah atribut awal dan r adalah jumlah komponen yang dipilih.

$$V_r = \frac{\sum_j^r \lambda j}{\sum_j^d \lambda j} \ .100\% \qquad (6)$$

5. *Cluster*

Cluster atau *clustering* adalah teknik untuk mengelompokkan data menjadi beberapa grup atau kluster berdasarkan kesamaan antar data. Menurut Zulfa Nabila, Auliya Rahman Isnain, Permata, Zaenal Abidin *Cluster* adalah kumpulan dari record yang memiliki kemiripan satu sama lain, dan berbeda dengan record di klaster lain. *Clustering* mencoba untuk membagi seluruh kumpulan data menjadi kelompok-kelompok yang relatif memiliki kemiripan, di mana kemiripan record dalam satu kelompok akan bernilai maksimal, sedangkan kemiripan dengan record dalam kelompok lain akan bernilai minimal. *Clustering* dalam data *mining* berguna untuk menemukan pola distribusi di dalam sebuah data set yang berguna untuk proses analisa data. Kesamaan objek biasanya diperoleh dari kedekatan nilai-nilai atribut yang menjelaskan objek-objek data, sedangkan objek-objek data biasanya direpresentasikan sebagai sebuah titik dalam ruang multidimensi [23].

6. Phyhton

Python merupakan salah satu bahasa pemrograman yang banyak digunakan oleh perusahaaan besar maupun para *developer* untuk mengembangkan berbagai macam aplikasi berbasis desktop, web dan mobile. Python diciptakan oleh Guido van Rossum di Belanda pada tahun 1990 dan namanya diambildari acara televisi kesukaan Guido Monty Python's Flying Circus. Van Rossum mengembangkan Python sebagaihobi, kemudian Python menjadi bahasa pemrograman yang dipakai secara luas dalam industri dan pendidikan karena sederhana, ringkas, sintak sintuitif dan memiliki pustaka yang luas [24]. Bahasa pemrograman Python ditandai dengan sintaksis yang mudah dibaca dan dipahami, yang memungkinkan pemrogram untuk mengembangkan kode dengan cepat dan

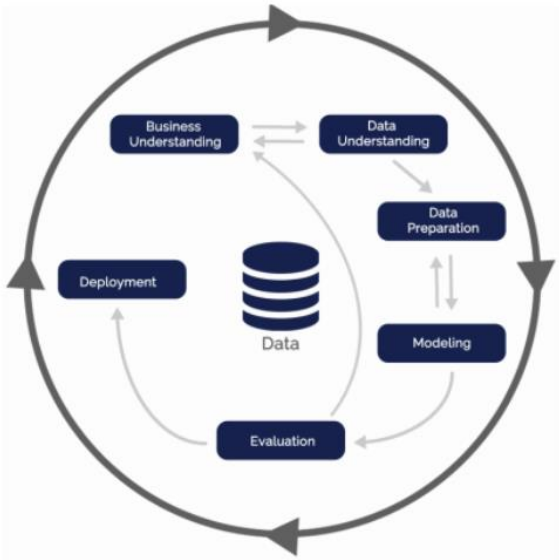
efisien. Python juga menyediakan berbagai library dan alat bantu yang kuat untuk analisis data, seperti Pandas, NumPy dan Matplotlib [25].

7. Streamlit

Streamlit adalah kerangka kerja web yang ditujukan untuk menyebarkan model dan visualisasi dengan mudah menggunakan bahasa Python, yang cepat dan minimalis tetapi juga memiliki tampilan yang cukup baik serta ramah pengguna. Streamlit merupakan aplikasi yang tidak berbayar dan pengguna tidak perlu memiliki pengetahuan pengembangan *front-end* yang mahir untuk mengoperasikannya. Streamlit dapat dijalankan pada editor Anaconda serta bahasa Python seri 3.7 ke atas, tetapi tidak mendukung pada editor *Jupyter Notebook*, sehingga harus dikonversi ke editor *Pycharm* atau *Visual Code*. Tampilan beranda pada aplikasi Streamlit dapat dipisahkan menjadi dua bagian, yaitu buttons, untuk pemilihan menu, serta tampilan *visual chart*. Hal ini menyebabkan dibutuhkannya library NumPy serta Pandas untuk menampilkan grafik. Keluaran grafik sejalan dengan hasil olah data metode pembelajaran mesin menggunakan kombinasi lapisan tersembunyi LSTM dan GRU. Buttons berfungsi untuk memilih dataset dari kategori negara, jenis hewan, arsitektur lapisan tersembunyi, optimizer, serta pilihan untuk epoch dan prediksi dalam beberapa tahun ke depan [26].

8. Metodologi Penelitian CRISP-DM

Metodologi penelitian yang digunakan adalah CRISP-DM. CRISP-DM (*Cross Industry Standard Process For Data Mining*) adalah standar proses data *mining* yang akan digunakan pada penelitian. Proses penelitian ini mengacu pada enam tahap yaitu pemahaman bisbis, pemahaman data, persiapan data, pemodelan, evaluasi dan penyebaran [27]. Pada gambar 2.2 di bawah ini merupakan tahapan-tahapan CRISP-DM:



Gambar 2. 2 Tahapan CRISP-DM

a. *Business Understanding*

Pada tahap memahami perusahaan ini, penulis perlu memahami dan memahami model bisnis bank. Ada juga puluhan fitur canggih, seperti manajemen bisnis, untuk membantu perusahaan mendigitalkan bisnisnya dan bersaing dengan kompetitor di industri yang bergerak cepat [28].

b. *Data Understanding*

Tahap pemahaman data ini berawal dari mengumpulkan informasi, kemudian mengidentifikasi kualitas informasi yang digunakan, mencari informasi tersembunyi yang dapat membentuk hipotesis baru. Data mining dilakukan pada database perusahaan, dimana pada database terdapat 15 yaitu tabel Id_Nasabah, Usia, Jenis_Kelamin, Jumlah_Tanggungan, Pendidikan, Status_Pernikahan, Pendapatan, Kategori_Kartu, Lama Menjadi Nasabah, Jumlah_Layanan, Limit_Kredit, Saldo_Revolting_Total, Total_Transaksi, Jumlah_Melakukan_Transaksi, Rasio_Penggunaan_Rata-Rata.

c. *Data Preparation*

Ada beberapa hal yang melatarbelakangi langkah data preparation:

1) Data *Selection*

Data Selection adalah data yang dipilih untuk diolah berdasarkan kolom yang tersedia dan sesuai kebutuhan. Data yang tidak terkait dengan metode yang digunakan kemudian akan dihapus.

2) Data *Preprocessing*

Setelah tahap data selection, data dilanjutkan sehingga data yang disiapkan bersih dari data *noise* dan *missing value*. Data diurutkan berdasarkan dengan nilai data yang ada pada setiap kolom

3) Data *Transformation*

Jika data yang dipilih bebas dari data *noise* dan *missing value*. Data tersebut kemudian dapat melakukan *min-max normalization* untuk mengubah data menjadi nilai yang bermakna. Normalisasi digunakan untuk meningkatkan akurasi proses komputasi numerik pada skala data antara 0 dan 1.

d. *Modelling*

Sesudah data yang dapat digunakan melewati proses pengolahan data, data tersebut diproses dengan bobot sesuai nilai dan tujuan untuk memudahkan pengolahan data. Berdasarkan data penjualan, pengolahan data dilakukan dengan algoritma pengelompokan data mining berbasis bagian yang disesuaikan dengan atribut yang dibuat dalam kumpulan data, dan menentukan jenis pelanggan berdsarkan *cluster* yang dihasilkan dari berbagai macam segmentasi.

e. *Evaluation*

Tahap evaluasi merupakan tahap lanjutan dimana tujuan data mining dievaluasi secara menyeluruh untuk mendapatkan pemodelan yang diinginkan. Beberapa hal yang dilakukan dalam fase ini antara lain mengevaluasi hasil, seberapa jauh pemodelan telah mencapai tujuan yang diinginkan, proses pengecekan atau pengecekan ulang untuk memastikan bahwa semua langkah yang dilakukan tidak meleset, dan

pendefinisian langkah-langkah yang dilakukan selanjutnya yaitu melanjutkan ke tahap *deployment* atau kembali ke tahap awal yaitu business understanding. Pada tahap evaluasi, hasil pengelompokan dari tahap pemodelan dievaluasi menggunakan metode PCA untuk mendapatkan jumlah cluster yang optimal. Pada tahap ini dilakukan pemodelan dengan PCA adalah metode analisis untuk mengidentifikasi sikap pelanggan dan menyajikan sikap pelanggan berdasarkan data yang sudah di preprosessing sehinggan menghasilkan jumlah PC1, PC2, PC3.

f. Deployment

Pada tahap diseminasi ini, informasi dan pengetahuan yang diperoleh disajikan dalam bentuk yang lebih mudah dipahami oleh masyarakat umum. Pada langkah ini juga digunakan website Streamlit untuk memvisualisasikan hasil yang diperoleh. Dengan diperkenalkannya situs web Streamlit, memudahkan untuk membaca data, berbagi data, berkolaborasi, dan bahkan mengekspor data. Hasil penelitian ini berupa data nasabah, informasi data, visualisasi data, serta visualisasi klaster berdasarkan data yang dihasilkan untuk mengetasuhi jumlah klaster yang disegmentasikan.

9. Blackbox Testing

Metode *BlackboxTesting* merupakan salah satu metode yang mudah digunakan karena hanya memerlukan batas bawah dan batas atas dari data yang di harapkan. Estimasi banyaknya data uji dapat dihitung melalui banyaknya field data entri yang akan diuji, aturan entri yang harus dipenuhi serta kasus batas atas dan batas bawah yang memenuhi. Dan dengan metode ini dapat diketahui jika fungsionalitas masih dapat menerima masukan data yang tidak diharapkan maka menyebabkan data yang disimpan kurang valid.

Pengujian adalah satu set aktifitas yang direncanakan dan sistematis untuk menguji atau mengevaluasi kebenaran yang diinginkan. Pengujian perangkat lunak dari segi spesifikasi fungsional tanpa menguji desain dan kode program untuk mengetahui apakah fungsi, masukan dan keluaran dari

perangkat lunak sesuai dengan spesifikasi yang dibutuhkan. Pengujian pada sistem menggunakann metode *Black Box*, tujuannya mengetahui kelemahan dari sistem agar data yang dihasilkan sesuai dengan data yang dimasukkan setelah data dieksekusi dan menghindari kekurangan dan kesalahan pada aplikasi sebelum digunakan oleh user [29].

10. *Whitebox Testing*

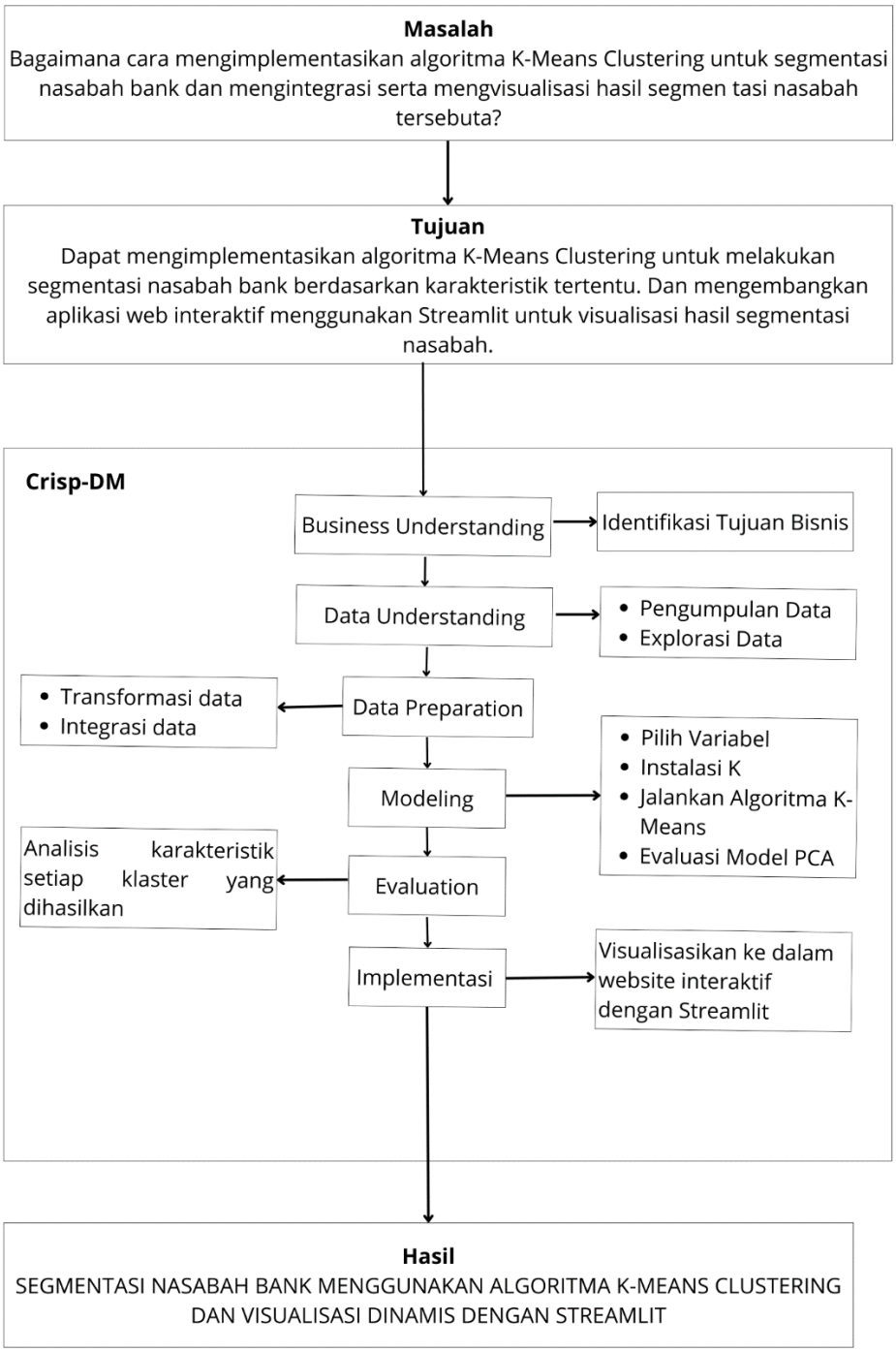
Pengujian *White Box*, adalah suatu metode pengujian aplikasi yang menggunakan penjelasan struktur kontrol sebagai bagian dari component-level design untuk membuat test cases. *White Box* sendiri mempunyai beberapa teknik di dalam pengujiannya, seperti Data *Flow Testing*, *Control Flow Testing*, Basic Path / Path Testing, dan *Loop Testing*. Dalam Pengujian *White Box* para penguji perlu mengetahui secara dalam source code yang akan diuji. Pengujian *White Box* dapat mengungkap kesalahan implementasi dari sebuah aplikasi. Pengujian ini dapat diterapkan pada tingkatan integrasi, unit dan sistem [30].

Ada beberapa kelebihan dan kekurangan dalam pengujian menggunakan metode *White Box* antara lain:

a. Kelebihan
    1) Metode *White Box* dapat memperlihatkan galat pada kode yang dibuat dengan menghapus baris yang tidak diperlukan.
    2) Maksimalnya cakupan pengujian aplikasi saat uji coba sebuah skenario.
b. Kekurangan
    1) Biaya pengujian menggunakan metode *White Box* sangatlah tinggi karena dibutuhkan penguji yang berpengalaman dalam bidang ini.
    2) Beberapa alur program akan dibiarkan tidak diuji karena tidak memungkinkan untuk menguji setiap baris kode untuk menemukan kesalahan.

C. Kerangka Berfikir

Kerangka berpikir dibuat untuk mempermudah proses penelitian karena telah mencakup pemecahan permasalahan yang telah dirumuskan. Dalam permasalahan pada penelitian ini adanya ketidak pahaman mengintegrasi pelanggan untuk mengelompokan pelanggan bank dalam beberapa klaster. Dengan permasalahan tersebut penulis bertujuan untuk membangun sebuah segmentasi yang dapat mengklastering pelanggan menggunakan Algoritma K-Means *Clustering*. Untuk kerangka berpikir pada penelitian ini dapat dilihat pada Gambar 2.3.



Gambar 2. 3 Kerangka Berfikir

BAB III
METOD PENELITIAN

A. Metode Penelitian

Metode penelitian adalah pendekatan atau prosedur sistematis yang digunakan untuk mengumpulkan, menganalisis, dan menafsirkan data dalam rangka menjawab pertanyaan penelitian atau mencapai tujuan penelitian tertentu. Metode penelitian mempunyai maanfaat dan memfasilitasi atau membantu menjawab atau menentukan bagaimana rumusan masalah. Untuk penelitian ini penulis menggunakan metode Crisp-DM sebagai metode penelitian. Crisp-DM merupakan pendekatan sistematis yang terdiri dari serangkaian tahapan yang bertujuan untuk menghasilkan model segmentasi pelanggan yang efektif dan dapat diimplementasikan dengan baik. Berikut adalah langkah-langkah dalam Metode Penelitian Crisp-DM:

1. Pemahaman Bisnis (*Business Understanding*)
2. Pemahaman Data (*Data Understanding*)
3. Persiapan Data (*Data Preparation*)
4. Modeling
5. Evaluasi (*Evaluation*)
6. Implementasi

B. Fokus Penelitian

Fokus penelitian ini adalah untuk mengidentifikasi dan mengelompokkan nasabah bank ke dalam segmen-segmen yang berbeda berdasarkan pola transaksi, karakteristik demografis, atau faktor-faktor lainnya yang relevan. Penelitian ini juga bertujuan untuk mengembangkan alat visualisasi dinamis menggunakan platform Streamlit untuk menganalisis dan memahami hasil segmentasi dengan lebih interaktif.

1. Segmentasi Nasabah

   Penelitian ini akan memusatkan perhatian pada penggunaan algoritma K-Means *clustering* untuk membagi nasabah bank menjadi

segmen-segmen yang homogen berdasarkan pola transaksi, profil demografis, atau variabel lainnya yang relevan. Tujuan utamanya adalah untuk mengidentifikasi kelompok nasabah yang memiliki karakteristik serupa sehingga bank dapat menyusun strategi pemasaran yang lebih efektif dan layanan yang lebih personal.

2. Algoritma K-Means *Clustering*

   Penelitian akan mengeksplorasi implementasi algoritma K-Means *clustering* dalam konteks segmentasi nasabah bank. Hal ini mencakup pemilihan variabel yang tepat, penentuan jumlah *cluster* yang optimal, dan interpretasi hasil *clustering* untuk pemahaman yang lebih baik tentang karakteristik setiap segmen nasabah.

3. Visualisasi Dinamis dengan Streamlit

   Salah satu fokus utama penelitian ini adalah pengembangan aplikasi visualisasi dinamis menggunakan platform Streamlit. Aplikasi ini akan memungkinkan pengguna, seperti manajer bank atau analis data, untuk menjelajahi hasil segmentasi dengan lebih interaktif, menganalisis distribusi nasabah dalam setiap segmen, dan memahami tren atau pola yang muncul dengan lebih baik.

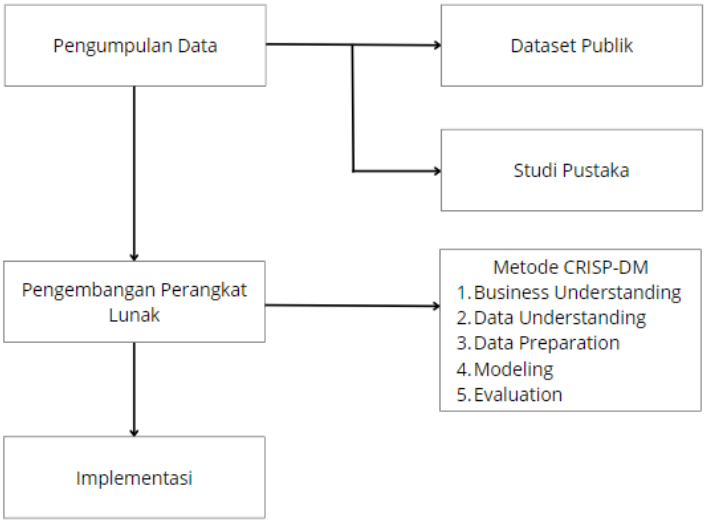
C. Desain Penelitian

Desain penelitian untuk segmentasi nasabah menggunakan Algoritma K-Mean *Clustering* berbasis web interaktif menggunakan streamlit dengan pendekatan CRISP-DM dengan tahapan-tahapan berikut:

1) **Tahap Pemahaman Bisnis**: Pada tahap ini, peneliti mengidentifikasi tujuan bisnis dari data nasabah. Peneliti menententukan kriteria keberhasilan yang dapat diukur. Kemudian penetiti memahami situasi bisnis, tantangan, dan peluang yang dihadapi. Selain itu peneliti akan menganalisis tentang karakter nasabah yang akan implementasikan.
2) **Tahap Pemahaman atau Pengenalan Data**: Pada tahap ini, peneliti akan menginisiasi koleksi data. Setelah itu, dilakukan analisis untuk meningkatkan pemahaman data, mengidentifikasi kualitas data,

menemukan pengetahuan awal di dalam data, atau mendeteksi subset yang menarik untuk membangun hipotesis mengenai informasi yang tersembunyi. Langkah-langkah yang terdapat pada tahapan ini adalah mengumpulkan data, mendeskripsikan data, dan memastikan kembali kualitas data..

3) **Tahap Preparation**: Pada tahap ini, peneliti melakukan pembersihan data dari nilai yang hilang, duplikat, atau noise dengan mengubah format data, memilih fitur yang relevan dan menggabungkan data dari berbagai sumber apabila diperlukan. Kemudian peneliti mentrasformasikan data ke format untuk dianalisis ke model.
4) **Tahap Modeling**: Pada tahap ini, Peneliti memilih teknik pemodelan yang sesuai dengan tujuan bisnis dan karakteristik data. Peneliti memilih Algoritma K-Means *Clustering* dan PCA. Peneliti kemudian melakukan standarisasi data menggunakan `StandardScaler()`. Selanjutnya peneliti pengujian model *inersia* dan *silhouette score* untuk menentukan nikai klaster. Kemudian peneliti memvisualisasikan klaster hasil K-Means pada komponen PCA.
5) **Tahap Evaluasi**: Pada tahap ini, peneliti melakukan evaluasi kinerja model dari hasil pengujian model yang telah ditetapkan sebelumnya. Selanjutnya peneliti memvisualisasikan hasil klaster kedalam *boxplot* dan pairplot untuk variabel terhadap klaster.
6) **Tahap Implementasi**: Pada tahap ini, peneliti mengimplementasikan model dan hasil kedalam sistem yang telah dibuat menggunakan streamlit. Peneliti mengvisualisasikan kedalam website interaktif yang menampilkan hasil pemodelan dan hasil yang sudah di analisis.
7) **Tahap Pengujian**: Tahap ini akan dilakukan untuk memastikan bahwa sistem klastering yang dikembangkan sesuai dengan desainnya dan berfungsi dengan baik. Sistem akan diuji dengan menggunakan dataset yang relevan dan diverifikasi oleh tim penguji untuk memeriksa akurasi dan keandalannya dalam mengklasterkan nasabah bank.



Gambar 3.1 Skema Desain Penelitian

D. Teknik Pengumpulan Data

1. Dataset Publik

Pada penelitian ini, peneliti mengumpulkan data berupa csv dari dataset Data Nasabah yang tersedia secara public di *Kaggle.com* yang dipublikasi oleh Tarisha Mazaya. Dataset tersebut terdiri dari 1050 nasabah pada sebuah bank. Dataset ini berisi informasi seperti Id Nasabah, Usia,Jenis Kelamin, Jumlah Tanggungan, Pendidikan, Status Pernikahan, Pendapatan, Kategori Kartu, Lama Menjadi Nasabah, Jumlah Layanan, Limit Kredit, Saldo Revolting Total, Total Transaksi, Jumlah Melakukan Transaksi, Rasio Penggunaan Rata-Rata. Setelah mengunduh dataset, peneliti menimpannya di *Google Drive* sebagai penyimpanan yang aman.

Selanjutnya, peneliti melakukan pra-pemprosesan data, untuk mengubah nama kolom menjadi huruf capital diawal kata, kemudian menampilkan bentuk data, informasi umum, deskripsi statistic, dan jumlah nilai unik. Data kemudain drop untuk menghapus kolom yang tidak diperlukan dan menghapus baris yang menggandung nilai kosong dan memeriksa jumlah data duplikat. Selanjutnya peneliti menganalisis statistic mean median untuk menghitung rata-rata median kolom numeric, kemudian

peneliti memvisualisasi data untuk menampilkan distribusi jenis kelamin, menmpilkan variable numeric dan histogram untuk kolom numeric. Peneliti kemudian menggunakan StandardScaler untuk nyiapkan dan melakukan standarisasi data. Selanjutnya peneliti melakukan visualisasi K-Mean *Clustering* dan PCA untuk melakukan klstering pada data yang mana data direduksi dimensinya terlebih dahulu. Selanjutnya data di evalusasi untuk menmpilkan scatter plot 3D dari klaster yang terbentuk, lalu menampilkan *boxplot* dari *variable numeric* berdasarkan klaster serta menampilkan pairplot dari variable numeric dengan warna berdasarkan klaster. Untuk contoh data nasabah bank dapat dilihat pada Gambar 3.2.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	ID_Nasabah	Usia	Jenis_Kelamin	Jumlah_Tang	Pendidikan	Status_Pernikah	Pendapatan	Kategori_Kartu	Lama_Mei	Jumlah_La	Limit_Kredit	Saldo_Revolt	Total_Transak	Jumlah_Mela	Rasio_Penggunaan_	
2	717574683	53	Perempuan	3		Menikah	44280	Biru	46	6	2192	1146	4756	85	0.523	
3	769662033	45	Perempuan	2	Kuliah		9982	Biru	29	2	1438.3	0	5025	85	0	
4	719377383	40	Laki-Laki	2	Kuliah	Belum Menikah	64988	Biru	28	1	10880	2025	4957	90	0.186	
5	820657083	53	Laki-Laki	4	Kuliah	Belum Menikah	117549	Perak	48	5	34516	800	3819	72	0.023	
6	709836258	45	Perempuan	4	Sarjana	Menikah	21579	Biru	41	5	1927	1337	5127	81	0.694	
7	779833908		Perempuan	2	Sarjana	Bercerai		Biru	23	6	18332	1376	4095	67	0.075	
8	719976558	47	Laki-Laki	2	Sarjana		69152	Biru	34	3	4249	1562	1881	49	0.368	
9	818502708		Perempuan	1	Tidak Berpend	Belum Menikah	43872	Biru	46	5	2284	1674	4885	86	0.733	
10	786509358	44	Laki-Laki	3		Menikah	84471	Biru	38	3	4198	1782	2051	36	0.424	
11	789226683	44	Laki-Laki	1		Menikah	62013	Biru	33	6	4969	1187	3878	82	0.239	
12	719912808	52	Laki-Laki	2	SMA	Belum Menikah	77519	Biru	39	3	14973	0	4288	72	0	
13	714816258	37	Laki-Laki	2	Sarjana	Belum Menikah	73950	Biru	36	4	7638	947	2336	58	0.124	
14	721402233		Laki-Laki	3	Tidak Berpend	Menikah	159371	Biru	28	2	19192	1507	15615	125	0.079	
15	708445608	48	Laki-Laki	3	Kuliah	Menikah	16963	Biru	37	4	3305	2517	2027	57	0.762	
16	721112508	48	Perempuan	3	Sarjana	Menikah	35598	Biru	40	2	5298	0	4834	80	0	
17	714920208	37	Perempuan	1	SMA		4531	Biru	30	3	3003	1810	3148	68	0.603	
18	716399583	34	Laki-Laki	1	Tidak Berpend	Menikah	171327	Biru	36	6	19630	1251	2286	39	0.064	
19	718500633	45	Perempuan	4		Menikah	38366	Biru	36	5	7924	1896	3270	58	0.239	
20	717733158		Perempuan	3	Doktor	Menikah	39302	Biru	29	4	3186	2517	4197	84	0.79	
21	716866308	38	Laki-Laki	2	SMA	Menikah	73964	Biru	33	6	5940	1101	1872	34	0.185	
22	708447858	59	Perempuan	2	Tidak Berpend	Menikah	19012	Biru	36	6	3850	0	15107	117	0	

data_nasabah

Gambar 3.2 Contoh Data Segmentasi Nasabah

2. Studi Pustaka

Studi pustaka merupakan fase krusial dalam proses penelitian. Dimana peneliti memperdalam pemahaman mereka tentang topik yang diteliti serta memperkokoh dasar teoritis yang telah ada sebelumnya. Dalam penelitian ini melibatkan pencarian, pengumpulan, analisis literatur atau karya ilmiah yang relevan dengan topik penelitian yang sedang diteliti. Pada teknik penelitian ini membantu peneliti untuk memperoleh wawasan yang mendalam tentang topik penelitian dan memperkuat dasar teoritis dari

penelitian terdahulu. Beberapa sumber yang digunakan dalam studi pustaka mencakup:

a) Menggunakan buku yan berkaitan dengan penelitian.
b) Menggunakan jurnal sebagai bahan penelitian yang terkait.
c) Meggunakan *website*/internet guna mencari informasi terkait.

E. Teknik Analisis Data

Teknik analisis data yang telah dilakukan setelah semua data yang dibutuhkan terkumpul. Berikut adalah teknik analisis data yang dilakukan dalam penelitian ini adalah:

1. Mengumpulkan keseluruhan data nasabah bank yang diperlukan dengan cara mengunduh dari website penyedia datanya yaitu *Kaggle*.
2. Melakukan analisis kebutuhan kebutuhan pengguna yang diperlukan untuk pembuatan website interaktif dengan streamlit.
3. Mengolah data yang sudah terkumpul, kemudian diproses dengan metode atau algoritma yang sudah ditentukan menjadi suatu *website* interaktif yang sudah direncanakan menggunakan streamlit.
4. Membuat simpulan akhir

BAB IV
HASIL DAN PEMBAHSAN

A. Hasil

Hasil dari Web Streamlit Segmentasi nasabah bank menggunakan algoritma K-Means *Clustering* menggunakan metode Crisp-DM dalam proses penganalisis datanya. Berikut adalah tahapan yang digunakan:

1. Business Understanding

Pada penelitian ini, peneliti melakukan pemahaman kebutuhan bisnis yang dimiliki dan target yang hendak dicapai. Data nasabah bank yang dipilih sebagai studi kasus ini memiliki permaslahan dalam melakukan identifikasi segmen. Pada data nasabah ini memiliki pemahaman yang lebih terperinci mengenai karakter nasabah bukan lagi barang yang bagus untuk dimiliki, namun merupakan keharusan yang strategis dan kompetitif bagi penyedia perbankan. Segmentasi nasbah menjadi tujuan dalam penentuan karteristik nasabah untuk dikelompokan menjadi beberapa klaster. Data nasabah yang telah dikumpulkan akan di prosesing menggunakan algoritma K-Means *clustering* untuk mengidentifikasi kelompok-kelompok nasabah yang memiliki kesamaan karakteristik. Kemudian di analisis menggunakan PCA untuk memperkuat pengelompokan nasabah, mengidentifikasi kebutuhan dan preferensi spesifik dari setiap segmen. Selanjutnya mengevaluasi hasil dari pemprosesan data dengan algoritma yang diterapkan dan melakukan evaluasi berkala untuk melihat perubahan dalam preferensi dan perilaku pelanggan. Dan menyesuaikan dan mengoptimalkan strategi berdasarkan hasil analisis dan feedback nasabah.

2. Data Understanding

Pada tahap ini, Data yang digunakan adalah data dari *kaggle* yaitu data nasabah bank yang di publis oleh peneliti bernama Tarisha Mazaya. Data ini disimpan dalam bentuk excel. Berikut temuan yang diperoleh dari evaluasi dan eksplorasi terhadap data tersebut. Untuk melakukan eksplorasi, penulis mengganti nama header sehingga lebih mudah diakses. Penamaan

header dapat dilihat pada Tabel 2 dan informasi dataset dapat dilihat pada Table 4.1.

Tabel 4. 1 Penamaan Header

Nama Header	Penjelasan
Id_Nasabah	Data kategori yang berisi ID unik untuk setiap nasabah
Usia	Data kategori yang beris usia nasbah
Jenis_Kelamin	Data kategori yang berisi jenis kelamin nasabah
Jumlah_Tanggungan	Data kategori yang berisi jumlah tanggungan dalam keluarga nasabah
Pendidikan	Data kategori yang berisi pendidikan terakhir nasabah
Status_Pernikahan	Data kategori yang berisi status nasbah apakan sudah menikah atau belum menikah
Pendapatan	Data kategori yang berisi pendapat nasabah setaip bulannya
Kategori_Kartu	Data kategori yang berisi jenis warna kartu kredit yang nasabah gunakan
Lama_Menjadi_Nasabah	Data kategori yang berisi berapa lama nasabah menjadi pengguna bank tersebut
Jumlah_Layanan	Data kategori yang berisi berapa banyak layanan nasabah yang gunakan

Limit_Kredit	Data kategori yang berisi berapa limit keredit nasabah
Saldo_Revolting_Total	Data kategori yang berisi jumlah total saldi revolting nasabah
Total_Transaksi	Data kategori yang berisi jumlah total transaksi nasabah
Jumlah_Melakukan_Transaksi	Data kategori yang berisi jumlah nasabah melakukan transaksi
Rasio_Penggunaan_Rata-rata	Data kategori yang berisi rata-rata penggunaan transaksi nasabah

Tabel 4. 2 Informasi Dataset

No	**Kolom**	**Jumlah *Non-Null Value***	**Tipe Data**
0	Id_Nasabah	1050	Int64
1	Usia	1000	Float64
2	Jenis_Kelamin	1050	Object
3	Jumlah_Tanggungan	1050	Int64
4	Pendidikan	883	Object
5	Status_Pernikahan	976	Object
6	Pendapatan	932	Float64
7	Kategori_Kartu	1050	Object
8	Lama_Menjadi_Nasabah	1050	Int64
9	Jumlah_Layanan	1050	Int64
10	Limit_Kredit	1050	Float64
11	Saldo_Revolting_Total	1050	Int64
12	Total_Transaksi	1050	Int64
13	Jumlah_Melakukan_Transaksi	1050	Int64
14	Rasio_Penggunaan_Rata-rata	1050	Float64

Pada Table 4.2 menampilkan informasi dataset tiap kolom dengan jumlah *Non-Null Value* dan tipe datanya. *Non-Nul Value* merupakan elemen atau entri data set yang memiliki nilai yang terdefinisi. Table diatas berisi 1050 entri dengan 15 kolom yang memuat informasi tentang transaksi, nasabah, dan detail terkait lainnya. Meskipun sebagian besar kolom memiliki data lengkap, ada beberapa nilai yang hilang di kolom Pendidikan (167 nilai hilang), Status_Pernikahan (74 nilai hilang), dan Pendapatan (118 nilai hilang). Tipe data yang digunakan beragam, mulai dari string (object), angka decimal (float64), hingga angka bulat (int64).

3. Data Preparation

Penelitian menggunakan *Google Colab* dalam pemprosesan datasetnya, untuk memudahkan perhitungan nilai-nilai dalam proses perhitungan. Maka menggunakan tahapan pemprosesan dari tranformasi data dan memvisualisasi data.

Tabel 4. 3 Infromasi Data setelah penghapus baris atau kolom dari DataFrame

No	**Kolom**	**Jumlah *Non-Null Value***	**Tipe Data**
0	Id_Nasabah	1050	Int64
1	Usia	1000	Float64
2	Jenis_Kelamin	1050	Object
3	Jumlah_Tanggungan	1050	Int64
4	Pendidikan	883	Object
5	Status_Pernikahan	976	Object
6	Pendapatan	932	Float64
7	Kategori_Kartu	1050	Object
8	Lama_Menjadi_Nasabah	1050	Int64
9	Jumlah_Layanan	1050	Int64
10	Limit_Kredit	1050	Float64

11	Saldo_Revolting_Total	1050	Int64
12	Total_Transaksi	1050	Int64
13	Jumlah_Melakukan_Transaksi	1050	Int64

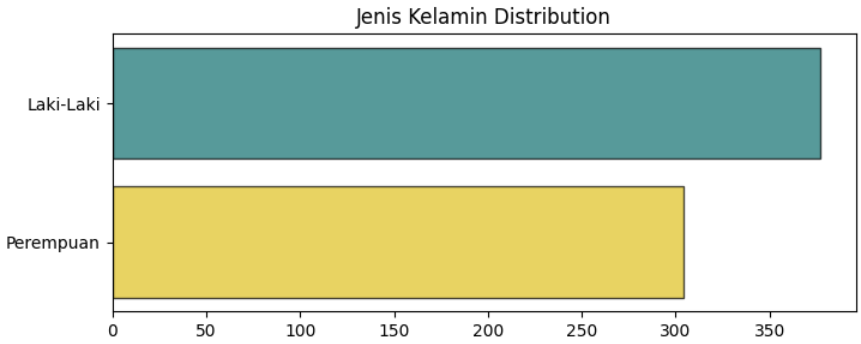
Pada Table 4.3 dilakukan penghapusan kolom Jumlah_Melakukan_Transaksi dan Rasio_Penggunaan_Rata-Rata. Hal ini dilakukan karna kolom tersebut tidak diperlukan untuk analisis lebih lanjut, karena tidak memiliki nilai hilang, atau tidak relevan dengan tujuan analisis.

Tabel 4. 4 Rata-rata Mean dan Median

	Mean	**Median**
Id_Nasabah	7.399795e+08	717569283.0
Usia	4.628634e+01	46.0
Jumlah_Tanggungan	2.318649e+00	2.0
Pendapatan	5.866732e+04	51357.0
Lama_Menjadi_Nasabah	3.625698e+01	36.0
Jumlah_Layanan	3.950073e+00	4.0
Limit_Kredit	8.244363e+03	4416.0
Saldo_Revolting_Total	1.244843e+03	1367.0
Total_Transaksi	4.724887e+03	4092.0

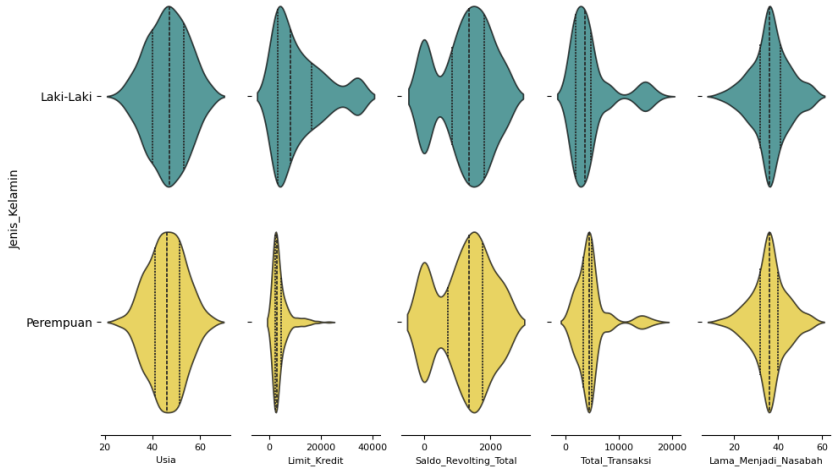
Pada table 4.4 menampilkan hasil statistik deskriptif dari data nasabah, meliputi rata-rata (mean) dan nilai tengah (median) dari beberapa variabel. Fungsi dari mean dan median adalah pengukuran pemusatan dalam analisis statistic dan data. Keduanya memberikan informasi tentang titik tengah atau pusat dari distribusi data. Mean sendiri adalah jumlah total dari semua nilai dataset dibagi dengan nilai tersebut, sedangkan median adalah nilai tengah dari dataset yang diurutkan, apabila jumlah nilai dalam dataset ganjil median adalah nilai yang berbeda di posisi tengah. Pada Id_Nasabah menunjukkan rata-rata dan median nilai ID nasabah. Karena ID nasabah biasanya berupa angka acak atau berurutan, nilai ini tidak memiliki arti yang

signifikan secara statistik. Usia rata-rata nasabah adalah sekitar 46.3 tahun, dengan usia tengah (median) juga 46 tahun, menunjukkan distribusi usia yang cukup simetris. Rata-rata jumlah tanggungan per nasabah adalah sekitar 2.3, dengan nilai median 2, menunjukkan bahwa sebagian besar nasabah memiliki sekitar 2 tanggungan. Rata-rata pendapatan nasabah adalah sekitar 58,667.32, dengan median 51,357. Ini menunjukkan bahwa ada beberapa nasabah dengan pendapatan yang sangat tinggi yang mempengaruhi rata-rata (mean) menjadi lebih tinggi daripada median. Rata-rata lama waktu menjadi nasabah adalah sekitar 36.3 tahun, dengan median juga 36 tahun, menunjukkan distribusi yang cukup seimbang. Rata-rata jumlah layanan yang digunakan oleh nasabah adalah sekitar 3.95, dengan median 4, menunjukkan sebagian besar nasabah menggunakan sekitar 4 layanan. Rata-rata limit kredit yang diberikan adalah sekitar 8,244.36, dengan median 4,416. Ini menunjukkan ada beberapa nasabah dengan limit kredit yang sangat tinggi, yang mempengaruhi rata-rata menjadi lebih tinggi daripada median. Rata-rata saldo revolving total adalah sekitar 1,244.84, dengan median 1,367. Ini menunjukkan distribusi saldo yang cukup seimbang. Rata-rata total transaksi adalah sekitar 4,724.89, dengan median 4,092. Ini menunjukkan adanya beberapa nasabah dengan total transaksi yang sangat tinggi yang mempengaruhi rata-rata menjadi lebih tinggi daripada median. Secara keseluruhan, data ini memberikan gambaran mengenai profil rata-rata dan distribusi nasabah berdasarkan beberapa variabel kunci. Median yang lebih rendah dari mean pada beberapa variabel (seperti Pendapatan dan Limit_Kredit) menunjukkan adanya beberapa nilai ekstrim yang tinggi dalam dataset.



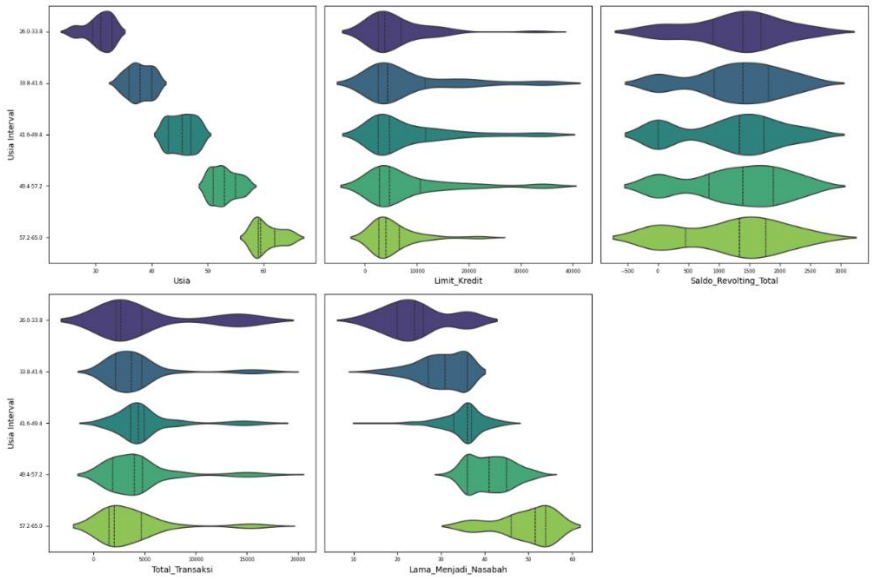
Gambar 4. 1 Distribusi Jenis Kelamin

Pada Gambar 4.1 menampilkan grafik batang distibusi jenis kelamin yang menunjukan jumlah atau frekuensi jenis kelamin laki-laki dan perempuan. Pada grafik menampilkan sumbu Y untuk jenis kelamin dan sumbu X untuk jumlah atau frekuensi setiap kategori jenis kelamin dengan skala yang berjalan dari 0 hingga sekitar 360. Pada batang warna hijau untuk jenis kelamin laki-laki terdapat sekitar 360 individu dalam kategori ini. Sedangkan pada batang warna kuning terdapat 340 individu dalam kategori ini. Jadi perbandingan populasi laki-laki sedikit lebih besar dibandingkan populasi perempuan.



Gambar 4. 2 Pebandingan Distribusi beberapa Variabel berdasarkan Jenis Kelamin

Gambar 4.2 menunjukan beberapa variable berdasarkan jenis kelamin. Pada variabel usia laki-laki dan perempuan memiliki distribusi yang sama yaitu disekitar usia 30 hingga 40 tahun. Pada variabel limit kredit laki-laki memiliki distribusi di sekitar 20.000 hingga 40.000 dan pada perempuan memiliki distribusi di sekitar 20.000 hingga 30.000. Pada variabel saldo revolving total distribusi untuk laki-laki cukup variatif, namun sebagian besar nilai berada dibawah 2.000 dan pada perempuan menunjukan bahwa sebagian besar nilai juga berada dibawah 2.000, dengan distribusi yang sedikit lebih sempit dibandingkan laki-laki. Pada variabel total transaksi distribusi laki-laki cukup bervariasi dengan puncak distribusi di sekitar 5.000 hingga 10.000 dan pada perempuan menunjukan rentang yang lebih sempit dengan puncak distribusi di sekitar 5.000 hingga 10.000. Pada variabel lama menjadi nasabah distribusi untuk laki-laki dan perempuan sama menunjukan puncak distribusi disekitar 40 hingga 50 tahun.



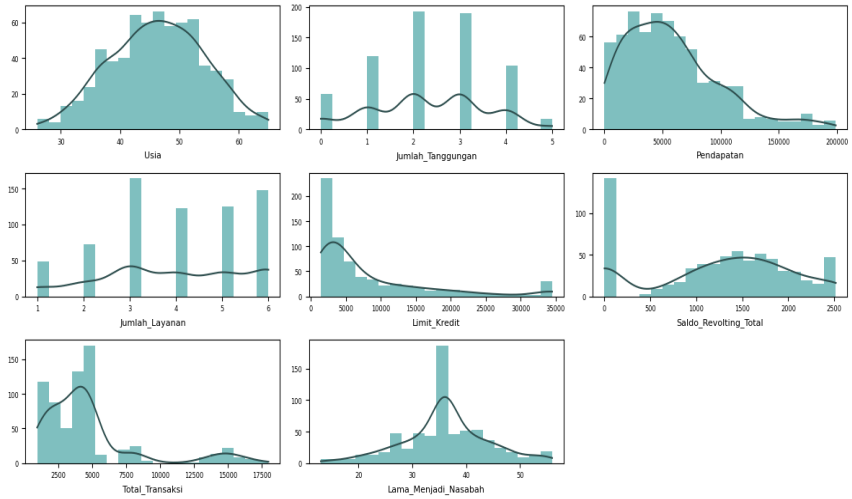
Gambar 4. 3 Bar Chat beberapa Variabel berdasarkan Usia

Gambar 4.3 menunjukan bar chat beberapa variabel berdasarkan interval usia. Pada usia menunjukan distribusi usia dalam beberapa

intervalnya yang berbeda, dengan rentang yang semakin lebar seiring dengan bertambahnya usia dengan puncak distribusi terlihat lebih merata disetiap intervalnya. Pada limit kredit interval usia 26.0-33.8 memiliki rentan yang cukup luas dengan puncak distribusi di sekitar 10.000 hingga 20.000, interval usia 33.8-41.6 memiliki distribusi lebih merata dengan puncak distribusi di sekitar 10.000 hingga 20.000, interval usia 41.6-49.4 distribusi limit kreditnya memiliki puncak distribusi yang serupa dengan interval usia sebelumnya, yaitu disekitar 10.000 hingga 20.000, interval usia 49.4-57.2 distribusi limit kreditnya memiliki puncak yang lebih besar, namun measih berkisar di sekitar 10.000 hingga 20.000, interval usia 57.2-65.0 distribusi limitnya memiliki puncak distribusi yang lebih besar dengan rentang yang lebih luas.

Pada saldo revolving total interval usia 26.0-33.8 menunjukan distribusi interval usianya memiliki puncak distribusi disekitar 500 hingga 1.500, interval usia 33.8-41.6, 41.6-49.4, 49.4-57.2 menunjukan distribusi saldo revolving total memiliki puncak distribusi disekitar 1.000 hingga 1.500, interval usia 57.2-65.0 menunjukan distribusi saldo revilving total memiliki puncak distribusi di sekitar 1.000 hingga 1.500 dengan rantang yang lebih luas diabnadingkan interval usia lainnya. Pada total transaksi interval usia 26.0-33.8 menunjukan distribusi total transaksi memiliki rentang yang yang luas dengan puncak distribusi di sekitar 5.000 hingga 10.000, interval usia 33.8-41.6, 41.6-49.4, 49.4-57.2 menunjukan distribusi total transaksi memiliki rentang yang cukup luas dengan puncak distribusi sama di sekitar 5.000 hingga 10.000, sedangkan interval usia 57.2-65.0 menunjukan distribusi total transaksi memiliki rentang yang lebih sempit dengan puncak distribusi yang lebih rendah. Pada garafik lama menjadi nasbah interval usia 26.0-33.8 dan 33.8-41.6 menunjukan distribusi lama menjadi nasabah memiliki puncak distribusi di sekitar 10 hingga 20 tahun, interval usia 41.6-49.4 dan 49.4-57.2 menunjukan distribusi lama menjadi nasabah memiliki puncak distribusi di sekitar 20 hingga 30 tahun,

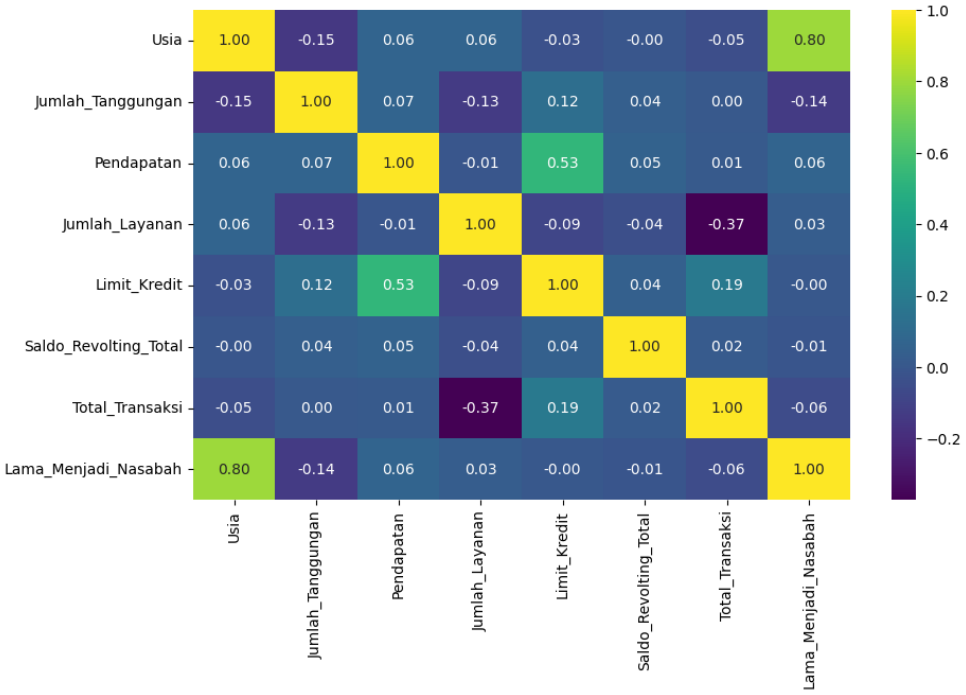
sedangkan interval usia 57.2-65.0 menunjukan distribusi lama menjadi nasabah memiliki puncak distribusi di sekitar 20 hingga 30 tahun dengan rentang yang lebih luas.



Gambar 4. 4 Histogram Distribusi dari beberapa Variabel

Gambar 4.4 menampilkan beberapa histogram dari hasil visualisasi pendistribusian dari kolom-kolom numeric dalam dataset. Fungsi dari histogram di atas untuk melihat pola distribusi, kemencengan, dan potensi *outliers* dalam data. Pada histrogram usia menunjukan simetris dengan puncak distribusi di sekitar usia 45-50 tahun, sebagian besar nasabah berada dalam rentang usia 30-60 tahun. Pada histogram jumlah tanggungan cenderung multimodal dengan puncak diatribusi pada nilai 0, 2 dan 4 yang menunjukkan bahwa sebagaian besar nasabah memilki 0, 2 dan 4 tanggungan. Pada grafik pendapatan menunjukkan puncak distribusi di sekitar 50.000 hingga 75.000 dengan penurunan tajam setelahnya, ini menunjukan bahwa sebagian besar nasabah memiliki pendapatan dalam rentan tersebut da nada beberapa nasabah dengan pendapatan yang sangat tinggi. Pada grafik jumlah layanan menunjukkan beberapa puncak distribusi yang berbeda dengan puncak utama di angka 3, ini menunjukkan bahwa sebagian besar nasabah menggunakan sekitar 3 layanan yang ditawarkan

oleh bank. Pada grafik limit kredit menunjukkan distribusi limit kredit cenderung miring ke kanan (*positively skewed*) dengan puncak distribusi di sekitar 5.00 hingga 10.000, sebagian beasr nasabah memiliki limit kredit yang relatif rendah, namun ada beberapa nasabah dengan limit kredit yang sangat tinggi. Pada grafik saldo revilving total menunjukkan bahwa distribusi saldo revilving total cenderung miring ke kanan dengan puncak distribusi di sekitar 0 hingga 500, ada penurunan jumlah nasabah dengan saldo yang tinggi, namun ada beberapa nasabah dengan saldo yang sangat tinggi. Pada grafik total transaksi menunjukan bahwa puncak distribusi di sekitar 5.000 hingga 7.500, sebagian besar nasabah memiliki total transaksi dalam rentang ini, namun ada beberapa nasabah dengan total transaksi yang sangat tinggi. Pada grafik lama menjadi nasabah distribusinya cenderung simetris dengan puncak distribusinya di sekitar 30 hingga 40 tahun, senagian besar nasabahnya telah menjadi nasabah dalam rentang waktu ini.



Gambar 4. 5 Matriks Korelasi

Gambar 4.5 adalah matriks korelasi yang menunjukkan hubungan linier antara pasangan variabel dalam dataset. Fungsi dari matriks korelasi ini adalah untuk mengukur kekuatan dan arah hubungan kinier atara dua variabel, membatu dalam mengidentifikasi pasangan variabel yang memiliki hubungan yang kuat baik positif maupun negatif, visualisasi korelasi membatu dalam memberikan wawasan cepat tentang data. Matriks korelasi ini menggunakan skala dari -1 hingga 1 yang di mana 1 menunjukkan korelasi positif sempurna, -1 menunjukkan korelasi negatif sempurna.0 menunjukkan tidak ada korelasi.

Berikut penjelasan hasil matrik diatas:

a. Usia:
    1) Lama_Menjadi_Nasabah: Korelasi positif sangat kuat (0.80), menunjukkan bahwa semakin lama seseorang menjadi nasabah, semakin tua usianya.
    2) Jumlah_Tanggungan: Korelasi negatif lemah (-0.15), menunjukkan sedikit kecenderungan bahwa semakin tua usia seseorang, semakin sedikit tanggungannya.

b. Korelasi dengan variabel lainnya sangat lemah (mendekati 0), menunjukkan tidak ada hubungan linier yang signifikan.
   Jumlah_Tanggungan:

    1) Limit_Kredit: Korelasi positif lemah (0.12), menunjukkan sedikit kecenderungan bahwa semakin banyak tanggungan seseorang, semakin tinggi limit kreditnya.
    2) Pendapatan: Korelasi positif lemah (0.07), menunjukkan sedikit kecenderungan bahwa semakin tinggi pendapatan seseorang, semakin banyak tanggungannya.
    3) Korelasi dengan variabel lainnya sangat lemah (mendekati 0), menunjukkan tidak ada hubungan linier yang signifikan.

c. Pendapatan:
    1) Limit_Kredit: Korelasi positif sedang (0.53), menunjukkan bahwa semakin tinggi pendapatan seseorang, semakin tinggi limit kreditnya.
    2) Korelasi dengan variabel lainnya sangat lemah (mendekati 0), menunjukkan tidak ada hubungan linier yang signifikan.

d. Jumlah_Layanan:

    1) Total_Transaksi: Korelasi negatif sedang (-0.37), menunjukkan bahwa semakin banyak layanan yang digunakan oleh nasabah, semakin sedikit total transaksinya.
    2) Korelasi dengan variabel lainnya sangat lemah (mendekati 0), menunjukkan tidak ada hubungan linier yang signifikan.

e. Limit_Kredit:

    1) Pendapatan: Korelasi positif sedang (0.53), seperti yang sudah dijelaskan sebelumnya.
    2) Total_Transaksi: Korelasi positif lemah (0.19), menunjukkan sedikit kecenderungan bahwa semakin tinggi limit kredit seseorang, semakin banyak total transaksinya.
    3) Korelasi dengan variabel lainnya sangat lemah (mendekati 0), menunjukkan tidak ada hubungan linier yang signifikan.

f. Saldo_Revolving_Total:

    1) Korelasi dengan semua variabel sangat lemah (mendekati 0), menunjukkan tidak ada hubungan linier yang signifikan.

g. Total_Transaksi:

    1) Jumlah_Layanan: Korelasi negatif sedang (-0.37), seperti yang sudah dijelaskan sebelumnya.

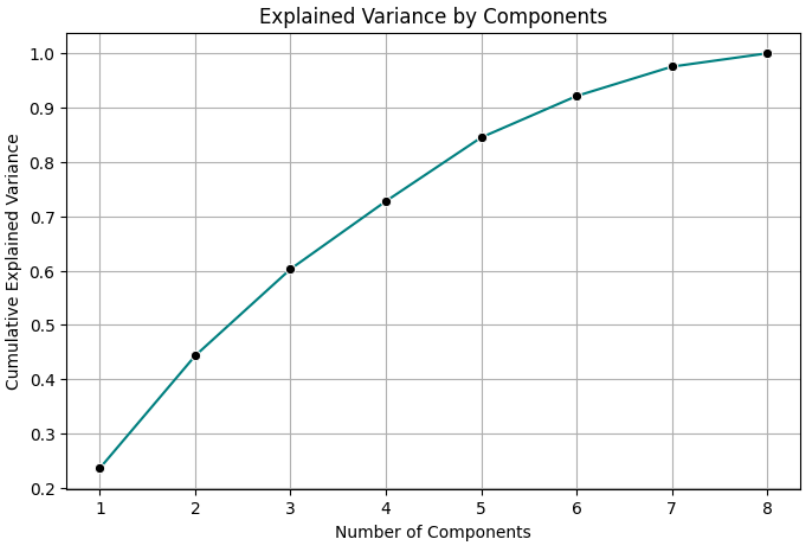
2) Limit_Kredit: Korelasi positif lemah (0.19), seperti yang sudah dijelaskan sebelumnya.
3) Korelasi dengan variabel lainnya sangat lemah (mendekati 0), menunjukkan tidak ada hubungan linier yang signifikan.

h. Lama_Menjadi_Nasabah:

1) Usia: Korelasi positif sangat kuat (0.80), seperti yang sudah dijelaskan sebelumnya.
2) Korelasi dengan variabel lainnya sangat lemah (mendekati 0), menunjukkan tidak ada hubungan linier yang signifikan.

4. Data *Modelling*

Metode yang digunakan untuk pemodelan data mining dalam penelitian ini adalah algoritma K-means *Clustering* untuk menentukan jumlah klaster. Langkah awal yang dilakukan adalah melakukan standarisasi data untuk memastikan bahwa semua fitur memiliki standar deviasi 1 dan rata-rata 0. Selanjutnya, digunakan metode PCA untuk menentukan variabel yang penting dalam data dengan mengambil jumlah komponen yang menyumbang 70% sampai 80% [22]. Gambar 4.6 menunjukkan hasil dari penerapan metode PCA.



Gambar 4. 6 Kumulatif Varian Variabel dengan PCA

Seperti yang terlihat pada Gambar 4.6 adalah grafik kumulatif varian variabel dengan PCA. Garfik di atas digunakan untuk menmpilkan hasil analisis komponen utama PCA pada data yang telah diskalakan (X_scaled), dan kemudian memvisualisasikan varians yang dijelaskan secara kumulatif oleh setiap komponen utama. Komponen pertama memiliki varian sebesar 0,21, diikuti komponen kedua sebesar 0,45 dan komponen ketiga sebesar 0,6. Sebanyak 3 komponen di perlukan untuk mempertahankan 70% dari varian data. Berdasarkan hasil analisis, terdapat 3 komponen variabel yang dianggap penting dan akan digunakan dalam algoritma K-means. PCA sendiri adalah teknik reduksi dimensi yang sering digunakan sebelum menerapkan algoritma clustering, PCA juga berfungsi untuk pengurangan jumlah filter dalam dataset dengan pengubahan fitur asdli menjadi serangkaian fitur baru yang disebut *principal components*.

Tabel 4. 5 Tabel Hasil Transformasi PCA

PC1	PC2	PC3
6576.96	1925.54	-51.23
61449.16	19175.99	-3777.44

-37561.50	-2041.59	754.98
19484.12	4461.83	-1117.77
15115.25	-2633.89	-2054.87

Perhitungan k-means clustering setelah transformasi PCA dengan 4 klaster:

a) Inisialisasi Centroid

Inisialisasi centroid untuk 4 klaster (k = 4) sebagai berikut:

Centroid 1 = [6576.96, 1925.54, -51.23]

Centroid 2 = [61449.16, 19175.99, -3777.44]

Centroid 3 = [-37561.50, -2041.59, 754.98]

Centroid 4 = [19484.12, 4461.83, -1117.77]

Centroid 4 = [15115.25, -2633.89, -2054.87]

Penugasan:

Data 0 = Centroid 1

Data 1 = Centroid 2

Data 2 = Centroid 3

Data 3 = Centroid 4

Data 4 = Centroid 4

b) Assign Cluster

Menghitung jarak Euclidean dari setiap titik ke centroid dan assign ke cluster terdekat.

Rumus jarak Euclidean antara dua titik A = ($a1$, $a2$, $a3$) dan B = ($b1$, $b2$, $b3$):

Distance = $\sqrt{(a1 - b1)^2 + (a2 - b2)^2 + (a3 - b3)^2}$

Perhitungan jarak:

1) Jarak dari data 0 ke centroid

Jarak ke centroid 1:

$$d(0,1) = \sqrt{\begin{matrix}(6576.96 - 6576.96)^2 + \\ (1925.54 - 1925.54)^2 \\ +(-51.23 + 51.23)^2\end{matrix}} = 0$$

Jarak ke centroid 2:

$$d(0,2) = \sqrt{\begin{matrix}(6576.96 - 61449.16)^2 + \\ (1925.54 - 19175.99)^2 \\ +(-51.23 + 3777.44)^2\end{matrix}} = 62325.68$$

Jarak ke centroid 3:

$$d(0,3) = \sqrt{\begin{matrix}(6576.96 - (-37561.52))^2 + \\ (1925.54 - (-2041.60))^2 \\ +(-51.23 - 754{,}98)^2\end{matrix}} = 44600.37$$

Jarak ke centroid 4:

$$d(0,4) = \sqrt{\begin{matrix}(6576.96 - 19484.12)^2 + \\ (1925.54 - 4461.48)^2 \\ +(-51.23 + 1117.77)^2\end{matrix}} = 13880.23$$

2) Jarak dari data 1 ke centroid

Jarak ke centroid 1:

$$d(1,0) = \sqrt{\begin{matrix}(61449.16 - 6576.97)^2 + \\ (19175.99 - 1925.54)^2 \\ +(-3777.45 - (-3777.44))^2\end{matrix}} = 62325.68$$

Jarak ke centroid 2:

$$d(1,2) = \sqrt{\begin{matrix}(61449.16 - 61449.16)^2 + \\ (19175.99 - 19175.99)^2 \\ +(-33777.44 - (-3777.44))^2\end{matrix}} = 0$$

Jarak ke centroid 3:

$$d(1,3) = \sqrt{\begin{matrix}(61449.16 - (-37561.51))^2 + \\ (19175.99 - (-2041.60))^2 \\ +(-3777.45 - 754.98)^2\end{matrix}} = 108903.23$$

Jarak ke centroid 4:

$$d(1,4) = \sqrt{\begin{matrix}(61449.16 - 19484.12)^2 + \\ (19175.99 - 4461.84)^2 \\ +(-3777.45 - (-1117.77))^2\end{matrix}} = 42687.99$$

3) Jarak dari data 2 ke centroid

Jarak ke centroid 1:

$$d(2,1) = \sqrt{\begin{matrix}(-37561.51 - 6576.97)^2 + \\ (-2041.60 - 1925.54)^2 \\ +(754.98 - (-51.24))^2\end{matrix}} = 44600.37$$

Jarak ke centroid 2:

$$d(2,2) = \sqrt{\begin{matrix}(-37561.51 - 61449.16)^2 + \\ (-2041.60 - 19175.99)^2 \\ +(754.98 - (-3777.45))^2\end{matrix}} = 108903.23$$

Jarak ke centroid 3:

$$d(2,3) = \sqrt{\begin{matrix}(-37561.51-(-37561.51)^2+\\(-2041.60-(-2041.60))^2\\+(754.98-754.98)^2\end{matrix}} = 0$$

Jarak ke centroid 4:

$$d(2,4) = \sqrt{\begin{matrix}(-37561.51-19484.12)^2+\\(-2041.60-4461.84)^2\\+(754.98-(-1117.77))^2\end{matrix}} = 75795.66$$

4) Jarak dari data 3 ke centroid

Jarak ke centroid 1:

$$d(3,1) = \sqrt{\begin{matrix}(19484.12-6576.97)^2+\\(4461.84-1925.54)^2\\+(-1117.77-(-51.24))^2\end{matrix}} = 13880.23$$

Jarak ke centroid 2:

$$d(3,2) = \sqrt{\begin{matrix}(19484.12-61449.16)^2+\\(4461.84-19175.99)^2\\+(-1117.77-(-3777.45))^2\end{matrix}} = 42687.99$$

Jarak ke centroid 3:

$$d(3,3) = \sqrt{\begin{matrix}(19484.12-(-37561.51))^2+\\(4461.84-(-2041.60))^2\\+(-1117.77-754.98)^2\end{matrix}} = 75795.66$$

Jarak ke centroid 4:

$$d(3,4) = \sqrt{\begin{matrix}(19484.12 - 19484.12)^2 + \\ (4461.84 - 4461.84)^2 \\ +(-1117.77 - (-1117.77))^2\end{matrix}} = 0$$

5) Jarak dari data 4 ke centroid

Jarak ke centroid 1:

$$d(4,1) = \sqrt{\begin{matrix}(15115.26 - 6576.97)^2 + \\ (-2633.89 - 1925.54)^2 \\ +(-2054.88 - (-51.24))^2\end{matrix}} = 9394.38$$

Jarak ke centroid 2:

$$d(4,2) = \sqrt{\begin{matrix}(15115.26 - 61449.16)^2 + \\ (-2633.89 - 19175.99)^2 \\ +(-2054.88 - (-3777.45))^2\end{matrix}} = 64219.32$$

Jarak ke centroid 3:

$$d(4,3) = \sqrt{\begin{matrix}(15115.26 - (-37561.51))^2 + \\ (-2633.89 - (-2041.60))^2 \\ +(-2054.88 - 754.98)^2\end{matrix}} = 9394.38$$

Jarak ke centroid 4:

$$d(4,4) = \sqrt{\begin{matrix}(15115.26 - 19484.12)^2 + \\ (-2633.89 - 4461.84)^2 \\ +(-2054.88 - (-1117.77))^2\end{matrix}} = 7062.98$$

c) Pembaruan Centroid

Centroid untuk cluster 1 (data 0):

- PC1 : 6576.97

- PC2 : 1925.54
- PC3 : -51.24

Centroid untuk cluster 2 (data 1):

- PC1 : 61449.16
- PC2 : 19 175.99
- PC3 : -3777.45

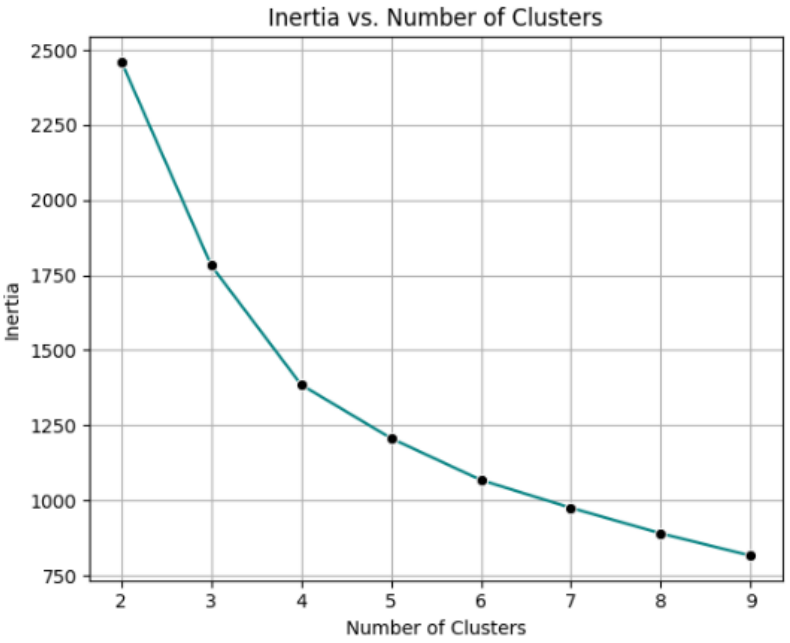
Centroid untuk cluster 3 (data 2):

- PC1 : -3756.51
- PC2 : -2041.60
- PC3 :754.98

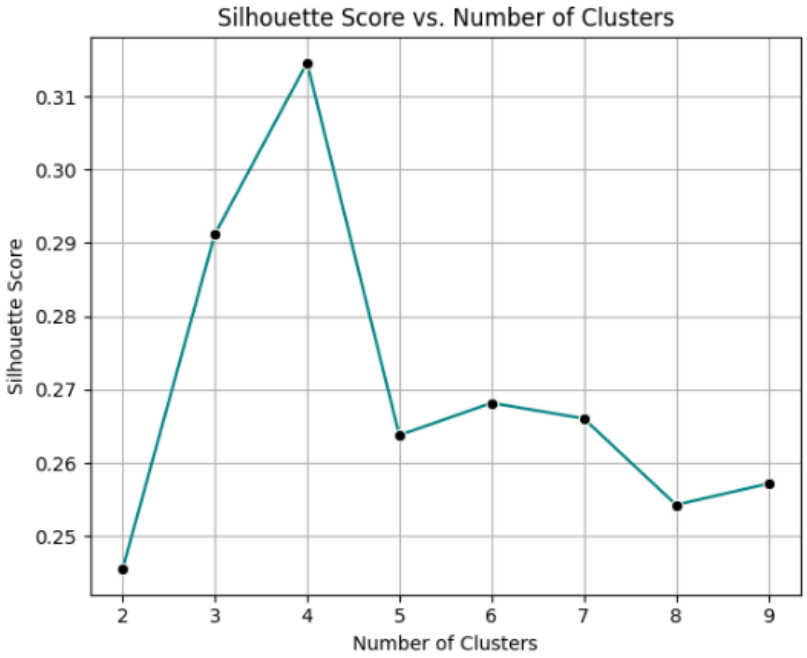
Centroid untuk cluster 4 (data 3 dan data 4):

- PC1 : 17399.69
- PC2 : 919.97
- PC3 : -1606.32

Selanjutnya, dua model yang digunakan untuk menentukan pusat cluster atau tingkat akurasi klaster yang optimal dalam K-Means, yaitu *Inersia* dan *Silhouette Score.* Inersia merupakan total dari jarak kuadrat antara setiap titik data dengan centroid klasternya. Semakin kecil inersia, semakin baik Clusteringnya. Sedangkan *Silhouette Score* digunakan untuk mengukur seberapa mirip satau data dengan klasternya sendiri dibandingkan dengan klasterlainnya. Hasil dari kedua medel tersebut dapat terlihat pada Gambar 4.7.



Gambar 4. 7 *Inersia*



Gambar 4. 8 *Silhouette Score*

Berdasarkan kedua hasil visualisasi model *Inersia* dan *Silhouette Score* di atas, didapatkan bahwa model *Inersia* menghasilkan nilai *k* antara

3 dan 4 dan model *Silhouette Score* mencapai hasil dengan nilai $k = 4$. Maka dapat diambil nilai yang sama yaitu $k = 4$. Kedua metode evaluasi klaster yang digunakan menunjukkan bahwa klaster dengan jumlah 4 adalah yang paling optimal.

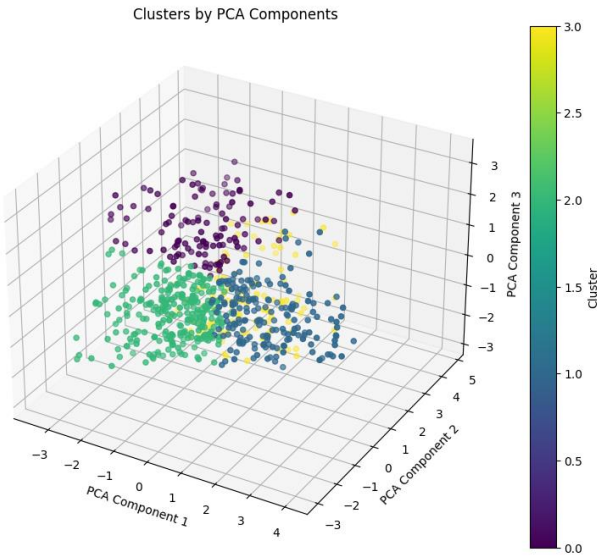
Tabel 4. 6 Sampel Jarak Setiap Data dengan Titik Pusat Klaster

PCA1	PCA2	PCA3	Cluster
-1.579408	0.585301	0.825344	0
0.959798	3.033119	-1.716245	3
0.131718	-0.820958	0.193267	2
0.618807	0.850520	0.088852	1
-0.563672	-0.343593	-0.677228	2

Pada Tabel 4.6 menunjukkan komponen utama (PCA1, PCA2, PCA3) dari enam sampel data beserta klaster yang sesuai berdasarkan titik pusat klaster. Pada tabel menunjukkan nilai dari tiga komponen utama (PCA1, PCA2, PCA3) yang dihasilkan dari proses PCA. Nilai-nilai ini adalah representasi dari data asli yang telah ditransformasi ke dalam ruang komponen utama. Setiap baris merepresentasikan sebuah sampel data, dan nilai-nilai dalam kolom PCA1, PCA2, dan PCA3 menunjukkan posisi sampel tersebut dalam ruang tiga dimensi komponen utama. Selanjutnya kolom klaster menunjukkan klaster yang dihasilkan dari algoritma K-Means *Clustering* untuk setiap sampel data. Nilai dalam kolom ini mengindikasikan ke klaster mana sampel tersebut ditempatkan.

5. Evaluation

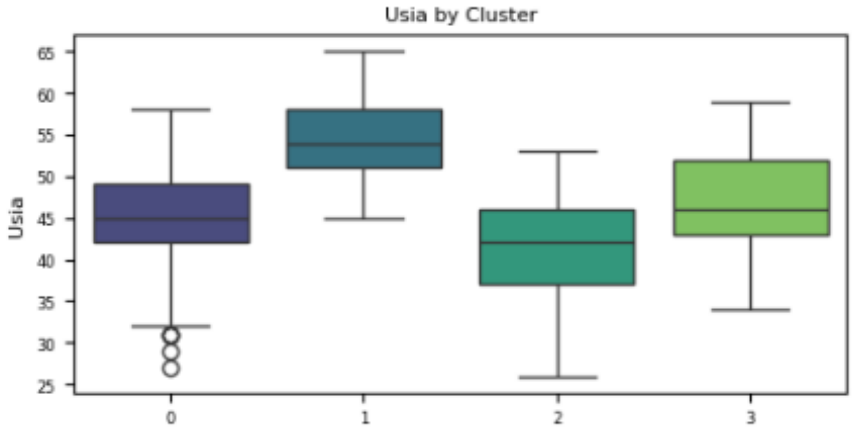
Hasil dari analisis K-Means dapat divisualisasikan menggunakan scatter plot agar dapat dengan mudah melihat pola atau struktur yang muncul dari pengelompokan tersebut. Gambar 4.9 adalah visualisasi hasil clustering menggunakan algoritma K-Means dan PCA.



Gambar 4. 9 Hasil Clustering Menggunakan K-Means dan PCA

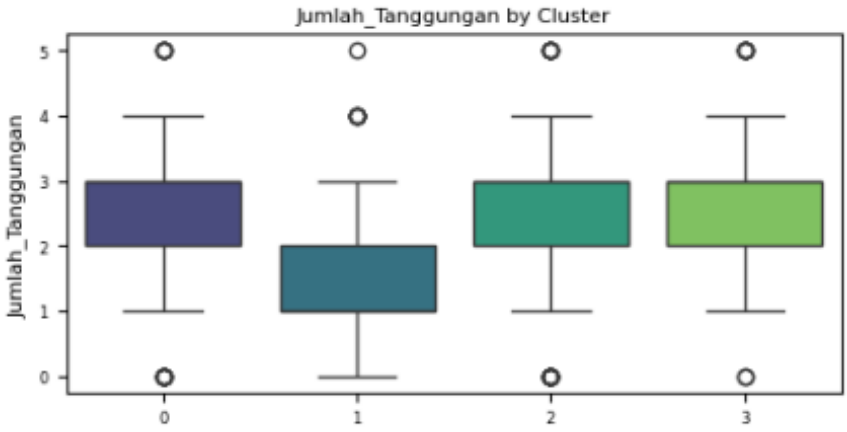
Gambar 4.9 merupakan hasil *clustering* menggunakan K-Means dan PCA yang terbagi menjadi 4 klaster, menunjukkan bahwa data telah berhasil dikelompokkan menjadi empat kelompok yang berbeda berdasarkan kesamaan fitur. Setiap titik pada *scatter plot* mewakili sebuah data dalam dataset yang telah dikelompokkan ke dalam salah satu klaster. Warna atau symbol yang berbeda dapat digunakan untuk membedakan setiap klaster. Hal ini dapat dilakukan karena telah dilakukan reduksi dimensi pada data dengan menggunakan PCA. Tanpa menggunakan PCA, karakteristik dari setiap klaster akan sulit dibedakan karena setiap klasternya memiliki dimensi data yang tinggi.

Warna dari titik-titik tersebut menunjukkan klaster mana. Skala warna di sebelah kanan (dari biru ke kuning) menunjukkan nomor klaster. Dari analisis hasil, kita bisa melihat bahwa ada 4 klaster yang berbeda (0, 1, 2, dan 3). Klaster 0 diwakili oleh warna biru, klaster 1 diwakili warna hijau, klaster 2 diwakili oleh warna unggu, klaster 3 diwakili oleh warna kuning. Titik-titik yang memiliki warna yang sama cenderung berkumpul di daerah tertentu, menunjukkan bahwa data dalam klaster tersebut memiliki kesamaan fitur yang signifikan yang dipertahankan bahkan setelah reduksi dimensi.



Gambar 4. 10 Klaster Berdasarkan Umur

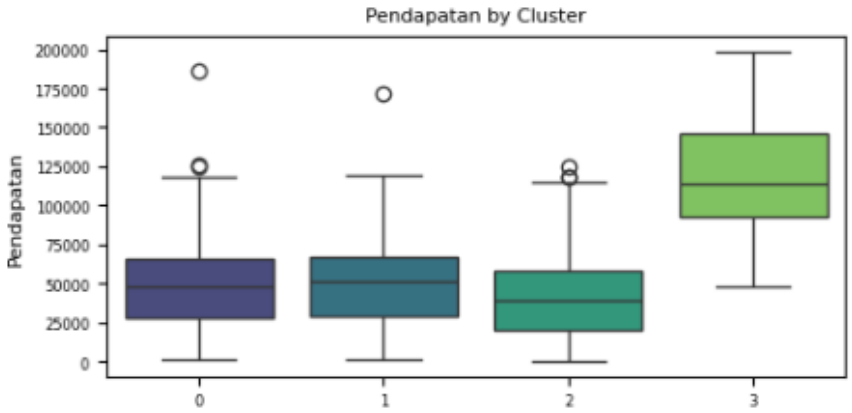
Pada Gambar 4.10 memberikan informasi mengenai setiap klaster berdasarkan usia. Terlihat bahwa pusat data di klaster nol berada pada rentang usia 43 tahun sampai 49 tahun. Selanjutnya pusat data di klaster satu berada pada rentang usia 52 tahun sampai 58 tahun. Pusat data klaster dua berada pada rentang usia 38 tahun sampai 47 tahun. Lalu yang terakhir klaster tiga di mana pusat data berada pada rentang usia 48 tahun sampai 53 tahun.



Gambar 4. 11 Klaster Berdasarkan Jumlah Tanggungan

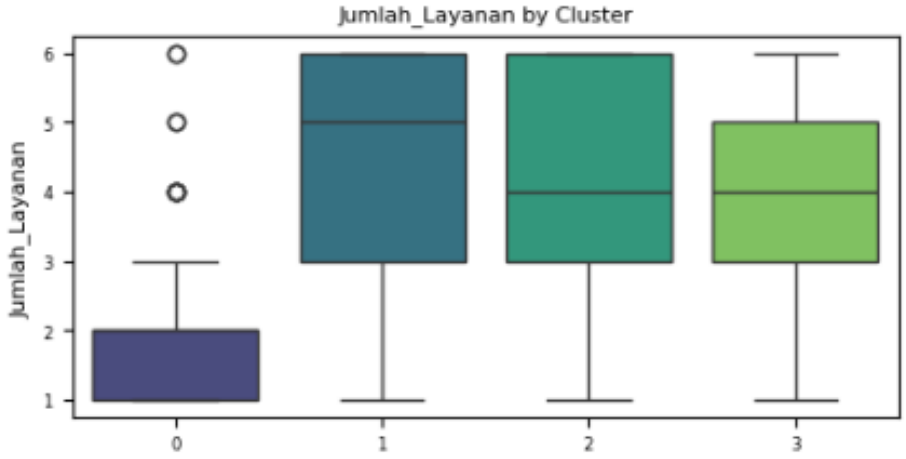
Gambar 4.11 memberikan informasi mengenai setiap klaster berdasarkan jumlah tanggungan. Pusat klaster nol dengan jumlah tanggungan 2 sampai 3 tanggungan. Selanjutnya pusat klaster satu dengan

jumlah tanggungan 1 sampai 2 tanggungan. Sedangakan pusat klaster dua dan tiga sama dengan jumlah tanggungan 2 sampai 3 tanggungan.



Gambar 4. 12 Klaster Berdasarkan Pendapatan

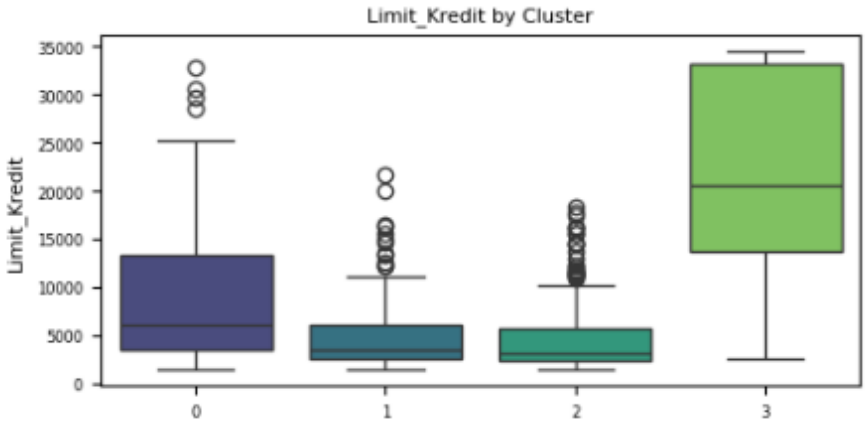
Gambar 4.12 memberikan informasi mengenai setiap kaster pendapatan nasabah. Terlihat pusat klaster nol dan satu pada pendapatan berkisar dari 30.000 sampai 60.000. Selanjutnya pusat klaster dua pada pendapatan berkisar dari 25.000 sampai 55.000. Terakhir pada pusat klaster tiga pada pendapatan berkisar dari 100.000 sampai 150.000.



Gambar 4. 13 Klaster Berdasarkan Jumlah Layanan

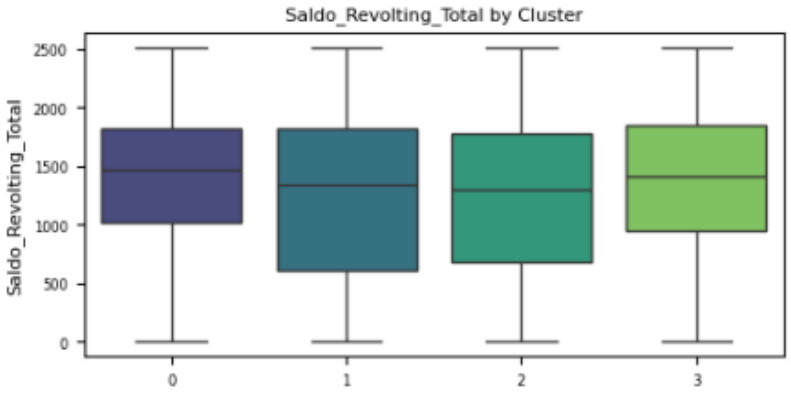
Gambar 4.13 memberikan informasi mengenai setiap klaster berdasarkan jumlah layanan. Terlihat bahwa pusat data di klaster nol berada pada jumlah layanan ke 1 sampai 2. Selanjutnya pusat data di klaster satu

dan dua berada pada jumlah layanan ke 3 sampai 6. Pusat data di klaster tiga berada pada jumlah layanan ke 3 sampai 5.



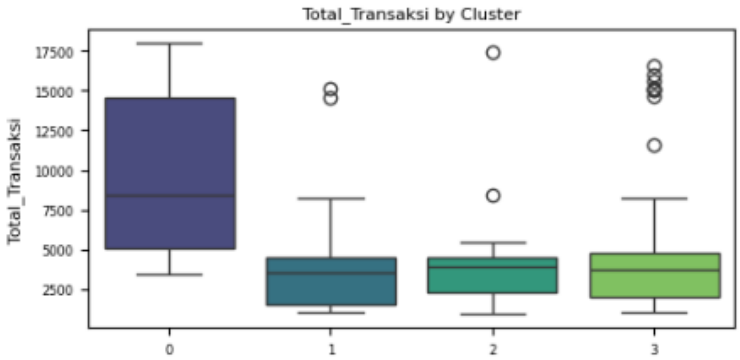
Gambar 4. 14 Klaster Berdasarkan Limit Kredit

Gambar 4.14 memberikan informasi mengenai setiap klaster berdasarkan limit kredit. Terlihat pada gambar bahwa pusat data di klaster nol berada pada limit kredit 3.000 sampai 14.000. Selanjutnya pusat data diklaster satu dan dua sama berada pada limit kredit 2.500 sampai 6.000. Terakhir pada pusat data di klaster tiga berada pada limit kredit 15.000 sampai 34.000.



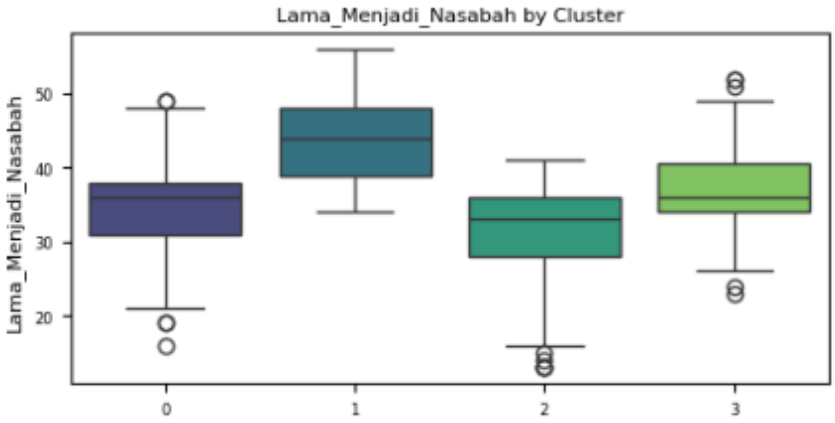
Gambar 4. 15 Klaster Berdasarkan Saldo Revolting Total

Gambar 4.15 memberikan informasi mengenai setiap klaster berdasarkan saldo revolting total. Terlihat pada gambar bahwa pusat data di klaster nol berada pada saldo revolting total 1.000 sampai 1.800. Selanjutnya pusat data di klaster satu berada pada saldo revolting total 600 sampai 1.800. Pusat data di klaster dua berada pada saldo revolting total 700 sampai 1.700. Terakhir pada pusat data di klaster tiga beada di saldo revolting total 1.000 sampai 1.900.



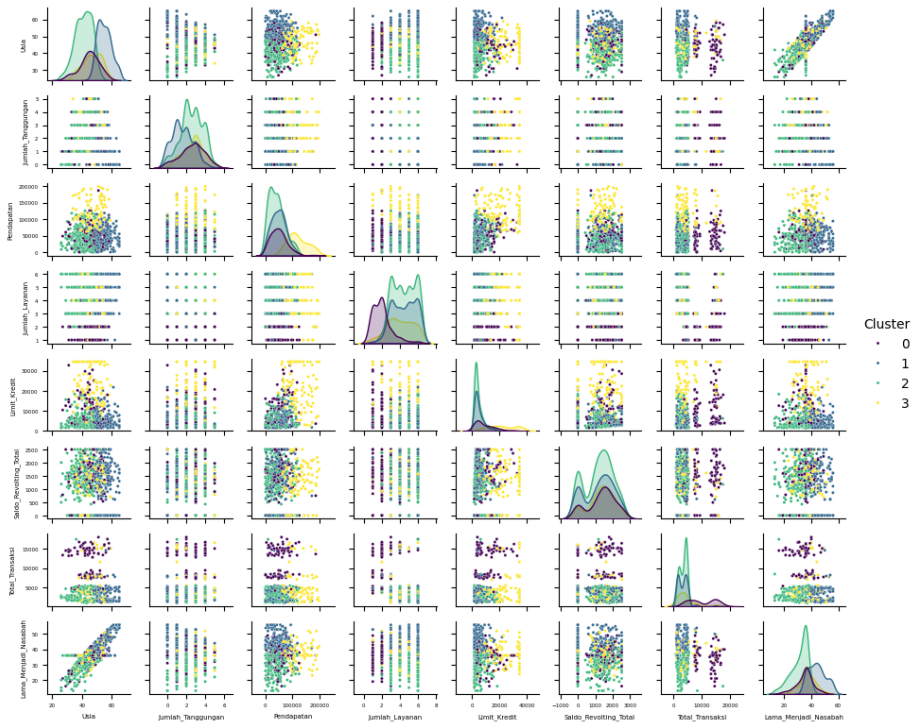
Gambar 4. 16 Klaster Berdasarkan Total Transaksi

Gambar 4.16 memberikan informasi mengenai setiap klaster berdasarkan total transaksi. Terlihat pada gambar bahwa pusat data di klaster nol berada pada total transaksi 5.000 sampai 14.000. Selajutnya pada pusat data di klaster satu berada pada total transaksi 1.500 sampai 4.500. Pusat data di klaster dua berada pada total trasaksi 2.500 sampai 4.500. Terakhir pada pusat data di klaster tiga berada pada total transaksi 2.000 sampai 5.000.



Gambar 4. 17 Klaster Bedasarkan Lama Menjadi Nasabah

Gambar 4.17 memberikan infomasi mengenai setiap klaster berdasarkan lama menjadi nasabah. Terlihat pada gambar bahwa pusat data di klaster nol berada pada lama menjadi nasbah selama 7 tahun. Selanjutnya pusat data di klaster satu berada pada lama menjadi nasabah selama 8 tahun. Pusat data di klaster dua beradaa pada lama menjadi nasabah selama 10 tahun. Terakhir pada pusat data di klaster tiga berada pada lama menjadi nasabah selama 4 tahun.



Gambar 4. 18 Pairplot Klaster

Gambar 4.18 adalah pairplot klaster yang menampilkan distribusi dan hubungan antara berbagai fitur dalam dataset, yang telah dikelompokkan menjadi beberapa klaster. Pairplot klaster di atas menunjukkan hubungan antara 8 fitur yang berbeda yaitu Usia, Jumlah_Tanggungan, Pendapatan, Jumlah_Layanan, Limit_Kredit, Saldo_Revolving_Total, Total_Transaksi, Lama_Menjadi_Nasabah. Dan klaster data telah dikelompokkan menjadi 4 klaster, masing-masing diwakili oleh warna yang berbeda:

a) Klaster 0: Ungu
b) Klaster 1: Hijau
c) Klaster 2: Biru
d) Klaster 3: Kuning

Diagonal dari pairplot menampilkan distribusi (density plot) dari masing-masing fitur untuk setiap klaster. Pada bagian off-diagonal

menampilkan scatter plots yang menunjukkan hubungan antara pasangan fitur yang berbeda.

Berikut adalah analisis detail setiap fitur:

a) Usia: Klaster 0 dan 2 cenderung memiliki distribusi yang berbeda dibandingkan klaster lainnya.
b) Jumlah_Tanggungan: Klaster-klaster memiliki jumlah tanggungan yang tersebar merata, dengan sedikit perbedaan antar klaster.
c) Pendapatan: Klaster 3 (kuning) memiliki rentang pendapatan yang lebih luas dibandingkan klaster lain.
d) Jumlah_Layanan: Klaster 1 (hijau) dan klaster 3 (kuning) tampaknya memiliki jumlah layanan yang lebih bervariasi.
e) Limit_Kredit: Klaster 0 (ungu) cenderung memiliki limit kredit yang lebih tinggi.
f) Saldo_Revolving_Total: Klaster 2 (biru) memiliki saldo revolving yang lebih tinggi.
g) Total_Transaksi: Klaster 3 (kuning) memiliki total transaksi yang lebih tinggi.
h) Lama_Menjadi_Nasabah: Klaster 1 (hijau) cenderung memiliki nasabah yang lebih lama menjadi nasabah.

Hubungan Antar Fitur:

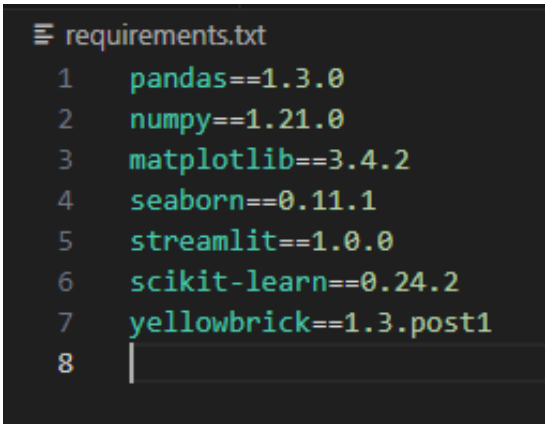
a) Usia vs. Limit_Kredit: Terdapat hubungan positif antara usia dan limit kredit, terutama pada klaster 0 (ungu).
b) Pendapatan vs. Limit_Kredit: Klaster 3 (kuning) menunjukkan hubungan yang lebih bervariasi antara pendapatan dan limit kredit.
c) Jumlah_Layanan vs. Total_Transaksi: Klaster 3 (kuning) memiliki variasi yang lebih besar dalam jumlah layanan dan total transaksi.
d) Saldo_Revolving_Total vs. Total_Transaksi: Klaster 2 (biru) menunjukkan variasi yang lebih besar dalam saldo revolving dan total transaksi.

6. Implementasi Sistem

a) Pengembang Aplikasi dengan Framework Streamlit

Aplikasi akan dikembangkan menggunakan *framework* Streamlit. Antarmuka pengguna yang sederhana dan interktif akan dibangun. Dalam konteks ini, Streamlit berfungsi sebagai web interaktif, karena mempercepat pembuatan aplikasi yang dapat digunakan untuk visualisasi data, analisis data, dan demo model machine learning. Sebelum pembuatan aplikasi streamlit, penulis mempersiapkan berbagai kebutuhan yang memdukung. Berikut adalah langkah-langkah untuk membuat aplikasi sederhana menggunakan streamlit:

1) Menyiapkan lingkuan pengembangan dengan instalasi python dan pastikan python sudah terinstall.
2) Instalisasi Streamlit dengan menggunakan pip install streamlit.
3) Mendefinisikan Dependencies dengan membuat file requirements.txt yang berisi semua dependensi proyek yang diperlukan.



Gambar 4. 19 Requiremenst.txt

Berdasarkan Gambar 4.19, terdapat beberapa requirements yang digunakan untuk mengelola dependensi dalam proyek ini.

Pandas adalah pustaka open-source yang menyediakan struktur data dan alat analisis data yang mudah digunakan dan cepat untuk bahasa pemrograman python, pandas digunakan untuk manipulasi data, analisis data dan struktur data yang efisien (DataFrame dan Series). Sementara numpy adalah pustaka dasar untuk komputasi ilmiah di python yang menyediakan array multidimensi, fungsi matematika, dan aljabar linier. Matplotlib adalah pustaka plotting 2D yang menghasilkan gambar berkualitas tinggi dlam berbagai format, matplotlib digunakan untuk membuat visualisasi data seperti grafik garis, grafik batang dan histogram. Seaborn adalah pustaka visualisai data berbasis matplotlib yang menyediakan antarmuka tingkat tinggi untuk menggambar grafik statistik yang atraktif dan informatif, yang mana seaborn digunakan untuk membuat grafik statistic dengan lebih mudah dan lebih menarik secara visual dibandingkan dengan matplotlib. Streamlit digunakan untuk membuat aplikasi web untuk visualisasi data, dashboard, dan demo model machine learning dengan cepat dan mudah. Scikit-learn adalah pustaka untuk machine learning di python yang menyediakan berbagai algoritma dan alat untuk klasifikasi, regresi dan clustering, yang mana digunakan untuk membangun dan menevaluasi model *mechine learning*. *Yellowbrick* digunakan untk membuat visualisasi diagnostic dan evaluasi untuk model machine learning.

Tahap selanjutnya adalah penulisan kode untuk pembuatan aplikasi dengan streamlit. Dalam proses pembuatan kode juga mengubah format label yang awalnya dari *encoding* yang tipe datanya integer saat pelatihan menggunakan model K-Means diubah menjadi *string*. Berikutnya kode streamlit segmentasi nasabah dapat dilihat pada kode dibawah ini.

```
import streamlit as st
import numpy as np
import pandas as pd
from sklearn.cluster import KMeans
from sklearn.metrics import silhouette_score
from sklearn.preprocessing import StandardScaler
from sklearn.decomposition import PCA
import matplotlib.pyplot as plt
import seaborn as sns
from mpl_toolkits.mplot3d import Axes3D

# Streamlit app
st.title('Segmentation Nasabah Menggunakan
Algoritma KMeans Clustering dan PCA')
st.write('This application performs customer
segmentation using KMeans clustering and PCA for
dimensionality reduction.')

# Load dataset
@st.cache_data
def load_data():
    data = pd.read_csv('data_nasabah.csv')
    df = pd.DataFrame(data)
    df.columns = [i.title() for i in df.columns]
           df.drop(['Jumlah_Melakukan_Transaksi',
'Rasio_Penggunaan_Rata-Rata'],            axis=1,
inplace=True)
    df.dropna(inplace=True)
    return df

df = load_data()

# Define KMeans functions
def kmeans_inertia(num_clusters, x_vals):
    inertia = []
    for num in num_clusters:
```

```
                    kms   =   KMeans(n_clusters=num,
random_state=42, n_init=10)
        kms.fit(x_vals)
        inertia.append(kms.inertia_)
    return inertia

def kmeans_sil(num_clusters, x_vals):
    sil_score = []
    for num in num_clusters:
        kms = KMeans(n_clusters=num,
random_state=42, n_init=10)
        kms.fit(x_vals)
        sil_score.append(silhouette_score(x_vals
, kms.labels_))
    return sil_score

# Sidebar
st.sidebar.title("Data Preprocessing")

# Display dataset
st.title("Data Nasabah")
st.write(df.head())

# Data Info
st.subheader("Data Information")
st.write(f"Shape of data: {df.shape}")
st.write(f"Number of rows: {df.shape[0]}")
st.write(f"Number of columns: {df.shape[1]}")

# Data Visualization
st.subheader("Data Visualization")
custom_palette = {'Laki-Laki': 'teal',
'Perempuan': 'gold'}
fig, ax = plt.subplots(figsize=(8, 3))
sns.countplot(data=df, y='Jenis_Kelamin',
alpha=0.7, palette=custom_palette,
edgecolor='black', ax=ax)
```

```
ax.set_yticklabels(['Laki-Laki', 'Perempuan'])
ax.set_title('Jenis Kelamin Distribution')
st.pyplot(fig)

# Data Preprocessing
numeric_columns = ['Usia', 'Jumlah_Tanggungan',
'Pendapatan', 'Jumlah_Layanan', 'Limit_Kredit',
'Saldo_Revolting_Total',
                   'Total_Transaksi',
'Lama_Menjadi_Nasabah']
X = df[numeric_columns]
X_scaled = StandardScaler().fit_transform(X)

# PCA
pca = PCA(n_components=3)
scores_pca = pca.fit_transform(X_scaled)

# KMeans Clustering
num_clusters = range(2, 10)
inertia = kmeans_inertia(num_clusters,
scores_pca)
sil_score = kmeans_sil(num_clusters, scores_pca)

# Violin Plots
st.subheader("Violin Plots")
fig, axes = plt.subplots(1, 5, figsize=(20, 6))
variables = ['Usia', 'Limit_Kredit',
'Saldo_Revolting_Total', 'Total_Transaksi',
'Lama_Menjadi_Nasabah']
for i, var in enumerate(variables):
    sns.violinplot(data=df, x=var,
y='Jenis_Kelamin', palette=custom_palette,
inner='quartile', ax=axes[i])
st.pyplot(fig)

# Histogram
st.subheader("Histograms")
```

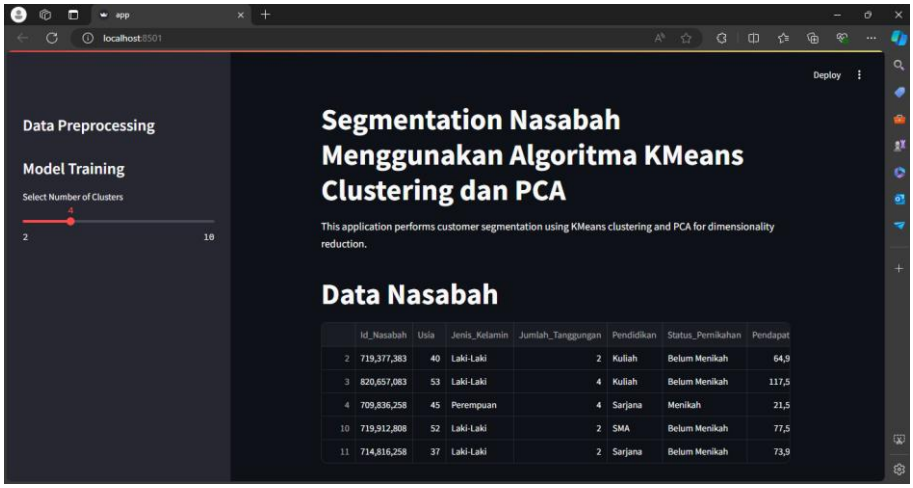
```
fig, axes = plt.subplots(3, 3, figsize=(12, 12))
axes = axes.flatten()
for i, var in enumerate(numeric_columns):
    sns.histplot(data=df, x=var, bins=20,
kde=True, color='Teal', alpha=0.5, linewidth=0,
ax=axes[i])
st.pyplot(fig)

# Correlation Matrix
st.subheader("Correlation Matrix")
corr_matrix = df.loc[:, numeric_columns].corr()
fig, ax = plt.subplots(figsize=(10, 6))
sns.heatmap(corr_matrix, annot=True,
annot_kws={'size':10}, fmt='.2f',
cmap='viridis', ax=ax)
st.pyplot(fig)

# Inertia and Silhouette Score
st.write('## Inertia and Silhouette Score
Model')
fig, ax = plt.subplots(1, 2, figsize=(12, 5))

sns.lineplot(x=num_clusters, y=inertia,
marker='o', ax=ax[0], color='teal',
markerfacecolor='black')
ax[0].set_title('Inertia vs. Number of
Clusters')
ax[0].set_xlabel('Number of Clusters')
ax[0].set_ylabel('Inertia')
ax[0].grid(True)

sns.lineplot(x=num_clusters, y=sil_score,
marker='o', ax=ax[1], color='teal',
markerfacecolor='black')
ax[1].set_title('Silhouette Score vs. Number of
Clusters')
ax[1].set_xlabel('Number of Clusters')
```

```
ax[1].set_ylabel('Silhouette Score')
ax[1].grid(True)

st.pyplot(fig)

# Fit final model with optimal clusters
optimal_clusters = 4
kmeans_pca = KMeans(n_clusters=optimal_clusters,
init='k-means++', n_init=10, random_state=42)
kmeans_pca.fit(scores_pca)

# Create DataFrame with cluster labels
df_pca_kmeans =
pd.concat([X.reset_index(drop=True),
pd.DataFrame(scores_pca)], axis=1)
df_pca_kmeans.columns.values[-3:] = ['PCA1',
'PCA2', 'PCA3']
df_pca_kmeans['Cluster'] = kmeans_pca.labels_

st.write('## Clustered Data')
st.write(df_pca_kmeans.head())

# Data Preprocessing
st.sidebar.title("Model Training")

# Standardize the data
X = df.loc[:, numeric_columns].copy()
X_scaled = StandardScaler().fit_transform(X)

# PCA
pca = PCA(n_components=3)
pca.fit(X_scaled)
scores_pca = pca.transform(X_scaled)

# KMeans Clustering
num_clusters = st.sidebar.slider("Select Number
of Clusters", 2, 10, 4)
```

```
kmeans_pca = KMeans(n_clusters=num_clusters,
init='k-means++', n_init=10, random_state=42)
kmeans_pca.fit(scores_pca)

# Adding clusters to the dataframe
df_pca_kmeans =
pd.concat([X.reset_index(drop=True),
pd.DataFrame(scores_pca)], axis=1)
df_pca_kmeans.columns.values[-3:] = ['PCA1',
'PCA2', 'PCA3']
df_pca_kmeans['Cluster'] = kmeans_pca.labels_

# Cluster Visualization
st.subheader("Cluster Visualization")
fig = plt.figure(figsize=(10, 8))
ax = fig.add_subplot(111, projection='3d')
scatter = ax.scatter(df_pca_kmeans['PCA1'],
df_pca_kmeans['PCA2'], df_pca_kmeans['PCA3'],
                     c=df_pca_kmeans['Cluster'],
cmap='viridis')
ax.set_xlabel('PCA Component 1')
ax.set_ylabel('PCA Component 2')
ax.set_zlabel('PCA Component 3')
ax.set_title('Clusters by PCA Components')
legend = plt.colorbar(scatter)
legend.set_label('Cluster')
st.pyplot(fig)

# Box Plots
st.subheader("Box Plots by Cluster")
fig, axes = plt.subplots(4, 2, figsize=(15, 15))
axes = axes.flatten()
for i, column in enumerate(numeric_columns):
    sns.boxplot(data=df_pca_kmeans, x='Cluster',
y=column, palette='viridis', ax=axes[i])
    axes[i].set_title(f'{column} by Cluster',
fontsize=8)
```

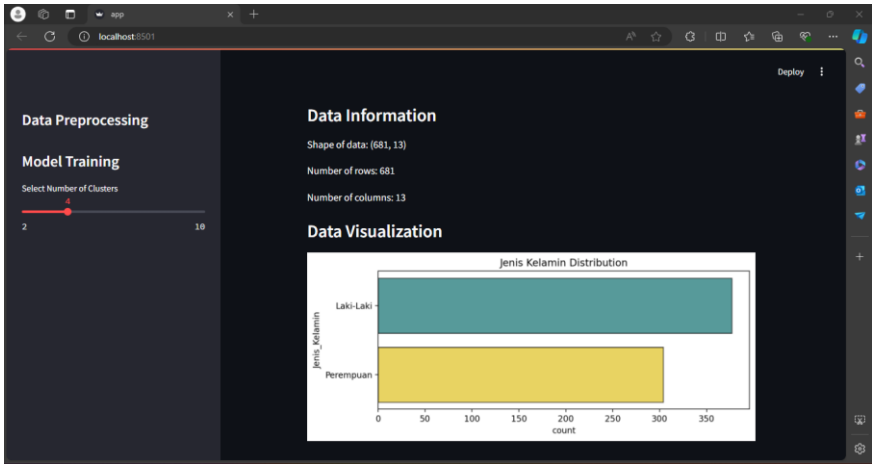
```
st.pyplot(fig)

# Pairplot
st.subheader("Pairplot by Cluster")
pairplot = sns.pairplot(df_pca_kmeans,
vars=numeric_columns, hue='Cluster',
palette='viridis', plot_kws={'s': 5}, height=1,
aspect=1.2)
for ax in pairplot.axes.flat:
    plt.setp(ax.get_xticklabels(), fontsize=6)
    plt.setp(ax.get_yticklabels(), fontsize=6)
st.pyplot(pairplot)
```

b) Hasil Web Streamlit



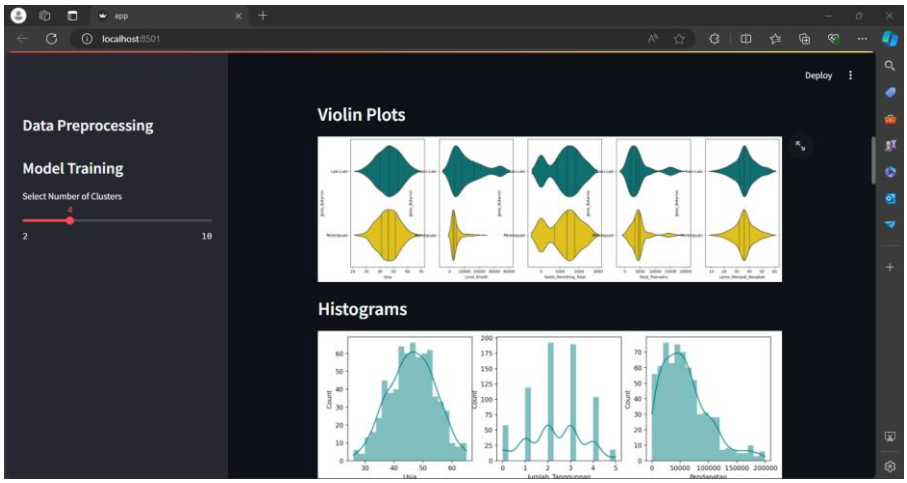
Gambar 4. 20 Dasboard Segmentasi Nasabah

Pada Gambar 4.20 menampilkan dashboard streamlit segmentasi nasabah. Gambar diatas terdapat tampilan informasi mengenai data nasabah bank, dan tampilan data *preprosesing* untuk mengatu jumlah klaster.



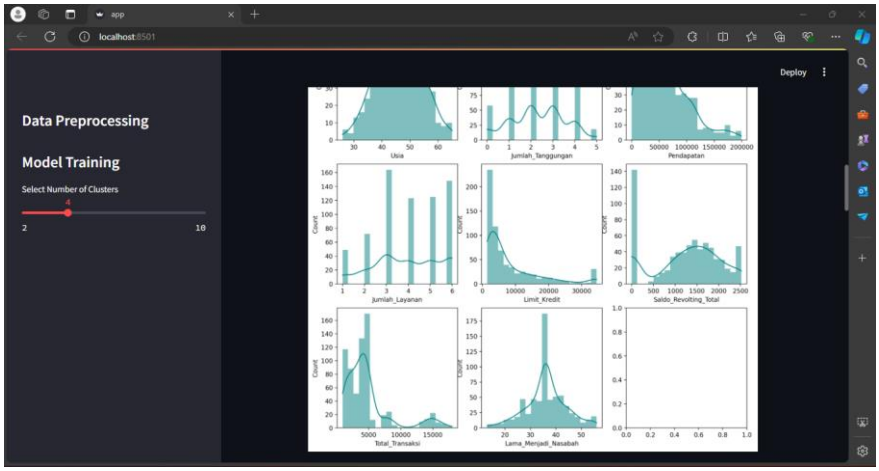
Gambar 4. 21 Data Informasi dan Data Visualisasi

Gambar 4.21 menampilkan dashboard jumlah informasi data dan diagram visualisasi data berdasarkan jenis kelamin, yang mana jenis kelamin laki-laki berwarna biru dan jenis kelamin perempuan berwarna kuning.



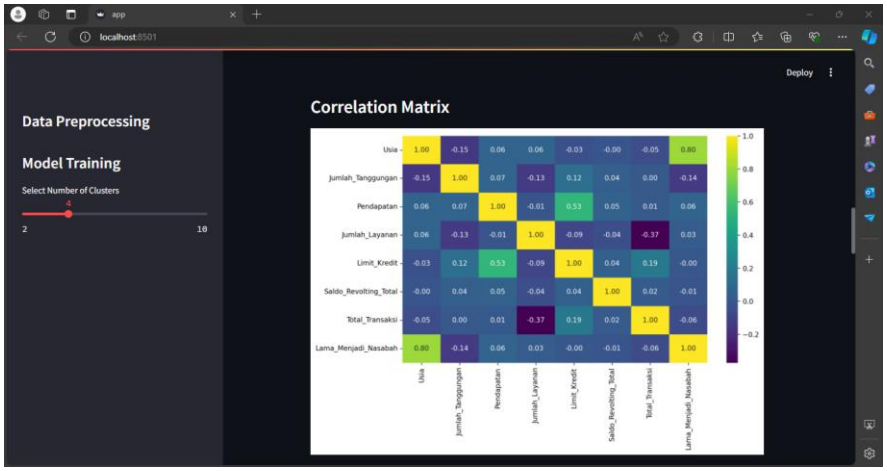
Gambar 4. 22 Tampilan *Violin Plots*

Gambar 4.22 menampilkan dashboard violin plots jenis kelamin dan juga histogram data nasabah.



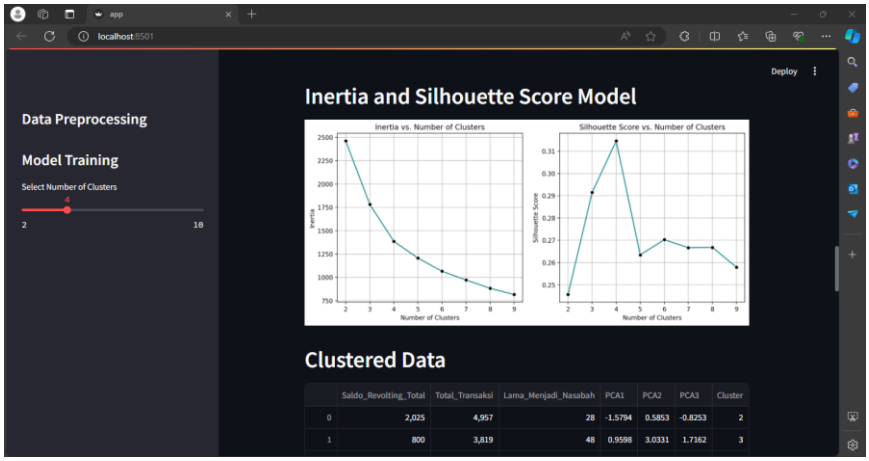
Gambar 4. 23 Tampilan Histogram

Gambar 4.23 menampilkan dashboard histogram usia, jumlah tanggungan, pendapatan, jumlah layanan, limit kredit, saldo revolting total, total transaksi dan lama memjadi nasabah.



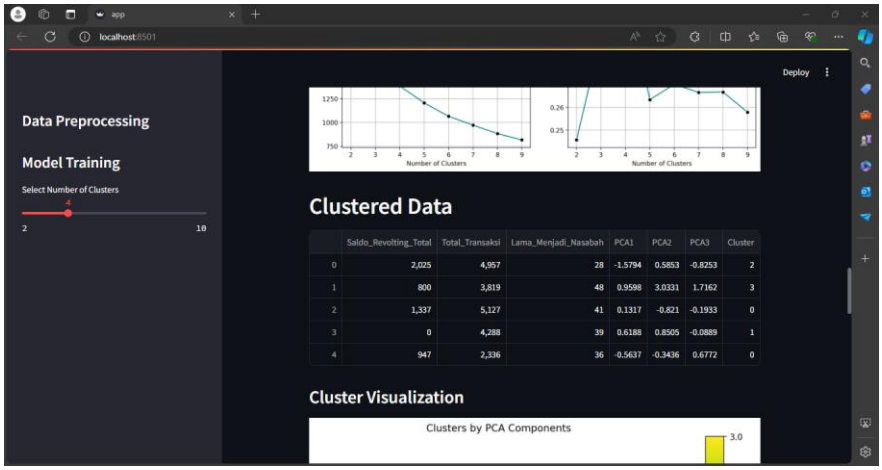
Gambar 4. 24 Tampilan *Correlation Matrix*

Gambar 4.24 menampilkan dashboard *correlation matrix* berdasarkan usia, jumlah tanggungan, pendapatan, jumlah layanan, limit kredit, saldo revolting total, total transaksi dan lama memjadi nasabah.



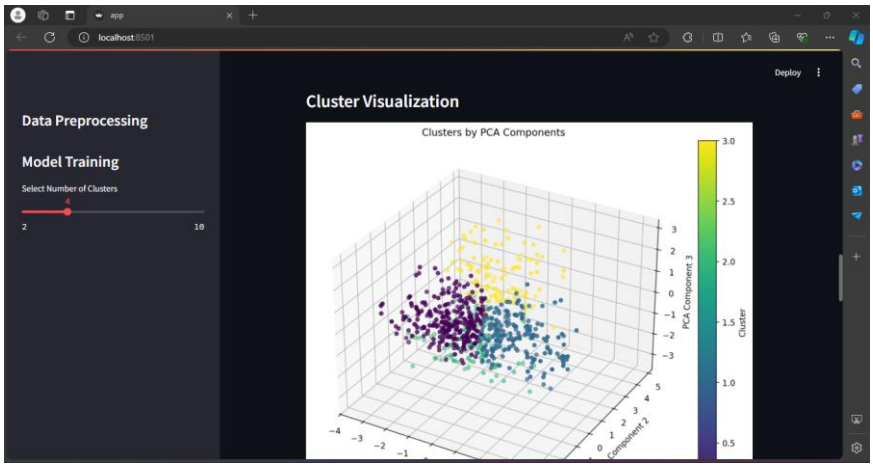
Gambar 4. 25 *Inertia and Silhouette Score Model*

Gambar 4.25 menampilkan dashboard *Inertia* dan *Silhouette Score* model dengan titik nilai pusatnya berada pada nilai 4. Gambar diatas menunjukan grafik *Inertia vs. Number of Clusters* untuk berbagai jumlah kluster. *Inertia* adalah ukuran seberapa baik titik data dalam kluster dikompaksi. Semakin rendah inertia, semakin baik klusteringnya. Grafik ini membantu pengguna untuk memilih jumlah kluster optimal dengan mencari "elbow point" di mana penurunan inertia mulai melambat. Grafik *Silhouette Score vs. Number of Clusters* menunjukkan nilai *silhouette score* untuk berbagai jumlah kluster. *Silhouette score* mengukur seberapa mirip objek dengan kluster mereka sendiri dibandingkan dengan kluster lain. Skor yang lebih tinggi menunjukkan kluster yang lebih baik. Grafik ini membantu pengguna untuk memilih jumlah kluster yang menghasilkan klustering dengan kualitas terbaik.



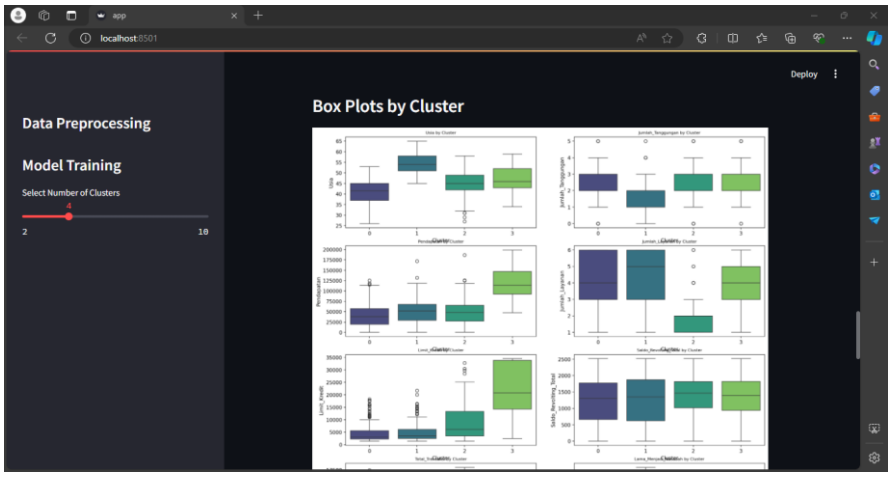
Gambar 4. 26 Tabel Cluster Data

Gambar 4.26 menampilkan table *cluster* data yang menunjukkan beberapa baris data yang telah dikelompokkan menggunakan model KMeans. Kolom terakhir ("Cluster") menunjukkan label kluster yang dihasilkan oleh model KMeans. Kolom PCA1, PCA2, dan PCA3 adalah komponen utama dari data yang telah direduksi dimensinya menggunakan PCA.



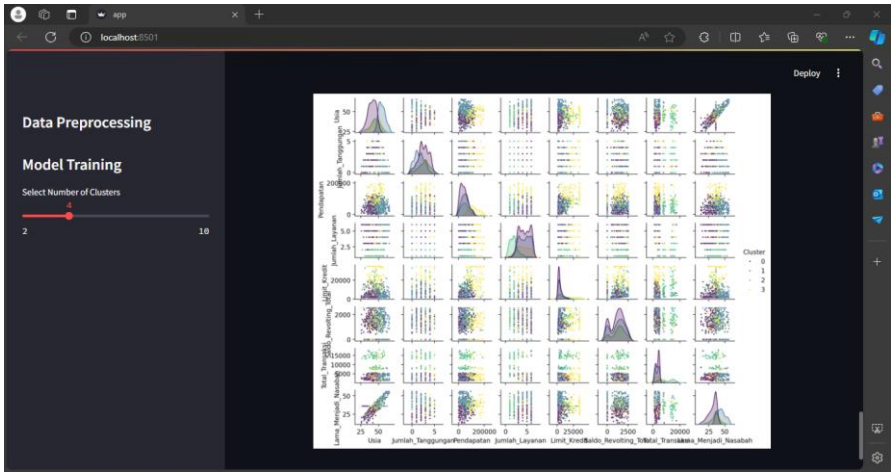
Gambar 4. 27 Tampilan *Visualisasi Cluster*

Gambar 4.27 menampilkan dashboard visualisasi klaster yang merupakan plot sebar 3D yang menggambarkan kelompok titik data berdasarkan tiga komponen analisis PCA.



Gambar 4. 28 Tampilan *Box Plots by Cluster*

Gambar 4.28 merupakan dashboard yang menapilkan box plots klaster berdasarkan usia, jumlah tanggungan, pendapatan, jumlah layanan, limit kredit, saldo revolting total, total transaksi dan lama memjadi nasabah.



Gambar 4. 29 Tampilan *Pairplot by Cluster*

Gambar 4.29 menampilkan dashboard *Pairplot by Cluster* , gambar diatas merupakan matriks plot sebar (disebut juga plot berpasangan). Matriks ini menunjukkan plot sebar berpasangan antara berbagai fitur kumpulan data. Setiap subplot membandingkan dua fitur berbeda, dan titik-titiknya diberi kode warna berdasarkan penetapan cluster. Sumbu x dan y dari plot sebar mewakili fitur berbeda dari kumpulan data, diberi label dengan nama seperti Usia, Total_Liabilitas, Pendapatan, Total_Layanan, Batas_Kredit, Revolving, Tingkat_Transaksi, Pengisian_Pelanggan. Kode warna pada plot sebar menunjukkan cluster berbeda yang ditetapkan oleh algoritma clustering. Sebelah kanan setiap plot menunjukkan nomor cluster (misalnya 0, 1, 2, 3).

7. Pengujian

Pengujian dilakukan untuk memastikan bahwa aplikasi berfungsi dengan baik dan memenuhi standar kualitas. Dalam penelitian ini, dua jenis pengujian digunakan, yaitu *white box testing* dan *black box testing* adalah memastikan setiap bagian dari kode berfungsi sesuai dengan spesifikasi yang telah ditetapkan.

a) Pegujian *White Box Testing*

White box testing adalah teknik pengujian perangkat lunak yang melibatkan pemeriksaan kode sumber dari program untuk mendeteksi adanya kesalahan. Tujuan dari pengujian ini adalah memastikan setiap bagian dari kode berfungsi sesuai dengan spesifikasi yang telah ditetapkan. Pengujian *white box testing* dapat dilihat pada Table 4.7.

Tabel 4. 7 White Box Testing

No	Code	Pengertian
1	`data = pd.read_csv('/content/drive/MyDrive/data nasabah/data_nasabah.csv')` `df = pd.DataFrame(data) df`	Memastikan dataset dimuat dengan benar

2	`df.columns = [i.title() for i in df.columns]`
	
	`df.columns`
		Memastikan nama kolom sesuai format

3	`df.info()`	Menampilkan info dataset untuk memahami struktur dan tipe data

4	`print("Shape of data: {}".format(df.shape))`
	
	`print("Number of rows: {}".format(df.shape[0]))`
	
	`print("Number of columns: {}".format(df.shape[1]))`
		Menampilkan ukuran dataset

5	`df = df.drop(['Jumlah_Melakukan_Transaksi', 'Rasio_Penggunaan_Rata-Rata'], axis=1)`	Menghapus kolom yang tidak diperlukan
6	`df.describe()`	Menampilkan deskripsi statistik dasar dari data
7	`df.nunique()`	Menghitung jumlah nilai unik di setiap kolom

8	`df.dropna(inplace=True)`
	
	`df.isnull().sum()`
		Menghapus baris dengan nilai yang hilang

9	`df_duplicated = df[df.duplicated()]`
	
	`df_duplicated.shape[0]`
		Menghitung jumlah baris duplikat

10	mean_median = pd.concat([df.mean(numeric_only=True), df.median(numeric_only=True)], axis=1) mean_median.columns = ['Mean', 'Median'] mean_median	Menghitung mean dan median dari setiap kolom numerik
11	sns.countplot(data=df, y='Jenis_Kelamin', alpha=0.7, palette=custom_palette, edgecolor='black')	Membuat plot distribusi jenis kelamin
12	sns.violinplot(data=df, x=var, y='Jenis_Kelamin', palette=custom_palette, inner='quartile')	Membuat violin plot untuk variabel terhadap jenis kelamin
13	df['GroupUsia'] = pd.cut(df['Usia'], bins=5) df['LabelGroupUsia'] = df['GroupUsia'].apply(format_interval) sns.violinplot(data=df, x=var, y='LabelGroupUsia', palette='viridis', inner='quartile')	Membuat interval usia dan visualisasinya
14	sns.histplot(data=df, x=var, bins=20, kde=True, color='Teal', alpha=0.5, linewidth=0)	Membuat histogram untuk variabel numerik
15	sns.heatmap(corr_matrix, annot=True, annot_kws={'size':10}, fmt='.2f', cmap='viridis')	Membuat heatmap untuk korelasi antar variabel numerik
16	X_scaled = StandardScaler().fit_transform(X)	Melakukan standarisasi data

17	def kmeans_inertia(num_clusters, x_vals):
	
	for num in num_clusters:
	
	kms = KMeans(n_clusters=num, random_state=42, n_init=10)
	
	kms.fit(x_vals)
	
	inertia.append(kms.inertia_)
	
	return inertia
		Menghitung inersia untuk berbagai jumlah kluster

18	def kmeans_sil(num_clusters, x_vals):
	
	for num in num_clusters:
	
	kms = KMeans(n_clusters=num, random_state=42, n_init=10)
	
	kms.fit(x_vals)
	sil_score.append(silhouette_score(x_vals, kms.labels_))
	
	return sil_score
		Menghitung silhouette score untuk berbagai jumlah kluster

19	pca = PCA() pca.fit(X_scaled) sns.lineplot(x=range(1, num_components + 1), y=pca.explained_variance_ratio_.cumsum(), marker='o', color='teal', markerfacecolor='black')	Melakukan PCA dan menentukan jumlah komponen
20	kmeans_pca = KMeans(n_clusters=4, init='k-means++', n_init=10, random_state=42) kmeans_pca.fit(scores_pca)	Melakukan KMeans pada hasil PCA

21	`scatter = ax.scatter(x_axis, y_axis, z_axis, c=df_pca_kmeans['Cluster'], cmap='viridis')`
	
	`ax.set_xlabel('PCA Component 1')`
	
	`ax.set_ylabel('PCA Component 2')`
	
	`ax.set_zlabel('PCA Component 3')`
	
	`ax.set_title('Clusters by PCA Components')`
		Memvisualisasikan kluster hasil KMeans pada komponen PCA

22	`sns.boxplot(data=df_pca_kmeans, x='Cluster', y=column, ax=ax[i], palette='viridis')`	Membuat boxplot untuk variabel terhadap kluster
23	`pairplot = sns.pairplot(df_pca_kmeans, vars=columns, hue='Cluster', palette='viridis', plot_kws={'s': 5}, height=1, aspect=1.2)`	Membuat pairplot untuk variabel terhadap kluster

1) Penjelasan :
   a) Note 1: Memuat dataset dari Google Drive dan memastikan data telah dimuat dengan benar.
   b) Note 2: Memastikan semua nama kolom dalam format title case untuk konsistensi.
   c) Note 3: Memastikan tipe data dan jumlah kolom dan baris sesuai dengan yang diharapkan.
   d) Note 4: Mengecek ukuran dataset untuk memastikan jumlah baris dan kolom sesuai.
   e) Note 5: Menghapus kolom yang tidak diperlukan untuk analisis lebih lanjut.
   f) Note 6: Memeriksa statistik dasar dari data untuk mendapatkan wawasan awal.

g) Note 7: Menghitung jumlah nilai unik di setiap kolom untuk memahami distribusi data.
h) Note 8: Membersihkan data dari nilai yang hilang dan memastikan tidak ada missing values yang tersisa.
i) Note 9: Memastikan tidak ada baris duplikat dalam data.
j) Note 10: Menghitung rata-rata dan median untuk memahami distribusi data numerik.
k) Note 11: Memvisualisasikan distribusi jenis kelamin untuk memahami komposisi data.
l) Note 12: Memvisualisasikan distribusi variabel numerik terhadap jenis kelamin.
m) Note 13: Membuat interval usia dan memvisualisasikan distribusi variabel numerik terhadap interval usia.
n) Note 14: Membuat histogram untuk memvisualisasikan distribusi variabel numerik.
o) Note 15: Membuat heatmap untuk memahami korelasi antar variabel numerik.
p) Note 16: Melakukan standarisasi data untuk menghilangkan pengaruh skala variabel.
q) Note 17: Menghitung inersia untuk berbagai jumlah kluster untuk menentukan jumlah kluster optimal.
r) Note 18: Menghitung silhouette score untuk berbagai jumlah kluster untuk menentukan jumlah kluster optimal.
s) Note 19: Melakukan PCA untuk mengurangi dimensi data dan menentukan jumlah komponen yang optimal.
t) Note 20: Melakukan clustering KMeans pada data yang telah dikurangi dimensinya dengan PCA.
u) Note 21: Memvisualisasikan hasil clustering pada komponen PCA untuk memahami distribusi kluster.
v) Note 22: Memvisualisasikan distribusi variabel terhadap kluster untuk memahami karakteristik kluster.

w) Note 23: Membuat pairplot untuk memvisualisasikan hubungan antar variabel terhadap kluster.

2) Basis Path untuk Fungsi Prediksi

Basis path (jalur dasar) untuk pengujian white box dari fungsi prediksi yang diimplementasikan di Visual Studio Code. Basis path testing adalah teknik yang digunakan untuk memastikan bahwa semua jalur eksekusi yang mungkin dalam kode telah diuji setidaknya sekali.

a) Cyclomatic Complexity

Cyclomatic Complexity (CC) dapat dihitung menggunakan formula:

CC = $E - N + 2P$

Dimana:

E = jumlah edge.

N = jumlah node.

P = komponen terhubung (biasanya 1 untuk graf yang terhubung tunggal).

Dari langkah-langkah di atas, kita dapat melihat bahwa ada 23 node (N = 23).

b) Edges (E)

Setiap langkah terhubung satu sama lain dalam urutan. Jadi, jumlah edge adalah $N - 1$ dalam kasus ini, yaitu 22 (E = 22).

c) Menghitung Cyclomatic Complexity

Dengan $P = 1$:

CC = 22 – 23 + 2

CC = 1

Namun, karena ada beberapa fungsi terpisah seperti kmeans_inertia dan kmeans_sil, kita perlu mempertimbangkan jalur di dalam fungsi tersebut juga.

d) Rincian Fungsi:

i. kmeans_inertia: Ada satu loop untuk setiap jumlah kluster.
ii. kmeans_sil: Ada satu loop untuk setiap jumlah kluster.

Jika kita asumsikan setiap fungsi menambah satu jalur independen, maka kita harus menambahkan ini ke perhitungan CC.

e) Tambahan Jalur

Setiap fungsi menambah satu jalur independen. Jadi kita tambahkan 2 jalur lagi.

CC total = 1 + 2 = 3

f) Identifikasi Jalur Independen

i. Path 1: Muat Data -> Format Nama Kolom -> Tampilkan Informasi Dataset -> Tampilkan Ukuran Dataset -> Hapus Kolom yang Tidak Diperlukan -> Tampilkan Deskripsi Statistik -> Hitung Nilai Unik -> Hapus Baris dengan Nilai Hilang -> Hitung Jumlah Baris Duplikat -> Hitung Mean dan Median -> Plot Distribusi Jenis Kelamin -> Violin Plot Jenis Kelamin -> Buat Interval Usia dan Visualisasinya -> Histogram Variabel Numerik -> Heatmap Korelasi -> Standarisasi Data -> Hitung Inersia untuk Berbagai Jumlah Kluster -> Hitung Silhouette Score untuk Berbagai Jumlah Kluster -> PCA dan Tentukan Jumlah Komponen -> KMeans pada Hasil PCA -> Visualisasikan Kluster pada Komponen PCA -> Boxplot untuk Variabel terhadap Kluster -> Pairplot untuk Variabel terhadap Kluster

ii. Path 2: kmeans_inertia -> Loop (Setiap jumlah kluster)

iii. Path 3: kmeans_sil -> Loop (Setiap jumlah kluster)

Kompleksitas siklomatik dari kode ini adalah 3, yang berarti ada 3 jalur independen yang perlu diuji untuk memastikan setiap jalur dalam kode dieksekusi setidaknya sekali. Jalur-jalur ini mencakup alur utama dari proses analisis data, serta dua fungsi terpisah untuk menghitung *inersia* dan *silhouette score*.

b) Pengujian *Black Box Testing*

Pengujian *black box testing* dan hasil pengujian *black box testing* dapat dilihat pada Tabel 9 dan Tabel 10.

Tabel 4. 8 Pengujian *Black Box Testing*

Nama Pengujian	Test Case	Hasil yang diharapkan	Hasil yang didapatkan	Keterangan	
				Diterima	Ditolak
Streamlit	Menjalankan terminal untuk menjalankan web streamlit	Website dapat menampilkan dassborad segmentasi nasabah	Website akan menampilkan dassborad segmentasi nasabah		
Sidebar Model Training Cluster	Memilih jumlah kluster dengan slider	Website dapat menjalankan K-means dengan jumlah cluster baru dan	Website akan menjalankan K-Means dengan jumlah cluster baru dan		

		menampilkan hasil	menampilkan hasil		
Tampilan data nasabah	Menampilkan informasi tentang data yang dimuat	Website dapat menampilkan informasi data yang dimuat	Website menampilkan informasi yang benar tentang jumlah baris dan kolom, serta informasi lainnya tentang data		
Tampilan grafik model inertia dan silhouette score	Menampilkan informasi mengenai model inertia dan silhouette score untuk pemodelan klaster	Website dapat menampilkan grafik model	Website menampilkan grafik model inertia dan silhouette score muncul dengan benar sesuai dengan jumlah kluster yang dipilih		
Tampilan Cluster by PCA	Menampilkan visualisas	Website dapat menampilka	Website menampilkan grafik 3D		

Komponen	i cluster dari komponen PCA	n grafik 3D cluster by PCA Component	menunjukkan distribusi data berdasarkan komponen PCA dan kluster		
Tampilan Box Plots	Visualisasi box plot berdasarkan kluster	Website dapat menampilkan grafik box plot dengan benar	Website menampilkan grafik box plot ditampilkan dengan benar sesuai dengan kluster		
Tampilan Pairplot	Visualisasi pairplot berdasarkan kluster	Website dapat menampilkan grafik pairplot yang sesuai	Website menampilkan pairplot ditampilkan dengan benar sesuai dengan kluster		

Tabel 4. 9 Hasil Pengujian *black box testing*

Nama pengujian	Text Case	Hasil Yang Diharapkan	Hasil yang Didapatkan	Keterangan					
				Diterima			Ditolak		
				1	2	3	1	2	3

Streamlit	Menjalankan terminal untuk menjalankan web streamlit	Website dapat menampilkan dassborad segmentasi nasabah	Website akan menampilkan dassborad segmentasi nasabah	√	√	√			
Sidebar Model Training Cluster	Memilih jumlah kluster dengan slider	Memilih jumlah kluster dengan slider	Website akan menjalankan K-Means dengan jumlah cluster baru dan menampilkan hasil	√	√	√			
Tampilan data nasabah	Menampilkan informasi tentang data yang dimuat	Website dapat menampilkan informasi data yang dimuat	Website menampilkan informasi yang benar tentang jumlah baris dan kolom, serta informasi lainnya	√	√	√			

			tentang data						
Tampilan grafik model inertia dan silhouette score	Menampilkan informasi mengenai model inertia dan silhouette score	Website dapat menampilkan grafik model	Website menampilkan grafik model inertia dan silhouette score muncul dengan benar sesuai dengan jumlah kluster yang dipilih	√	√	√			
Tampilan Cluster by PCA Komponen	Menampilkan visualisasi cluster dari komponen PCA	Website dapat menampilkan grafik 3D cluster by PCA Component	Website menampilkan grafik 3D menunjukkan distribusi data berdasarkan komponen PCA dan kluster	√	√	√			

Tampilan Box Plots	Visualisasi box plot berdasarkan kluster	Website dapat menampilkan grafik box plot dengan benar	Website menampilkan grafik box plot ditampilkan dengan benar sesuai dengan kluster	√	√	√			
Tampilan Pairplot	Visualisasi pairplot berdasarkan kluster	Website dapat menampilkan grafik pairplot yang sesuai	Website menampilkan pairplot ditampilkan dengan benar sesuai dengan kluster	√	√	√			

Kesimpilan hasil *black box testing* :

Bedasarkan *black box testing* dari 7 pengujian pada web streanlit yang didapat dari 3 responden, berikut ini hasil *black box* testing :

1) Pengujian Pertama

Tercapai : $\frac{7}{7}\ x\ 100\% = 100\%$

Gagal : $\frac{0}{7}\ x\ 100\% = 0\%$

2) Pengujian Kedua

Tercapai : $\frac{7}{7}\ x\ 100\%$ = 100%

Gagal : $\frac{0}{7}\ x\ 100\%$ = 0%

3) Pengujian Ketiga

Tercapai : $\frac{7}{7}\ x\ 100\%$ = 100%

Gagal : $\frac{0}{7}\ x\ 100\%$ = 0%

Jumlah presentase rata-rata yang tercapai = $\frac{300\%}{3}$ = 100%

Jumlah presentase rata-rata gagal = $\frac{0\%}{3}$ = 0%

Berdasarkan hasil perhitungan tersebut, dari tujuh pengujian yang dilakukan oleh 3 responden, semua pengujian *black box* berhasil mencapai tingkat keberhasilan 100%, sementara yang gagal mencapai tingkat kegagalan 0%. Dari hasil ini, dapat disimpulkan bahwa aplikasi berfungsi sesuai dengan fungsionalitas yang diharapkan.

c) *User Acceptance Testing* (UAT)

User acceptance testing dilakukan untuk memverifikasi bahwa *website* streamlit segmentasi nasabah memenuhi persyaratan dan ekspektasi pengguna. Proses pengujian ini memberikan kesempatan bagi pengguna untuk memberikan umpan balik dan memastikan bahwa aplikasi siap untuk digunakan dalam lingkungan produksi atau operasional. Pengujian *user acceptance* ini melibatkan 5 responden yang diminta untuk mengisi kuesioner dengan skala likert dari 1 hingga 5. Hasil kelayakan aplikasi dapat ditemukan dalam Tabel 4.10.

Tabel 4. 10 *User Acceptance Testing* (UAT)

No	Pertanyaan	Skor				
		Tidak Setuju	Kurang Setuju	Cukup Setuju	Setuju	Sangat Setuju
Aspek Kegunaan						
1	Apakah website segmentasi nasabah dapat bermanfaat bagi pengguna?					
2	Apakah website segmentasi nasabah memberikan informasi tentang jumlah klaster nasabah?					
Aspek Kemudahan Pengguna						
3	Apakah website segmentasi nasabah mudah dipahami?					

4	Apakah website segmentasi nasabah sesuai yang diharapkan?					
5	Apakah website segmentasi nasabah berisi informasi yang dibutuhkan?					
6	Apakah website segmentasi nasabah sesuai dengan keperluan anda?					
Aspek *User Interface* (UI)						
7	Apakah website segmentasi nasabah memiliki tampilan					

	yang mudah dipahami?					
8	Apakah website sebmentasi nasabah memiliki tampilan yang menarik?					
9	Apakah website segmentasi nasabah memiliki tema yang menarik?					
10	Apakah website segmentasi nasabah perlu di kembangkan lagi?					

Keterangan :

1 = Tidak Setuju

2 = Kurang Setuju

3 = Cukup Setuju

4 = Setuju

5 = Sangat Setuju

Berikut merupakan hasil dari kuesioner *user acceptance testing* yang telah disebarkan kepada 3 responden. Hasil *user acceptance testing* dapat dilihat pada Tabel 4.11.

Tabel 4. 11 Hasil *user acceptance testing* (UAT)

Pernyataan	Hasil pengujian		
	Responden 1	Responden 2	Responden 3
1	5	4	4
2	4	4	4
3	4	4	4
4	4	4	4
5	4	4	4
6	4	4	5
7	4	4	4
8	4	4	4
9	4	4	4
10	5	5	5
Jumlah Skor	42	41	42
Presentase	84%	82%	84%
Total	250%		

Dari hasil presentase yang diberikan oleh 3 responden untuk setiap pertanyaan yang mencakup aspek kegunaan, kemudahan penggunaan dan antar muka pengguna (UI), kemudian mencari nilai rata-ratanya. Tujuannya adalah untuk mengetahui tingkat penerimaan website streamlit yang dikembangan oleh responden. Nilai rata-rata ini dapat dihitung menggunakan persamaan berikut :

$$\text{Presentase rata-rata} = \frac{Jumlah\ total\ presentase}{Jumlah\ responden}$$

$$\text{Presentase rata-rata} = \frac{250\%}{3} = 83{,}3\%$$

Dengan daftar kategori presentase sebagai berikut :

0% - 20% = Sangat Kurang

21% - 40% = Kurang

41% - 60% = Cukup Baik

61% - 80% = Baik

81% - 100% = Sangat Baik

Dari perhitungan tersebut, diperoleh presentase rata-rata dari ketiga aspek sebesar 83,3%. Dengan demikian, dapat disimpulkan bahwa pengujuan UAT pada web streamlit ini mendapat kategori “Sangat Baik”.

B. Pembahasan

Dari hasil penelitian menggunakan metode Crisp-DM, website streamlit segmentasi nasabah bank dibuat melalui 7 tahapan, yaitu pemahaman bisnis, pemahaman data, persiapan data, modeling, evaluasi, implementasi dan pengujian.

1. Business Understanding

Pada tahap pertama, yaitu tahap pemahaman bisnis dengan proses bisnis yang perlu dipahami dalam penelitian yang berkaitan dengan segmentasi nasabah. Adapun fokus pada pemahaman yang lebih terperinci mengenai karakter nasabah bukan lagi barang yang bagus untuk dimiliki, namun keharusan yang strategis dan kompetitif bagi penyedia perbankan. Pada penelitian ini mengambil data nasabah bank yang memiliki jumlah nasabah 1050 dengan kurang lebih 10 ribu transaksi. Kemudian data diolah dengan diprosesing menggunakan algoritma K-Means clustering dan dianalisis menggunakan PCA untuk

mengidentifikasi kelompok-kelompok nasabah yang memiliki kesamaan karakteristik dan memperkuat pengelompokan nasabah, mengidentifikasi kebutuhan dan preferensi spesifik dari setiap segmen.

2. Data Understanding

Pada tahap kedua, yaitu tahap pemahaman data. Data yang digunakan dalam penelitian ini adalah data nasabah bank dengan jumlah nasabah 1050 nasabah yang diambil dari *Kaggle*. Kemudian data dikumpulkan lalu disimpan dalam bentuk excel atau csv. Selanjutnya data dievaluasi dan dieksplorasi untuk memahami dan mengganti nama *header* agar lebih mudah diakses. Kemudian data diolah untuk menghilangkan atribut yang tidak di pakai.

3. Data Preparation

Pada tahap persiapan data dilakukan pemilihan data yang telah dikumpulkan pada tahap pemahaman data. Selanjutnya peneliti menggunakan *goole colab* untuk mempermudah pemprosesan data untuk mempermudah perhitungan datanya. Kemudian peneliti memproses data mentah menjadi bentuk yang sesuai untuk modeling. Dengan melakukan pembersihan data seperti mengatasi missing values, outliers, transformasi data dengan menormalisasi data, menghapus nilai yang tidak diperlukan, menghitung nilai rata-rata mean median. Kemudian data di visualisasikan berdasarkan jenis kelamin dan didistribusikan berdasarkan usia, pendapatan, jumlah tanggungan, jumlah layanan, limit kredit, saldo revilting total, total transaksi dan lama menjadi nasbah.

4. Modelling

Pada tahap modelling, peneliti menggunakan algoritma K-Means Clustering dan PCA untuk menentukan jumlah klaster. Pada tahap ini dilakukan standarisasi data dengan StandardScaler(), kemudian mengevaluasi performa model K-Means pada berbagai klaster. Pengevalusaian model menggunakan fungsi kmeans_inertia untuk mengevaluasi kinerja model KMeans berdasarkan nilai inertia. Nilai

inertia yang lebih rendah menunjukkan klaster yang lebih kompak. Fungsi kmeans_sil digunakan untuk mengevaluasi kinerja model KMeans berdasarkan silhouette score. Nilai *silhouette score* yang lebih tinggi menunjukkan klaster yang lebih baik. Selanjutnya, digunakan metode PCA untuk menentukan variabel yang penting dalam data. Berikutnya dilakukan pengujian modelan dengan dua model yang digunakan untuk menentukan pusat cluster yang optimal dalam K-Means, yaitu *Inertia* dan *Silhouette Score*. Pada penerapan pemodelan dihasilkan jumlah klaster yang diinginkan dengan memunculkan nilai k = 4. Selanjutnya penerapan algoritma K-Means pada data yang telah direduksi dimensinya menggunakan PCA dan menggabungkan hasilnya kedalan dataframe untuk di analisis lebih lanjut.

Pada tahap ini K-Means dan PCA digunakan untuk mengelompokkan data menjadi beberapa kalster berdasarkan kesamaan fitur dan mereduksi dimensi data dengan mempertahankan sebanyak mungkin variabel data. Dengan tahapannya yaitu mulai dari persiapan data, normalisasi, analisis komponen utama, hingga klasterisasi, diikuti dengan evaluasi hasil menggunakan metrik seperti inertia atau silhouette score dengan contoh data nasabah bank yang diambil dari *kaggle* sebelumnya.

5. Evaluation

Pada tahap evaluasi, peneliti melakukan pemvisualisasian kebentuk scatter plot agar dapat dengan mudah melihat pola atau struktur yang muncul dari pengelompokan tersebut. Pada evaluasi ini memunculkan grafik klaster berdasarkan usia, pendapatan, jumlah tanggungan, jumlah layanan, limit kredit, saldo revilting total, total transaksi dan lama menjadi nasbah.

6. Implementasi

Pada tahap implementasi ini, peneliti peneliti menggunakan web streamlit. Dalam pembuatan streamlit menggunakan visual code. Dalam konteks ini, Streamlit berfungsi sebagai web interaktif, karena

mempercepat pembuatan aplikasi yang dapat digunakan untuk visualisasi data, analisis data, dan demo model machine learning.

7. **Pengujian**

   Tahap terakhir yaitu pengujian untuk memastikan kualititas website yang berfungsi dengan baik. Penulis melakukan 3 jenis pengujian yaitu Black Box Testing, White Box Testing dan User Acceptance Testing (UAT). Dalam pengujian White Box Testing memberikan hasil memuaskan dengan jumlah path 3. Pengujian Black Box Testing, berhasil mencapai presentase 100% keberhasilan, sedangkan kegagalan mendapat presentase 0% dari tiga responden dan 7 pengujian. Pengujian User Acceptance Testing (UAT) juga berhasil dngan 83,3% dari 3 responden dengan 10 pertanyaan.

BAB V
PENUTUP

A. Kesimpulan

Berdasarkan penelitian yang telah dilakukan penulisan, maka dapat ditarik beberapa kesimpulan sebagai berikut:

1. Hasil penelitian menunjukkan bahwa telah dilakukan segmentasi nasabah dengan menggunakan algoritma K-Means Clustering dengan PCA sebagai pengurangi jumlah dimensi (fitur) dalam dataset, yang dapat membantu menyederhanakan data tanpa kehilangan informasi penting. Penentuan jumlah klaster terbaik menggunakan *Inersia* dan *Silhouette Score* dengan nilai *Inersia* berpusat diantara nilai 3 dan 4, sedangkan *Silhouette Score* berpusat pada nilai 4, maka dapat ditentukan jumlah klaster yang sama yaitu klaster k = 4. Dengan klaster berdasarkan karateristik usia, pendapatan, jumlah tanggungan, jumlah layanan, limit kredit, saldo revolving total, total transaksi, lama menjadi nasabah.
2. Pada penelitian ini menggunakan 3 tahap pengujian pada pengenbangan web streamlit, yaitu *white box testing, black box testing, user acceptance test* (UAT). Untuk pengujian *white box testing* memberikan hasil memuaskan dengan jumlah path 3. Pengujian *black box testing*, berhasil mencapai presentase 100% keberhasilan, sedangkan kegagalan mendapat presentase 0% dari tiga responden dan 7 pengujian. Pengujian *User Acceptance Testing* (UAT) juga berhasil dngan 83,3% dari 3 responden dengan 10 pertanyaan.
3. Penelitian ini menunjukkan bahwa penggunaan K-Means Clustering, bersama dengan PCA dan visualisasi dinamis menggunakan Streamlit, adalah pendekatan yang efektif untuk segmentasi nasabah bank. Hasil analisis memberikan dasar yang solid untuk pengembangan strategi pemasaran yang lebih terarah dan layanan yang lebih personal bagi nasabah. Implementasi dan pengembangan lebih lanjut dari sistem ini diharapkan

dapat memberikan manfaat yang signifikan bagi industri perbankan dalam memahami dan melayani nasabah mereka dengan lebih baik.

B. Saran

1. Hasil dari proses analisis k-means clustering ini dapat digunakan dalam mengambil keputusan lebih lanjut dalam penetapan pelanggan potensial kedepannya.
2. Penelitian ini juga dapat dikembangkan dengan menggunakan tools yang lain nya selain dengan menggunakan *Google Colab* untuk membantu mencari hasil yang yang lebih optimal lagi.
3. Fitur-fitur interaktif tambahan dalam Streamlit dapat lebih dikembangkan untuk memberikan analisis yang lebih mendalam. Misalnya, tambahkan opsi untuk filter data berdasarkan kriteria tertentu, atau fitur drill-down yang memungkinkan pengguna untuk melihat rincian kluster secara lebih detail.

DAFTAR PUSTAKA

[1] M. A. Erista Lutfi Ervina, "STRATEGI SEGMENTASI PASAR DALAM MENINGKATKAN JUMLAH NASABAH PADA PRODUK TABUNGANKU DI BANK MUAMALAT KANTOR CABANG PEMBANTU MADIUN," *Journal of Islamic Banking and Finance,* vol. 1, 2022.

[2] S. L. A. Agus Wahyu Irawan, "Analisis Metode SMART Dalam Strategi Segmentasi Pasar (Studi Produk Tabungan Simitra Mikro Di Bank Mitra Syariah Kantor Cabang," *Jurnal Ilmiah Ekonomi Syariah,* vol. 5, 2022.

[3] G. S. W. Moch Rizky Wijaya, "Customer Segmentation berdasarkan Usia, Jumlah Kredit dan LamaKredit Nasabah di Bank XYZ menggunakan Model K-Means Clustering," *Prosiding Seminar Nasional UMC,* vol. 1, 2021.

[4] A. F. Nurin Fadilah Adani, "Implementasi Data Mining Untuk Pengelompokan Data Penjualan Berdasarkan Pola Pembelian Menggunakan Algoritma K-Means Cluster Pada Toko Syihan," *Jurnal CyberTech,* vol. 10, 2020.

[5] S. F. F. A. S. Satria Ardi Perdana, "Analisis Pelanggan Menggunakan K-Means Clustering Studi Kasus Aplikasi Alfagift," *Sebatik,* vol. 26, 2022.

[6] F. Dinda Giantika Utami, "Penerapan Algoritma K-Means untuk Pengelompokan Produksi Telur Ayam ras Petelur di Indonesia," *E-PROSIDING SISTEM INFORMASI,* vol. 3, 2022.

[7] Widyawati, "Penerapan Agglomerative Hierarchical Clustering Untuk Segmentasi Pelanggan," *Jurnal Ilmiah Sinus(JIS),* vol. 18, 2020.

[8] T. S. S. R. T. Nita Mirantika, "Implementasi Algoritma K-Medoids Clustering Untuk Menentukan Segmentasi Pelanggan," *Jurnal Nuansa Informatika,* vol. 17, 2023.

[9] R. N. N. L. A. J. H. Ira Ariati, "SEGMENTASI PELANGGAN MENGGUNAKAN K-MEAS CLUSTERING STUDI KASUS PELANGGAN UHT MILK GREENFIELD," *Cerdika: Jurnal Ilmiah Indonesia,* vol. 3, 2023.

[10] M. A. Y. D. Gustientiedina, "Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan Pada RSUD Pekanbaru," *Jurnal Nasional teknologi dan Sistem Informasi,* vol. 5, 2019.

[11] A. T. F. D. A. Nita Mirantikka, "PENERAPAN ALGORITMA K-MEANS CLUSTERING UNTUK PENGELOMPOKAN PENYEBARAN COVID-19 DI PROVINSI JAWA BARAT," *JURNAL NUANSA INFORMATIKA,* vol. 15, 2021.

[12] E. S. B. Qorik Indah Mawarni, "Implementasi Algoritma K-Means Clustering Dalam Penilaian Kedisiplinan Siswa," *Jurnal Sistem Komputer dan Informatika(JSON),* vol. 3, 2022.

[13] I. S. A. E. P. Beta Estria Adiana, "Analisis Segmentasi Pelanggan Menggunakan Kombinasi RFM Model dan Teknik Clustering," *Jurnak Teknik Elektro dan Teknologi Informasi ,* vol. 2, 2018.

[14] H. Nanang Khoirul Ahmadi, "Analisis Segmentasi Terhadap Keputusan Pembelian Produk Eiger Di Bandar Lampung," *Jurnal Manajemen Magister,* vol. 3, 2017.

[15] D. A. H. J. E. Hendra Di Kesuma, "Implementasi Data Mining Prediksi Mahasiswa Baru Menggunakan Algoritma Regresi Linear Berganda," *Jurnal Ilmiah Binary STMIK Bina Nusantara Jaya,* vol. 4, 2022.

[16] F. Marisa, "Educational Data Mining (Konsep Dan Penerapan)," *Jurnal Teknologi Informasi,* vol. 4, 2019.

[17] B. I. N. A. Sekar Setyaningtyas, "TINJAUAN PUSTAKA SISTEMATIS PADA DATA MINING:STUDI KASUS ALGORITMA K-MEANS CLUSTERING," *Jurnal Teknoif Teknik Informatika Institut Teknologi Padang,* vol. 10, 2022.

[18] S. H. A. Rozzi Kesuma Dinata, "Analisis K-MeansClustering pada DataSepeda Motor," *Informatics Journal,* vol. 5, 2020.

[19] M. Z. Nasution, "Penerapan Principal Component Analisis (PCA) dalam Penentuan Faktor Dominan Yang Mempengaruhi Prestasi Belajar Siswa," *Jurnal Teknologi Infromasi,* vol. 3, 2019.

[20] I. M. S. Dyah Hediyati, "Penerapan Principal Component Analysis(PCA) Untuk Reduksi Dimensi Pada Proses Clustering Data Produksi Pertanian Di Kabupaten Bojonegoro," *Journal Information Engineering and Educational Technology,* vol. 5, 2021.

[21] N. H. G. B. S. Elly Muningsih, "Penerapan Metode Principle Component Analysis (PCA) untuk Clustering Data Kunjungan Wisatawan Mancanegara ke Indonesia," *Bianglala Informatika,* vol. 8, 2020.

[22] I. S. Dewi Sartika, "Penerapan Metode Principal Component Analysis (PCA) Pada Klasifikasi Status Kredit Nasabah Bank Sumsel Babel Cabang KM 12 Palembang Menggunakan Metode Decision Tree," *Jurnal Ilmu Komputer dan Teknologi Informasi,* vol. 14, 2022.

[23] A. R. I. P. Z. A. Zulfa Nabila, "Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means," *JTSI,* vol. 2, 2021.

[24] B. K. Muhammad Romzi, "PEMBELAJARAN PEMROGRAMAN PYTHON DENGAN PENDEKATAN LOGIKA ALGORITMA," *JTIM: Jurnal Teknik Informatika Mahakarya,* vol. 3, 2020.

[25] D. R. W. O. R. M. T. R. Y. N. Kairos Abinaya Susanto, "Implementasi Bahasa Python Dalam Menganalisis Pengaruh Rokok Terhadap Risiko Pasien Terkena Penyakit Stroke," *Jurnal Publikasi Teknik Informatika (JUPTI),* vol. 2, pp. 48-58, 2023.

[26] N. A. A. S. B. K. R. R. Widi Hastomo, "Metode Pembelajaran Mesin untuk Memprediksi Emisi Manure Management," *Jurnal Nasional Teknik Elektro dan Teknologi Informasi ,* vol. 11, 2022.

[27] I. K. N. Yogasetya Suhanda, "Penerapan Metode Crisp-DM Dengan Algoritma K-Means Clustering Untuk Segmentasi Mahasiswa Berdasarkan Kualitas Akademik," *Jurnal Teknologi Informatika dan Komputer MH Thamrin* , vol. 6, 2020.

[28] A. T. D. S. Muhammad Jordy, "Analisis Segmentasi Recency dan Customer Value Pada AVANA Indonesia Dengan Algoritma K-Means dan Model RFM (Recency,Frequency and Monetary)," *Journal of Information System Research (JOSH),* vol. 4, 2023.

[29] N. M. D. Febriyant, "Implementasi Black Box Testing Pada Sistem Informasi Manajemen Dosen," *JITTER,* vol. 2, 2021.

[30] M. F. Londjo, "Implementasi White Box Testing dengan Teknik Basis Path Pada Pengujian Form Login," *Jurnal Siliwangi,* vol. 7, 2021.

[31] F. F. F. R. Nuril Huda Ahsina, "ANALISIS SEGMENTASI PELANGGAN BANK BERDASARKAN PENGAMBILAN KREDIT DENGAN MENGGUNAKAN METODE K-MEANS CLUSTERING," *Jurnal Ilmiah Teknologi Informasi Terapan,* vol. 8, 2022.

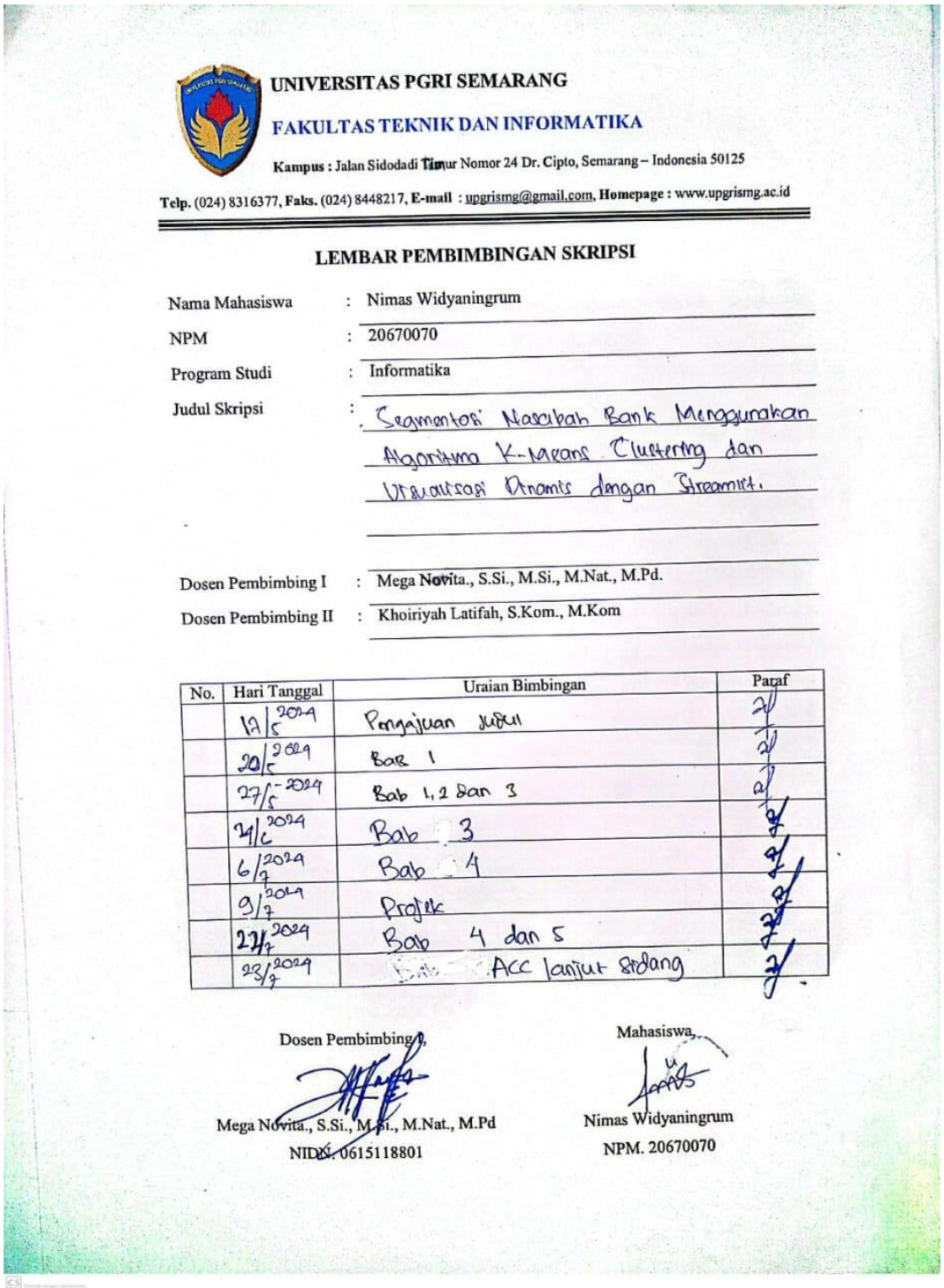
[32] H. M. Apip Pramudiansyah, "SEGMENTASI PELANGGAN MENGGUNAKAN ALGORITMA K-MEANS BERDASARKAN MODEL RECENCY FREQUENCY MONETARY," *Jurnal Ilmiah Ilmu Komputer,* vol. 7, 2021.

[33] K. O. Y. R. Q. Yefta Christian, "Penerapan K-Means pada Segmentasi Pasar untuk Riset Pemasaran padaStartup Early Stage dengan Menggunakan CRISP-DM," *JURIKOM (Jurnal Riset Komputer),* vol. 9, 2022.

[34] M. Z. Nasution, "PENERAPAN PRINCIPAL COMPONENT ANALYSIS (PCA) DALAM PENENTUAN FAKTOR DOMINAN YANG MEMPENGARUHI PRESTASI BELAJAR SISWA (Studi Kasus : SMK Raksana 2 Medan)," *Jurnal Teknologi Informasi*, vol. 3, 2019.

LAMPIRAN

Lampiran 1. Lembar Bimbingan Dosen Pembimbing 1



UNIVERSITAS PGRI SEMARANG

FAKULTAS TEKNIK DAN INFORMATIKA

Kampus : Jalan Sidodadi Timur Nomor 24 Dr. Cipto, Semarang – Indonesia 50125

Telp. (024) 8316377, Faks. (024) 8448217, E-mail : upgrismg@gmail.com, Homepage : www.upgrismg.ac.id

LEMBAR PEMBIMBINGAN SKRIPSI

Nama Mahasiswa : Nimas Widyaningrum

NPM : 20670070

Program Studi : Informatika

Judul Skripsi : Segmentasi Nasabah Bank Menggunakan Algoritma K-Means Clustering dan Visualisasi Dinamis dengan Streamlit.

Dosen Pembimbing I : Mega Novita., S.Si., M.Si., M.Nat., M.Pd.

Dosen Pembimbing II : Khoiriyah Latifah, S.Kom., M.Kom

No.	Hari Tanggal	Uraian Bimbingan	Paraf
	17/5 2024	Pengajuan Judul	
	20/5 2024	Bab 1	
	27/5 2024	Bab 1, 2 dan 3	
	24/6 2024	Bab 3	
	6/7 2024	Bab 4	
	9/7 2024	Prototype	
	22/7 2024	Bab 4 dan 5	
	23/7 2024	Bab 5 ACC lanjut sidang	

Dosen Pembimbing I,

Mega Novita., S.Si., M.Si., M.Nat., M.Pd

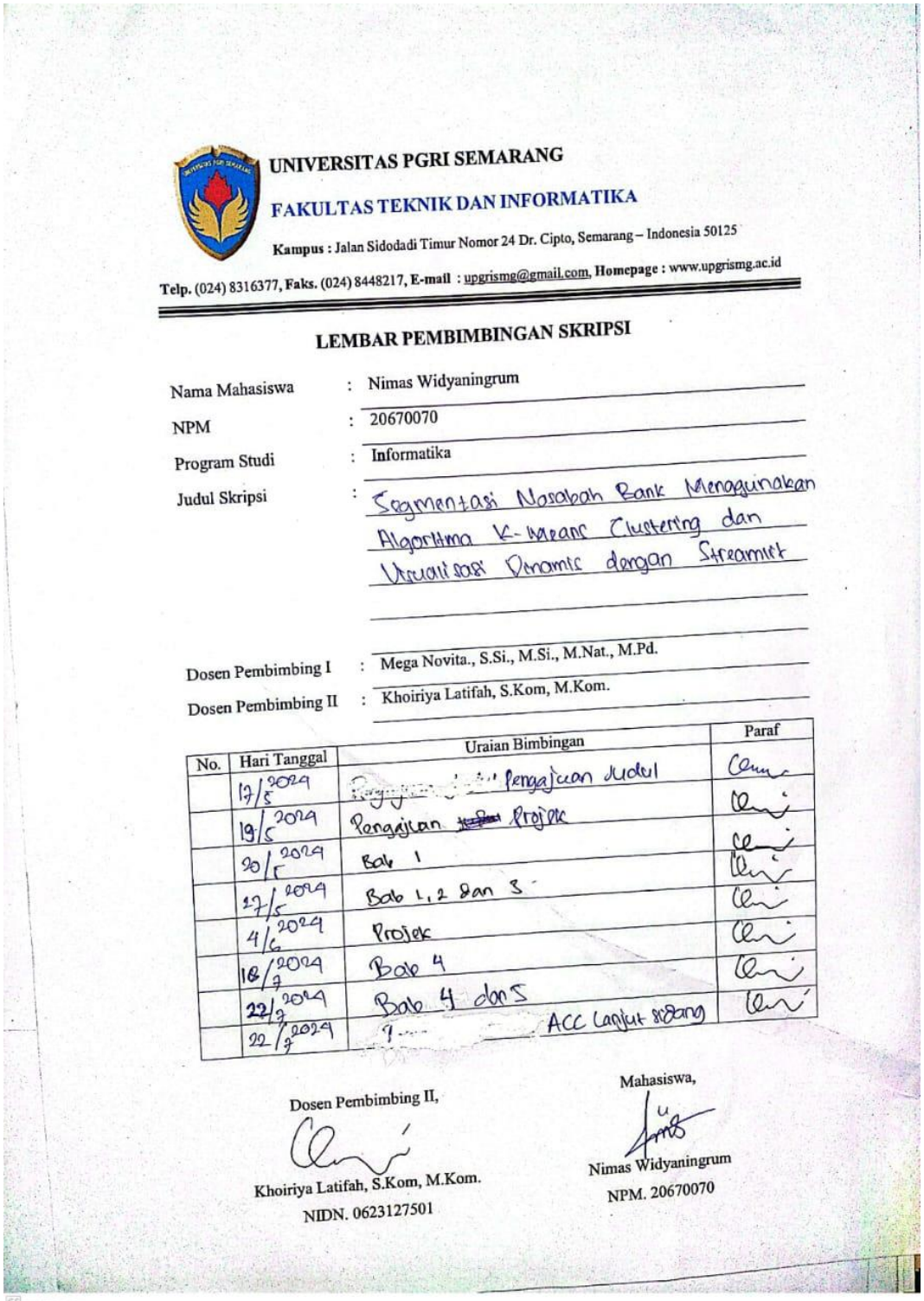
NIDN. 0615118801

Mahasiswa,

Nimas Widyaningrum

NPM. 20670070

Lampiran 2. Lembar Bimbingan Dosen Pembimbing 2



UNIVERSITAS PGRI SEMARANG

FAKULTAS TEKNIK DAN INFORMATIKA

Kampus : Jalan Sidodadi Timur Nomor 24 Dr. Cipto, Semarang – Indonesia 50125

Telp. (024) 8316377, Faks. (024) 8448217, E-mail : upgrismg@gmail.com, Homepage : www.upgrismg.ac.id

LEMBAR PEMBIMBINGAN SKRIPSI

Nama Mahasiswa : Nimas Widyaningrum

NPM : 20670070

Program Studi : Informatika

Judul Skripsi : Segmentasi Nasabah Bank Menggunakan Algoritma K-means Clustering dan Visualisasi Dinamis dengan Streamlit

Dosen Pembimbing I : Mega Novita., S.Si., M.Si., M.Nat., M.Pd.

Dosen Pembimbing II : Khoiriya Latifah, S.Kom, M.Kom.

No.	Hari Tanggal	Uraian Bimbingan	Paraf
	17/5 2024	Pengajuan Judul	
	19/5 2024	Pengajuan Projek	
	20/5 2024	Bab 1	
	27/5 2024	Bab 1, 2 dan 3	
	4/6 2024	Projek	
	18/7 2024	Bab 4	
	22/7 2024	Bab 4 dan 5	
	22/7 2024	ACC Lanjut sidang	

Dosen Pembimbing II,

Khoiriya Latifah, S.Kom, M.Kom.

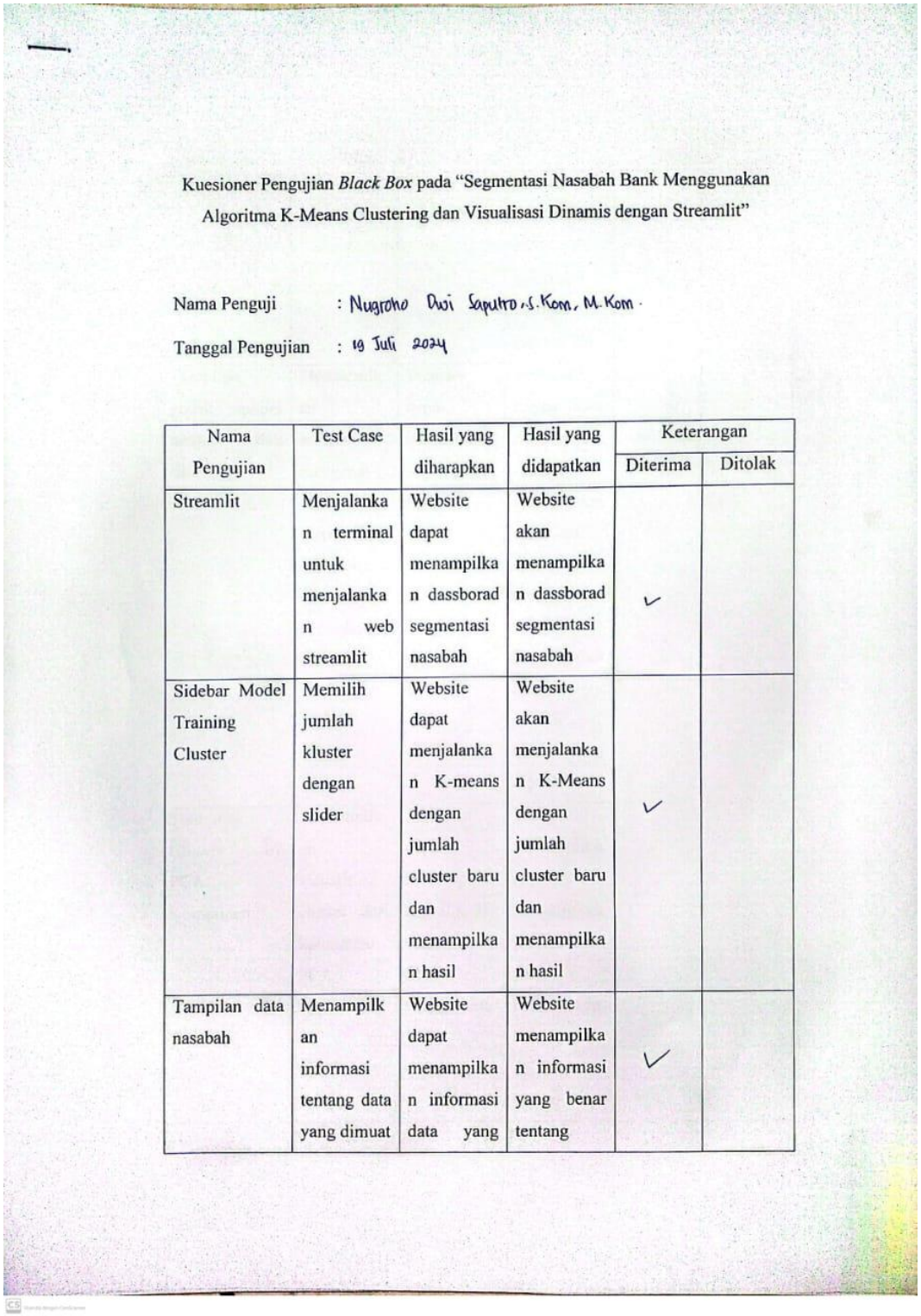
NIDN. 0623127501

Mahasiswa,

Nimas Widyaningrum

NPM. 20670070

Lampiran 3. Lembar Pengujian Black Box Penguji 1



Kuesioner Pengujian *Black Box* pada "Segmentasi Nasabah Bank Menggunakan Algoritma K-Means Clustering dan Visualisasi Dinamis dengan Streamlit"

Nama Penguji : Nugroho Dwi Saputro, S.Kom., M.Kom.

Tanggal Pengujian : 19 Juli 2024

Nama Pengujian	Test Case	Hasil yang diharapkan	Hasil yang didapatkan	Keterangan	
				Diterima	Ditolak
Streamlit	Menjalankan terminal untuk menjalankan web streamlit	Website dapat menampilkan dassborad segmentasi nasabah	Website akan menampilkan dassborad segmentasi nasabah	✓	
Sidebar Model Training Cluster	Memilih jumlah kluster dengan slider	Website dapat menjalankan K-means dengan jumlah cluster baru dan menampilkan hasil	Website akan menjalankan K-Means dengan jumlah cluster baru dan menampilkan hasil	✓	
Tampilan data nasabah	Menampilkan informasi tentang data yang dimuat	Website dapat menampilkan informasi data yang	Website menampilkan informasi yang benar tentang	✓	

		dimuat	jumlah baris dan kolom, serta informasi lainnya tentang data		
Tampilan grafik model inertia dan silhouette score	Menampilkan informasi mengenai model inertia dan silhouette score untuk pemodelan klaster	Website dapat menampilkan grafik model	Website menampilkan grafik model inertia dan silhouette score muncul dengan benar sesuai dengan jumlah kluster yang dipilih	✓	
Tampilan Cluster by PCA Komponen	Menampilkan visualisasi cluster dari komponen PCA	Website dapat menampilkan grafik 3D cluster by PCA Component	Website menampilkan grafik 3D menunjukkan distribusi data berdasarkan komponen PCA dan kluster	✓	

Tampilan Box Plots	Visualisasi box plot berdasarkan kluster	Website dapat menampilkan grafik box plot dengan benar	Website menampilkan grafik box plot ditampilkan dengan benar sesuai dengan kluster	✓	
Tampilan Pairplot	Visualisasi pairplot berdasarkan kluster	Website dapat menampilkan grafik pairplot yang sesuai	Website menampilkan pairplot ditampilkan dengan benar sesuai dengan kluster	✓	

Saran :

1. Penggunaan Stopwatch
2. Informasi prosedur penggunaan
3. Informasi hasil buat modalh di pahami!

Penguji,

Nugroho Dwi S.

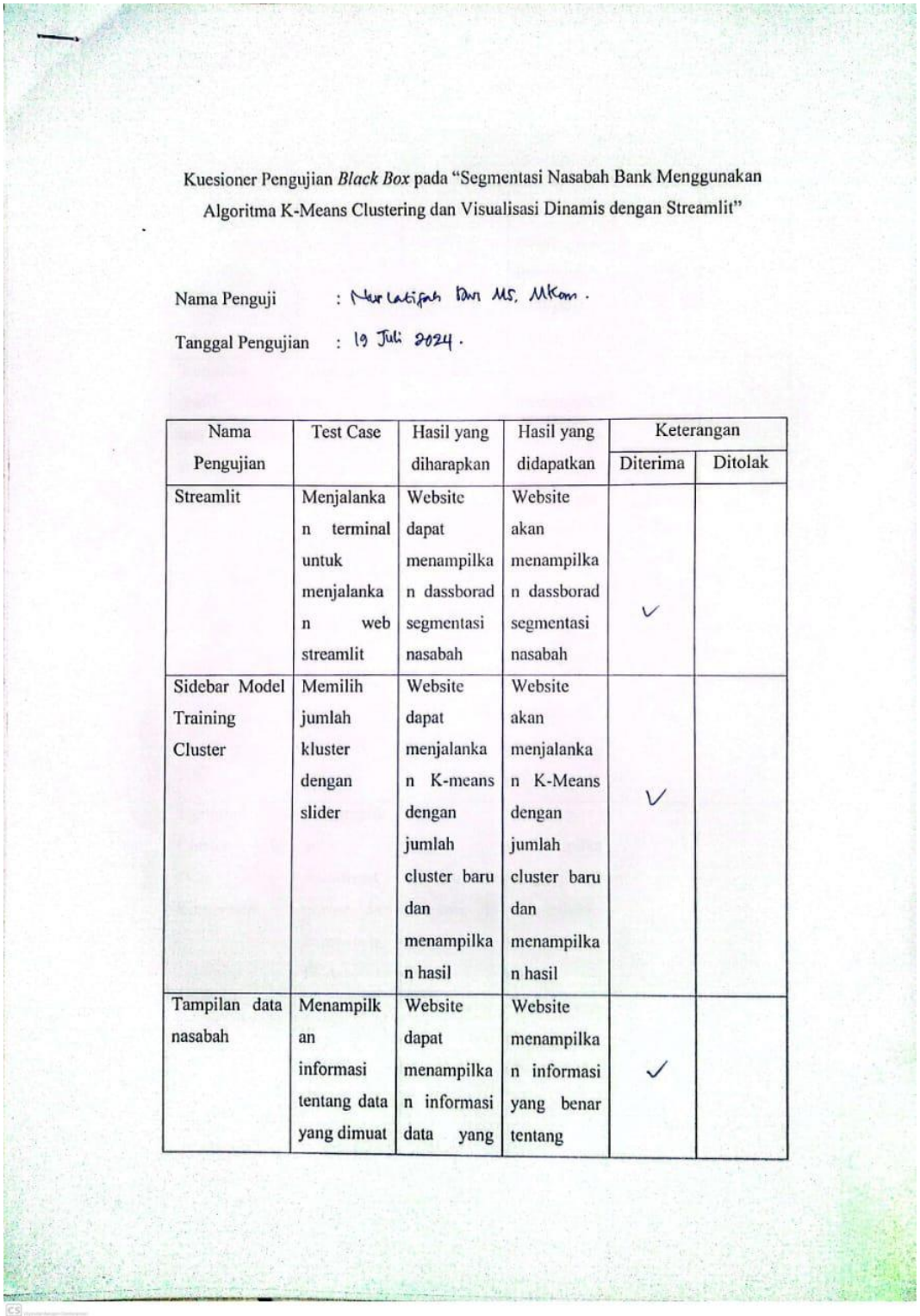
Kuesioner Pengujian *Black Box* pada "Segmentasi Nasabah Bank Menggunakan Algoritma K-Means Clustering dan Visualisasi Dinamis dengan Streamlit"

Nama Penguji : Nur Latifah Dwi MS, MKom.

Tanggal Pengujian : 19 Juli 2024.

Nama Pengujian	Test Case	Hasil yang diharapkan	Hasil yang didapatkan	Keterangan	
				Diterima	Ditolak
Streamlit	Menjalankan terminal untuk menjalankan web streamlit	Website dapat menampilkan dassborad segmentasi nasabah	Website akan menampilkan dassborad segmentasi nasabah	✓	
Sidebar Model Training Cluster	Memilih jumlah kluster dengan slider	Website dapat menjalankan K-means dengan jumlah cluster baru dan menampilkan hasil	Website akan menjalankan K-Means dengan jumlah cluster baru dan menampilkan hasil	✓	
Tampilan data nasabah	Menampilkan informasi tentang data yang dimuat	Website dapat menampilkan informasi data yang	Website menampilkan informasi yang benar tentang	✓	

Lampiran 4. Lembar Pengujian Black Box Penguji 2



Kuesioner Pengujian *Black Box* pada "Segmentasi Nasabah Bank Menggunakan Algoritma K-Means Clustering dan Visualisasi Dinamis dengan Streamlit"

Nama Penguji : Nur Latifah Dwi MS, MKom.

Tanggal Pengujian : 19 Juli 2024.

Nama Pengujian	Test Case	Hasil yang diharapkan	Hasil yang didapatkan	Keterangan	
				Diterima	Ditolak
Streamlit	Menjalankan terminal untuk menjalankan web streamlit	Website dapat menampilkan dassborad segmentasi nasabah	Website akan menampilkan dassborad segmentasi nasabah	✓	
Sidebar Model Training Cluster	Memilih jumlah kluster dengan slider	Website dapat menjalankan K-means dengan jumlah cluster baru dan menampilkan hasil	Website akan menjalankan K-Means dengan jumlah cluster baru dan menampilkan hasil	✓	
Tampilan data nasabah	Menampilkan informasi tentang data yang dimuat	Website dapat menampilkan informasi data yang	Website menampilkan informasi yang benar tentang	✓	

		dimuat	jumlah baris dan kolom, serta informasi lainnya tentang data		
Tampilan grafik model inertia dan silhouette score	Menampilkan informasi mengenai model inertia dan silhouette score untuk pemodelan klaster	Website dapat menampilkan grafik model	Website menampilkan grafik model inertia dan silhouette score muncul dengan benar sesuai dengan jumlah kluster yang dipilih	✓	
Tampilan Cluster by PCA Komponen	Menampilkan visualisasi cluster dari komponen PCA	Website dapat menampilkan grafik 3D cluster by PCA Component	Website menampilkan grafik 3D menunjukkan distribusi data berdasarkan komponen PCA dan kluster	✓	

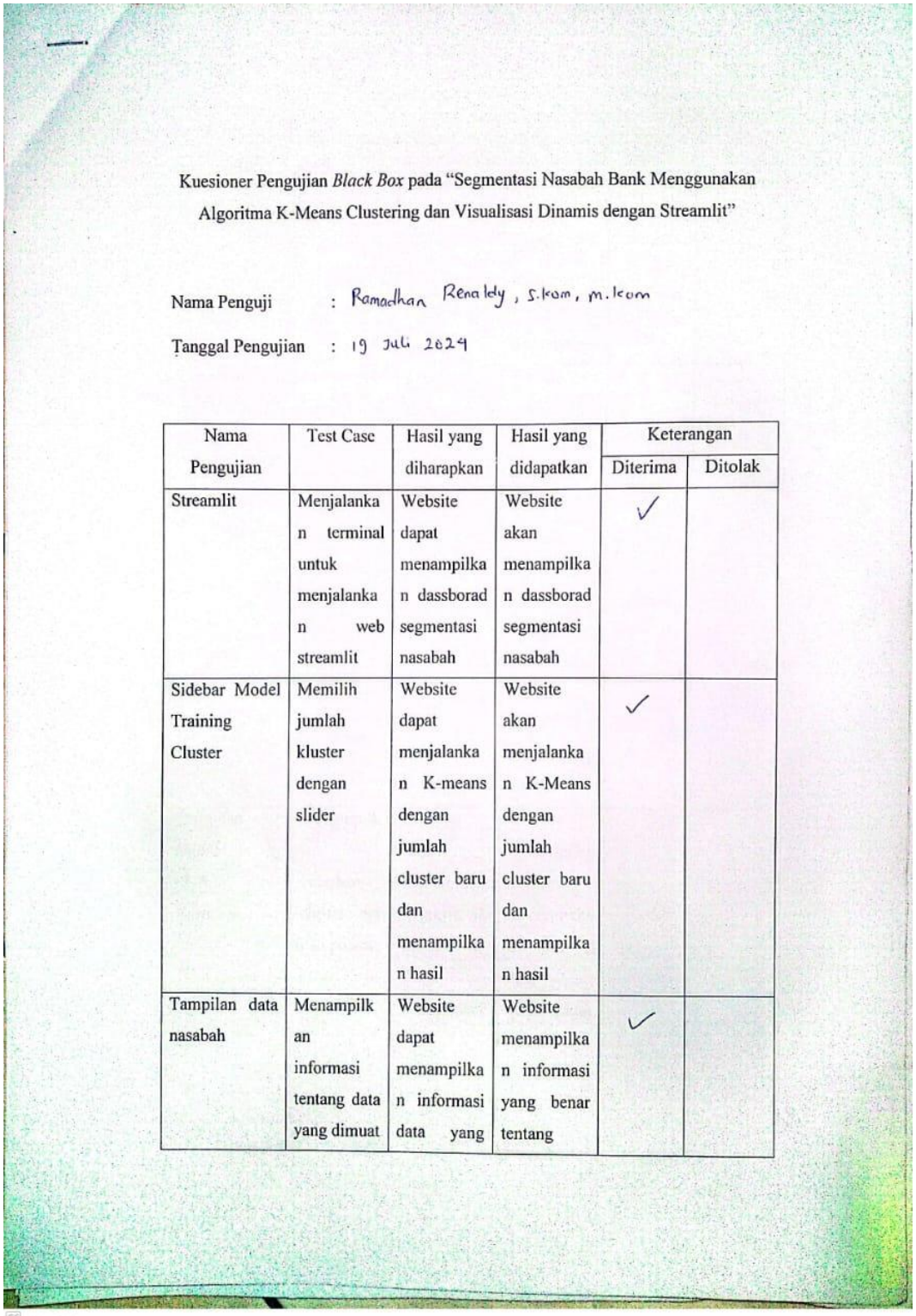
Tampilan Box Plots	Visualisasi box plot berdasarkan kluster	Website dapat menampilkan grafik box plot dengan benar	Website menampilkan grafik box plot ditampilkan dengan benar sesuai dengan kluster	✓	
Tampilan Pairplot	Visualisasi pairplot berdasarkan kluster	Website dapat menampilkan grafik pairplot yang sesuai	Website menampilkan pairplot ditampilkan dengan benar sesuai dengan kluster	✓	

Saran:
1. User interfacenya dibuat semenarik mungkin agar pengguna mengerti.

Penguji,

Nur Latifah Dwi MS, M.Kom

Lampiran 5. Lembar Pengujian Black Box Penguji 3



Kuesioner Pengujian *Black Box* pada "Segmentasi Nasabah Bank Menggunakan Algoritma K-Means Clustering dan Visualisasi Dinamis dengan Streamlit"

Nama Penguji : Ramadhan Renaldy, S.Kom, M.Kom

Tanggal Pengujian : 19 Juli 2024

Nama Pengujian	Test Case	Hasil yang diharapkan	Hasil yang didapatkan	Keterangan	
				Diterima	Ditolak
Streamlit	Menjalankan terminal untuk menjalankan web streamlit	Website dapat menampilkan dassborad segmentasi nasabah	Website akan menampilkan dassborad segmentasi nasabah	✓	
Sidebar Model Training Cluster	Memilih jumlah kluster dengan slider	Website dapat menjalankan K-means dengan jumlah cluster baru dan menampilkan hasil	Website akan menjalankan K-Means dengan jumlah cluster baru dan menampilkan hasil	✓	
Tampilan data nasabah	Menampilkan informasi tentang data yang dimuat	Website dapat menampilkan informasi data yang	Website menampilkan informasi yang benar tentang	✓	

		dimuat	jumlah baris dan kolom, serta informasi lainnya tentang data		
Tampilan grafik model inertia dan silhouette score	Menampilkan informasi mengenai model inertia dan silhouette score untuk pemodelan klaster	Website dapat menampilkan grafik model	Website menampilkan grafik model inertia dan silhouette score muncul dengan benar sesuai dengan jumlah kluster yang dipilih	✓	
Tampilan Cluster by PCA Komponen	Menampilkan visualisasi cluster dari komponen PCA	Website dapat menampilkan grafik 3D cluster by PCA Component	Website menampilkan grafik 3D menunjukkan distribusi data berdasarkan komponen PCA dan kluster	✓	

Tampilan Box Plots	Visualisasi box plot berdasarkan kluster	Website dapat menampilkan grafik box plot dengan benar	Website menampilkan grafik box plot ditampilkan dengan benar sesuai dengan kluster	✓	
Tampilan Pairplot	Visualisasi pairplot berdasarkan kluster	Website dapat menampilkan grafik pairplot yang sesuai	Website menampilkan pairplot ditampilkan dengan benar sesuai dengan kluster	✓	

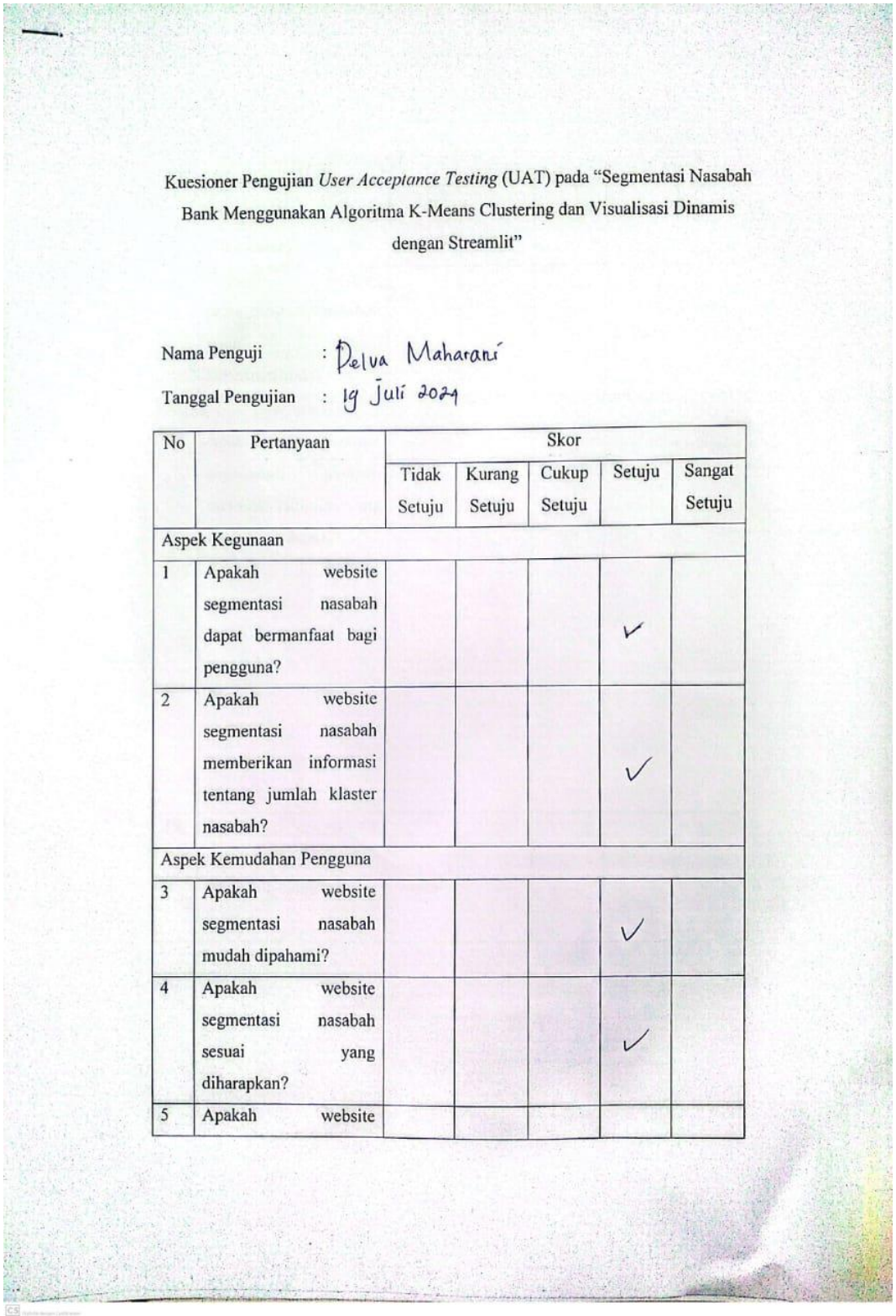
Saran :

1. Buat webnya lebih memusiacwi agar dapat dipahami
2. Untuk grafik tampilkan dengan grafik asli bukan dengan hanya mengganti gambar grafiknya Jika N of clusternya diubah

Penguji,

Ramadhan Kenaldy

Lampiran 6. Lembar Pengujian User Acceptance Testing (UAT) Penguji 1



Kuesioner Pengujian *User Acceptance Testing* (UAT) pada "Segmentasi Nasabah Bank Menggunakan Algoritma K-Means Clustering dan Visualisasi Dinamis dengan Streamlit"

Nama Penguji : Delva Maharani

Tanggal Pengujian : 19 Juli 2024

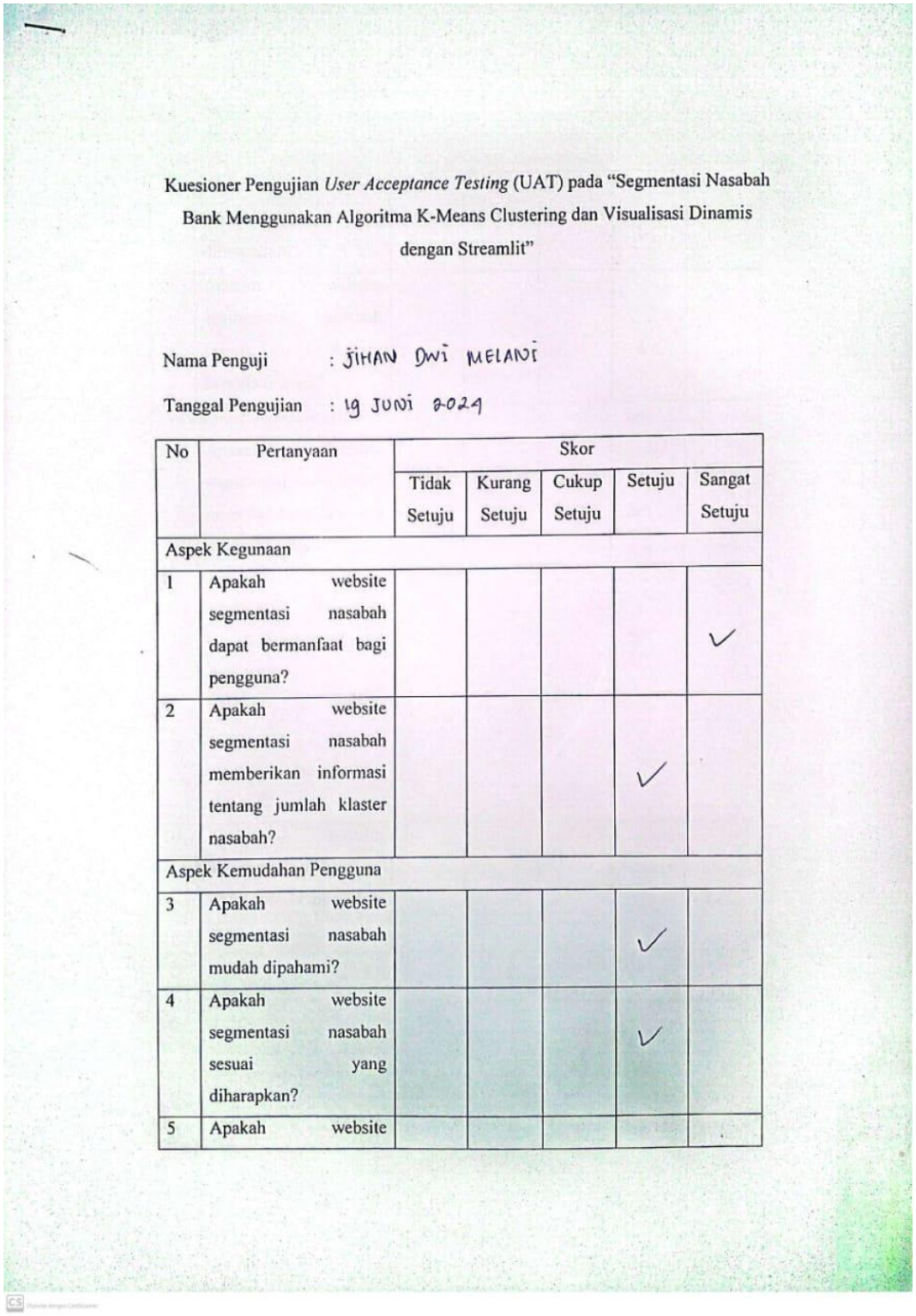
No	Pertanyaan	Skor				
		Tidak Setuju	Kurang Setuju	Cukup Setuju	Setuju	Sangat Setuju
Aspek Kegunaan						
1	Apakah website segmentasi nasabah dapat bermanfaat bagi pengguna?				✓	
2	Apakah website segmentasi nasabah memberikan informasi tentang jumlah klaster nasabah?				✓	
Aspek Kemudahan Pengguna						
3	Apakah website segmentasi nasabah mudah dipahami?				✓	
4	Apakah website segmentasi nasabah sesuai yang diharapkan?				✓	
5	Apakah website					

	segmentasi nasabah berisi informasi yang dibutuhkan?				✓	
6	Apakah website segmentasi nasabah sesuai dengan keperluan anda?				✓	
Aspek *User Interface* (UI)						
7	Apakah website segmentasi nasabah memiliki tampilan yang mudah dipahami?				✓	
8	Apakah website sebmentasi nasabah memiliki tampilan yang menarik?				✓	
9	Apakah website segmentasi nasabah memiliki tema yang menarik?				✓	
10	Apakah website segmentasi nasabah perlu di kembangkan lagi?					✓

Semarang, 19 Juli 2024.

Delva Maharani

Lampiran 7. Lembar Pengujian User Acceptance Testing (UAT) Penguji 2



Kuesioner Pengujian *User Acceptance Testing* (UAT) pada "Segmentasi Nasabah Bank Menggunakan Algoritma K-Means Clustering dan Visualisasi Dinamis dengan Streamlit"

Nama Penguji : JIHAN DWI MELANI

Tanggal Pengujian : 19 JUNI 2024

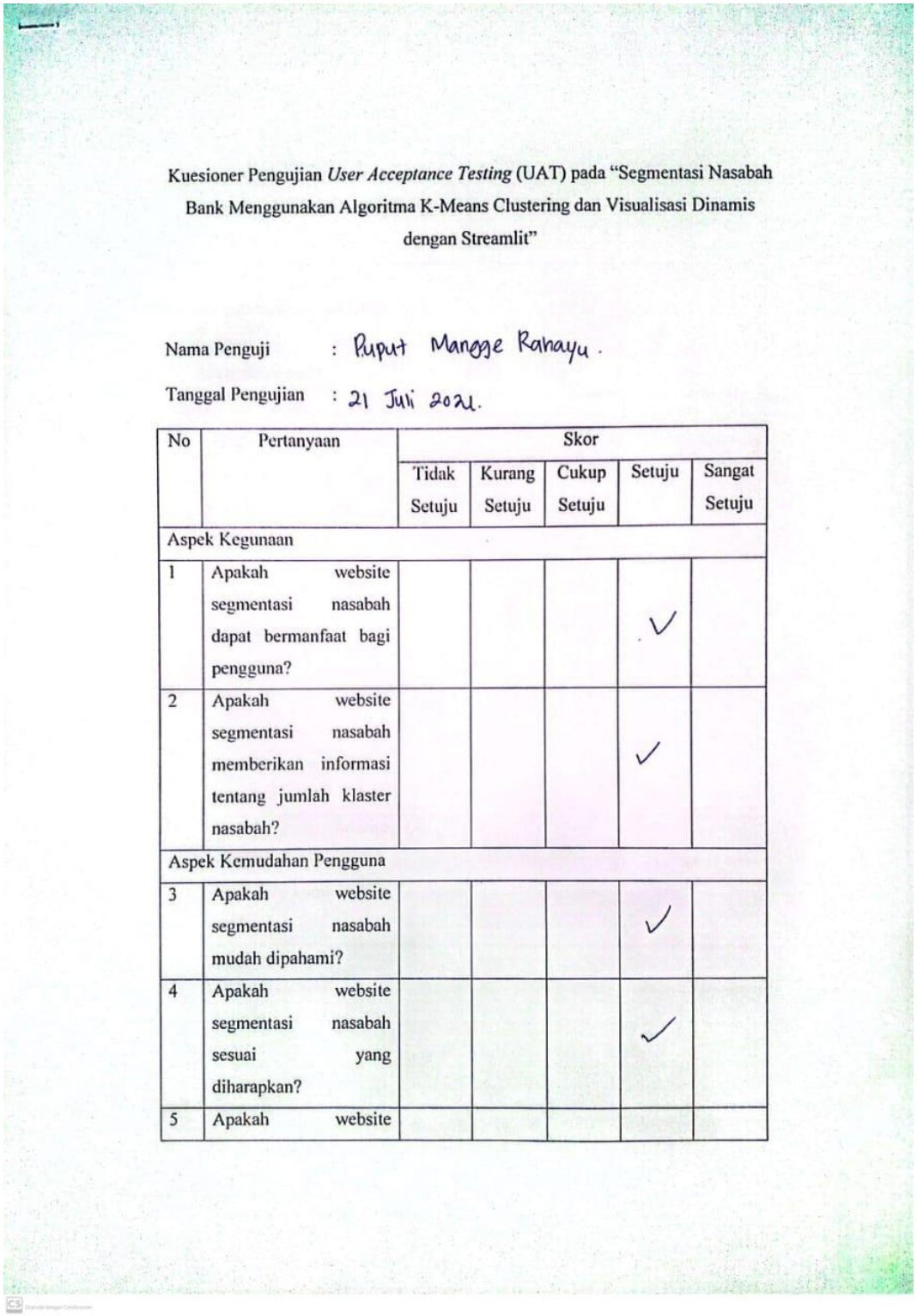
No	Pertanyaan	Skor				
		Tidak Setuju	Kurang Setuju	Cukup Setuju	Setuju	Sangat Setuju
Aspek Kegunaan						
1	Apakah website segmentasi nasabah dapat bermanfaat bagi pengguna?					✓
2	Apakah website segmentasi nasabah memberikan informasi tentang jumlah klaster nasabah?				✓	
Aspek Kemudahan Pengguna						
3	Apakah website segmentasi nasabah mudah dipahami?				✓	
4	Apakah website segmentasi nasabah sesuai yang diharapkan?				✓	
5	Apakah website					

	segmentasi nasabah berisi informasi yang dibutuhkan?				✓	
6	Apakah website segmentasi nasabah sesuai dengan keperluan anda?				✓	
Aspek *User Interface* (UI)						
7	Apakah website segmentasi nasabah memiliki tampilan yang mudah dipahami?				✓	
8	Apakah website sebmentasi nasabah memiliki tampilan yang menarik?				✓	
9	Apakah website segmentasi nasabah memiliki tema yang menarik?				✓	
10	Apakah website segmentasi nasabah perlu di kembangkan lagi?					✓

SEMARANG, 19 JULI 2024

JIHAN DWI MELANI

Lampiran 8. Lembar Pengujian User Acceptance Testing (UAT) Penguji 3



Kuesioner Pengujian *User Acceptance Testing* (UAT) pada "Segmentasi Nasabah Bank Menggunakan Algoritma K-Means Clustering dan Visualisasi Dinamis dengan Streamlit"

Nama Penguji : Puput Mangge Rahayu.

Tanggal Pengujian : 21 Juli 2021.

No	Pertanyaan	Skor				
		Tidak Setuju	Kurang Setuju	Cukup Setuju	Setuju	Sangat Setuju
Aspek Kegunaan						
1	Apakah website segmentasi nasabah dapat bermanfaat bagi pengguna?				✓	
2	Apakah website segmentasi nasabah memberikan informasi tentang jumlah klaster nasabah?				✓	
Aspek Kemudahan Pengguna						
3	Apakah website segmentasi nasabah mudah dipahami?				✓	
4	Apakah website segmentasi nasabah sesuai yang diharapkan?				✓	
5	Apakah website					

	segmentasi nasabah berisi informasi yang dibutuhkan?					
6	Apakah website segmentasi nasabah sesuai dengan keperluan anda?					✓
Aspek *User Interface* (UI)						
7	Apakah website segmentasi nasabah memiliki tampilan yang mudah dipahami?				✓	
8	Apakah website sebmentasi nasabah memiliki tampilan yang menarik?				✓	
9	Apakah website segmentasi nasabah memiliki tema yang menarik?				✓	
10	Apakah website segmentasi nasabah perlu di kembangkan lagi?					✓

Semarang, 21 Juli 2024.

PUPUT MR.