



KERJA PRAKTIK - SA 234702

PT APLIKANUSA LINTASARTA

PERBANDINGAN HASIL SEGMENTASI PELANGGAN
DENGAN RFM *SCORING* DAN RFM *BASED K-MEANS
CLUSTERING*

NATANAEL HADI MULIABANTA (5006201107)

Dosen Pembimbing
Ulil Azmi, S.Si, M.Si, M.Sc.

PROGRAM STUDI SARJANA
DEPARTEMEN AKTUARIA
FAKULTAS SAINS DAN ANALITIKA DATA
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA
2023



KERJA PRAKTIK - SA 234702

PT APLIKANUSA LINTASARTA

**PERBANDINGAN HASIL SEGMENTASI PELANGGAN
DENGAN RFM *SCORING* DAN RFM *BASED K-MEANS
CLUSTERING***

NATANAEL HADI MULIABANTA (5006201107)

**Dosen Pembimbing
Ulil Azmi, S.Si, M.Si, M.Sc.**

**PROGRAM STUDI SARJANA
DEPARTEMEN AKTUARIA
FAKULTAS SAINS DAN ANALITIKA DATA
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA
2024**

LEMBAR PENGESAHAN I

LAPORAN KERJA PRAKTIK

**Program Studi Sarjana Departemen Aktuaria
Fakultas Sains dan Analitika Data
Institut Teknologi Sepuluh Nopember Surabaya**

Surabaya, 3 Juli 2024

Menyetujui,


Dosen Pembimbing



Ulil Azmi, S.Si, M.Si, M.Sc
NIP. 1990201912069

Mengetahui,

**Kepala Departemen Aktuaria
FSAD ITS**



Dr. Drs. Soehardjoepri, M.Si.
NIP. 19620504 198701 1 00

LEMBAR PENGESAHAN II
LAPORAN KERJA PRAKTIK
PT APLIKANUSA LINTASARTA

Jakarta, 3 Juli 2024

Menyetujui,

Pembimbing Lapangan



Putri Christiani
Tax Officer

Mengetahui,



Sayyid Malik Alfattah
Junior Tax Manager

KATA PENGANTAR

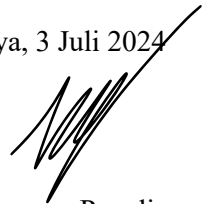
Puji dan syukur kami panjatkan kepada Tuhan Yang Maha Esa yang telah memberikan kami kekuatan serta rahmat-Nya sehingga laporan kerja praktik berjudul “Perbandingan Hasil Segmentasi Pelanggan dengan RFM *Scoring* DAN RFM *Based K-Means Clustering*” dapat diselesaikan untuk memenuhi persyaratan akademik di Departemen Aktuaria.

Penulis menyampaikan terima kasih kepada pihak – pihak yang telah memberikan bantuan, dukungan, dan bimbingan selama pelaksanaan kerja praktik ini, khususnya kepada:

1. Bapak Dr. Drs. Soehardjoepri, M.Si., selaku Kepala Departemen Aktuaria ITS.
2. Ibu Ulil Azmi, S.Si, M.Si, M.Sc., selaku dosen pembimbing II kerja praktik Departemen Aktuaria ITS.
3. Pihak PT Aplikanusa Lintasarta yang telah memberikan kesempatan dan wadah untuk melakukan kerja praktik.

Dalam penulisan laporan kerja praktik, penulis menyadari bahwa laporan ini belum sepenuhnya sempurna baik dalam pemilihan diksi maupun cara penulisannya. Oleh karena itu, penulis mohon maaf dan terbuka untuk menerima kritik serta saran yang membangun. Penulis berharap laporan kerja praktik ini dapat bermanfaat untuk kedepannya.

Surabaya, 3 Juli 2024



Penulis

DAFTAR ISI

LEMBAR PENGESAHAN I	i
LEMBAR PENGESAHAN II	ii
KATA PENGANTAR	iii
DAFTAR ISI	iv
DAFTAR TABEL	vii
DAFTAR LAMPIRAN	viii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Dasar Pemikiran.....	2
1.3 Tujuan	2
1.3.1 Tujuan Umum	2
1.3.2 Tujuan Khusus	3
1.4 Manfaat	3
1.4.1 Manfaat Bagi PT Aplikanusa Lintasarta	3
1.4.2 Manfaat Bagi Mahasiswa.....	3
1.4.3 Manfaat Bagi Departemen Aktuaria ITS	4
BAB II GAMBARAN UMUM INSTANSI	5
2.1 PT Aplikanusa Lintasarta	5
2.2 Visi dan Misi PT Aplikanusa Lintasarta	6
2.2.1 Visi PT Aplikanusa Lintasarta.....	6
2.2.2 Misi PT Aplikanusa Lintasarta.....	6
2.3 Struktur Organisasi PT Aplikanusa Lintasarta	6
BAB III PELAKSANAAN KERJA PRAKTIK	9
3.1 Pelaksanaan Kerja Praktik	9
3.2 Metode Penyelesaian Tugas Khusus.....	9
3.2.1 Analisis RFM.....	10
3.2.1 Analisis Statistika Deskriptif	11

3.2.3 IQR Method	12
3.2.4 Elbows Method	12
3.2.5 Silhouette Coefficient Method	13
3.2.6 K-Means Clustering	14
3.2.7 Uji Statistik	15
3.2.8 Sumber Data dan Variabel Penelitian.....	15
3.2.9 Langkah Analisis.....	17
BAB IV PELAKSANAAN KERJA PRAKTIK	19
4.1 Karakteristik Data	19
4.2 Analisis RFM	20
4.3 Analisis Deskriptif	21
4.4 <i>Preprocessing</i> Data	22
4.5 Segmentasi RFM <i>Scoring</i>	24
4.6 Penentuan Nilai K.....	27
4.7 Analisis Signifikansi <i>Cluster</i>	30
4.8 Analisis <i>Cluster K-Means</i>	30
4.9 Perbandingan Hasil K-Means dan RFM <i>Scoring</i>	33
BAB V PENUTUP.....	38
5.1 Kesimpulan	39
5.2 Saran	39
DAFTAR PUSTAKA	41
LAMPIRAN	42

DAFTAR GAMBAR

Gambar 2.1 Logo PT Aplikanusa Lintasarta.....	5
Gambar 2.2 Struktur Organisasi PT Aplikanusa Lintasarta	7
Gambar 3.1 Flowchart Analisis Data	18
Gambar 4.1 <i>Box Plot</i> Variabel <i>Recency</i> , <i>Frequency</i> , dan <i>Monetary</i>	22
Gambar 4.2 <i>Box Plot</i> Variabel <i>Recency</i> , <i>Frequency</i> , dan <i>Monetary</i> tanpa <i>outlier</i>	23
Gambar 4.3 <i>Pie Chart</i> Segmentasi RFM <i>scoring</i>	26
Gambar 4.4 Grafik <i>Elbows Method</i>	28
Gambar 4.5 Grafik <i>Sillhouette Plot</i>	29
Gambar 4.6 <i>Pie Chart</i> <i>K-Means Clustering</i>	32
Gambar 4.7 <i>Pie Chart</i> hasil segmentasi dari RFM <i>score</i> (a), <i>K-Means</i> (b)	34
Gambar 4.8 3D scatterplot hasil segmentasi RFM <i>score</i> tanpa <i>outlier</i> (a), <i>K-Means</i> (b)	35
Gambar 4.9 <i>Box Plot</i> <i>Perbandingan Segmentasi Recency(a)</i> , <i>Frequency(b)</i> , <i>Monetary(c)</i>	36

DAFTAR TABEL

Tabel 3.1	Indikator Penilaian RFM Score	11
Tabel 3.2	Struktur Data untuk Analisis RFM	16
Tabel 3.3	Struktur Data untuk Segmentasi	16
Tabel 4.2	Analisis RFM Data Transaksi	20
Tabel 4.3	Analisis Deskriptif Data RFM	21
Tabel 4.4	Data RFM Tanpa <i>Outlier</i>	23
Tabel 4.5	Tabel RFM dengan Skor	24
Tabel 4.6	Tabel RFM dengan Klasifikasi	25
Tabel 4.7	Jumlah Pelanggan dan Rata-Rata Variabel Setiap Segmen	26
Tabel 4.8	Nilai RFM setelah <i>rescaling</i>	27
Tabel 4.9	Nilai <i>Silhouette Score</i>	29
Tabel 4.10	Tabel ANOVA	30
Tabel 4.11	Nilai <i>Centroid Cluster</i>	31
Tabel 4.12	Jumlah Pelanggan dan Rata-Rata Variabel Setiap Segmen	33
Tabel 4.13	Jumlah Pelanggan Tiap Segmen Dari Kedua Metode	34

DAFTAR LAMPIRAN

Lampiran 1	Surat Penerimaan Kerja Praktik dari Perusahaan	43
Lampiran 2	Form F-5	44
Lampiran 3	Form F-6	45
Lampiran 4	Bukti Kerja Praktik Menggunakan GPS Map Camera	47
Lampiran 5	Syntax Python.....	54
Lampiran 6	<i>Raw</i> Data	66

BABI

PENDAHULUAN

1.1 Latar Belakang

Dalam persiapan memulai karir, mahasiswa perlu dipersiapkan untuk berpartisipasi dalam dunia kerja. Selain mempelajari pengetahuan yang diperoleh dalam perkuliahan, mahasiswa dapat meningkatkan kemampuannya dengan memperoleh keterampilan aplikasi, koneksi pribadi yang luas, keterampilan pengambilan keputusan, keterampilan komunikasi, dan kemampuan untuk merespon perubahan demi perubahan. kebutuhan siswa ini, Institut Teknologi Sepuluh Nopember (ITS) menyediakan wadah bagi mahasiswa sebagai bentuk penerapan ilmu khususnya melalui kerja praktek. Kerja langsung ini juga mendukung ITS dalam mencapai goals 1 yaitu mempersiapkan dan menghasilkan sumber daya manusia yang berkualitas melalui pengajaran dan penerapan ilmu. Salah satu bentuk penerapan ilmu adalah praktik. Kerja praktik ini bertujuan untuk meningkatkan pemahaman mahasiswa tentang materi yang diajarkan dalam perkuliahan dan untuk memenuhi syarat mereka untuk masuk ke masyarakat melalui aplikasi mereka di perusahaan.

Kerja Praktik sendiri memberikan manfaat bagi mahasiswa diantaranya mahasiswa dapat memperoleh pengalaman mengenai persoalan di dunia kerja, tanggung jawab pekerjaan dan menjaga komunikasi serta koordinasi antar rekan kerja sehingga dapat membantu mahasiswa untuk terjun menghadapi dunia kerja secara langsung. Selain itu, Kerja Praktik dapat memberikan manfaat bagi perusahaan terkait dimana Kerja Praktik dilaksanakan. Bagi perusahaan, manfaat yang didapatkan yaitu dapat membantu menyelesaikan permasalahan perusahaan dengan menggunakan mata kuliah Aktuaria yang sesuai. Oleh Karena itu, kami mahasiswa Departemen Aktuaria FSAD ITS bermaksud mengajukan Proposal Kerja Praktik ini untuk mengaplikasikan ilmu Aktuaria

di PT Aplikanusa Lintasarta yang akan dilakukan selama satu bulan, dimulai dari Januari 2024 sampai dengan Februari 2024.

1.2 Dasar Pemikiran

Dasar Pemikiran yang digunakan dalam pelaksanaan Kerja Praktik diantaranya sebagai berikut.

1. Tri Dharma Perguruan Tinggi, yaitu: pendidikan, penelitian, dan pengabdian masyarakat.
2. Tujuan pendidikan ITS Surabaya, yaitu kepemimpinan, keahlian, berpikir ilmiah dan sikap hidup bermasyarakat.
3. Meningkatkan kerja sama dengan berbagai bidang dan bermasyarakat yang bertujuan untuk meningkatkan relevansi mutu pendidikan dan penelitian.
4. Mampu menerapkan teori - teori yang sudah didapatkan selama kuliah secara praktikal di tempat kerja.

1.3 Tujuan

Kerja Praktik yang dilakukan oleh Mahasiswa Departemen Aktuaria FSAD ITS memiliki beberapa tujuan diantaranya sebagai berikut.

1.3.1 Tujuan Umum

Adapun tujuan umum yang ingin dicapai dalam pelaksanaan Kerja Praktik ini adalah sebagai berikut.

1. Sebagai salah satu syarat kelulusan Program Studi Sarjana (S-1).
2. Mengetahui penerapan ilmu aktuaria yang didapat dari perkuliahan di tempat kerja secara nyata.
3. Mendapatkan pengalaman bersosialisasi dan berkomunikasi di dunia kerja.
4. Mendapatkan kesempatan untuk berlatih menyelesaikan permasalahan dalam dunia kerja.
5. Memahami sistem kerja di PT Aplikanusa Lintasarta.

1.3.2 Tujuan Khusus

Adapun tujuan khusus yang ingin dicapai dalam pelaksanaan Kerja Praktik ini adalah sebagai berikut.

1. Memberikan alternatif penyelesaian dari permasalahan yang diberikan oleh PT Aplikanusa Lintasarta.
2. Mendapatkan pengalaman bersosialisasi dan berkomunikasi di PT Aplikanusa Lintasarta.

1.4 Manfaat

Dalam Kerja Praktik yang akan dilaksanakan mahasiswa Departemen Aktuaria ITS terdapat beberapa manfaat yang akan diperoleh perusahaan maupun mahasiswa, sepanjang dan setelah pelaksanaannya, antara lain:

1.4.1 Manfaat Bagi PT Aplikanusa Lintasarta

Berikut manfaat yang bisa didapatkan perusahaan dengan terlaksananya kuliah praktik ini:

1. Adanya kerjasama yang terjalin antara PT Aplikanusa Lintasarta dengan Insitut Teknologi Sepuluh Nopember, khususnya dengan Departemen Aktuaria ITS.
2. Dapat mengenali dan menyelesaikan masalah yang terjadi di PT Aplikanusa Lintasarta yang berhubungan dengan ilmu yang dipelajari di departemen Sains Aktuaria.

1.4.2 Manfaat Bagi Mahasiswa

Berikut manfaat yang bisa didapatkan Mahasiswa dengan terlaksananya kuliah praktik ini:

1. Menambah pengalaman dan wawasan mengenai dunia industri kerja dan penerapan ilmu aktuaria di secara nyata, serta mendapat kesempatan untuk lebih mengetahui PT Aplikanusa Lintasarta.

2. Menyelesaikanelesaikan mata kuliah Kerja Praktik sebagai salah satu syarat kelulusan dari mahasiswa Departemen Aktuaria ITS.
3. Menambah relasi dengan berbagai pihak ketika sedang melaksanakan Kerja Praktik.
4. Membuat mahasiswa menjadi lulusan yang berkualitas dan profesional di bidang ilmu aktuaria yang dimiliki.
5. Mampu menghasilkan lulusan yang profesional dan berkualitas dalam bidang yang dikuasai serta dapat menjalin kerja sama yang baik antara lingkungan akademik dengan dunia kerja khususnya pada perusahaan pemerintah atau perusahaan yang bersangkutan.

1.4.3 Manfaat Bagi Departemen Aktuaria ITS

Berikut manfaat yang bisa didapatkan Departemen Aktuaria ITS dengan terlaksananya kuliah praktik ini:

1. Terjalinnnya kerjasama antara perusahaan PT Aplikanusa Lintasarta dan Departemen Aktuaria ITS.
2. Menghasilkan lulusan yang berkompetensi di dunia kerja dan dikenal baik oleh perusahaan PT Aplikanusa Lintasarta secara langsung.
3. Memberikan informasi penerapan ilmu aktuaria di dunia kerja yang dapat digunakan sebagai bahan penelitian .

BAB II GAMBARAN UMUM INSTANSI

2.1 PT Aplikanusa Lintasarta



Gambar 2.1 Logo PT Aplikanusa Lintasarta

PT Aplikanusa Lintasarta merupakan penyedia Komunikasi Data, Internet dan IT Services untuk berbagai sektor industri. Perusahaan tersebut dimiliki mayoritas oleh Indosat dan sisanya dimiliki oleh sebagian institusi seperti yayasan dan koperasi. Perusahaan ini menyediakan berbagai ragam layanan solusi teresterial (Wireline, Wireless) dan VSAT dengan berbagai platform seperti Clear Channel, Frame Relay dan IP dan juga layanan Internet dedicated serta Data center.

PT Aplikanusa Lintasarta, atau biasa dikenal sebagai Lintasarta, didirikan pada 4 April 1988, saat teknologi komunikasi sedang berkembang pesat. Misi awal didirikannya Lintasarta adalah untuk membangun industri perbankan nasional dalam bidang teknologi informasi.

Pada mulanya Lintasarta bekerja sama dengan Bank Indonesia dalam mengembangkan dan mengimplementasikan suatu sistem otomatis untuk Laporan Perbankan pada Maret 1989. Sebelumnya, semua bank memberikan laporannya (termasuk dari cabang-cabangnya diseluruh Indonesia) ke Bank Indonesia melalui jasa pos. Tentu saja hal ini memakan waktu yang tidak sedikit. Dengan sistem yang dibangun in, semua bank dapat mengirim laporannya ke Bank Indonesia secara elektronis.

Selanjutnya, Lintasarta juga mengembangkan jaringan Automated Teller Machine (ATM) yang menghubungkan banyak bank. Jaringan ATM antar bank yang pertama adalah ATM-Bersama, yang diluncurkan pada Mart 1990. Dengan sistem ini, setiap terminal ATM-Bersama yang dimiliki oleh suatu bank dapat digunakan oleh seluruh pelanggan bank lainnya yang termasuk dalam jaringan ATM-Bersama. Pendekatan yang unik ini membuat bank-bank mampu memperluas jaringannya secara bersama-sama dengan biaya relatif rendah.

2.2 Visi dan Misi PT Aplikanusa Lintasarta

Berdasarkan laman resmi PT Aplikanusa Lintasarta, visi dan misi yang dimiliki PT Aplikanusa Lintasarta adalah sebagai berikut:

2.2.1 Visi PT Aplikanusa Lintasarta

Visi dari PT Aplikanusa Lintasarta yaitu Menjadi pemimpin dalam solusi informasi dan komunikasi bisnis di Indonesia

2.2.2 Misi PT Aplikanusa Lintasarta

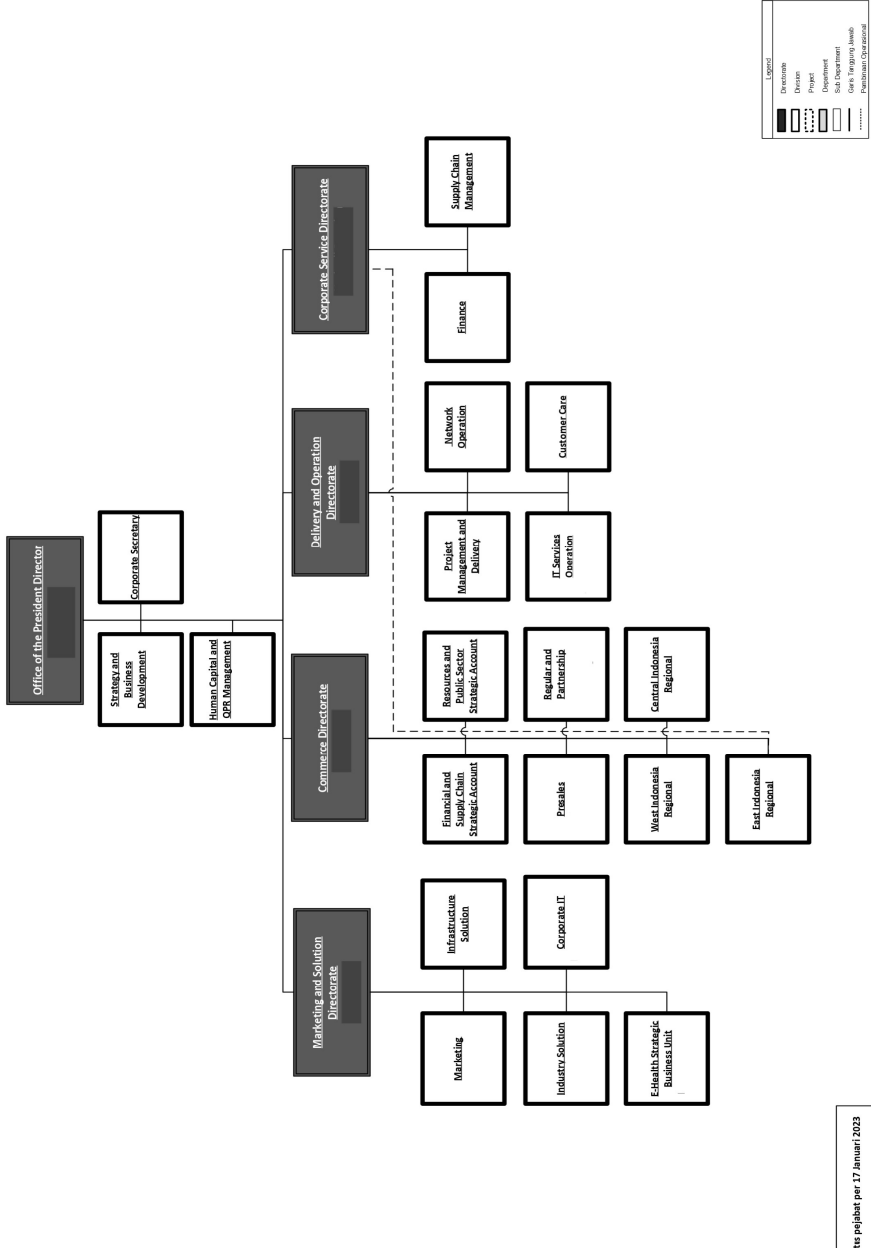
Misi dari PT Aplikanusa Lintasarta yaitu Membuat bisnis pelanggan lebih mudah dan bernilai tambah melalui solusi teknologi informasi dan komunikasi yang inovatif.

2.3 Struktur Organisasi PT Aplikanusa Lintasarta

PT Aplikanusa Lintasarta terdiri dari berbagai divisi dan subdivisi demi mencapai efektifitas dan efisiensi yang optimal. Bagan struktur organisasi PT Aplikanusa Lintasarta tersusun seperti pada halaman berikut:

STRUKTUR ORGANISASI PT. APLIKANUSA LINTASARTA

Lampiran 1
 Keputusan Direksi
 No : 016/DIREKSI/0000/2022
 Tgl Efektif : 1 Januari 2023



Legend:	
 	Directorate
 	Division
 	Project
 	Department
 	Unit
 	Core Business Unit
 	Partisipatif Operational

State pejabat per 17 Januari 2023

Gambar 2.2 Struktur Organisasi PT Aplikanusa Lintasarta

“Halaman ini sengaja dikosongkan.”

BAB III

PELAKSANAAN KERJA PRAKTIK

3.1 Pelaksanaan Kerja Praktik

Kerja Praktik berlangsung selama 1 (satu) bulan dengan rincian satu bulan pertama pelaksanaan kerja praktik bersama PT Aplikanusa Lintasarta dengan total 22 hari kerja. Adapun waktu dan tempat pelaksanaan kerja praktik adalah sebagai berikut.

Tanggal : 2 Januari 2024 – 2 Februari 2024
Tempat : Kantor Pusat PT Aplikanusa Lintasarta
Alamat : Menara Thamrin Lt.12, JL. MH. Thamrin, Jakarta Pusat, 10250

Selama kegiatan kerja praktik, penulis ditempatkan pada divisi *Finance* lebih tepatnya ditempatkan dalam tim pajak selama 23 hari kerja tersebut. Pada saat Kerja Praktik, diberikan penugasan seperti berikut:

1. Memproses data transaksi bukti potong untuk laporan keuangan mingguan.
2. Bekerja sama dengan pembimbing dalam automasi penginputan data menggunakan microsoft power automate demi meningkatkan efisiensi kerja

3.2 Metode Penyelesaian Tugas Khusus

Pada kegiatan Kerja Praktik di PT Aplikanusa Lintasarta kami diberikan oleh perusahaan. Data yang diperoleh telah didapat melalui persetujuan pembimbing lapangan pada kerja praktik dengan syarat bahwa nama pelanggan disensor. Data ini kemudian diolah membentuk model RFM yang kemudian akan dilakukan penelitian secara kuantitatif dengan metode analisis statistika deskriptif, *Elbows Method*, *Silhouette Coefficient Method*, dan *K-Means Clustering*.

3.2.1 Analisis RFM

Analisis yang umum digunakan dalam pengelompokan pelanggan adalah analisis model *Recency, Frequency, Monetary* (RFM). Model ini memaparkan pelanggan berdasarkan interval waktu kunjungan terakhir pelanggan, frekuensi kunjungan, dan besaran nilai yang dikeluarkan sebagai royalti perusahaan (Aggelis dkk., 2005 dan Chen dkk., 2009).

Meliput jurnal Cheng dan Chen(2009)dijelaskan bahwa semakin besar nilai R dan F maka kemungkinan pelanggan akan melakukan transaksi kembali dengan perusahaan tersebut. Selain itu semakin besar nilai M, maka kecenderungan pelanggan dalam memberikan respon kepada produk dan layanan perusahaan.

Penulis Handoyo (2023) juga mempublikasikan langkah-langkah RFM *scoring* yang mengkonversikan nilai normalisasi tiap variabel RFM menjadi suatu metrik *rating* dengan skala 0 sampai 1. RFM *score* dapat dihitung dengan pertama-tama mendapatkan nilai masing-masing variabel RFM yang sudah dinormalisasi. Nilai tiap variabel dinormalisasikan dengan rumus yang dituliskan sebagai berikut:

$$\hat{R} = 1 - \left(\frac{(R_i - R_{min})}{(R_{max} - R_{min})} \right) \quad (3.1)$$

$$\hat{F} = \left(\frac{(F_i - F_{min})}{(F_{max} - F_{min})} \right) \quad (3.2)$$

$$\hat{M} = \left(\frac{(M_i - M_{min})}{(M_{max} - M_{min})} \right) \quad (3.3)$$

Dimana :

R = Nilai variabel *recency*

F = Nilai variabel *frequency*

M = Nilai variabel *monetary*

\hat{R} = Nilai variabel *recency* Normalisasi

\hat{F} = Nilai variabel *frequency* Normalisasi

\hat{M} = Nilai variabel *monetary* Normalisasi

Kemudian nilai setiap variabel yang sudah dinormalisasi tersebut dikalikan dengan bobot penilaian khusus untuk setiap variabel. Nilai RFM *score* dapat dihitung dengan rumus sebagai berikut:

$$Score = (\hat{r}_i \times WR_j) + (\sum_{j=1}^n \hat{f}_{ij} \times WF_j) + (\sum_{j=1}^n \hat{m}_{ij} \times WM_j) \forall i, j \quad (3.4)$$

Dimana :

i = Nomor Pelanggan

j = Range waktu periode

WR_j = Bobot variabel *recency*

WF_j = Bobot variabel *frequency*

WM_j = Bobot variabel *monetary*

Nilai skor RFM yang didapat kemudian dapat dikategorikan menjadi 5 segmen. Indikator penilaian tiap segmen dapat dilihat pada tabel 3.1

Tabel 3.1 Indikator Penilaian RFM Score

Skor RFM	Keterangan
>0,7	Pelanggan Teratas
0,69 ≥ x > 5	Pelanggan Nilai Tinggi
0,49 ≥ x > 3	Pelanggan Nilai Sedang
0,29 ≥ x > 0,1	Pelanggan Nilai Rendah
≤0,09	Pelanggan Hilang

3.2.1 Analisis Statistika Deskriptif

Metode analisis deskriptif adalah statistik yang digunakan untuk menganalisis data dengan cara mendeskripsikan atau menggambarkan data yang telah terkumpul sebagaimana adanya tanpa bermaksud membuat kesimpulan yang berlaku untuk umum atau generalisasi (Soegiyono, 2011). Tujuan utama analisis statistik deskriptif adalah untuk memberikan gambaran mengenai variabel- variabel yang digunakan, seperti nilai minimum, nilai maksimum, rata-rata dan standar deviasi pada

masing-masing penelitian. Sehingga dengan penggambaran tersebut diharapkan dapat memberikan penjelasan secara umum mengenai variabel yang akan dianalisis agar pembaca lebih mudah untuk memahaminya.

3.2.3 IQR Method

Metode *Inter Quantile Range* merupakan metode penanganan *outlier*; Metode ini berguna untuk mengurangi pengaruh nilai ekstrim dalam sebuah data. Metode IQR lebih baik digunakan karena IQR hanya mempertimbangkan *absolute deviation* dari titik tengah yang menyebabkan ketahanannya atas asimetri data (Jones, 2019). Perhitungan IQR dilakukan dengan mengurangi nilai kuartil ketiga dengan kuartil pertama yang dituliskan sebagai berikut:

$$IQR = Q3 - Q1 \quad (3.5)$$

kemudian nilai IQR tersebut dikalikan dengan *threshold* untuk mendapatkan batas atas dan batas bawahnya. Berikut merupakan rumus *outlier* dari metode IQR:

$$|x_i| > \tilde{x} + \lambda IQR \quad (3.6)$$

Dimana :

x_i = *Outlier*

λ = Nilai *threshold*

3.2.4 Elbows Method

Elbows method adalah metode yang sering dipakai untuk menentukan jumlah *cluster* pada *k-means clustering*. Metode *elbows* merupakan salah satu metode untuk menentukan jumlah *cluster* yang tepat melalui persentase hasil perbandingan antara jumlah *cluster* yang akan membentuk siku pada suatu titik. Jika nilai *cluster* pertama dengan nilai *cluster* kedua memberikan sudut dalam grafik atau nilainya mengalami penurunan paling besar, maka jumlah nilai *cluster* tersebut yang tepat. Untuk mendapatkan perbandingannya adalah dengan menghitung *sum*

of square error (SSE) dari masing-masing nilai *cluster*. Karena semakin besar jumlah nilai *cluster* K , maka nilai SSE akan semakin kecil (Putu dkk., 2020). Berikut merupakan rumus dari SSE:

$$SSE = \sum_{K=1}^K \sum_{x_i \in S_k} ||x_i - c_k|| \quad (3.7)$$

Dimana :

K = *Cluster* ke- c

x_i = Jarak data objek ke- i

c_k = Pusat *Cluster* ke- i

3.2.5 Silhouette Coefficient Method

Silhouette coefficient merupakan metode yang digunakan untuk melihat kualitas dan kekuatannya dari *cluster*. Sehingga dapat terlihat seberapa baik atau buruknya suatu objek ditempatkan dalam suatu *cluster* tertentu. Metode *silhouette coefficient* merupakan gabungan dari dua metode yaitu metode *cohesion* yang berfungsi untuk mengukur seberapa dekat relasi antara objek dalam sebuah *cluster*, dan metode *separation* yang berfungsi untuk mengukur seberapa jauh sebuah *cluster* terpisah dengan *cluster* lain (Handoyo dkk., 2014). Untuk menghitung nilai *silhouette coefficient*, diperlukan perhitungan nilai *silhouette index* dari sebuah data ke- i . Nilai *silhouette coefficient* didapatkan dengan mencari nilai maksimal dari nilai *silhouette index* global dari jumlah *cluster* 2 sampai jumlah kluster $n-1$ seperti pada persamaan:

$$SC = \text{Max}_k SI(k) \quad (3.8)$$

Dimana :

SC = *Silhouette Coefficient*

SI = *Silhouette Index Global*

k = Jumlah *Cluster*

3.2.6 K-Means Clustering

Clustering merupakan proses mengelompokkan objek berdasarkan informasi yang diperoleh dari data untuk menjelaskan hubungan antar objek dengan tujuan memaksimalkan kesamaan antar anggota satu kelas. Tujuan dilakukannya *clustering* agar ditemukannya *cluster* yang berkualitas. Metode *clustering* dapat dibagi menjadi dua yaitu metode hirarki dan non hirarki atau disebut dengan algoritma *K-Means* (Somantri dkk., 2016). Metode hirarki akan membentuk sebuah tingkatan dari data yang digunakan. Sedangkan pada metode non hirarki jumlah *cluster* yang diinginkan ditentukan terlebih dahulu. Dengan menggunakan metode ini jumlah data yang digunakan dapat lebih besar daripada data yang digunakan pada metode hirarki.

Algoritma *K-means* merupakan metode *clustering* yang paling sederhana dan umum. Hal ini dikarenakan *K-means* mempunyai kemampuan mengelompokkan data dalam jumlah cukup besar dengan waktu komputasi yang cepat dan efisien (Khan & Ahmad, 2013). Suatu data menjadi anggota dari *cluster* ke-*K* apabila jarak data tersebut ke pusat *cluster* ke- *k* bernilai paling kecil jika dibandingkan dengan jarak ke pusat *cluster* lainnya. Selanjutnya, kelompokkan data-data yang menjadi anggota pada setiap *cluster*. Nilai pusat *cluster* yang baru dapat dihitung dengan cara mencari nilai rata-rata dari data-data yang menjadi anggota pada *cluster* tersebut dengan menggunakan rumus:

$$\mu_k = \frac{1}{N_k} \sum_{t=1}^{N_k} x_t \quad (3.9)$$

Dimana :

- μ_k = Titik centroid dari *cluster* ke-k
- N_k = Banyaknya data pada *cluster* ke-k
- x_t = data ke-t pada *cluster* ke-k

3.2.7 Uji Statistik

Dalam uji statistik ini dapat diketahui berdasarkan nilai F dan nilai signifikansi yang terdapat pada tabel ANOVA. Ketentuan penggunaan nilai F dalam analisis cluster adalah bahwa semakin besar nilai F maka semakin besar perbedaan antara ketiga cluster yang terbentuk. Nilai F didapatkan dari rumus :

$$F = \frac{\text{Between cluster mean}}{\text{within cluster mean}} \quad (3.10)$$

Begitu pula dengan nilai signifikansi, jika tingkat signifikansi $< \alpha_{0,05}$ maka semakin besar pula tingkat perbedaan antara ketiga cluster. Hipotesis uji signifikansi adalah sebagai berikut.

Hipotesis:

H_0 : Cluster tidak mempunyai perbedaan yang signifikan.

H_1 : Cluster mempunyai perbedaan yang signifikan.

Jika nilai signifikansi lebih besar dari $\alpha_{0,05}$ maka gagal tolak H_0 dan H_1 ditolak. Jika angka signifikan kurang dari $\alpha_{0,05}$ maka H_0 ditolak dan gagal tolak H_1 .

3.2.8 Sumber Data dan Variabel Penelitian

Variabel-variabel yang digunakan pada penelitian ini terdiri dari *Recency* dimana variabel ini menandakan rentan akhir transaksi dilakukan oleh pelanggan, *Frequency* yang merupakan jumlah transaksi dalam satu periode per pelanggan, dan *Monetary* yang merupakan nilai uang yang dikeluarkan pelanggan pada periode tersebut. Semua variabel ini akan dibentuk dari data kasir pajak PT Aplikanusa Lintasarta yang didapatkan dari persetujuan pembimbing perusahaan dan manajer divisi tax, dengan catatan nama pelanggan akan

disensor. Pada penelitian ini dibentuk variabel RFM dari data transaksi pelanggan dalam jangka waktu 4 tahun, Data yang digunakan untuk pembentukan variabel RFM terdapat dilihat pada Tabel 3.2

Tabel 3.2 Variabel Data untuk Analisis RFM

Variabel	Keterangan	Skala
DPP	Dasar Pemotongan Pajak adalah nilai pokok pemotongan pajak dari transaksi yang dilakukan dengan pelanggan dengan satuan ribuan rupiah	Rasio
<i>Invoice Date</i>	Tanggal <i>Invoice</i> adalah tanggal terjadinya transaksi yang dilakukan dengan pelanggan.	Interval

Variabel RFM yang didapat kemudian digunakan dalam proses segmentasi dimana data tersebut akan dijadikan alat ukur perilaku pelanggan untuk metode RFM *scoring* maupun *K-Means Clustering*. Struktur data yang digunakan pada proses segmentasi dijelaskan pada Tabel 3.3.

Tabel 3.3 Variabel Data untuk Segmentasi

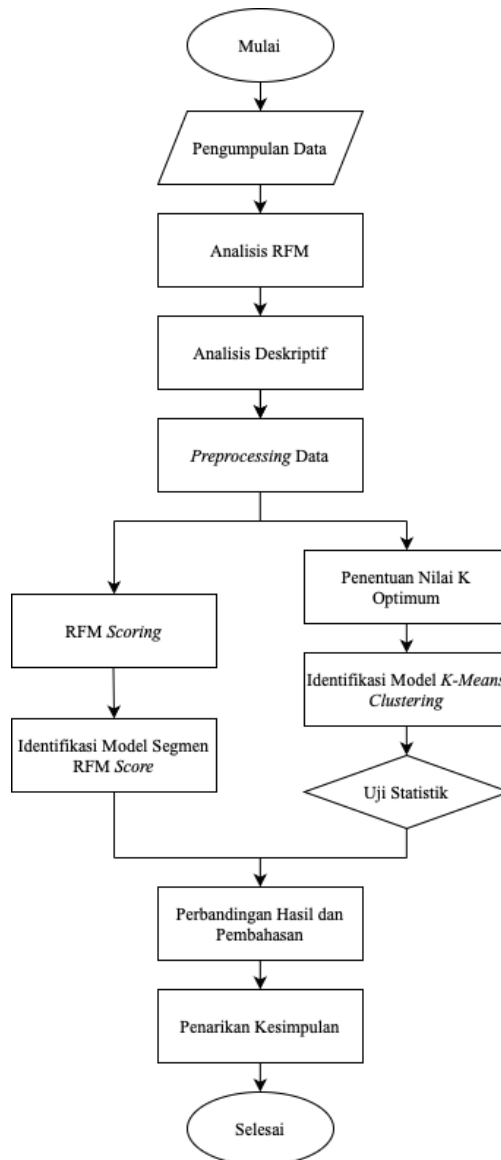
Variabel	Keterangan	Skala
Recency	Jarak waktu (bulanan) tanggal pembayaran pelanggan dengan tanggal titik terkini.	Rasio
<i>Frequency</i>	Jumlah transaksi yang dilakukan tiap pelanggan dalam rentan waktu penelitian	Rasio
<i>Monetary</i>	Total nilai transaksi yang dilakukan tiap pelanggan dalam rentan waktu penelitian	Rasio

3.2.9 Langkah Analisis

Langkah-langkah analisis yang dilakukan oleh peneliti pada penelitian dan laporan ini adalah

1. Melakukan studi literatur
2. Mengumpulkan data-data yang digunakan dalam analisis.
3. Melakukan Analisis RFM.
4. Menentukan karakteristik variabel pada data dengan melihat statistika deskriptif.
5. Melakukan *preprocessing* data
6. Melakukan segmentasi pelanggan menggunakan RFM *scoring*
7. Penentuan nilai K optimum untuk K-Means Clustering menggunakan *Elbows Method* dan *Silhouette Coefficient Method*.
8. Penerapan K-Means Clustering terhadap data berdasarkan jumlah K yang telah ditentukan.
9. Melakukan uji signifikansi pada *cluster* data
10. Membandingkan hasil segmen *K- Means* dengan RFM *score*.
11. Penarikan kesimpulan.

Dari langkah-langkah analisis yang dilakukan maka dapat dibentuk diagram alur untuk proses pengerjaan Laporan Kerja Praktik ini, seperti pada Gambar 3.1 berikut.



Gambar 3.1 Flowchart Analisis Data

BAB IV PELAKSANAAN KERJA PRAKTIK

4.1 Karakteristik Data

Data mentah yang berupa data pemotongan pajak dari transaksi para pelanggan dapat menjelaskan perilaku pengeluaran tiap pelanggan itu sendiri, mulai dari seberapa besar nilai nominal pembayaran, jumlah pembayaran, dan juga seberapa baru transaksi yang dilakukan tiap pelanggan. Semua perilaku tersebut dibentuk dari data transaksi yang diberikan untuk penelitian ini. Bentuk data mentah merupakan data input transaksi masuk yang dipaparkan sebagai berikut.

Tabel 4. 1 Data Transaksi Pelanggan

Transaksi ke-i	Tanggal	Pelanggan	No. Invoice	DPP(ribu)
1	2020-01-01	Customer_a	20/0008565	27.200
2	2020-01-01	Customer_b	20/0007703	1.000
3	2020-01-01	Customer_b	20/0005073	1.000
⋮	⋮	⋮	⋮	⋮
100.264	2024-02-28	Customer_cc	24/0001708	6.100
100.266	2024-02-28	Customer_cd	24/0008992	15.000

Data transaksi yang dipaparkan pada Tabel 4.1 memberikan informasi yang cukup jelas, dimana pembayaran tersebut dapat dianalisa dengan membuat analisis RFM. Variabel yang dibentuk dari analisis RFM seperti *recency*, *frequency*, dan *monetary* dapat mendeskripsikan perilaku dari masing-masing pelanggan yang bisa membantu untuk melakukan segmentasi dengan RFM *scoring* maupun *K-Means clustering*.

4.2 Analisis RFM

Analisis RFM dilakukan menggunakan python melalui *software* jupyter notebook dengan 100.266 data transaksi periode awal 2020 – awal 2024. Dalam analisis ini dibentuk tabel yang berisikan variabel penelitian dengan nilai *Frequency*, nilai *recency*, dan nilai *Monetary* dari setiap pelanggan.

Variabel RFM seperti nilai *recency* dihitung dengan menghitung jarak waktu (bulan) pembayaran terakhir setiap pelanggan dengan titik terkininya dimana pada penelitian ini yaitu bulan Februari 2024, nilai *frequency* dihitung dengan jumlah kuantitas transaksi yang dilakukan tiap pelanggan selama 4 tahun, dan *monetary* dengan menghitung total besaran pembayaran dari tiap pelanggan selama 4 tahun. Ketiga variabel RFM yang telah dihitung dapat dianalisa pada nilai yang dipaparkan pada Tabel 4.2.

Tabel 4.2 Analisis RFM Data Transaksi

Pelanggan	<i>Recency</i>	<i>Frequency</i>	<i>Monetary(ribu)</i>
Customer_1	9	1	30.000
Customer_2	1	143	1.582.000
Customer_3	0	14	49.000
⋮	⋮	⋮	⋮
Customer_3051	39	1	30.000
Customer_3052	1	20	114.000

Tabel 4.2 memaparkan nilai variabel RFM masing-masing pelanggan berdasarkan data transaksinya, seperti pada *Customer_1* yang memiliki nilai *recency* sebesar 9, nilai *frequency* sebesar 1 dan nilai *monetary* sebesar Rp 30.000.000 yang berarti *Customer_1* hanya pernah berkontrak 1 kali sebesar 30 juta, 9 bulan yang lalu, beda dengan *Customer_3* yang nilai *recency*, *frequency*, dan *monetary*-nya masing-masing 0, 14, dan 49 juta rupiah yang dapat diartikan *Customer_3* baru saja

memperpanjang kontraknya, dengan total kontrak sebanyak 14 kali yang bernilai 49 juta rupiah. Persebaran data variabel RFM tiap pelanggan dapat diamati lebih lanjut sifatnya dengan melakukan analisis statistika deskriptif.

4.3 Analisis Deskriptif

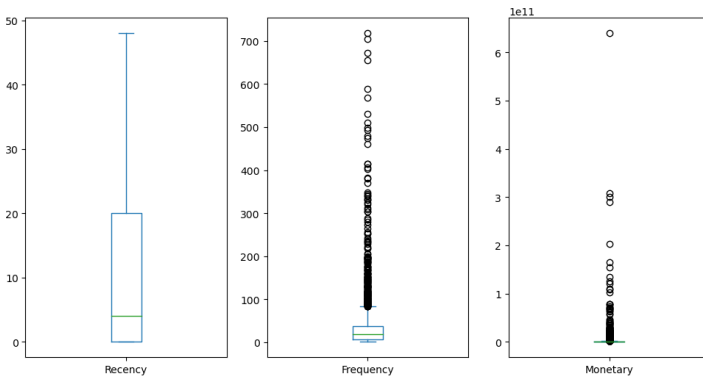
Analisis statistika deskriptif dilakukan dengan python menggunakan *software* Jupyter Lab dengan variabel RFM yang telah dibentuk. Pada analisis deskriptif yang dilakukan, didapatkan nilai minimum, maksimum, rata-rata untuk masing-masing variabel kontinu.

Tabel 4.3 Analisis Deskriptif Data RFM

<i>Variable</i>	<i>N</i>	<i>Mean</i>	<i>Min</i>	<i>Max</i>	<i>var</i>
<i>Recency</i>	3.052	10,722805	0,00000	48,0000	150,07
<i>Frequency</i>	3.052	32,693971	1,00000	718,0000	3.170,3
<i>Monetary</i>	3.052	25.740(juta)	0,0668(juta)	639.875(juta)	3.12×10^{20}

Tabel 4.3 menunjukkan terdapat data 3052 pelanggan dengan besaran *mean*, *minimum* serta *maximum* yang bervariasi. Salah satunya diketahui untuk nilai maksimum *recency* sebesar 48. Hal ini menunjukkan bahwa terdapat pelanggan yang sudah tidak lagi melakukan transaksi selama 48 bulan terakhir. Nilai tertinggi *frequency* berada pada 718 yang mengimplikasikan pelanggan tersebut memiliki lebih dari satu kontrak yang dibayarkan per-termin, dan dapat dilihat rata-rata nilai *monetary* sebesar 25 milyar dengan nilai paling kecil 668 ribu hingga 639 milyar.

perbedaan yang jauh ini diperjelas dengan nilai varians yang tinggi yang menandakan bahwa ada dispersi tingkat tinggi antar perilaku pembayaran tiap pelanggan. Hal ini juga memberikan petunjuk kemungkinan adanya *outlier* pada data, maka dari itu persebaran data pun divisualisasikan dengan menggunakan *box plot* pada Gambar 4.1.

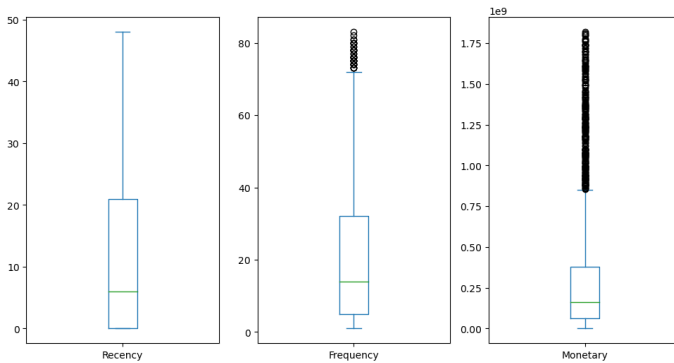


Gambar 4.1 Box Plot Variabel *Recency*, *Frequency*, dan *Monetary*

Dari Gambar 4.1 terlihat data variabel *recency* tersebar tanpa adanya *outlier*. Namun dapat dilihat juga bahwa terdapat *outlier* pada persebaran data variabel *frequency* dan *monetary*. *Outlier* pada data ditandai dengan titik-titik yang berjarak secara signifikan dari titik yang lainnya. Keberadaan *outlier* ini dapat mempengaruhi model *clustering* secara signifikan, maka baiknya dilakukan penanganan sebelum dilakukannya proses K-Means *clustering*.

4.4 Preprocessing Data

Outlier pada data dapat memberi pengaruh pada model *clustering* secara signifikan dikarenakan metode K-Means *Clustering* cenderung sensitif terhadap keberadaan outlier, Maka dilakukan penanganan terlebih dahulu menggunakan metode Interquartile Range(IQR) untuk mengurangi pelanggan dengan variabel RFM yang ekstrim. Setelah *outlier* pada data sudah ditangani, maka dilakukan lagi visualisasi terhadap persebaran data variabel RFM menggunakan *box plot* pada Gambar 4.2.



Gambar 4.2 Box Plot Variabel *Recency*, *Frequency*, dan *Monetary* tanpa outlier

Dari Gambar 4.2 terlihat persebaran titik lebih rapat tanpa adanya jarak yang signifikan antar satu titik ke titik lainnya. Data yang sudah dibersihkan dapat diamati lebih lanjut pada tabel statistika deskriptif berikut.

Tabel 4.4 Data RFM Tanpa *Outlier*

<i>Variable</i>	<i>N</i>	<i>Mean</i>	<i>Min</i>	<i>Max</i>	<i>var</i>
<i>Recency</i>	2.512	11,758360	0,00000	48,0000	158,22
<i>Frequency</i>	2.512	19,422373	1,00000	83,0000	293,52
<i>Monetary</i>	2.512	302,04(juta)	0,0668(juta)	1.818,18(juta)	$1,36 \times 10^{20}$

Dapat dilihat dari tabel 4.4 didapatkan pengurangan yang signifikan pada nilai *maximum* variabel *frequency* dan *monetary*. Nilai varians pada variabel *frequency* dan *monetary* berkurang secara signifikan dimana terjadi sebaliknya pada nilai varians pada variabel *recency* naik menjadi 158,22.

Data awal sebelum dilakukan penanganan outlier berjumlah 3.052, maka berdasarkan informasi pada Tabel 4.4 dapat diimplikasikan bahwa dihapuskan outlier pada *frequency* dan *monetary* sebanyak 540 data. Data bersih tanpa outlier adalah sebanyak 2.512 data, dan data yang sudah dibersihkan ini akan digunakan untuk proses segmentasi selanjutnya.

4.5 Segmentasi RFM Scoring

Data RFM yang terbentuk kemudian dinormalisasi untuk menghitung RFM *score* tiap pelanggan. Perhitungan skor dihitung dengan pembobotan 20% untuk *recency*, 40% untuk *frequency*, dan 40% untuk *monetary* menggunakan rumus 3.4. Nilai skor yang terhitung kemudian ditampilkan untuk peninjauan segmen pelanggannya.

Tabel 4.5 Tabel RFM dengan Skor

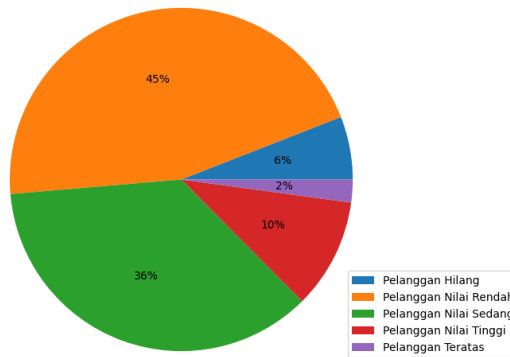
Pelanggan	<i>Recency</i>	<i>Frequency</i>	<i>Monetary(ribu)</i>	RFM <i>score</i>
Customer_1	9	1	30.000	0,169086
Customer_2	1	143	1.582.420	0,274180
Customer_3	0	14	40.000	0,283624
⋮	⋮	⋮	⋮	⋮
Customer_2511	39	1	30.000	0,044086
Customer_2512	1	20	114.000	0,313582

Dapat dilihat pada Tabel 4.5 telah terhitung skor RFM tiap pelanggan, skor ini kemudian digunakan untuk mengklasifikasi tiap pelanggan berdasarkan skor RFM-nya. Seperti pada Customer_1 didapatkan nilai *monetary* sebesar 30 juta, *frequency* sebesar 1 dan *recency* di angka 9 yang memberikan pelanggan ini skor RFM sebesar 0,169 dikarenakan angka *recency*, *frequency*, *monetary* yang relatif rendah. Nilai skor RFM yang telah terhitung kemudian dikonversikan menurut ketentuan yang tertulis pada Tabel 3.1 dan hasil segmentasinya disajikan pada Tabel 4.6.

Tabel 4.6 Tabel RFM dengan Klasifikasi

Pelanggan	<i>Recency</i>	<i>Frequency</i>	<i>Monetary(ribu)</i>	Segmen
Customer_1	9	1	30.000	Pelanggan Nilai Rendah
Customer_11	0	70	849.468	Pelanggan Teratas
Customer_33	0	47	1.206.476	Pelanggan Nilai Tinggi
⋮	⋮	⋮	⋮	⋮
Customer_2511	39	1	30.000	Pelanggan Hilang
Customer_2512	1	20	114.000	Pelanggan Nilai Sedang

Tabel 4.6 memaparkan klasifikasi beberapa pelanggan berdasarkan skor RFM yang telah dihitung. Pada Customer_1 terdapat nilai *recency* sebesar 9, *frequency* 1, dan *monetary* sebesar 30 juta, dapat dilihat juga pada Tabel 4.5 sebelumnya Customer_1 yang sama mendapatkan RFM *score* sebesar 0,169086 sehingga pelanggan tersebut masuk ke segmen Pelanggan Nilai Rendah. Terdapat 2512 pelanggan dengan proporsi tiap segmennya digambarkan menggunakan *pie chart* yang bisa dilihat pada gambar 4.3.



Gambar 4.3 Pie Chart Segmentasi RFM scoring

Populasi pelanggan terbagi dari yang paling banyak Pelanggan Nilai Rendah sebanyak 45%, Pelanggan Nilai Sedang sebanyak 36%, Pelanggan Nilai tinggi sebanyak 10%, Pelanggan nilai hilang sebanyak 6%, dan 2% sisanya merupakan pelanggan teratas. Proporsi ini diperjelas dengan jumlah pelanggan tiap segmen yang disajikan pada tabel berikut.

Tabel 4.7 Jumlah Pelanggan dan Rata-Rata Variabel Setiap Segmen

Pelanggan	<i>N</i>	<i>Recency</i>	<i>Frequency</i>	<i>Monetary(ribu)</i>
Hilang	149	37,03	2,47	29.857
Nilai Rendah	1.141	16,83	8,39	124.459
Nilai Sedang	907	4,33	27,18	350.736
Nilai Tinggi	261	3,18	41,61	862.674
Teratas	54	0,98	61,50	1.277.931

Didapat dari Tabel 4.7 terdapat 149 pelanggan yang masuk ke segmen Pelanggan Hilang, 1141 pelanggan ke segmen Nilai Rendah, 907 pelanggan ke segmen Nilai Sedang, 261 ke segmen Nilai Tinggi, dan sisanya ke segmen Teratas sebanyak 54

pelanggan. Didapat juga nilai rata-rata *frequency* dan *monetary* yang semakin tinggi seiring dengan naiknya kelas segmen dan kebalikannya untuk variabel *recency* dimana pelanggan teratas memiliki rata-rata *recency* terkecil sebesar 0.98 yang berarti mayoritas populasi segmen Pelanggan Teratas merupakan pelanggan yang rutin membayar atau baru saja membayarkan kontraknya.

4.6 Penentuan Nilai K

Nilai K merupakan jumlah kelompok atau *cluster* data yang akan dibentuk. Penentuan nilai K yang optimal dapat dilakukan menggunakan metode elbows dan metode silhouette, kedua metode tersebut sebaiknya digunakan bersamaan dalam menentukan hasil K yang paling optimal, Namun kedua metode ini sangat terpengaruh akan perbedaan skala antar variabel.

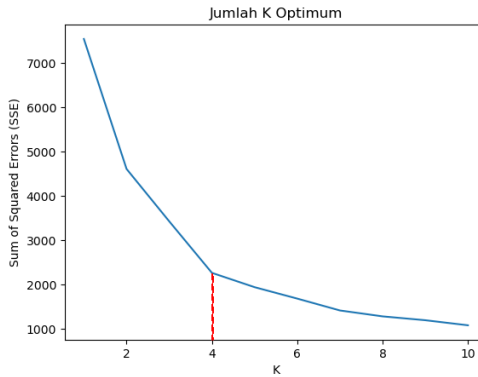
Dapat dilihat kembali pada Tabel 4.2 bahwa nilai *range* antar variabel berada di satuan yang berbeda, hal ini juga dapat mempengaruhi hasil *clustering*, maka sebelum itu dilakukan normalisasi terhadap semua variabel terlebih dahulu. Proses *rescaling* dilakukan menggunakan *library* StandardScaler pada python dengan hasil yang disajikan pada tabel berikut.

Tabel 4.8 Nilai RFM setelah *rescaling*

<i>Recency</i>	<i>Frequency</i>	<i>Monetary</i>
-0.219332	-1.075494	-0.735921
-0.934971	-0.316557	-0.684523
0.178245	0.384000	-0.570908
1.768554	-0.725215	-0.735840
1.370977	-0.199797	-0.300395
⋮	⋮	⋮
2.166131	-1.07549	-0.73592
-0.85546	0.033722	-0.50869

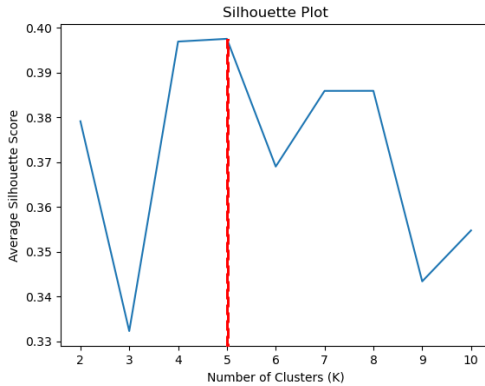
Dari tabel 4.8 dapat dilihat *range* nilai antar variabel sudah seragam setelah dilakukannya *rescaling*. Hal ini berguna agar hasil *clustering* tidak bias dikarenakan satuan variabel yang berbeda dan bisa dilanjutkan ke penentuan jumlah K.

Kedua metode penentuan nilai K optimal dilakukan menggunakan K-Means dari *library* sklearn dan divisualisasikan menggunakan *library* matplotlib dengan *random_state=42* untuk memastikan *output* yang konsisten.



Gambar 4.4 Grafik *Elbows Method*

Dari gambar 4.4 dapat dilihat titik penurunan terbesar berada pada jumlah $K = 4$, namun untuk memastikan dilakukan lagi pengujian lebih lanjut menggunakan *Silhouette Method*.



Gambar 4.5 Grafik *Sillhouette Plot*

Dari gambar 4.6 dapat diketahui *silhouette score* tertinggi berada pada jumlah $K = 5$ yang menunjukkan jumlah K paling optimal, hal ini diperjelas pada tabel 4.5 berikut.

Tabel 4.9 Nilai *Sillhouette Score*

K	<i>Silhouette Score</i>
2	0.379129
3	0.332253
4	0.396929
5	0.397542
6	0.368999
7	0.385906
8	0.385919

Tabel 4.9 memeperjelas bahwa nilai *silhouette score* tertinggi berada di jumlah $K = 5$ dengan nilai sebesar 0.397542, maka dari itu dapat disimpulkan bahwa jumlah *cluster* terbaik adalah 5.

4.7 Analisis Signifikansi Cluster

Uji signifikansi dilakukan pada variabel-variabel RFM tiap *cluster* sebagai alat ukur performa *cluster* yang dihasilkan. Uji signifikansi bertujuan untuk memastikan bahwa terdapat perbedaan yang signifikan pada variabel antar *cluster*.

Uji signifikansi *cluster* dilakukan menggunakan tabel ANOVA dengan python menggunakan *library* stasmodels. Tabel anova akan memaparkan nilai F dan juga *p-value* setiap *cluster* dimana semakin besar nilai F maka semakin besar perbedaan tiap *cluster* dan dimana *p-value* lebih kecil dari $\alpha_{0,05}$ maka bisa disimpulkan bahwa terdapat perbedaan yang signifikan antar *cluster*.

Tabel 4.10 Tabel ANOVA

	F	P-Value
<i>Recency</i>	1921.291864	0.0
<i>Frequency</i>	1567.486208	0.0
<i>Monetary</i>	1983.153043	0.0

Didapat dari Tabel 4.10 ketiga variabel *recency*, *frequency*, dan *monetary* yang memiliki *p-value* $< \alpha_{0,05}$ hingga ketiga fitur tersebut berhasil tolak h_0 yang menyimpulkan *cluster* memiliki perbedaan *recency*, *frequency* dan *monetary* yang cukup signifikan untuk membedakan antar *cluster*. Nilai F terbesar juga diperoleh fitur *recency* yang menyimpulkan variabel *recency* merupakan variabel utama yang menunjukkan perbedaan antar pelanggan.

4.8 Analisis Cluster K-Means

Terdapat lima *cluster* yang terbentuk dengan metode *K-Means Clustering*. Masing-masing *cluster* kemudian dipaparkan menggunakan nilai *centroid* masing-masing fitur dari tiap *cluster*.

Tabel 4.11 Nilai *Centroid Cluster*

<i>Cluster</i>	<i>Centroid</i>		
	<i>recency</i>	<i>frequency</i>	<i>monetary</i>
0	-0.79226461	1.71977546	2.18020764
1	0.2263241	-0.4808347	2.24855526
2	-0.42138271	-0.43296921	-0.44621916
3	-0.78536798	1.09803659	-0.01717165
4	1.32745113	-0.73083009	-0.46631614

Didapatkan nilai *centroid* dari setiap *cluster* seperti pada Tabel 4.11 dan dilakukan intepretasi pada setiap *cluster*, serta dilakukan pencocokkan *cluster* terhadap *label* klasifikasi RFM :

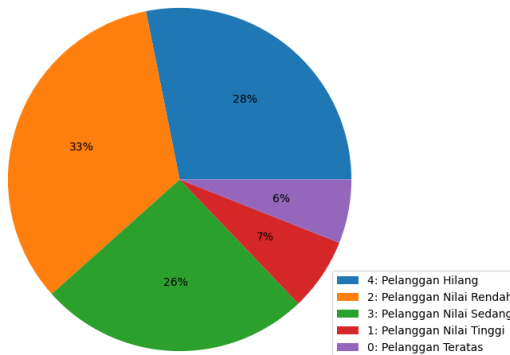
- *Cluster 0*: Pelanggan Teratas
adalah kelompok pelanggan yang memiliki nilai *recency* rendah, frekuensi pembelian tinggi, dan nilai moneter relatif tinggi. Mereka adalah pelanggan yang sering memperpanjang kontrak memiliki banyak kontrak dengan nilai transaksi per-kontrak yang besar.
- *Cluster 1*: Pelanggan Nilai Tinggi
adalah kelompok pelanggan yang memiliki nilai *recency* rendah, frekuensi pembelian rendah, dan nilai moneter tinggi. Mereka adalah pelanggan yang melakukan pembelian dengan nilai transaksi besar, tetapi tidak memiliki banyak kontrak atau membayar kontrak per-termin.
- *Cluster 2*: Pelanggan nilai Rendah
adalah kelompok pelanggan yang memiliki nilai *recency* rendah, frekuensi pembelian rendah, dan nilai moneter rendah. Mereka adalah pelanggan yang tidak terlalu aktif dan nilai transaksi per-kontraknya juga kecil.
- *Cluster 3*: Pelanggan Nilai Sedang

adalah kelompok pelanggan yang memiliki nilai *recency* rendah, frekuensi pembelian tinggi, dan nilai moneter rendah. Mereka adalah pelanggan yang aktif dan memiliki banyak kontrak yang nilai transaksinya relatif kecil.

- *Cluster 4: Pelanggan Hilang*

adalah kelompok pelanggan yang memiliki nilai *recency* tinggi, frekuensi pembelian rendah, dan nilai moneter rendah. Mereka adalah pelanggan yang sudah lama tidak memperpanjang kontrak dan nilai transaksi per-kontraknya kecil.

Lima *cluster* yang telah dibentuk ini dapat diobservasi lebih lanjut dengan dengan melihat proporsi setiap *cluster*-nya seperti pada gambar 4.7.



Gambar 4.6 Pie Chart K-Means Clustering

Proporsi *cluster* terbagi dari yang paling banyak *cluster 4* sebesar 30%, *cluster 1* sebanyak 29%, *cluster 0* sebanyak 26%,

cluster 2 sebanyak 10%, dan 5% sisanya adalah *cluster 3*. Nilai proporsi ini dapat dilihat dengan lebih jelas pada tabel berikut.

Tabel 4.12 Jumlah Pelanggan dan Rata-Rata Variabel Setiap Segmen

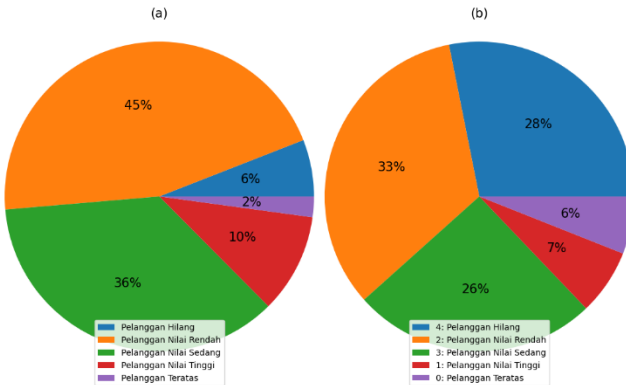
Pelanggan	<i>N</i>	<i>Recency</i>	<i>Frequency</i>	<i>Monetary(ribu)</i>
Hilang	716	28.453	6.904	129664
Nilai Rendah	764	6.459	12.006	137093
Nilai Sedang	662	1.881	38.231	295698
Nilai Tinggi	241	14.605	11.186	1133266
Teratas	129	1.795	48.881	1108000

Didapat dari Tabel 4.12 terdapat 716 pelanggan yang masuk ke segmen Pelanggan Hilang, 764 pelanggan ke segmen Nilai Rendah, 662 pelanggan ke segmen Nilai Sedang, 241 ke segmen Nilai Tinggi, dan sisanya ke segmen Teratas sebanyak 129 pelanggan.

Proporsi yang dihasilkan metode K-Means bisa dibilang berbeda dengan proporsi klasifikasi RFM *scoring*, maka akan dilakukan perbandingan hasil segmentasi lebih lanjut dari dua metode tersebut.

4.9 Perbandingan Hasil K-Means dan RFM *Scoring*

Hasil *cluster* dari metode *K-Means Clustering* akan dibandingkan dengan hasil segmentasi dari RFM *scoring* yang sudah ditangani *outliernya* menggunakan metode IQR, hal ini dilakukan agar jumlah sampel perbandingan sama dan sebanding. Proporsi segmen dari kedua metode dibandingkan menggunakan *Pie Chart*, dimana label pada *cluster* K-means akan di-*label* ulang dengan *label* RFM *scoring* yang sudah dicocokkan.



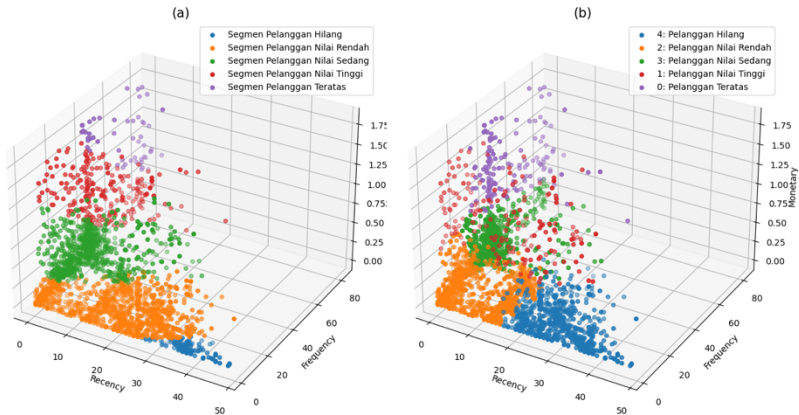
Gambar 4.7 Pie Chart hasil segmentasi dari RFM *score* (a), K-Means (b)

Dapat dilihat perbedaan proporsi masing-masing metode pada Gambar 4.7, diketahui bahwa perbedaan perbedaan paling besar ada pada jumlah Pelanggan Hilang dimana segmentasi hasil *K-Means Clustering* menghasilkan porsi sebesar 28% dibandingkan RFM *scoring* sebesar 6%. Persebaran Pelanggan Nilai Sedang hingga Pelanggan Teratas memiliki urutan porsi yang identik dimana Pelanggan Nilai Sedang lebih banyak dibandingkan Pelanggan Nilai Tinggi. Porsi paling kecil pada segmentasi dari kedua metode tetap sama, yaitu Pelanggan Teratas sebesar 2% dari RFM *scoring* dan 5% dari *K-Means clustering*, dengan pemaparan lebih jelas pada tabel jumlah pelanggan tiap segmen.

Tabel 4.13 Jumlah Pelanggan Tiap Segmen Dari Kedua Metode

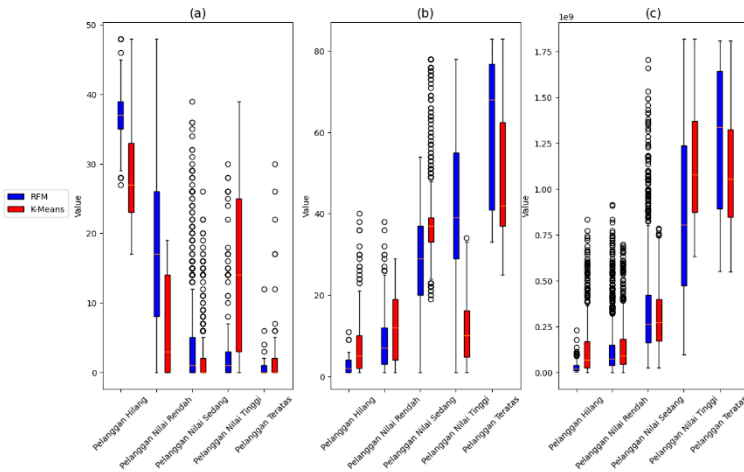
	<i>Scoring</i>	K-Means
Pelanggan Hilang	149	716
Pelanggan Nilai Rendah	1141	764
Pelanggan Nilai Sedang	907	662
Pelanggan Nilai Tinggi	261	241
Pelanggan Teratas	54	129

Berdasarkan Tabel 4.13 jumlah pelanggan tiap segmen hasil *K-Means Clustering* lebih banyak pada segmen Pelanggan Hilang dan Pelanggan Teratas, namun segmen sisanya memiliki jumlah lebih kecil dibandingkan hasil *RFM Scoring*. Distribusi pembagian pelanggan tiap segmen dapat divisualisasikan lagi menggunakan grafik 3D *scatterplot*.



Gambar 4.8 3D scatterplot hasil segmentasi RFM score tanpa outlier (a), K-Means (b)

Gambar 4.8 memaparkan perbedaan pada hasil segmentasi dimana dengan metode *RFM scoring* pembagian lebih simetris dibandingkan hasil *K-Means Clustering*, namun pembagian yang simetris tersebut belum tentu mengelompokkan pelanggan dengan perilaku pembayaran yang sama. Dapat dilihat lagi pada segmen pelanggan hilang dimana segmen hasil *RFM scoring* hanya mengelompokkan pelanggan dengan nilai *recency* sekitar 30 keatas dikelompokkan sebagai Pelanggan Hilang dan Metode *K-Means clustering* mengelompokkan pelanggan dengan nilai *recency* sekitar 20 keatas, untuk menangkap lebih jelas lagi perilaku pelanggan tiap segmen, dibuatkan *boxplot* perbandingan untuk visualisasi.



Gambar 4.9 Box Plot Perbandingan Segmentasi Recency(a), Frequency(b), Monetary(c)

Berdasarkan informasi yang dipaparkan pada Gambar 4.9 hasil segmentasi pelanggan oleh metode RFM *scoring* terdapat banyak tumpang tindih antar segmen pada nilai *recency* pada *boxplot* (a), terutama pada segmen Pelanggan Hilang dan Pelanggan Nilai Rendah dengan beberapa pelanggan dengan nilai *recency* yang tinggi masuk ke kelompok Pelanggan Nilai Rendah dimana informasi ini cukup krusial untuk mendefinisikan pelanggan yang sudah tidak lagi memberikan *profit* pada perusahaan.

Namun untuk *recency* segmen Pelanggan Nilai Tinggi dan Pelanggan Teratas hasil segmentasi *scoring* lebih optimal dikarenakan hanya terisi oleh pelanggan dengan nilai *recency* dibawah 5 yang berarti segmen tersebut berisikan pelanggan

dengan rentan pembayaran terakhir dari september 2023 hingga februari 2024.

Nilai *frequency* antar segmen dapat dilihat pada *box plot* (b) dimana *box* segmen hasil *scoring* terlihat saling tumpang tindih dibanding dengan segmen hasil *clustering* yang berisikan nilai *frequency* yang lebih variatif, hal ini menandakan bahwa metode *K-Means Clustering* berhasil mensegmentasi pelanggan dengan perilaku pembayaran yang berbeda.

Nilai *monetary* antar segmen pelanggan hasil RFM *scoring* lebih variatif daripada segmen *K-Means Clustering* dimana banyak terjadi tumpang tindih pada *box* antar Pelanggan Hilang dan Pelanggan Nilai Rendah dan antara Pelanggan Nilai Tinggi dan Pelanggan Teratas dengan nilai *median*-nya yang juga bertumpang tindih.

Nilai yang tumpang tindih tersebut juga dapat dipengaruhi oleh *frequency* pembayaran tiap segmen pada hasil RFM *scoring*, dimana semakin tinggi kuantitas pembayaran semakin tinggi juga nilai *monetary* yang didapat. Hal ini diperjelas pada *box plot* (b) dan (c) dimana nilai *monetary* tiap segmen hasil *scoring* naik bersamaan dengan nilai *frequency* yang semakin tinggi, berbeda halnya dengan segmen hasil *K-Means* dimana nilai *frequency* tiap *cluster* berbeda, yang mengindikasikan nilai *monetary* pada segmen yang dihasilkan tidak berkorelasi dengan hasil kuantitas pembayaran yang lebih tinggi.

“Halaman ini sengaja dikosongkan.”

BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan analisis yang telah dilakukan, segmen pelanggan yang dihasilkan oleh *K-Means Clustering* lebih bervariasi daripada segmen RFM *scoring*. Segmen Hasil *K-Means Clustering* juga berhasil mengelompokkan pelanggan dengan pembayaran besar namun frekuensi pembayaran yang lebih sedikit yang tidak berkorelasi dengan kuantitas pembayaran pelanggan tersebut.

Segmen Pelanggan Hilang yang didapat dari *K-Means Clustering* juga lebih akurat tanpa adanya pelanggan dengan tanggal pembayaran terakhir yang sudah terlalu lama masuk ke segmen Pelanggan Nilai Rendah. Namun terdapat beberapa Pelanggan Nilai Tinggi hasil *K-Means Clustering* dengan rentan pembayaran yang sudah lama, yang berarti pelanggan tersebut belum lagi melakukan pembayaran tetapi dimasukkan ke segmen tersebut karena kuantitas pembayaran dan *profitability* dari pelanggan tersebut.

Dari hasil penelitian ini dapat disimpulkan bahwa metode *K-Means Clustering* lebih baik dalam mengelompokkan pelanggan berdasarkan perilaku dan *profitability* masing-masing pelanggan, dibandingkan dengan RFM *scoring* yang mengelompokkan pelanggan secara bertahap dengan pertimbangan kuantitas pembayaran yang mengecualikan pelanggan dengan *profitability* besar namun kuantitas pembayaran kecil.

5.2 Saran

Berdasarkan hasil penelitian pada kegiatan Kerja Praktik ini, terdapat beberapa saran yang dapat ditujukan untuk penelitian lebih lanjut.

1. Pengerjaan Kerja Praktik memiliki keterbatasan pada data yang digunakan. Oleh karena itu, pada penelitian selanjutnya dapat melengkapi data pada tahun tertentu untuk mendapatkan hasil segmentasi yang lebih baik.
2. Hasil segmentasi yang didapatkan memiliki nilai *silhouette coefficient* kurang dari nol karena sifat *skewness* dari data, hal ini dapat ditangani dengan penggunaan metode analisis clustering yang lebih robust seperti Fuzzy C-Means, DBScan, *Gaussian Mixture Models*, dll.
3. Segmentasi RFM *scoring* pelanggan dapat diulangi lagi menggunakan bobot RFM yang lebih sesuai mengikuti keinginan dan objektif masing-masing perusahaan sebagai poin perbandingan antar metode.

DAFTAR PUSTAKA

- Aggelis, V., & Christodoulakis, D. (2005). *Customer Clustering using RFM analysis*.
- Cheng, C. H., & Chen, Y. S. (2009). Classifying the segmentation of customer value via RFM model and RS theory. *Expert Systems with Applications*, 36(3 PART 1), 4176–4184. <https://doi.org/10.1016/j.eswa.2008.04.003>
- Handoyo, R., Rumani M, R., & Nasution, S. M. (2014). *PERBANDINGAN METODE CLUSTERING MENGGUNAKAN METODE SINGLE LINKAGE DAN K-MEANS PADA PENGELOMPOKAN DOKUMEN* (Vol. 15, Nomor 2).
- Jones, P. R. (2019). A note on detecting statistical outliers in psychophysical data. Dalam *Attention, Perception, and Psychophysics* (Vol. 81, Nomor 5, hlm. 1189–1196). Springer New York LLC. <https://doi.org/10.3758/s13414-019-01726-3>
- Khan, S. S., & Ahmad, A. (2013). Cluster center initialization algorithm for K-modes clustering. *Expert Systems with Applications*, 40(18), 7444–7456. <https://doi.org/10.1016/j.eswa.2013.07.002>
- Putu, N., Merliana, E., & Santoso, A. J. (2020). *ANALISA PENENTUAN JUMLAH CLUSTER TERBAIK PADA METODE K-MEANS CLUSTERING*.
- Soegiyono. (2011). *Metode Penelitian Kuantitatif, Kualitatif dan R&D*.
- Somantri, O., Wiyono, S., Teknik Informatika, J., & Harapan Bersama Tegal, P. (2016). Metode K-Means untuk Optimasi Klasifikasi Tema Tugas Akhir Mahasiswa Menggunakan Support Vector Machine (SVM). *Scientific Journal of Informatics*, 3(1). <http://journal.unnes.ac.id/nju/index.php/sji>

“Halaman ini sengaja dikosongkan.”

LAMPIRAN

Lampiran 1 Surat Penerimaan Kerja Praktik dari Perusahaan



Wed, Jan 3, 9:34 AM



Informasi PKL Lintasarta inbox x

Hermi Tri Hartanti <hermi.hartanti@lintasarta.co.id>
to me, Sayid, Putri, Annisa, Rekrutmen ▾

Translate to English

X

Dear **Natanael Hadi Muliyabanta**,

Berdasarkan permohonan kerja praktik yang Anda ajukan, kami menyatakan bahwa Anda dapat mulai aktif melakukan kerja praktik di Lintasarta pada:

Periode : 02 Januari 2024 – 02 Februari 2024

Lokasi Kerja : WFA & WFO - Kantor Lintasarta, Gu Arcadia, Menara Thamrin (Jl. M.H. Thamrin, RT.2/RW.1, Kb. Silih, Kec. Menteng, Kota Jakarta Pusat 10250)
Pembimbing : Mba Putri Christiani

Silahkan menghubungi Mba Putri, selaku pembimbing lapangan Anda ybs untuk mengetahui lebih lanjut terkait tugas selama periode magang.

Mohon agar bersedia mengirimkan kembali form data diri beserta surat pernyataan terlampir yang sudah dilengkapi melalui e-mail dan mengirimkan laporan kehadiran yang sudah dilengkapi setiap awal bulan kepada kami.

Jika ada yang kurang jelas silahkan ditanyakan yaaa.

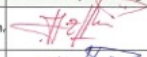



Demikian disampaikan, terma kasih 🙏

Regards,












Tanti

Hermi Tri Hartanti,
HUMAN CAPITAL & QPR MANAGEMENT

Lampiran 2 Form F-5

		PROGRAM STUDI SARJANA, DEPARTEMEN AKTUIARIA, FSAD-ITS <i>Undergraduate Program, Department of Actuarial Science, FSAD-ITS</i>		F-5						
BUKTI BIMBINGAN LAPORAN KERJA PRAKTIK <i>Evidence of Practical Work Supervising at The Company</i>										
KP-51-05	2024	Kode/code: SA234702	SKS/Credit: 2 SKS							
Nama/Name : <u>Nasanael Haer Mulikbanta</u> NRP/Student Identity Number : <u>502620109</u> Nama Instansi/Company Name : <u>PT Aplanusa Lingsarera</u> Unit Kerja/Work Unit : <u>Tax - Finance</u> Nama Pembimbing/Supervisor Name : <u>Ibu Ulil Azmi, S.Si, M.Si</u> Periode KP di Perusahaan/Time Period of PW : <u>2 Januari - 2 Februari 2024</u>										
No	Tanggal Date	Materi yang dibahas Proposal Component Discussion	Tanda tangan Dosen Pembimbing Lecturer Supervisor Sign							
1	24 Juni 2024	Kelompok bridging tiap sebelum tabel dan gambar, Variabel, penciltop, penerapan program, boxplot pada visualisasi perbandingan, recall rumus								
2	28 Juni 2024	formasi.								
3	1 Juli 2024	Revisi Minor & Bimbingan Poster								
4										
5										
6										
7										
Surabaya, 1 Juli 2024 Dosen Pembimbing KP I PW Lecturer Supervisor I  (Ibu Ulil Azmi, S.Si, M.Si) NIP. 1990 2019 12 069		Surabaya, 20 Dosen Pembimbing KP II PW Lecturer Supervisor II (.....) NIP.								
Form F-5 merupakan form bukti bahwa mahasiswa/i telah melakukan pembimbingan selama pembuatan laporan KP kepada dosen pembimbing KP. Form ini harus dilampirkan pada laporan KP. Proses pembelajaran di Departemen Aktuaria ITS meliputi: perkuliahan, Kerja praktik, dan Tugas Akhir. Berikut adalah beberapa dokumen yang digunakan pada proses Kerja Praktik, yaitu: 1) SOP KP (SOP), 2) Pedoman, 3) Formulir pengajuan Surat Permohonan KP (F-1), 4) Surat permohonan KP di Perusahaan (F-2), 5) Surat balasan dari perusahaan (F-3), 6) Formulir rekaman kegiatan (F-4, F-5, F-6), 7) Formulir penilaian (F-7, F-8, dan F-9). <i>The learning process in the Department of Actuarial Science ITS includes: Lectures, Practical Work (PW), and Final Project (FP). There are some documents in the process of PW, i.e.: 1) SOP of PW (SOP), 2) Manual, 3) Form of filing request letter (F-1), 4) Letter of PW request to the Company (F-2), 5) Letter reply from the company (F-3), 6) Form of recording activities (F-4, F-5, F-6), 7) Form of assessment (F-7, F-8, and F-9).</i>										
F-1	F-2	F-3	F-4	F-5	F-6	F-7	F-8	F-9		
SOP of PW	Practical work Reporting form	Form of filing request to the company	Letter of PW request to the company	Letter reply from the company	Rekaman supervising form	Form report supervising form	Activity form in the company	Form Company assessment form	Assessment report form	Simulation Questionnaire

Lampiran 3 Form F-6

 KP-S1-06		PROGRAM STUDI SARJANA, DEPARTEMEN AKTUARIA, FSAD-ITS Undergraduate Program, Department of Actuarial Science, FSAD-ITS			F-6
		BUKTI KEGIATAN DI PERUSAHAAN Evidence of Activity at The Company			
2024		Kode/code: SA234702	SKS/Credit: 2 SKS		
Nama/Name		Nataaniel Hadi Muliapanta			
NRP/Student Identity Number		500620107			
Nama Instansi/Company Name		PT APLIMANUS LOMBATA			
Unit Kerja/Work Unit		Tax Finance			
Nama Pembimbing/Supervisor Name		Puri Christiana			
Periode KP di Perusahaan/Time Period of PW		2 Januari - 2 Februari 2024			
No	Tanggal Date	Jam Kerja Working time		Kegiatan Activity	Tanda Tangan Pembimbing Lapangan CSS*)
		Mulai Start	Selesai Finish		
1	2 Jan	7:30	16:30	Pengenalan prosedur rekap bukti menggunakan power automate microsoft	
2	3 Jan	8:30	17:30	merekap bukti potong menggunakan microsoft power automate	
3	4 Jan	8:30	17:30	merekap bukti potong menggunakan microsoft power automate	
4	5 Jan	8:00	17:00	Filing bukti potong per pelanggan	
5	6 Jan	9:00	19:00	mempariasiapkan data untuk financial weekly reports	
6	9 Jan	8:30	17:30	Receive bukti potong pada oracle EBI	
7	10 Jan	9:00	17:00	Peningkatan efisiensi automate & rekap bukti potong	
8	11 Jan	8:30	17:30	Peningkatan efisiensi automate & rekap bukti potong	
9	12 Jan	8:30	17:30	Receive bukap uncollected EBI	
10	15 Jan	9:00	17:00	Receive bukap uncollected EBI	

11	16 Jan	8:30	17:30	Receive buypor di EBS	fb
12	17 Jan	8:00	17:04	reunap buypor otomatis	fb
13	18 Jan	8:30	17:30	Receive buypor di EBS	fb
14	18 Jan	8:30	17:30	Receive buypor di EBS	fb
15	22 Jan	9:00	17:00	reunap buypor dgn otomatis	fb
16	23 Jan	8:30	17:30	Receive buypor wils	fb
17	24 Jan	8:00	17:00	reunap buypor dgn otomatis	fb
18	25 Jan	8:50	17:30	receive buypor wils	fb
19	26 Jan	8:30	17:30	receive buypor wils	fb
20	29 Jan	8:00	17:00	reunap buypor dgn otomatis	fb
21	30 Jan	8:30	17:30	Receive buypor wils	fb
22	31 Jan	8:00	17:00	Reunap buypor dgn otomatis	fb
23	1 Feb	8:30	17:30	Receive buypor wils	fb
24	2 Feb	8:30	17:30	Receive buypor wils	fb

Catatan/Note: Salin berkas ini jika diperlukan/Copy this form if needed.
 *1 Tanda tangan pembimbing lapangan/CS: Company Supervisor Signature
 Setiap paraf harus disertai dengan stempel perusahaan/Each initial sign should be stamped with the company stamp.

Form F-6 merupakan bukti bahwa mahasiswa telah melaksanakan kerja praktik di perusahaan dan telah melakukan pembimbingan dengan pembimbing perusahaan. Form ini harus dilampirkan pada laporan KP.

Proses pembelajaran di Departemen Aktuaria ITS meliputi: perkuliahan, Kerja praktik, dan Tugas Akhir. Berikut adalah beberapa dokumen yang digunakan pada proses Kerja Praktik, yaitu: 1) SOP KP (SOP), 2) Pedoman, 3) Formulir peninjauan Surat Permohonan KP (F-1), 4) Surat permohonan KP di Perusahaan (F-2), 5) Surat balasan dari perusahaan (F-3), 6) Formulir rekaman kegiatan (F-4, F-5, F-6), 7) Formulir penilaian (F-7, F-8, dan F-9).


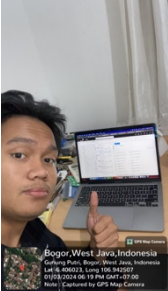

The learning process in the Department of Actuarial Science ITS includes: lectures, Practical Work (PW), and Final Project (FP). There are some documents in the process of PW, i.e.: 1) SOP of PW (SOP), 2) Manual, 3) Form of filing request letter PW (F-1), 4) Letter of PW request to the Company (F-2), 5) Letter reply from the company (F-3), 6) Form of recording activities (F-4, F-5, F-6), 7) Form of assessment (F-7, F-8, and F-9).

Jakarta, 02 Februari 2024
 Mengetahui,
 Pimpinan Perusahaan
 Company Leaders




[Signature]
 (...Syaiful Malik A.)
 NIP. 92302342





F-4	F-5	F-6	F-7	F-8	F-9	F-10	F-11	F-12	F-13	
SOP of PW	Practical Work report writing form	Form of filing request to the company	Letter of PW request to the company	Letter reply from the company	PW proposal/ supporting form	Activity/Item in the company	Supervising form	PW company assessment form	Assessment of report/Item	Supervisor Questionnaire




Lampiran 4 Bukti Kerja Praktik Menggunakan GPS Map Camera

Tanggal	Bukti
2 Januari 2024	
3 Januari 2024	
4 Januari 2024	

<p>5 Januari 2024</p>	
<p>8 Januari 2024</p>	
<p>9 Januari 2024</p>	
<p>10 Januari 2024</p>	

11 Januari 2024	 <p>Central Jakarta, Jakarta, Indonesia Jalan MH. Thamrin, Menteng, Central Jakarta, 10130, Jakarta, Indonesia Lat. S: 10.3317, Long: 106.822794 01/11/2024 08:29 AM GMT+07:00 Note: Captured by GPS Map Camera</p>
12 Januari 2024	 <p>Central Jakarta, Jakarta, Indonesia Jalan MH. Thamrin, Menteng, Central Jakarta, 10130, Jakarta, Indonesia Lat. S: 10.3316, Long: 106.822808 01/12/2024 03:45 PM GMT+07:00 Note: Captured by GPS Map Camera</p>
15 Januari 2024	 <p>Central Jakarta, Jakarta, Indonesia Jalan MH. Thamrin, Menteng, Central Jakarta, 10130, Jakarta, Indonesia Lat. S: 10.3316, Long: 106.822808 01/15/2024 10:01 AM GMT+07:00 Note: Captured by GPS Map Camera</p>

<p>16 Januari 2024</p>	
<p>17 Januari 2024</p>	
<p>19 Januari 2024</p>	
<p>22 Januari 2024</p>	

<p>23 Januari 2024</p>	
<p>24 Januari 2024</p>	
<p>25 Januari 2024</p>	

<p>26 Januari 2024</p>	 <p>Central Jakarta, Jakarta, Indonesia Jalan MH. Thamrin, Menteng, Central Jakarta, 10130, Jakarta, Indonesia Lat: 6.183374, Long: 106.822793 01/26/2024 08:10 AM GMT+07:00 Photo: Captured by XPERIA Map Camera</p>
<p>29 Januari 2024</p>	 <p>Bogor, West Java, Indonesia Gangway Rudi, Bogor, West Java, Indonesia Lat: 6.592213, Long: 106.883371 01/29/2024 02:35 PM GMT+07:00 Photo: Captured by XPERIA Map Camera</p>
<p>30 Januari 2024</p>	 <p>Central Jakarta, Jakarta, Indonesia Taman Duren, Duren Kaya, Menteng, Central Jakarta, Jakarta, Indonesia Lat: 6.185675, Long: 106.822793 01/30/2024 02:00 PM GMT+07:00 Photo: Captured by XPERIA Map Camera</p>
<p>1 Februari 2024</p>	 <p>Central Jakarta, Jakarta, Indonesia Jalan MH. Thamrin, Menteng, Central Jakarta, 10130, Jakarta, Indonesia Lat: 6.183374, Long: 106.822793 02/01/2024 03:19 PM GMT+07:00 Photo: Captured by XPERIA Map Camera</p>

2 Februari 2024



Lampiran 5 Syntax Python

Analisis RFM
<pre>import pandas as pd import datetime as dt import numpy as np import datetime import matplotlib.pyplot as plt import seaborn as sns import scipy.cluster.hierarchy as sch import statsmodels.api as sm import warnings warnings.filterwarnings("ignore")</pre>
<pre>df=pd.read_excel("/Users/natanmuliabanta/Downloads/DATA KP FINAL.xlsx") df.head()</pre>
<pre>#membuat tabel recency recency_df = df.groupby(by='Customer Name', as_index=False)['Invoice Date'].max() recency_df.columns = ['Pelanggan', 'Pembayaran Terakhir'] #menentukan titik terkini untuk recency recent_date = recency_df['Pembayaran Terakhir'].max() #menghitung recency berdasarkan bulan pembayaran recency_df['Recency'] = recency_df['Pembayaran Terakhir'].apply(lambda x: (recent_date.year - x.year) * 12 + recent_date.month - x.month) recency_df.head()</pre>
<pre>#menentukan frequency tiap pelanggan dari frekuensi pembayaran frequency_df = df.drop_duplicates().groupby(by=['Customer Name'], as_index=False)['Invoice Date'].count() frequency_df.columns = ['Pelanggan', 'Frequency'] frequency_df.head()</pre>

<pre>#menentukan monetary tiap pelanggan dari besaran pembayaran (DPP) monetary_df = df.groupby(by='Customer Name', as_index=False)['DPP'].sum() monetary_df.columns = ['Pelanggan', 'Monetary'] monetary_df.head()</pre>
<pre>#menyatukan Recency, Frequency, dan Monetary dalam satu tabel rf_df = recency_df.merge(frequency_df, on='Pelanggan') rfm_df = rf_df.merge(monetary_df, on='Pelanggan').drop(columns='Pembayaran Terakhir') rfm_df.head()</pre>
<pre>#save data ke csv untuk penelitian lebih lanjut rfm_df.to_csv('/Users/natanmuliabanta/Desktop/KP/PENELITIAN/ Data Final/rfm_data.csv', index=False)</pre>
<p>Analisis Deskriptif</p> <pre>rfm_describe_withvar = rfm_df[['Recency', 'Frequency', 'Monetary']].agg(['count', 'mean', 'std', 'var', 'min', 'median', 'max']) print(rfm_describe_withvar)</pre>
<pre>#Membuat boxplot persebaran data rfm_df.plot(kind='box', subplots=True, layout=(1, 3), figsize=(12, 6)) plt.show()</pre>
<p>Preprocessing</p> <pre>#IQR untuk frequency Q1_frequency = rfm_df['Frequency'].quantile(0.25) Q3_frequency = rfm_df['Frequency'].quantile(0.75) IQR_frequency = Q3_frequency - Q1_frequency lower_bound_frequency = Q1_frequency - 1.5 * IQR_frequency upper_bound_frequency = Q3_frequency + 1.5 * IQR_frequency #IQR untuk monetary Q1_monetary = rfm_df['Monetary'].quantile(0.25) Q3_monetary = rfm_df['Monetary'].quantile(0.75)</pre>

<pre> IQR_monetary = Q3_monetary - Q1_monetary lower_bound_monetary = Q1_monetary - 1.5 * IQR_monetary upper_bound_monetary = Q3_monetary + 1.5 * IQR_monetary outliers = rfm_df[(rfm_df['Frequency'] < lower_bound_frequency) (rfm_df['Frequency'] > upper_bound_frequency) (rfm_df['Monetary'] < lower_bound_monetary) (rfm_df['Monetary'] > upper_bound_monetary)] #drop row yang memiliki outlier rfm_df_clean = rfm_df.drop(outliers.index) print(rfm_df_clean.shape) </pre>
<pre> #Boxplot data tanpa outlier rfm_df_clean.plot(kind='box', subplots=True, layout=(1, 3), figsize=(12, 6)) plt.show() </pre>
<pre> #Statistika Deskriptif data tanpa outlier rfmclean_describe_withvar = rfm_df_clean[['Recency', 'Frequency', 'Monetary']].agg(['count', 'mean', 'std', 'var', 'min', 'median', 'max']) print(rfmclean_describe_withvar) </pre>
<pre> #save data bebas outlier ke csv untuk penelitian lebih lanjut rfm_df_clean.to_csv('/Users/natanmuliabanta/Desktop/KP/PENELI TIAN/Data Final/rfm_data_clean.csv', index=False) </pre>
<p>RFM Scoring</p> <pre> from sklearn.preprocessing import MinMaxScaler # Set kolom yang akan dinormalisasi (Recency, Frequency, Monetary) cols_to_scale = ['Recency', 'Frequency', 'Monetary'] # Normalisasi menggunakan MinMax scaler = MinMaxScaler() </pre>

```

# Normalisasi nilai variabel berdasarkan nilai tiap kolomnya untuk
perhitungan skor
rfm_df_norm = rfm_df_clean.copy()
for col in cols_to_scale:
    rfm_df_norm[[col]] = scaler.fit_transform(rfm_df_clean[[col]])

#Transformasi nilai Recency 1-x dikarenakan lebih kecil recency
lebih baik
rfm_df_norm['Recency'] = rfm_df_norm['Recency'].apply(lambda
x: 1 - x)

rfm_df_norm.head()

#Menghitung skor dengan bobot 20%(R) 40%(F) 40%(M)
rfm_df_norm['RFM_Score'] = (rfm_df_norm['Recency'] * 0.2) +
(rfm_df_norm['Frequency'] * 0.4) + (rfm_df_norm['Monetary'] *
0.4)
rfm_df_norm.head()

#Konversi skor RFM dengan segmen Pelanggan
rfm_df_segmen = rfm_df_norm.copy()
rfm_df_segmen["Segmen"] =
np.where(rfm_df_segmen['RFM_Score'] >
        0.69, "Pelanggan Teratas",
        (np.where(
            rfm_df_segmen['RFM_Score'] > 0.49,
            "Pelanggan Nilai Tinggi",
            (np.where(
                rfm_df_segmen['RFM_Score'] > 0.29,
                "Pelanggan Nilai Sedang",
                np.where(rfm_df_segmen['RFM_Score'] > 0.09,
                    'Pelanggan Nilai Rendah', 'Pelanggan Hilang'))))))

rfm_df_segmen.drop(columns=['RFM_Score'], inplace=True)
rfm_df_segmen.head()

segmentasi_rfm = pd.merge(rfm_df_clean,
rfm_df_segmen[['Segmen']], left_index=True, right_index=True)

```

<pre>segmentasi_rfm.head()</pre>
<pre>#save RFM Model ke excel untuk penelitian lebih lanjut segmentasi_rfm.to_excel('/Users/natanmuliabanta/Desktop/KP/PE NELITIAN/Data Final/RFM Model.xlsx', index=False)</pre>
<pre>#Pembuatan Pie Chart Segmen RFM counts = segmentasi_rfm.Segmen.value_counts() ordinal_order = ["Pelanggan Hilang", "Pelanggan Nilai Rendah", "Pelanggan Nilai Sedang", "Pelanggan Nilai Tinggi", "Pelanggan Teratas"] # Reorder the counts series according to the ordinal order sorted_counts = counts.reindex(ordinal_order) plt.figure(figsize=(7, 7)) plt.pie(sorted_counts, autopct='%0.0f%%') plt.legend(ordinal_order, loc="upper right", bbox_to_anchor=(1.3, 0.3)) plt.show()</pre>
<p>Penentuan Jumlah K Optimal</p>
<pre>#Standarisasi Data tanpa outlier from sklearn.preprocessing import StandardScaler scaler = StandardScaler() rfm_df_clean_scaled = scaler.fit_transform(rfm_df_clean[['Recency', 'Frequency', 'Monetary']]) rfm_df_clean_scaled = pd.DataFrame(rfm_df_clean_scaled, columns=['Recency', 'Frequency', 'Monetary']) rfm_df_clean_scaled.head()</pre>
<pre>from sklearn.cluster import KMeans import matplotlib.pyplot as plt K_values = range(1, 11) SSE_values = [] for K in K_values:</pre>

<pre> kmeans = KMeans(n_clusters=K, random_state=42) kmeans.fit(rfm_df_clean_scaled) SSE_values.append(kmeans.inertia_) # Membuat plot elbow plt.plot(K_values, SSE_values) plt.xlabel('K') plt.ylabel('Sum of Squared Errors (SSE)') plt.title('Jumlah K Optimum') plt.show() </pre>
<pre> from sklearn.metrics import silhouette_score k_values = [2, 3, 4, 5, 6, 7, 8, 9, 10] silhouette_scores = [] for k in k_values: kmeans = KMeans(n_clusters=k, random_state=42) cluster_labels = kmeans.fit_predict(rfm_df_clean_scaled) silhouette_avg = silhouette_score(rfm_df_clean_scaled, cluster_labels) silhouette_scores.append(silhouette_avg) # Membuat plot silhouette score plt.plot(k_values, silhouette_scores) plt.xlabel('Number of Clusters (K)') plt.ylabel('Average Silhouette Score') plt.title('Silhouette Plot') plt.show() </pre>
<pre> # Calculate silhouette scores for different values of K silhouette_scores = [] for k in range(2, 9): kmeans = KMeans(n_clusters=k, random_state=42) cluster_labels = kmeans.fit_predict(rfm_df_clean_scaled) silhouette_avg = silhouette_score(rfm_df_clean_scaled, cluster_labels) </pre>

<pre> silhouette_scores.append((k, silhouette_avg)) # tabel silhouette score silhouette_table = pd.DataFrame(silhouette_scores, columns=['K', 'Silhouette Score']) print(silhouette_table) </pre>
<p>K-Means Clustering</p> <pre> # Clustering dengan jumlah K=5 kmeans = KMeans(n_clusters=5, random_state=42) kmeans.fit(rfm_df_clean_scaled) labels = kmeans.labels_ # Membuat dataframe baru dengan cluster masing-masing pelanggan rfm_df_clean_cluster = rfm_df_clean.copy() rfm_df_clean_cluster['cluster'] = labels rfm_df_clean_cluster.head() </pre>
<pre> # Print nilai centroid tiap variabel dari masing-masing cluster kmeans.cluster_centers_ </pre>
<pre> #konversi segmen K-means berdasarkan centroid segmentasi_kmeans = rfm_df_clean_cluster.copy() cluster_labels = {0: "0: Pelanggan Teratas", 1: "1: Pelanggan Nilai Tinggi", 3: "3: Pelanggan Nilai Sedang", 2: "2: Pelanggan Nilai Rendah", 4: "4: Pelanggan Hilang"} segmentasi_kmeans['Cluster'] = segmentasi_kmeans['cluster'].map(cluster_labels) segmentasi_kmeans.head() </pre>
<pre> #Membuat pie chart segmen K-means counts = segmentasi_kmeans.Cluster.value_counts() ordinal_order = ["4: Pelanggan Hilang", "2: Pelanggan Nilai Rendah", "3: Pelanggan Nilai Sedang", "1: Pelanggan Nilai Tinggi", "0: Pelanggan Teratas"] </pre>

```
# Reorder the counts series according to the ordinal order
sorted_counts = counts.reindex(ordinal_order)

plt.figure(figsize=(7, 7))
plt.pie(sorted_counts, autopct='%0.0f%%')
plt.legend(ordinal_order, loc="upper right", bbox_to_anchor=(1.3,
0.3))
plt.show()
```

Uji Signifikansi

```
from statsmodels.formula.api import ols
#Uji ANOVA tiap Cluster
model = ols('Recency ~ C(cluster)',
data=rfm_df_clean_cluster).fit()
print(sm.stats.anova_lm(model, typ=2))
```

```
model = ols('Frequency ~ C(cluster)',
data=rfm_df_clean_cluster).fit()
print(sm.stats.anova_lm(model, typ=2))
```

```
model = ols('Monetary ~ C(cluster)',
data=rfm_df_clean_cluster).fit()
print(sm.stats.anova_lm(model, typ=2))
```

Perbandingan hasil segmentasi

```
#Pie Chart Perbandingan
import matplotlib.pyplot as plt

# Plot 1
counts1 = segmentasi_rfm.Segmen.value_counts()
ordinal_order1 = ["Pelanggan Hilang", "Pelanggan Nilai Rendah",
"Pelanggan Nilai Sedang", "Pelanggan Nilai Tinggi", "Pelanggan
Teratas"]
sorted_counts1 = counts1.reindex(ordinal_order1)

# Plot 2
counts2 = segmentasi_kmeans.Cluster.value_counts()
```

```

ordinal_order2 = ["4: Pelanggan Hilang", "2: Pelanggan Nilai
Rendah", "3: Pelanggan Nilai Sedang", "1: Pelanggan Nilai
Tinggi", "0: Pelanggan Teratas"]
sorted_counts2 = counts2.reindex(ordinal_order2)

# Create subplots
fig, (ax1, ax2) = plt.subplots(1, 2, figsize=(12, 7))

# Plot 1
ax1.pie(sorted_counts1, autopct='%0.0f%%', textprops={'fontsize':
15})
ax1.legend(ordinal_order1, loc="lower center")
ax1.set_title("(a)", fontsize=15, y=0.95)

# Plot 2
ax2.pie(sorted_counts2, autopct='%0.0f%%', textprops={'fontsize':
15})
ax2.legend(ordinal_order2, loc="lower center")
ax2.set_title("(b)", fontsize=15, y=0.95)

plt.subplots_adjust(wspace=-0.6)
plt.tight_layout()
plt.show()

```

```

#3D scatterplot perbandingan
# Plot 1
ordinal_order1 = ["Pelanggan Hilang", "Pelanggan Nilai Rendah",
"Pelanggan Nilai Sedang", "Pelanggan Nilai Tinggi", "Pelanggan
Teratas"]

fig, (ax1, ax2) = plt.subplots(1, 2, figsize=(13, 10),
subplot_kw={'projection': '3d'})

for Segmen in ordinal_order1:
    Segmen_data = segmentasi_rfm[segmentasi_rfm['Segmen'] ==
Segmen]

```



```

ax1.scatter(Segmen_data['Recency'], Segmen_data['Frequency'],
Segmen_data['Monetary'], label=f'Segmen {Segmen}')

ax1.set_xlabel('Recency')
ax1.tick_params(axis='x', pad=7)
ax1.set_ylabel('Frequency')
ax1.tick_params(axis='y', pad=7)
ax1.set_zlabel('Monetary', ha='center')
ax1.tick_params(axis='z', pad=7)
ax1.set_title('(a)', fontsize=15)
ax1.legend()

# Plot 2
ordinal_order2 = ["4: Pelanggan Hilang", "2: Pelanggan Nilai
Rendah", "3: Pelanggan Nilai Sedang", "1: Pelanggan Nilai
Tinggi", "0: Pelanggan Teratas"]

for cluster in ordinal_order2:
    cluster_data = segmentasi_kmeans[segmentasi_kmeans['Cluster']
== cluster]
    ax2.scatter(cluster_data['Recency'], cluster_data['Frequency'],
cluster_data['Monetary'], label=cluster)

ax2.set_xlabel('Recency')
ax2.tick_params(axis='x', pad=7)
ax2.set_ylabel('Frequency')
ax2.tick_params(axis='y', pad=7)
ax2.set_zlabel('Monetary', ha='center')
ax2.tick_params(axis='z', pad=7)
ax2.set_title('(b)', fontsize=15)
ax2.legend()

plt.tight_layout()
plt.show()
#BoxPlot Perbandingan

```

```

# Plot 1
fig, (ax1, ax2, ax3) = plt.subplots(1, 3, figsize=(14, 8))

ordinal_order1 = ["Pelanggan Hilang", "Pelanggan Nilai Rendah",
"Pelanggan Nilai Sedang", "Pelanggan Nilai Tinggi", "Pelanggan
Teratas"]
ordinal_order2 = ["4: Pelanggan Hilang", "2: Pelanggan Nilai
Rendah", "3: Pelanggan Nilai Sedang", "1: Pelanggan Nilai
Tinggi", "0: Pelanggan Teratas"]

for i, (Segmen, cluster) in enumerate(zip(ordinal_order1,
ordinal_order2)):
    Segmen_data = segmentasi_rfm[segmentasi_rfm['Segmen'] ==
Segmen]
    cluster_data = segmentasi_kmeans[segmentasi_kmeans['Cluster']
== cluster]

    bp1 = ax1.boxplot([Segmen_data['Recency']], positions=[i],
patch_artist=True, boxprops={'facecolor':'blue'})
    bp2 = ax1.boxplot([cluster_data['Recency']], positions=[i+0.3],
patch_artist=True, boxprops={'facecolor':'red'})

    bp3 = ax2.boxplot([Segmen_data['Frequency']], positions=[i],
patch_artist=True, boxprops={'facecolor':'blue'})
    bp4 = ax2.boxplot([cluster_data['Frequency']], positions=[i+0.3],
patch_artist=True, boxprops={'facecolor':'red'})

    bp5 = ax3.boxplot([Segmen_data['Monetary']], positions=[i],
patch_artist=True, boxprops={'facecolor':'blue'})
    bp6 = ax3.boxplot([cluster_data['Monetary']], positions=[i+0.3],
patch_artist=True, boxprops={'facecolor':'red'})

ax1.set_title('(a)', fontsize=15)
ax2.set_title('(b)', fontsize=15)
ax3.set_title('(c)', fontsize=15)
ax1.set_ylabel('Value')

```

```
ax2.set_ylabel('Value')
ax3.set_ylabel('Value')
ax1.set_xticks(range(len(ordinal_order1)))
ax2.set_xticks(range(len(ordinal_order1)))
ax3.set_xticks(range(len(ordinal_order1)))
ax1.set_xticklabels(ordinal_order1, rotation=45)
ax2.set_xticklabels(ordinal_order1, rotation=45)
ax3.set_xticklabels(ordinal_order1, rotation=45)

# Add legend
fig.legend([bp1['boxes'][0], bp2['boxes'][0]], ['RFM', 'K-Means'],
           fontsize=15, loc='center', bbox_to_anchor=(0.03, 0.5))

plt.show()
```

