

PENERAPAN *BUSINESS INTELLIGENCE* TERHADAP DATA PELANGGAN  
MENGUNAKAN METODE *CLUSTERING* MULTIDIMENSI UNTUK  
MENENTUKAN SEGMENTASI PELANGGAN  
(STUDI KASUS: PT GAMA INOVASI BERDIKARI)



SKRIPSI

Diajukan sebagai salah satu syarat untuk menyelesaikan  
Pendidikan Diploma Empat (D-4) Program Studi Teknik Komputer dan Jaringan  
Jurusan Teknik Elektro  
Politeknik Negeri Ujung Pandang

HAIRATUNNISA

425 19 034

PROGRAM STUDI D-4 TEKNIK KOMPUTER DAN JARINGAN  
JURUSAN TEKNIK ELEKTRO  
POLITEKNIK NEGERI UJUNG PANDANG  
MAKASSAR

2023

## HALAMAN PENGESAHAN

Skripsi dengan judul “PENERAPAN *BUSINESS INTELLIGENCE* TERHADAP DATA PELANGGAN MENGGUNAKAN METODE *CLUSTERING* MULTIDIMENSI UNTUK MENENTUKAN SEGMENTASI PELANGGAN (STUDI KASUS: PT GAMA INOVASI BERDIKARI)” oleh HAIRATUNNISA Nomor Induk Mahasiswa 425 19 034 telah diterima dan disahkan sebagai salah satu syarat untuk memperoleh gelar Diploma IV (D-4/S1 Terapan) pada Program Studi Teknik Komputer dan Jaringan Jurusan Teknik Elektro Politeknik Negeri Ujung Pandang.

Makassar, 16 Agustus 2023

### Mengesahkan,

Pembimbing I,



Zawiyah Saharuna, S.T., M.Eng.  
NIP. 198309032014042001

Pembimbing II,



Eddy Tungadi, S.T., M.T.  
NIP. 197908232010121001

### Mengetahui,

Koordinator Program Studi

Teknik Komputer dan Jaringan  
Politeknik Negeri Ujung Pandang



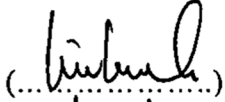
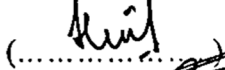
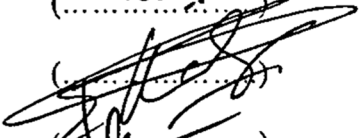

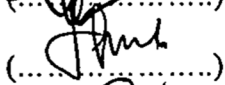
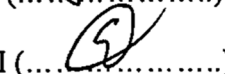
Eddy Tungadi, S.T., M.T.  
NIP. 197908232010121001

## HALAMAN PENERIMAAN

Pada hari ini Selasa, 22 Agustus 2023 Tim Penguji Ujian Sidang Skripsi telah menerima dengan baik skripsi oleh mahasiswa: Hairatunnisa (42519034) dengan judul "Penerapan *Business Intelligence* Terhadap Data Pelanggan Menggunakan Metode *Clustering* Multidimensi Untuk Segmentasi Pelanggan (Studi Kasus: PT Gama Inovasi Berdikari)".

Makassar, 22 Agustus 2023

Tim Penguji Ujian Skripsi:

- |   |               |   |
|---|---------------|---|
| 1. Iin Karmila Yusri, S.ST., M.Eng., Ph.D | Ketua         | (  )  |
| 2. Tantri Indrabulan, S.T., M.T.          | Sekretaris    | (  ) |
| 3. Muh. Fajri Raharjo, S.T., M.T.         | Anggota       | (  ) |
| 4. Drs. Kasim, M.T.                       | Anggota       | (  ) |
| 5. Zawiyah Saharuna, S.T., M.Eng.         | Pembimbing I  | (  ) |
| 6. Eddy Tungadi, S.T., M.T.               | Pembimbing II | (  ) |

## DAFTAR ISI

HALAMAN PENGESAHAN.....	ii
HALAMAN PENERIMAAN.....	iii
DAFTAR ISI.....	iv
KATA PENGANTAR.....	vii
DAFTAR TABEL.....	ix
DAFTAR GAMBAR.....	x
SURAT PERNYATAAN.....	xii
RINGKASAN.....	xiii
BAB I PENDAHULUAN.....	14
1.1 Latar Belakang.....	14
1.2 Rumusan Masalah.....	15
1.3 Ruang Lingkup Penelitian.....	16
1.4 Tujuan Penelitian.....	16
1.5 Manfaat Penelitian.....	16
BAB II TINJAUAN PUSTAKA.....	17
2.1 PT Gama Inovasi Berdikari.....	17
2.2 <i>Business Intelligence</i> .....	18
2.3 Segmentasi Pelanggan.....	20
2.4 Segmentasi Pasar.....	21
2.5 <i>Pre-Processing Data</i> .....	23

2.6	<i>Label Encoding</i> .....	23
2.7	Analisis Korelasi .....	24
2.8	Metode Elbow .....	25
2.9	Metode <i>Clustering</i> .....	26
2.10	<i>Clustering</i> Multidimensi.....	26
2.11	Algoritma K-Means.....	28
2.12	R Programming .....	28
2.13	<i>Interface</i> .....	29
2.14	<i>System Usability Scale</i> .....	31
BAB III METODE PENELITIAN.....		34
3.1	Tempat dan Waktu Penelitian .....	34
3.2	Prosedur Penelitian.....	34
BAB IV HASIL DAN PEMBAHASAN .....		43
4.1	Deskripsi Dataset.....	43
4.2	<i>Pre-Processing Data</i> .....	46
4.2.1	<i>Import Library</i> .....	46
4.2.2	<i>Data Cleaning</i> .....	46
4.3	Pemilihan Variabel.....	49
4.4	Tahapan Klasterisasi.....	52
4.4.1	<i>Load</i> Dataset pada Google Spreadsheet.....	52
4.4.2	Sinkronisasi Google Spreadsheets .....	53
4.4.3	Penentuan Jumlah <i>Cluster</i> .....	54

4.4.4	Proses <i>Clustering</i> .....	55
4.4.5	Validasi Hasil <i>Clustering</i> .....	59
4.4.6	Simpan Dataset.....	60
4.5	Integrasi Google Spreadsheets dan App Script .....	60
4.6	Visualisasi .....	62
4.7	Analisis Hasil <i>Clustering</i> .....	65
4.8	Implementasi Sistem.....	69
4.9	Hasil Pengujian <i>System Usability Scale</i> (SUS).....	75
BAB V KESIMPULAN DAN SARAN.....		77
5.1	Kesimpulan.....	77
5.2	Saran.....	77
DAFTAR PUSTAKA .....		78
LAMPIRAN.....		82

## KATA PENGANTAR

Segala puji syukur atas nikmat dan karunia yang diberikan oleh Allah SWT, sehingga Penulis dapat menyelesaikan tugas akhir ini. Tugas akhir ini disusun dalam rangka memenuhi salah satu persyaratan memperoleh gelar Diploma IV (D-4/S1 Terapan) pada Program Studi Teknik Komputer dan Jaringan Jurusan Teknik Elektro Politeknik Negeri Ujung Pandang.


Penulis menyadari bahwa keberhasilan penyusunan tugas akhir ini tidak lepas dari bantuan berbagai pihak baik secara langsung maupun tidak langsung. Oleh karenanya, Penulis menyampaikan apresiasi dengan mengucapkan terima kasih sebesar-besarnya kepada :

1. Orang tua penulis yang selalu memberikan dukungan dan doa.
2. Bapak Ir. Ilyas Mansur, M.T. selaku Direktur Politeknik Negeri Ujung Pandang.
3. Bapak Ahmad Rizal Sultan, S.T., M.T., Ph.D. selaku Ketua Jurusan Teknik Elektro Politeknik Negeri Ujung Pandang.
4. Bapak Eddy Tungadi, S.T., M.T. selaku Koordinator Program Studi Teknik Komputer dan Jaringan.
5. Ibu Zawayah Saharuna, S.T., M.Eng. selaku pembimbing I dan Bapak Eddy Tungadi, S.T., M.T. selaku pembimbing II yang senantiasa meluangkan waktu dalam membimbing, memotivasi, dan mengarahkan Penulis sehingga dapat menyelesaikan penelitian ini.

6. Seluruh dosen dan Staf Jurusan Teknik Elektro, khususnya Program Studi D4 Teknik Komputer dan Jaringan.
7. Teman-teman Angkatan 2019 Program Studi Teknik Komputer dan Jaringan yang selalu memberikan dukungan dalam penyusunan tugas akhir ini.
8. Semua pihak yang telah memberikan bantuan moril maupun materil yang tidak dapat disebutkan satu per satu.

Menyadari adanya keterbatasan pengetahuan, Penulis menyadari bahwa tugas akhir ini masih jauh dari kesempurnaan. Oleh karena itu, Penulis mengharapkan saran dan kritik yang membangun demi perbaikan di masa mendatang. Semoga tulisan ini dapat bermanfaat bagi para Pembaca.

Makassar, 16 Agustus 2023

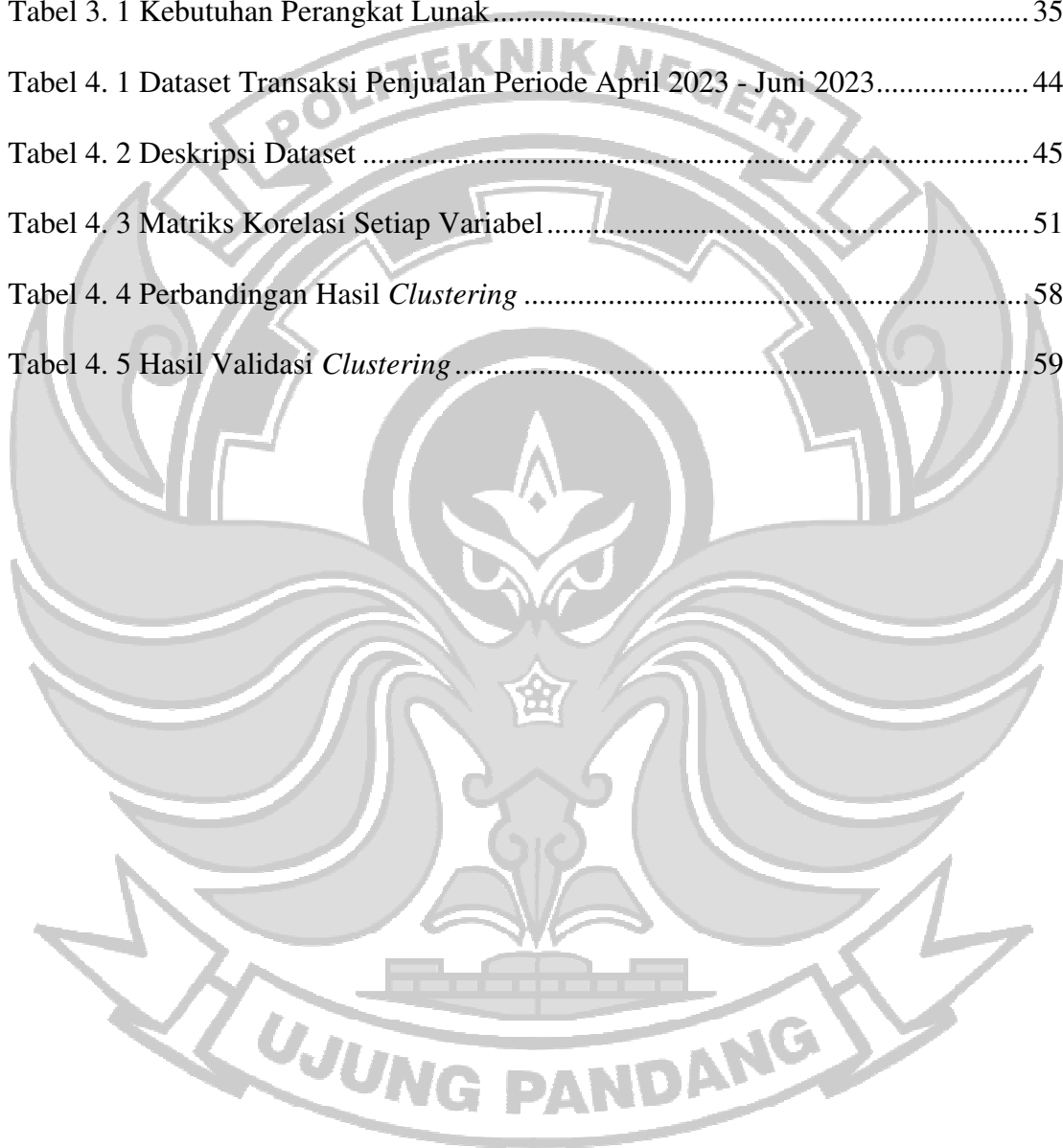


Penulis



## DAFTAR TABEL

Tabel 2. 1 Penilaian Skor SUS.....	33
Tabel 3. 1 Kebutuhan Perangkat Lunak.....	35
Tabel 4. 1 Dataset Transaksi Penjualan Periode April 2023 - Juni 2023.....	44
Tabel 4. 2 Deskripsi Dataset.....	45
Tabel 4. 3 Matriks Korelasi Setiap Variabel.....	51
Tabel 4. 4 Perbandingan Hasil <i>Clustering</i> .....	58
Tabel 4. 5 Hasil Validasi <i>Clustering</i> .....	59



## DAFTAR GAMBAR

Gambar 2. 1 <i>Precentile Rank</i> dan <i>Letter Grades</i> (Sauro, 2011) .....	32
Gambar 3. 1 Tahapan Prosedur Penelitian .....	34
Gambar 3. 2 <i>Flowchart Metode Elbow</i> .....	38
Gambar 3. 3 <i>Flowchart Algoritma K-Means</i> .....	39
Gambar 4. 1 Hasil Pengecekan Data yang Kosong .....	47
Gambar 4. 2 Hasil Pengecekan <i>Outlier</i> .....	48
Gambar 4. 3 Proses <i>Encoding</i> .....	49
Gambar 4. 4 Korelasi Setiap Variabel .....	50
Gambar 4. 5 Simpan Dataset ke Google Drive .....	53
Gambar 4. 6 Proses Autentikasi .....	53
Gambar 4. 7 Tampilan Dataset di RStudio .....	54
Gambar 4. 8 Grafik Metode Elbow .....	55
Gambar 4. 9 <i>Select Variable</i> untuk <i>Clustering</i> .....	56
Gambar 4. 10 Proses <i>Clustering</i> dengan KMeans .....	56
Gambar 4. 11 Hasil <i>Clustering</i> dengan 2 <i>Cluster</i> .....	56
Gambar 4. 12 Proses <i>Clustering</i> KMeans dengan 3 <i>Cluster</i> .....	57
Gambar 4. 13 Hasil <i>Clustering</i> dengan 3 <i>Cluster</i> .....	58
Gambar 4. 14 Mencari Nilai SSE Setiap <i>Cluster</i> .....	59
Gambar 4. 15 Upload Dataset Hasil <i>Clustering</i> .....	60
Gambar 4. 16 Fungsi <i>doPost</i> .....	61

Gambar 4. 17 <i>Function addDataToSheet</i> .....	62
Gambar 4. 18 Dashboard Reporting Customer Segmentation .....	63
Gambar 4. 19 Persentase <i>Cluster</i> .....	66
Gambar 4. 20 Grafik Perbandingan Penjualan Berdasarkan <i>Gender</i> .....	66
Gambar 4. 21 Grafik Perbandingan Harga Jual dan Pemberian diskon.....	67
Gambar 4. 22 Segmen Pelanggan Berdasarkan Geografis.....	68
Gambar 4. 23 Grafik <i>Customer by Total Payment</i> .....	69
Gambar 4. 24 Tampilan Bagian Header Pada Website.....	70
Gambar 4. 25 Tampilan Diagram Alur Penggunaan Sistem.....	70
Gambar 4. 26 Tampilan Menu Upload dan Klastering .....	71
Gambar 4. 27 Tampilan Kondisi Sebelum Pilih File .....	71
Gambar 4. 28 Upload File.....	72
Gambar 4. 29 Tambahkan File ke Dataset dan Klasterisasi.....	72
Gambar 4. 30 Hasil <i>Clustering (Data Discretization)</i> .....	73
Gambar 4. 31 Hasil <i>Clustering</i> Setelah Penambahan Data.....	74
Gambar 4. 32 Hasil <i>Clustering</i> Sebelum Penambahan Data.....	74
Gambar 4. 33 Tampilan Dashboard Reporting .....	75

## SURAT PERNYATAAN

Saya yang bertanda tangan dibawah ini:

Nama : Hairatunnisa

NIM : 425 19 034

Menyatakan dengan sebenar-benarnya bahwa segala pernyataan dalam skripsi ini yang berjudul Penerapan *Business Intelligence* Terhadap Data Pelanggan Menggunakan Metode *Clustering* Multidimensi Untuk Segmentasi Pelanggan (Studi Kasus: PT Gama Inovasi Berdikari) merupakan gagasan dan hasil karya saya sendiri dengan arahan komisi pembimbing, dan belum pernah diajukan dalam bentuk apapun pada perguruan tinggi dan instansi manapun.

Semua data dan informasi yang digunakan telah dinyatakan secara jelas dan dapat diperiksa kebenarannya. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan dari penulis lain telah disebutkan dalam naskah dan dicantumkan dalam skripsi ini.

Jika pernyataan saya tersebut diatas tidak benar, Saya siap menanggung resiko yang ditetapkan oleh Politeknik Negeri Ujung Pandang.

Makassar, 04 Oktober 2023



Hairatunnisa  
42519034

PENERAPAN *BUSINESS INTELLIGENCE* TERHADAP DATA PELANGGAN  
MENGUNAKAN METODE *CLUSTERING* MULTIDIMENSI UNTUK  
MENENTUKAN SEGMENTASI PELANGGAN  
(STUDI KASUS: PT GAMA INOVASI BERDIKARI)

**RINGKASAN**

Persaingan bisnis yang semakin ketat menuntut perusahaan untuk menghadapi berbagai tantangan ekonomi dalam mempersiapkan strategi bisnis yang tepat. PT Gama Inovasi Berdikari (GIB) sebagai perusahaan akselerator dan hilirisator juga menghadapi kendala dalam melakukan pemetaan strategi bisnis. Dimana hal tersebut dapat mempengaruhi kecepatan pengambilan keputusan dan perencanaan strategi pemasaran yang optimal. Oleh karena itu, penelitian ini memberikan solusi yaitu menerapkan *Business Intelligence* pada sistem dengan menggunakan metode *clustering* multidimensi untuk melakukan segmentasi pelanggan berdasarkan karakteristik dan perilaku penggunaan produk menggunakan algoritma K-Means.

Penelitian ini telah berhasil melakukan klusterisasi terhadap data transaksi penjualan *online* PT Gama Inovasi Berdikari menggunakan algoritma K-Means melalui RStudio. Tingkat akurasi perbedaan antar *cluster* yang terbentuk mencapai 75,5% yang menunjukkan efektivitas algoritma dalam pemetaan data. Selain itu, hasil pengujian menggunakan metode *System Usability Scale* (SUS) mengindikasikan kesuksesan dalam penerapan sistem *Business Intelligence* dengan skor 72,5. Sistem ini memungkinkan pengguna untuk menambahkan data baru dan melakukan *clustering* secara otomatis. Hasil *clustering* akan ditampilkan melalui dashboard yang dapat membantu dalam penentuan strategi *marketing* secara efisien.

**Kata Kunci:** *Business Intelligence*, K-Means, *clustering* multidimensi, Segmentasi Pelanggan, *System Usability Scale*

## BAB I PENDAHULUAN

### 1.1 Latar Belakang

Persaingan bisnis yang semakin meningkat menuntut perusahaan untuk meningkatkan daya saingnya secara terus menerus dalam waktu cepat agar mampu menanggapi kebutuhan pasar. Jika dikaitkan dengan aspek pemasaran, perusahaan yang memiliki pemasaran yang kuat akan mampu bersaing dalam persaingan bisnis yang ketat. Teknik *marketing* dan validasi produk merupakan kunci utama bagi sebuah bisnis dalam menjaga keberlanjutan sebuah usaha dan memasarkan produk atau jasanya kepada calon konsumen.

PT Gama Inovasi Berdikari (GIB) merupakan perusahaan yang bergerak dibidang akselerator dan hilirisator perusahaan berbasis inovasi. Sampai saat ini perusahaan ini sudah memiliki startup rintisan sebanyak 300 lebih yang tersebar di beberapa daerah di Indonesia. PT GIB dalam melakukan segmentasi pelanggan masih melalui pendekatan acak yang dimana ketika data yang diolah semakin banyak proses analisa data menjadi lambat karena pihak manajemen harus mengolah kembali data yang telah dihasilkan agar dapat divisualisasikan dengan baik. Proses tersebut berpotensi memperlambat kecepatan dalam pengambilan keputusan dari manajemen dan mempersulit pemetaan strategi bisnis dikarenakan teknik *marketing* yang masih kurang dalam hal penentuan strategi bisnis yang tepat. Hal tersebut berdasarkan wawancara yang telah dilakukan dengan Cathrine Andriani dan Dyah Ratri selaku *Human Resource Development* (HRD) di PT GIB.

Pada tahap rilis produk sering terjadi masalah di mana produk yang dirilis ternyata tidak diminati atau tidak dibutuhkan oleh masyarakat. Adapun produk yang ditawarkan perusahaan adalah produk herbal. Kendala yang terjadi dalam usaha ini adalah kesulitan dalam meningkatkan jumlah peminat atau pelanggan terhadap penggunaan produk herbal dan juga persaingan bisnis yang bergerak dibidang yang sama. Berdasarkan data dari *Center Economics and Development Studies (CEDS)* Fakultas Ekonomi dan Bisnis (FEB) Universitas Padjadjaran (UNPAD) yang menyatakan bahwa nilai indeks persaingan usaha pada tahun 2022 mengalami peningkatan ke angka 4,87 dari indeks tahun sebelumnya di angka 4,81.

Maka dari itu, agar dapat meningkatkan daya beli pelanggan serta pelanggan yang ada tidak beralih kepada pesaing dengan usaha yang sejenis maka perusahaan dituntut untuk mengembangkan strategi pemasaran dengan baik dan tepat sasaran. Persoalan tersebut dapat diselesaikan dengan menerapkan *Business Intelligence* menggunakan metode *clustering* mutidimensi untuk membuat segmentasi pelanggan berdasarkan karakteristik dan perilaku pelanggan dalam menggunakan produk yang dapat membantu perusahaan dalam mengembangkan strategi pemasaran yang lebih efektif dan melakukan pengembangan produk sesuai dengan kebutuhan pelanggan.

## **1.2 Rumusan Masalah**

Berdasarkan latar belakang yang telah diuraikan sebelumnya, maka rumusan masalah dari penelitian ini adalah:

1. Bagaimana melakukan *clustering* multidimensi menggunakan algoritma KMeans dalam menentukan strategi yang tepat untuk meningkatkan penjualan?
2. Bagaimana cara membangun sebuah sistem *Business Intelligence* untuk menampilkan hasil segmentasi pelanggan menggunakan metode *clustering* multidimensi pada PT GIB?

### **1.3 Ruang Lingkup Penelitian**

Adapun ruang lingkup penelitian ini sebagai berikut:

1. Dataset yang digunakan dalam penelitian ini yaitu data transaksi penjualan PT GIB periode April 2023 – Juni 2023.
2. Pemilihan variabel menggunakan metode *Pearson (Product Moment)*.

### **1.4 Tujuan Penelitian**

Adapun tujuan yang ingin dicapai dari penelitian ini adalah:

1. Melakukan *clustering* multidimensi untuk menentukan segmentasi pelanggan dengan menggunakan algoritma KMeans.
2. Membuat sebuah sistem *Business Intelligence* dalam menentukan segmen pasar di PT GIB.

### **1.5 Manfaat Penelitian**

Manfaat dari penelitian ini adalah sistem *Business Intelligence* dapat membantu PT GIB melakukan segmentasi pelanggan dengan mudah dan efektif.



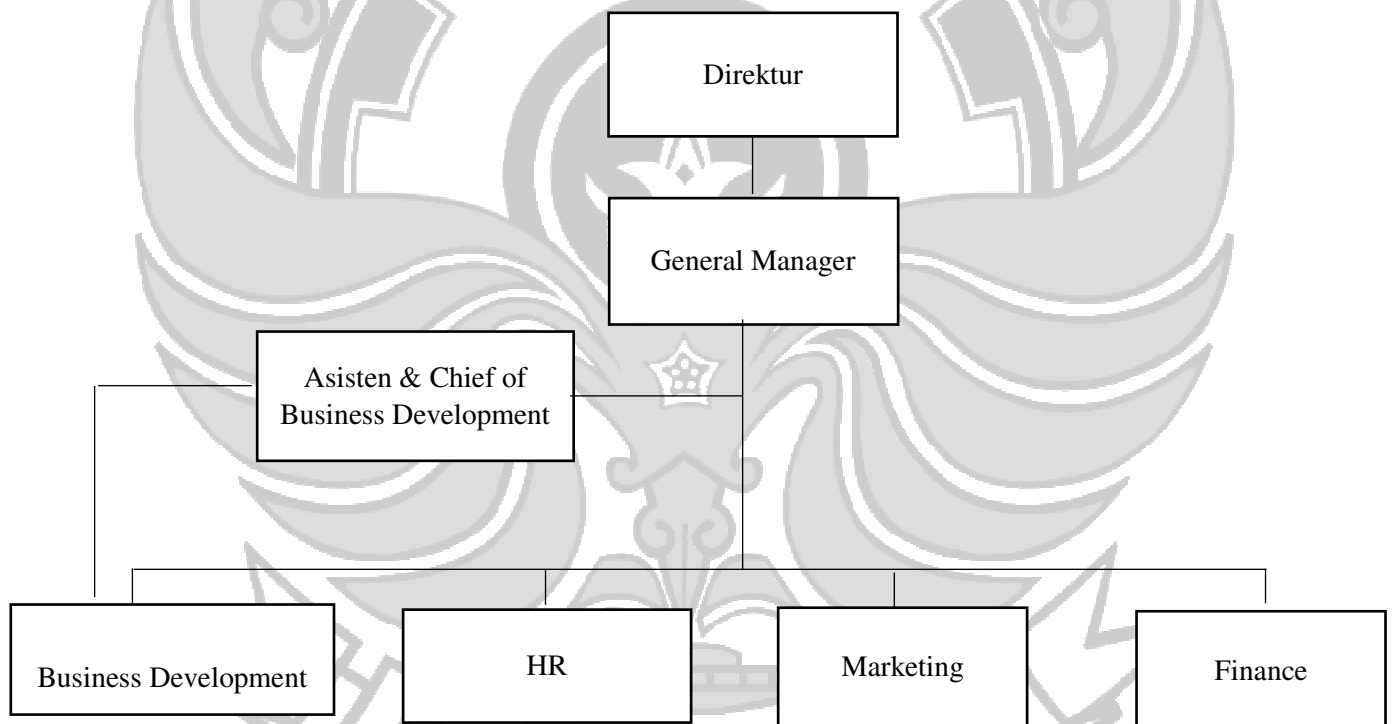
## BAB II TINJAUAN PUSTAKA

### 2.1 PT Gama Inovasi Berdikari

PT Gama Inovasi Berdikari (GIB) merupakan perusahaan yang bergerak dalam bidang akselerasi bisnis inovasi dan berada di bawah naungan Universitas Gadjah Mada. PT GIB menjadi pioner dalam akselerasi bisnis berbasis inovasi dengan ekosistem universitas sebagai dukungan kesiapan bangsa menuju industri 4.0. Berdiri pada tahun 2014 dengan visi membentuk kreativitas wirausahawan muda di Indonesia. Hingga saat ini PT GIB telah memiliki portofolio lebih dari 300 startup di beberapa daerah di Indonesia yang bergerak di berbagai sektor. Pengembangan telah dilakukan pada 8 strategi area bisnis diantaranya *Internet of Things (IoT)*, *Healt-Tech*, *Edu-Tech*, *Fin-Tech*, *E-Commerce*, *Agritech*, Logistik, dan Energi. Salah satu area bisnis pada PT GIB adalah bidang logistik dimana terdapat unit retail Gibmate yang menawarkan produk herbal, seperti produk Wedang GamaFood, Wedang Telang, Ubiekies, dan Si Daun.

PT GIB dalam melakukan pengolahan data khususnya pada konteks segmentasi pelanggan menggunakan metode acak atau belum menerapkan teknologi *Business Intelligence* dan metode *clustering*. Adapun metode acak yang diterapkan yaitu menentukan parameter untuk pengelompokan (*Gender*, *Buyer*, *TotalSales*, dll) secara random dan tidak menentukan jumlah data yang akan diolah yang dapat mengakibatkan *imbalance* dataset. Adapun pengolahan data tersebut dilakukan oleh tim pemasaran yang digambarkan pada struktur organisasi di Gambar 2.1.

Selain itu dalam pengolahan data PT GIB juga mengkategorikan pelanggan ke dalam dua kelompok, yaitu kelompok peminat dan kelompok pelanggan. Kelompok peminat merujuk pada individu yang menunjukkan minat awal atau kecenderungan tertarik terhadap produk yang ditawarkan melalui kampanye dan belum melakukan pembelian atau berlangganan tetapi menunjukkan minat potensial. Sedangkan, pelanggan adalah individu yang telah melakukan pembelian atau berlangganan terhadap produk tersebut.



Gambar 2. 1 Struktur Organisasi PT GIB

## 2.2 *Business Intelligence*

*Business Intelligence* adalah kategori yang umumnya digunakan pada aplikasi dan teknologi untuk mengumpulkan, menyimpan, menganalisa, dan menyediakan akses

pada data yang dapat membantu pengguna dari kalangan perusahaan agar dapat mengambil kebijakan yang lebih baik dan tepat sesuai dengan kebutuhan (Oktavian & Rumaisa, 2022). Secara umum, BI merupakan sebuah proses untuk melakukan ekstraksi data-data operasional perusahaan dan mengumpulkannya dalam sebuah data *warehouse* yang selanjutnya diproses menggunakan berbagai analisis statistik dalam proses data mining sehingga didapat berbagai kecenderungan atau pola dari data (Wijaya & Gantini, 2019).

Pada tahun 1989, Howard Dresner menggunakan istilah *business intelligence* untuk mendeskripsikan seperangkat konsep dan metode yang berguna untuk meningkatkan kemampuan pembuatan keputusan dengan bantuan sistem yang berbasis fakta atau realita yang terjadi. *Business intelligence* menjelaskan tentang suatu konsep dan metode bagaimana untuk meningkatkan kualitas pengambilan keputusan bisnis berdasarkan sistem yang berbasis data. Dengan ketatnya persaingan bisnis maka sangat diperlukan analisis data sebagai cara untuk tetap bertahan dalam dunia bisnis dan analisis data yang cermat dibutuhkan dalam proses pengambilan keputusan. Analisis data telah menjadi kebutuhan vital untuk meningkatkan persaingan pada dunia bisnis karena menggunakan intuisi dalam pengambilan keputusan sudah tidak lagi relevan dalam persaingan bisnis yang semakin ketat (Subroto & Endaryati, 2021).

Penelitian terkait yang telah dilakukan oleh Karami dan Widharto mengenai implementasi *business intelligence* dalam membuat keputusan strategi *marketing* terhadap hasil segmentasi pelanggan (Karami & Widharto, 2023) dan penelitian yang

dilakukan oleh Oktavian dan Rumaisa terkait penerapan *business intelligence* berdasarkan lokasi penjualan pada UKM Foodendez yang menggunakan algoritma apriori untuk menghasilkan data segmentasi pasar yang dapat membantu memberikan informasi prediksi minat konsumen terhadap suatu produk (Oktavian & Rumaisa, 2022).

Adapun pada penelitian ini akan diimplementasikan sebuah sistem *business intelligence* yang bertujuan dapat membantu PT GIB dalam penentuan strategi pemasaran berdasarkan hasil segmentasi pelanggan melalui *clustering* multidimensi dari data transaksi pelanggan.

### **2.3 Segmentasi Pelanggan**

Segmentasi pelanggan merupakan pembentukan kelompok data baru yang berfungsi dalam penilaian loyalitas suatu pelanggan pada perusahaan. Data yang dinilai dapat berupa jejak transaksi pelanggan pada perusahaan, rentan waktu sering mengunjungi perusahaan, dan frekuensi kedatangan pelanggan yang bertujuan untuk mengetahui sesering apakah seorang pelanggan berkunjung dan sebesar apakah seorang pelanggan melakukan transaksi (Kiat, 2020).

Berbagai penelitian terkait yang melakukan segmentasi pelanggan menggunakan teknik *clustering*. Penelitian yang dilakukan oleh Mirantika dkk mengenai segmentasi pelanggan perusahaan ritel dengan menerapkan teknik *clustering* dengan algoritma K-Medoids dengan menggunakan perwakilan data sebagai pusat *cluster* (Mirantika et al., 2023). Dan penelitian yang dilakukan oleh Sulistyawati dan Sadikin mengenai

segmentasi pelanggan pada Perum BULOG kantor wilayah DKI Jakarta & Banten dengan menggunakan algoritma K-Medoids diperoleh hasil jumlah *cluster* optimal adalah 3 *cluster* dengan nilai maksimum *Silhouette Index* adalah 0,375 dan nilai minimum *Davies Doublin Index* adalah 1,030 (Mirantika et al., 2023).

## 2.4 Segmentasi Pasar

Segmentasi pasar adalah proses untuk mengidentifikasi segmen pasar dan proses membagi basis pelanggan yang luas menjadi sub-kelompok konsumen yang terdiri dari pelanggan yang ada dan calon pelanggan. Tujuan segmentasi pasar adalah untuk menghasilkan volume penjualan yang tinggi dan dengan melakukan segmentasi akan mengetahui kebutuhan pelanggan (Aprilyani et al., 2020). Ada beberapa kategori segmentasi pasar, yaitu sebagai berikut.

### a. Segmentasi Pasar Secara Demografis

Segmentasi pasar secara demografis adalah strategi yang digunakan dalam pemasaran untuk membagi pasar menjadi kelompok-kelompok yang berdasarkan pada karakteristik demografis dari konsumen. Faktor-faktor demografis mencakup atribut-atribut seperti usia, jenis kelamin, pendapatan, tingkat pendidikan, pekerjaan, dan agama. Tujuan dari segmentasi pasar ini adalah untuk lebih memahami dan menargetkan kelompok-kelompok spesifik yang memiliki karakteristik serupa sehingga produk atau layanan dapat diadaptasi dan dipasarkan dengan lebih efektif sesuai dengan kebutuhan dan preferensi masing-masing kelompok.

b. Segmentasi Pasar Secara Geografis

Segmentasi pasar secara geografis adalah pendekatan dalam pemasaran di mana pasar dibagi menjadi kelompok-kelompok berdasarkan lokasi geografis atau wilayah geografis tertentu. Pendekatan ini mempertimbangkan perbedaan dalam karakteristik sosial, budaya, iklim, kebiasaan konsumen, serta preferensi yang mungkin muncul karena perbedaan geografis.

c. Behavioral atau Perilaku

Segmentasi pasar berdasarkan perilaku adalah pendekatan dalam pemasaran di mana pasar dibagi menjadi kelompok-kelompok berdasarkan perilaku pembelian, kebiasaan konsumen, preferensi, dan respons terhadap produk atau layanan. Pendekatan ini mengarah pada pemahaman lebih dalam tentang bagaimana konsumen berinteraksi dengan produk atau merek, bagaimana konsumen merespon kampanye pemasaran, dan bagaimana konsumen memutuskan untuk melakukan pembelian.

d. Psikografis

Segmentasi pasar berdasarkan psikografis adalah pendekatan dalam pemasaran di mana pasar dibagi menjadi kelompok-kelompok berdasarkan atribut-atribut psikologis dan sosial konsumen. Pendekatan ini mempertimbangkan faktor-faktor seperti kepribadian, nilai-nilai, sikap, minat, gaya hidup, dan motivasi yang mempengaruhi perilaku pembelian konsumen.

## **2.5 Pre-Processing Data**

*Pre-processing* data merupakan tahap awal dalam pengolahan data yang meliputi *cleaning*, transformasi, integrasi, dan reduksi data untuk mempersiapkan data dalam proses analisis. Pada tahap *cleaning* data dilakukan pengecekan data *outlier*, *missing value*, dan melakukan penghapusan atau penanganan pada data *outlier* dan *missing value* untuk memastikan kualitas data yang akan di *cluster*. Hal ini dapat dilakukan dengan menggunakan metode statistik, seperti mengatasi *outlier* dengan imputasi data atau mengganti *missing value* menggunakan nilai lain dari hasil imputasi data atau. Setelah *outlier* diidentifikasi selanjutnya dapat memutuskan apakah akan menghapus baris atau kolom yang mengandung *outlier*, melakukan imputasi nilai dengan menggantikan *outlier* dengan nilai yang lebih realistis, atau mengubah *outlier* menjadi nilai yang mendekati batas atas / bawah (Mishra et al., 2020). Menurut Walimbe (2018) dalam situs *datasciencecentral.com* menjelaskan bahwa *outlier* boleh tidak dihapus jika memiliki nilai bisnis yang tinggi dan nilai bisnis yang disampaikan sejalan dengan tujuan bisnis.

## **2.6 Label Encoding**

*Label Encoding* adalah suatu teknik dalam pemrosesan data yang digunakan untuk mengubah data kategorikal atau *string* menjadi numerik. Dalam analisis data atau pembuatan model *machine learning* sering kali diperlukan representasi numerik untuk variabel-variabel kategorikal agar bisa diolah oleh algoritma yang memerlukan input

numerik. Dalam *label encoding* setiap nilai unik dari suatu variabel kategorikal diberi label numerik berdasarkan urutan kemunculannya.

## 2.7 Analisis Korelasi

Analisis korelasi adalah perhitungan statistik yang dapat digunakan untuk mengetahui derajat hubungan linier antara variabel yang satu dengan yang lain. Koefisien korelasi dilambangkan sebagai “*r*” dan paling umum diukur menggunakan koefisien korelasi Pearson. Koefisien korelasi Pearson (*Product Moment*) mengukur sejauh mana dua variabel berkorelasi linear. Teknik analisis korelasi pearson termasuk teknik statistika parametik yang menggunakan data interval dan rasio dengan persyaratan tertentu. Berikut rumus yang digunakan untuk mencari nilai korelasi Pearson.

$$r = \frac{n \sum X_i Y_i - \sum X_i \sum Y_i}{\sqrt{n \sum X_i^2 - (\sum X)^2} \sqrt{n \sum Y_i^2 - (\sum Y)^2}} \quad (2.1)$$

Persamaan 2. 1 Rumus Korelasi *Pearson (Product Moment)*

Keterangan:

*r* : koefisien korelasi

$\sum X$  : penjumlahan variabel X

$\sum Y$  : penjumlahan variabel Y

$\sum X \sum Y$  : perkalian penjumlahan variabel X dan Y

Nilai koefisien korelasi berkisar antara -1 hingga 1. Nilai +1 menunjukkan korelasi positif sempurna, sedangkan nilai -1 menunjukkan korelasi negatif sempurna. Dan nilai 0 menunjukkan bahwa tidak ada korelasi linier antara variabel (Purba & Purba, 2022).



Menurut Chin (1998) nilai *R-Square* dikategorikan kuat jika lebih dari 0.67, moderat jika lebih dari 0.33 tetapi lebih rendah dari 0.67, dan lemah jika lebih dari 0.19 tetapi lebih rendah dari 0.33.

## 2.8 Metode Elbow

Metode elbow adalah teknik yang digunakan dalam analisis *clustering* untuk menentukan jumlah *cluster* ( $k$ ) yang optimal pada suatu dataset. Metode ini didasarkan pada plot variasi yang dijelaskan sebagai fungsi dari jumlah *cluster* dan memilih siku dari kurva sebagai jumlah *cluster* ( $k$ ) yang akan digunakan. Nilai  $k$  pada kombinasi siku dengan K-Means adalah grafik hubungan *cluster* dengan penurunan error (Merliana & Santoso, 2015).

Metode Elbow menentukan *cluster* optimal dengan cara mencari nilai *Sum of Squared Error* (SSE) dari tiap jumlah *cluster*. Nilai SSE merupakan jumlah kuadrat jarak dari setiap titik data dengan *centroid cluster* (Ekasetya & Jananto, 2020). Untuk menghitung SSE dapat menggunakan rumus sebagai berikut.

$$SSE = \sum (y - \Delta)^2 \quad (2.2)$$

Persamaan 2. 2 Rumus *Sum of Squares Error* (SSE)

Keterangan:

SSE : *Sum of Squared Error*

$y$  : Posisi data

$\Delta$  : *Centroid*

## 2.9 Metode *Clustering*

Konsep data mining yang mampu mengelompokkan data ke dalam kelompok yang serupa disebut dengan *clustering*. *Clustering* adalah teknik yang dapat digunakan untuk mengidentifikasi objek yang bekerja dengan mengkarakterisasi kelompok yang memiliki karakteristik yang hampir sama dari objek tersebut (Jumadi et al., 2021). *Clustering* mengelompokkan data berdasarkan *cluster*/kategori dan merupakan teknik untuk mengatur data yang tidak terstruktur menjadi struktur data yang informatif serta dapat dilakukan dalam dua dimensi atau dalam ruang multidimensi dengan jumlah dimensi yang lebih tinggi.

Algoritma *clustering* terdiri dari dua bagian yaitu hirarkis dan partitional. Algoritma hirarkis menemukan *cluster* secara berurutan dimana *cluster* ditetapkan sebelumnya, sedangkan algoritma partitional menentukan semua kelompok pada waktu tertentu (Murpratiwi et al., 2021). Pengelompokan data dapat dilakukan dengan dua dimensi dan multidimensi

## 2.10 *Clustering* Multidimensi

*Clustering* multidimensi adalah metode pengelompokan objek yang memiliki beberapa dimensi atau atribut (Cordeiro et al., 2021) (Gautam et al., 2015). Metode ini digunakan untuk memecahkan masalah klasifikasi data yang sangat kompleks dengan menggunakan algoritma pengelompokan. Ada beberapa teori terkait metode *clustering* multidimensi (Ahmed et al., 2019), yaitu :

### 1. Jenis-jenis metode *clustering*

Metode *clustering* multidimensi di antaranya adalah *hierarchical clustering* dan *non-hierarchical clustering*. *Hierarchical clustering* mengelompokkan objek berdasarkan jarak antara mereka, sedangkan *non-hierarchical clustering* menggunakan pembagian objek ke dalam kelompok yang sebelumnya sudah ditentukan.

### 2. Validasi *clustering*

Validasi *clustering* adalah proses evaluasi hasil *clustering* multidimensi. Beberapa metode validasi *clustering* yang umum digunakan adalah *internal validation* dan *eksternal validation*. *Internal validation* digunakan untuk mengevaluasi hasil *clustering* secara internal, sedangkan *eksternal validation* digunakan untuk membandingkan hasil *clustering* dengan hasil yang sebenarnya.

### 3. Analisis faktor dalam *clustering*

Analisis faktor dapat digunakan untuk mengurangi dimensi dari data sebelum melakukan *clustering*. Dengan mengurangi dimensi data proses *clustering* multidimensi menjadi lebih mudah dan efektif.

### 4. Evaluasi hasil *clustering*

Setelah melakukan proses *clustering* multidimensi, perlu dilakukan evaluasi terhadap hasil *clustering*. Evaluasi dapat dilakukan dengan menggunakan metode seperti tampilan grafik, analisis siluet, dan pengujian hipotesis.

## 2.11 Algoritma K-Means

K-Means merupakan algoritma *clustering* yang termasuk ke dalam *partition clustering* yakni metode *clustering* yang mengelompokkan data berdasarkan tingkat kemiripan. K-Means menggunakan rata-rata nilai fitur untuk menghitung jarak antara setiap titik data. Algoritma ini bertujuan untuk membagi data ke dalam beberapa kelompok (*cluster*) berdasarkan kemiripan di antara objek-objek data tersebut. Pada algoritma K-Means, setiap *cluster* memiliki sebuah pusat (*centroid*) yang dihitung berdasarkan rata-rata objek data pada *cluster* tersebut. Data yang berada pada *cluster* yang sama memiliki tingkat kesamaan yang tinggi sedangkan data yang berada pada *cluster* yang berbeda tingkat kemiripannya rendah. Penentuan kemiripan bisa dilakukan dengan banyak cara antara lain berdasarkan jarak, pola, dan kepadatan. *Euclidean distance* merupakan salah satu metode penghitungan tingkat kemiripan antar titik data yang berbasis pada jarak (Paembonan & Abduh, 2021).

## 2.12 R Programming

R programming adalah salah satu bahasa pemrograman untuk memanipulasi data, simulasi, kalkulasi, dan peragaan grafik. R memiliki kemampuan menganalisis data dengan sangat efektif dan dilengkapi dengan operator pengolahan array dan matriks. R memiliki kemampuan penampilan grafik untuk peragaan datanya (Rumahorbo & Sekarwati, 2020). Bahasa R dapat diimplementasikan menggunakan RStudio. RStudio merupakan *software Integrated Development Environment (IDE)* yang digunakan untuk menjalankan Bahasa pemrograman R. Salah satu keunggulan

pada RStudio adalah dapat dijalankan di browser. Sehingga pengguna tidak perlu melakukan installasi R kecuali instalasi paket (*package*) pemrograman yang sesuai dengan kebutuhan pengguna (Priharsari, 2020).

### 2.13 *Interface*

*Interface* adalah tampilan dari sebuah produk yang berfungsi menjembatani sistem dengan pengguna. Tampilan *interface* bisa berupa warna, bentuk serta tulisan yang menarik. Ada berbagai *tools* yang dapat digunakan untuk membuat *interface* salah satunya adalah *Google Sites*. *Google Sites* merupakan produk *open source* dari *Google* yang memungkinkan pembuatan situs web dan dapat diakses oleh siapa saja yang memiliki akun *Google*.

Dengan adanya *Google Sites* pengguna dapat menggabungkan berbagai informasi dalam satu tempat (termasuk video, presentasi, lampiran, teks, dan lainnya) yang dapat dibagikan sesuai kebutuhan pengguna (Jubaidah & Zulkarnain, 2020). Berbagai produk *Google* dapat diintegrasikan dengan *Google Sites* termasuk *Google Spreadsheets* dan *Looker Studio*. *Google Spreadsheets* adalah aplikasi berbasis web yang memungkinkan pengguna untuk membuat, memperbaharui, dan memodifikasi *Spreadsheets* dan berbagi data secara online. *Google Spreadsheets* juga menyediakan bahasa pemrograman yang dikenal sebagai GAS (*Google App Script*). Kode GAS dapat dieksekusi secara *remote* di dalam *Google Cloud* (Fernando, 2018).

*Google Apps Script* yang merupakan platform pengembangan aplikasi dan pengkodean yang dibangun di dalam *Google Apps* yang memungkinkan pengguna

untuk menambahkan fungsionalitas ke *Spreadsheets*, Gmail, Situs, dan layanan lainnya dari *Google*. *Google* mendefinisikan *Google Apps Script* sebagai bahasa pemrograman *Java Script* yang menyediakan cara mudah dalam mengotomatiskan tugas di seluruh produk *Google* dan layanan pihak ketiga (Rahman, 2020).

Dengan *Google Sites* sebagai penghubung, berbagai produk *Google* dapat diintegrasikan untuk menyatukan proses visualisasi melalui *Looker Studio*. *Looker Studio* (sebelumnya dikenal sebagai *Google Data Studio*) adalah sebuah program berbasis *cloud* yang dirancang untuk menganalisis data kompleks dengan cara yang menarik dan terorganisir (Nisa et al., 2023). Tujuannya adalah untuk membantu pengguna dalam membuat laporan yang dinamis dan menarik secara visual agar mudah dipahami dengan menghubungkan sumber data eksternal ke dalam platform yang mudah dinavigasi dan menghasilkan berbagai laporan berbasis data.

Dashboard adalah media untuk menampilkan informasi dalam berbagai bentuk, seperti tabel, grafik, dan indikator visual. Hal ini dapat membantu menyajikan informasi penting tentang suatu perusahaan dalam satu layar yang memungkinkan pemahaman yang cepat terhadap situasi. Dashboard juga berguna untuk mengawasi dan mengontrol bisnis yang sedang berjalan. Terdapat tiga lapisan informasi dalam dashboard, yaitu *monitoring*, analisis, dan manajemen (Maulida et al., 2020). Dalam konteks ini, hasil visualisasi dari *Looker Studio* dapat ditampilkan melalui *Google Sites* untuk mempermudah akses dan pemahaman terhadap informasi yang disajikan.

## 2.14 *System Usability Scale*

*System Usability Scale* (SUS) merupakan salah satu metode pengukuran atau penilaian tingkat kegunaan dari berbagai macam produk dan layanan seperti perangkat keras, perangkat lunak, perangkat seluler, situs web, dan aplikasi. Penelitian-penelitian sebelumnya telah mengindikasikan bahwa SUS menunjukkan tingkat reliabilitas yang tinggi dengan validitas yang baik serta kemampuan untuk disesuaikan dengan berbagai konteks. Dalam SUS responden memberikan respon dengan menggunakan skala 1 (sangat tidak setuju) hingga 5 (sangat setuju) untuk masing-masing item pertanyaan. Skor yang diberikan pada skala SUS memiliki rentang antara 0 hingga 100 yang artinya semakin tinggi skor maka semakin baik persepsi mengenai kegunaan (Pal & Vanijja, 2020).

Penilaian dengan metode SUS dilakukan dengan cara berikut (Buana & Sari, 2022).

1. Skala yang digunakan adalah sangat tidak setuju (*strongly disagree*) sampai sangat setuju (*strongly agree*) atau bernilai 1 sampai 5. “Tidak setuju”, “Netral”, “Setuju”, dan “Sangat setuju” atas 10 item pernyataan SUS sesuai dengan penilaian subyektifnya. Jika responden merasa tidak menemukan skala respon yang tepat responden harus mengisi titik tengah skala pengujian.
2. Untuk pernyataan bernomor ganjil dihitung dengan cara nilai dari respon pengguna dikurangi dengan nilai 1.
3. Untuk pernyataan bernomor genap dihitung dengan cara nilai 5 dikurangi dengan nilai dari respon pengguna.

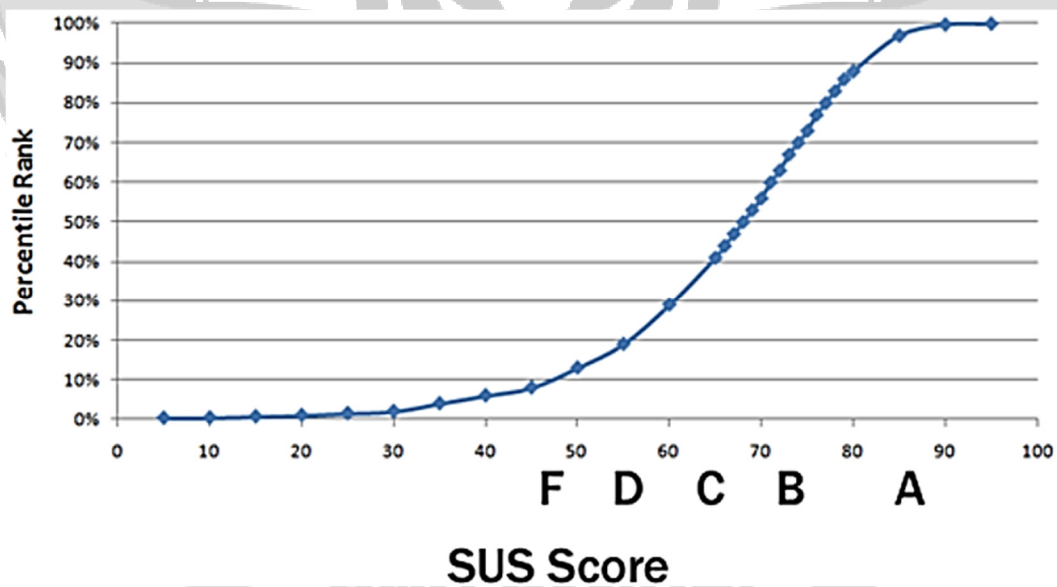
4. Jumlahkan nilai respon yang telah dihitung pada poin 2 dan 3 diatas dan dikalikan hasilnya dengan nilai 2.5. Hasil perhitungan ini akan mengkonversi rentang nilai menjadi antara 0–100.

Berikut rumus perhitungan skor SUS:

$$\text{Skor SUS} = ((R1 - 1) + (5 - R2) + (R3 - 1) + (5 - R4) + (R5 - 1) + (5 - R6) + (R7 - 1) + (5 - R8) + (R9 - 1) + (5 - R10)) * 2.5 \quad (2.3)$$

Persamaan 2. 3 Rumus Perhitungan skor SUS

Jeff Sauro menginterpretasikan nilai SUS dengan peringkat persentase (*percentile ranks*) dan kelas huruf (*letter grades*) dari A sampai dengan F, dimana A adalah kelas terbaik dan F adalah kelas terburuk.



Gambar 2. 2 Percentile Rank dan Letter Grades (Sauro, 2011)



Ketentuan untuk *percentile rank* dan *letter grades* dapat dilihat pada Tabel 2.1

(Buana & Sari, 2022).

Tabel 2. 1 Penilaian Skor SUS

SUS Score	Nilai	Penilaian
>80.3	A	Sangat Baik
68 – 80.3	B	Baik
68	C	Kurang Baik
51.68	D	Sangat Buruk
<51	F	Buruk



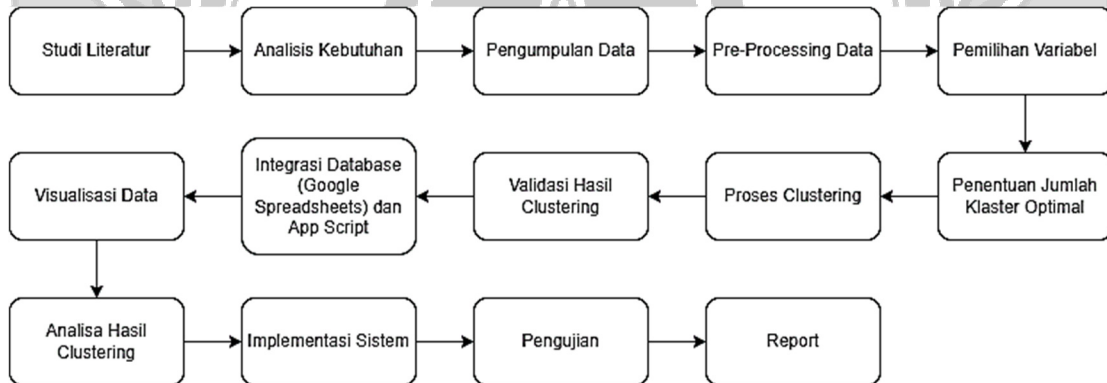
## BAB III METODE PENELITIAN

### 3.1 Tempat dan Waktu Penelitian

Penelitian ini dilaksanakan di Laboratorium CNAP Gedung Administrasi (AD) lantai 3 kampus 1 Politeknik Negeri Ujung Pandang. Dimulai pada bulan Maret 2023 sampai dengan Agustus 2023.

### 3.2 Prosedur Penelitian

Adapun tahapan dalam penelitian ini agar lebih terstruktur dan sistematis sehingga hasil yang diperoleh sesuai dengan tujuan dapat dilihat pada Gambar 3.1.



Gambar 3. 1 Tahapan Prosedur Penelitian

#### 3.2.1 Studi Literatur

Studi literatur adalah serangkaian kegiatan yang dilakukan untuk pendekatan penelitian dengan cara mencari referensi atas landasan teori yang relevan dengan kasus atau permasalahan yang sedang diteliti. Metode lain yang dapat dilakukan adalah

dengan melakukan pengumpulan data pustaka, membaca dan mencatat, serta mengolah bahan penelitian tersebut.

### 3.2.2 Analisis Kebutuhan

Tahap ini bertujuan untuk mengidentifikasi dan memahami kebutuhan dalam melakukan penelitian ini. Adapun analisis kebutuhan dalam melakukan penelitian ini adalah sebagai berikut.

a. Perangkat Keras

- Laptop dengan spesifikasi minimum sebagai berikut:

- RAM 4 GB
- Kapasitas HDD 20 GB

b. Perangkat Lunak

*Software* atau tools pendukung yang dibutuhkan dijelaskan pada Tabel 3.1.

Tabel 3. 1 Kebutuhan Perangkat Lunak

No.	Perangkat Lunak	Deskripsi
1	Sistem operasi windows	Sebagai sistem operasi yang digunakan dalam proses penelitian.
2	Google Collabs	Digunakan untuk pengolahan dataset menggunakan bahasa pemrograman Python
3	RStudio	Sebagai tools yang digunakan dalam proses klasterisasi

No.	Perangkat Lunak	Deskripsi
4	Python	Sebagai bahasa pemrograman yang digunakan untuk <i>preprocessing</i> data
5	R	Sebagai Bahasa pemrograman yang digunakan untuk klasterisasi
6	Spreadsheets	Sebagai database untuk menyimpan dataset dan hasil klastering
7	Looker Data Studio	Sebagai platform yang digunakan untuk visualisasi data hasil kluster
8	Google App Script	Sebagai bahasa pemrograman java yang akan digunakan untuk mengotomatisasi klasterisasi
9	Google Sites	Sebagai platform yang digunakan untuk menyajikan <i>web interface</i>

### 3.2.3 Pengumpulan Data

Penelitian ini menggunakan dataset transaksi penjualan yang bersumber dari PT GIB melalui persetujuan dari Kepala Pemasaran. Data yang telah didapatkan yaitu data transaksi penjualan pada periode April 2023 sampai dengan Juni 2023.

### 3.2.4 Pre-Processing Data

Pada tahapan ini akan dilakukan *cleaning* data, pemilihan data, dan transformasi data. Tahap pembersihan pada data awal untuk memastikan tidak ada data

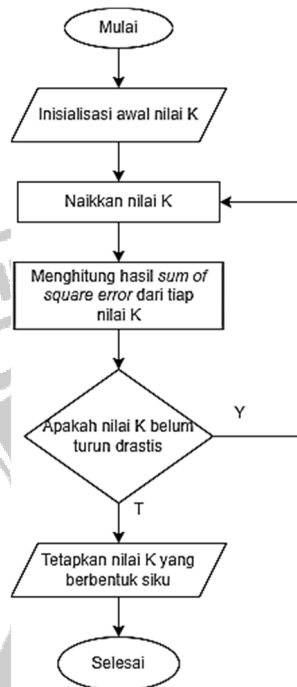
yang tidak lengkap atau *null* dan data redundant serta mempersiapkan data agar data siap diolah sesuai dengan kebutuhan.

### **3.2.5 Pemilihan Variabel**

Pada tahap ini merupakan proses pemilihan variabel yang dianggap relevan dan memiliki dampak signifikan dalam mempengaruhi segmentasi pelanggan. Pada penelitian ini untuk menentukan variabel yang berpengaruh menggunakan metode analisis multivariate yaitu metode korelasi *Product Moment (Pearson)*.

### **3.2.6 Penentuan Jumlah Kluster Optimal**

Sebelum melakukan tahap *clustering* terlebih dahulu menentukan jumlah kluster optimal. Salah satu metode yang digunakan untuk menentukan K optimal adalah metode elbow yang dilakukan menggunakan *tools* RStudio dengan bahasa pemrograman R. Hasil dari penentuan jumlah *cluster* menggunakan metode elbow akan dijadikan dasar dalam melakukan klasterisasi pada algoritma K-Means. Untuk algoritma metode elbow dalam menentukan nilai K dapat dilihat pada Gambar 3.2



Gambar 3. 2 *Flowchart Metode Elbow*

### 3.2.7 Proses Clustering

Pada tahap ini akan dilakukan segmentasi pelanggan berdasarkan dataset yang telah dibersihkan. Kemudian dilakukan pengelompokan dengan menggunakan algoritma K-Means. Adapun langkah-langkah dalam melakukan *clustering* menggunakan algoritma K-Means adalah sebagai berikut.

1. Tentukan jumlah K yang akan dibentuk.
2. Tentukan *centroid* pusat secara acak.
3. Hitung jarak antara setiap titik data dengan masing-masing *centroid*. Untuk menghitung jarak antara objek dengan *centroid* dapat menggunakan rumus *Euclidian Distance*. *Euclidean distance* dihitung sebagai jarak antara dua titik

dalam ruang *Euclidean* yaitu dengan menghitung akar kuadrat dari jumlah kuadrat selisih antara koordinat titik-titik tersebut.

$$d_{xy} = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_{kx} - y_{ky})^2} \quad (3.1)$$

Keterangan:

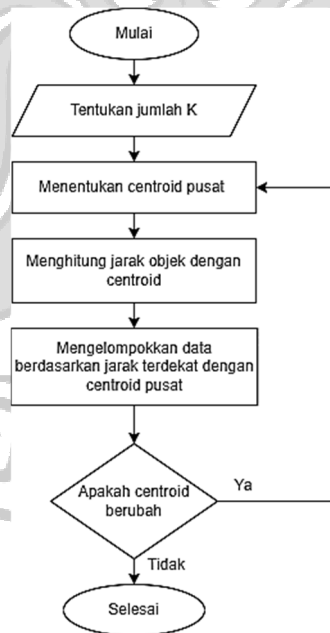
$d_{xy}$  : jarak data x ke centroid y

$k$  : banyaknya objek

$x_{kx}$  : data ke-x pada atribut data ke-k

$y_{ky}$  : data ke-y pada atribut data ke-k

4. Mengalokasikan data ke centroid terdekat.
5. Menentukan titik *center cluster* baru.
6. Verifikasi titik pusat *cluster*



Gambar 3. 3 Flowchart Algoritma *K-Means*

### **3.2.8 Validasi Hasil *Clustering***

Tahap ini dilakukan untuk melakukan validasi hasil *clustering* yang telah dilakukan. Validasi hasil *clustering* ini mencakup penilaian terhadap nilai SSE (*Sum of Squared Errors*). *Sum of Squared Errors* mengukur total jarak kuadrat antara setiap data point dengan pusat *cluster* tempatnya berada. Nilai SSE yang lebih rendah menunjukkan bahwa pengelompokan lebih homogen yang artinya titik data dalam setiap *cluster* lebih dekat ke titik pusat (*centroid*). Oleh karena itu, SSE dapat digunakan sebagai ukuran kualitas hasil *clustering*.

### **3.2.9 Integrasi *Google Spreadsheet* dan *App Script***

Pada tahap ini dilakukan integrasi database hasil *clustering* dengan *App Script*. Integrasi database hasil *clustering* dengan *App Script* dilakukan untuk memungkinkan proses *clustering* berjalan secara otomatis ketika ada penambahan data baru. *App Script* merupakan platform *Google* yang terintegrasi dengan *Google Workspace* dan dapat melakukan otomatisasi.

### **3.2.10 Visualisasi Data**

Pada tahap ini dilakukan visualisasi data dalam bentuk grafik yang bertujuan untuk menampilkan hasil segmentasi yang telah dibuat pada proses *cluster* di tahap sebelumnya. Visualisasi data pada penelitian ini menggunakan *Looker Studio* yang akan diintegrasikan dengan *interface Google Sites*.

### **3.2.11 Analisis Segmentasi Pelanggan dan *Business Intelligence***

Pada tahap ini dilakukan analisis atau identifikasi karakteristik pelanggan pada setiap segmentasi menggunakan algoritma K-Means. Serta analisis BI melalui



visualisasi data dalam bentuk dashboard yang dapat membantu dalam melakukan analisis tren bisnis. Diharapkan dari hasil analisis ini dapat dijadikan rekomendasi bagi perusahaan dalam mengidentifikasi segmentasi pelanggan yang memiliki potensi untuk pengembangan bisnis serta dapat membantu perusahaan untuk membuat keputusan yang lebih baik tentang strategi pemasaran, pengembangan produk, dan peningkatan layanan pelanggan. Dengan memahami karakteristik pelanggan, perusahaan dapat memperkuat hubungan dengan pelanggan dan meningkatkan loyalitas pelanggan.

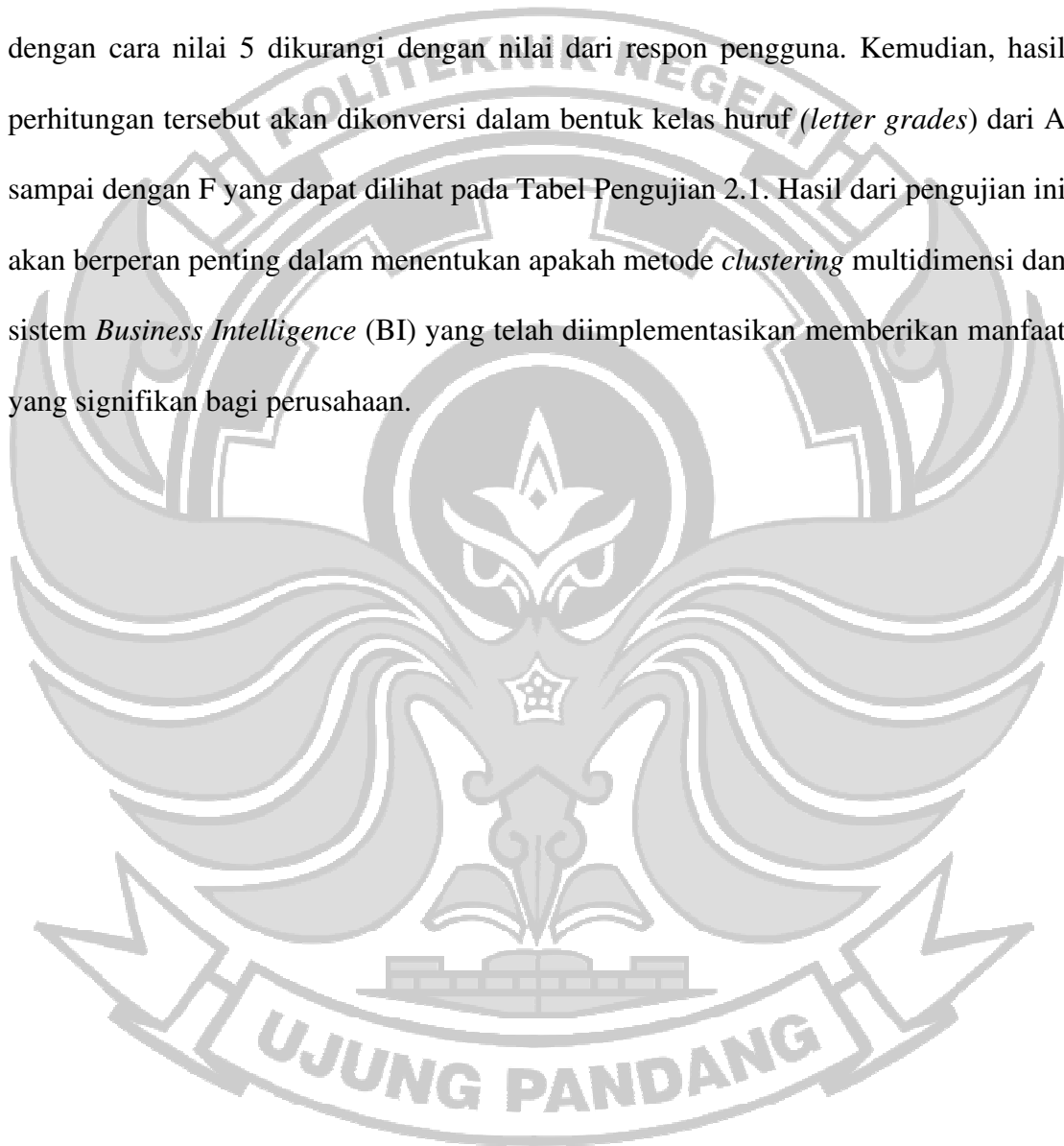
### **3.2.12 Implementasi Sistem**

Pada tahap ini dilakukan pembuatan *interface* sistem agar dapat mempermudah pengguna dalam mengakses dan menjalankan sistem. *Interface* ini akan menampilkan dashboard hasil visualisasi *clustering* yang dilakukan di *Looker Studio* dan menampilkan fitur *upload* dataset jika ada penambahan data baru. Untuk pembuatan *interface* ini menggunakan platform *Google Sites* yang bersifat publik.

### **3.2.13 Pengujian**

Sistem yang telah dibuat akan diuji tingkat kegunaannya menggunakan metode *System Usability Scale* (SUS). Pengujian ini melibatkan 1 responden, yaitu Kepala Pemasaran PT GIB untuk memberikan penilaian terhadap sistem dengan menjawab pertanyaan yang sudah disediakan dalam bentuk kuesioner dengan skala sangat tidak setuju (*strongly disagree*) sampai sangat setuju (*strongly agree*) atau bernilai 1 sampai 5 dan ketika responden tidak menemukan skala respon yang tepat responden harus mengisi titik tengah skala pengujian. Adapun daftar pertanyaan pengujian SUS dapat dilihat pada Lampiran 4.

Hasil dari penilaian yang dilakukan oleh responden akan dihitung menggunakan formula 2.3. Untuk pertanyaan bernomor ganjil dihitung dengan mengurangi nilai dari respon dengan angka 1 dan pertanyaan bernomor genap dihitung dengan cara nilai 5 dikurangi dengan nilai dari respon pengguna. Kemudian, hasil perhitungan tersebut akan dikonversi dalam bentuk kelas huruf (*letter grades*) dari A sampai dengan F yang dapat dilihat pada Tabel Pengujian 2.1. Hasil dari pengujian ini akan berperan penting dalam menentukan apakah metode *clustering* multidimensi dan sistem *Business Intelligence* (BI) yang telah diimplementasikan memberikan manfaat yang signifikan bagi perusahaan.



## BAB IV HASIL DAN PEMBAHASAN

Hasil dari penelitian ini berupa sistem visualisasi hasil segmentasi pelanggan PT Gama Inovasi Berdikari. Berdasarkan prosedur penelitian yang telah dipaparkan pada bab sebelumnya, penelitian ini diawali dengan kegiatan studi literatur sebagai dasar dalam melakukan penelitian serta analisis kebutuhan yang bertujuan untuk melakukan persiapan baik dari segi penyediaan *hardware* maupun *software* yang dapat mendukung proses penelitian berjalan dengan lancar dan baik. Suatu sistem pemantauan diperlukan agar pemantauan berjalan baik didukung oleh ketersediaan data yang akurat. Dengan adanya sistem *business intelligence* dalam pemantauan segmentasi pelanggan pengguna dapat menerima informasi segmen pasar untuk mendukung pengambilan keputusan.

### 4.1 Deskripsi Dataset

Dataset yang digunakan merupakan data transaksi penjualan PT GIB pada periode April 2023 sampai Juni 2023. Dataset tersebut memiliki 16 fitur dan 18 data transaksi. Tampilan dataset transaksi penjualan dapat dilihat pada Tabel 4.1

Tabel 4. 1 Dataset Transaksi Penjualan Periode April 2023 - Juni 2023

OrderNumber	Resi	ProductName	PricePeritem	Discount	TotalPrice	SalesQuantity	TotalPayment	Kota/Kabupaten	Provinsi	OrderPlacementTime	OrderCompletionTime	DeliveryOptions	OrderStatus	Buyer	Gender
230405B MVY7P	100070 074545	Wedang Telang	39.000	5.850	33.150	8	53.050	Kab. Bantul	DI Yogyakarta	2023-04- 05 20:26	2023-04- 09 04:28	Hemat- Anteraja	Selesai	chessyof ficial172	P
230405B MVY7P	100070 074545	Cookies Ubi Ubikies l	3.000	2.100	18.900	8	53.050	Kab. Bantul	DI Yogyakarta	2023-04- 05 20:26	2023-04- 09 04:28	Hemat- Anteraja	Selesai	chessyof ficial172	P
230416A DGHGB	100070 847858	Teh Jahe Herbal	30.000	0	150.000	6	233.900	Kab. Cilacap	Jawa Tengah	2023-04- 16 22:28	2023-04- 20 14:52	Reguler (Cashless)	Selesai	anggana nda_4	L
230416A DGHGB	100070 847858	Hampers/Pa rcel	83.000	0	83.000	6	233.900	Kab. Cilacap	Jawa Tengah	2023-04- 16 22:28	2023-04- 20 14:52	Reguler (Cashless)	Selesai	anggana nda_4	L
230417B 2AN2J2	100070 852671	Wedang Telang	39.000	5.850	33.150	12	160.550	Kab. Kebume	Jawa Tengah	2023-04- 17 04:41	2023-04- 22 04:20	Hemat- Anteraja	Selesai	chessyof ficial172	P
230417B 2AN2J2 O	100070 852671 46	Cookies Ubi Ubikies l Camilan	3.000	3.000	27.000	12	160.550	Kab. Kebume n	Jawa Tengah	2023-04- 17 04:41	2023-04- 22 04:20	Hemat- Anteraja Econom	Selesai	chessyof ficial172 0	P
230417B 2AN2J2 O	100070 852671 46	Hampers/Pa rcel Lebaran	83.000	0	83.000	12	160.550	Kab. Kebume n	Jawa Tengah	2023-04- 17 04:41	2023-04- 22 04:20	Hemat- Anteraja Econom	Selesai	chessyof ficial172 0	P
230501J DKHW6 DR	004203 427561	Wedang Uwuh Komplit dan	50.000	0	100.000	2	102.000	Kab. Kulon Progo	DI Yogyakarta	2023-05- 01 16:38	2023-05- 06 20:28	Hemat- SiCepat Halu	Selesai	verbysaj a	P
230609V Y1K2T6 II	004215 095364	Wedang Uwuh Komplit	31.500	6.000	57.000	2	58.000	Kota Jakarta Selatan	DKI Jakarta	2023-06- 09 14:36	2023-06- 16 09:24	Hemat- SiCepat Halu	Selesai	yulijaspa triana	P
230616H 04XY32 X	004216 266872	Wedang Telang Komplit	39.000	12.000	66.000	2	67.500	Kab. Bogor	Jawa Barat	2023-06- 16 10:11	2023-06- 18 11:43	Hemat- SiCepat Halu	Selesai	heryanif atmah	P

Adapun tabel yang mendeskripsikan setiap fitur dataset transaksi penjualan PT GIB dapat dilihat pada Tabel 4.2.

Tabel 4. 2 Deskripsi Dataset

<b>Nama Variabel</b>	<b>Deskripsi</b>
OrderNumber	Nomor pesanan
Resi	Nomor untuk paket yang akan dikirim
Date	Tanggal transaksi penjualan
ProductName	Nama setiap produk
Price Peritem	Harga peritem produk
Diskon	Potongan harga
TotalPrice	Harga produk setelah diskon
SalesQuantity	Jumlah produk yang dibeli
TotalPayment	Total harga yang dibayar
Kota/Kabupaten	Alamat <i>customer</i>
Provinsi	Wilayah tempat penjualan
OrderPlacementTime	Waktu pesanan dibuat
OrderCompletionTime	Waktu pesanan selesai
DeliveryOptions	Opsi pengantaran barang
OrderStatus	Status pesanan
Buyer	Account <i>customer</i>
Gender	Jenis kelamin <i>customer</i>
OrderStatus	Status pesanan (telah diproses atau dibatalkan)

## 4.2 Pre-Processing Data

### 4.2.1 Import Library

Hal pertama yang dilakukan sebelum *preprocessing* data adalah *import library* yang dibutuhkan. Adapun *library* yang digunakan adalah sebagai berikut.

- a. Pandas digunakan untuk memproses data yang meliputi pembersihan data, manipulasi data, analisis data, dan memuat dataset.
- b. Numpy (*Numerical Python*) digunakan untuk mengolah objek *array* berbentuk N-dimensional.
- c. *Sklearn.preprocessing* merupakan modul dalam *library Scikit-learn (Sklearn)* yang berfungsi untuk melakukan transformasi data seperti mengkodekan variabel kategorikal menjadi numerik, melakukan normalisasi data numerik, dan lain-lain.
- d. Seaborn sebagai visualisasi data untuk membantu dalam analisis data seperti *scatter plot*, *line plot*, *bar plot*, *heatmap*, dan lain-lain.

### 4.2.2 Data Cleaning

Proses data *cleaning* dilakukan untuk mengecek *outlier*, membersihkan *missing values*, dan mengubah format data pada dataset. Setelah diperiksa setiap *entri* dalam dataset yang digunakan memiliki nilai yang valid dan lengkap untuk setiap atributnya yang dapat dilihat pada Gambar 4.1.

```
missing_values = df.isnull().sum()
print("Jumlah missing value per kolom:")
print(missing_values)
```

```
Jumlah missing value per kolom:
OrderNumber      0
Resi              0
ProductName       0
PricePeritem     0
Discount          0
TotalPrice       0
SalesQuantity    0
TotalPayment     0
Kota/Kabupaten  0
Provinsi         0
OrderPlacementTime 0
OrderCompletionTime 0
DeliveryOptions  0
OrderStatus      0
Buyer            0
Gender           0
dtype: int64
```

Gambar 4. 1 Hasil Pengecekan Data yang Kosong

Pada Gambar 4.2 area yang ditandai dengan kotak merah menunjukkan hasil pengecekan data *outlier*. Terdapat 5 data *outlier* pada atribut *PricePeritem* dan masing-masing 1 data pada atribut *discount* dan *TotalPrice* yang terdeteksi sebagai *outlier*. *Outlier* yang terdeteksi tidak dihapus karena memiliki nilai bisnis dan sesuai dengan tujuan bisnis yang akan dianalisis.

```

q1 = df['PricePeritem'].quantile(0.25)
q3 = df['PricePeritem'].quantile(0.75)
iqr = q3 - q1
lower_bound = q1 - 1.5 * iqr
upper_bound = q3 + 1.5 * iqr
outliers = df[(df['PricePeritem'] < lower_bound) | (df['PricePeritem'] > upper_bound)]
print("Jumlah outlier PricePeritem:", len(outliers))

q1 = df['Discount'].quantile(0.25)
q3 = df['Discount'].quantile(0.75)
iqr = q3 - q1
lower_bound = q1 - 1.5 * iqr
upper_bound = q3 + 1.5 * iqr
outliers = df[(df['Discount'] < lower_bound) | (df['Discount'] > upper_bound)]
print("Jumlah outlier Discount:", len(outliers))

q1 = df['TotalPrice'].quantile(0.25)
q3 = df['TotalPrice'].quantile(0.75)
iqr = q3 - q1
lower_bound = q1 - 1.5 * iqr
upper_bound = q3 + 1.5 * iqr
outliers = df[(df['TotalPrice'] < lower_bound) | (df['TotalPrice'] > upper_bound)]
print("Jumlah outlier TotalPrice:", len(outliers))

q1 = df['SalesQuantity'].quantile(0.25)
q3 = df['SalesQuantity'].quantile(0.75)
iqr = q3 - q1
lower_bound = q1 - 1.5 * iqr
upper_bound = q3 + 1.5 * iqr
outliers = df[(df['SalesQuantity'] < lower_bound) | (df['SalesQuantity'] > upper_bound)]
print("Jumlah outlier SalesQuantity:", len(outliers))

q1 = df['TotalPayment'].quantile(0.25)
q3 = df['TotalPayment'].quantile(0.75)
iqr = q3 - q1
lower_bound = q1 - 1.5 * iqr
upper_bound = q3 + 1.5 * iqr
outliers = df[(df['TotalPayment'] < lower_bound) | (df['TotalPayment'] > upper_bound)]
print("Jumlah outlier TotalPayment:", len(outliers))

```

```

Jumlah outlier PricePeritem: 5
Jumlah outlier Discount: 1
Jumlah outlier TotalPrice: 1
Jumlah outlier SalesQuantity: 0
Jumlah outlier TotalPayment: 0

```

Gambar 4. 2 Hasil Pengecekan *Outlier*

Pada Tabel 4.1 menunjukkan tipe data pada atribut *ProductName*, *Kota/Kabupaten*, *Provinsi*, *Buyer*, dan *Gender* bukanlah tipe data numerik sehingga terlebih dahulu diubah menjadi numerik agar proses klusterisasi dapat optimal. Atribut tersebut akan ditransformasi menjadi bentuk numerik melalui proses *encoding* menggunakan *LabelEncoder* dari modul *Sklearn.preprocessing*. *Labelencoder* akan menetapkan nomor unik untuk setiap kelas data mulai dari 0 berdasarkan urutan abjad



dari data yang akan dilabeli. Dari hasil pelabelan yang dilakukan rincian pelabelan setiap data dapat dilihat pada Gambar 4.3 yang ditunjukkan dalam kotak merah.

```

from sklearn.preprocessing import LabelEncoder, MinMaxScaler
label_encoder = LabelEncoder()
df['ProductName'] = label_encoder.fit_transform(df['ProductName'])
print(label_encoder.classes_)
print(np.sort(df['ProductName'].unique()))
print('')

df['Kota/Kabupaten'] = label_encoder.fit_transform(df['Kota/Kabupaten'])
print(label_encoder.classes_)
print(np.sort(df['Kota/Kabupaten'].unique()))
print('')

df['Provinsi'] = label_encoder.fit_transform(df['Provinsi'])
print(label_encoder.classes_)
print(np.sort(df['Provinsi'].unique()))
print('')

df['Buyer'] = label_encoder.fit_transform(df['Buyer'])
print(label_encoder.classes_)
print(np.sort(df['Buyer'].unique()))
print('')

df['Gender'] = label_encoder.fit_transform(df['Gender'])
print(label_encoder.classes_)
print(np.sort(df['Gender'].unique()))
print('')

['(BELI 2 LEBIH HEMAT) WEDANG UMUH KOMPLIT INSTANT'
'Camilan Sehat | Keripik KASABI'
'Cookies Ubi UBIKIES | Camilan Sehat Gluten Free | Sachet'
'Hampers/Parcel Lebaran Idul Fitri | Hampers by Gamafood'
'Teh jahe herbal Gamafood | Ginger tea minuman jahe | detox tea'
'Teh serai herbal Gamafood | lemongrass tea | Teh Detox'
'WEDANG TELANG KOMPLIT ORGANIK tinggal seduh / Teh Telang Kering / Teh Bunga telang'
'Wedang Uuwuh Komplit Celup Gamafood | minuman herbal antioksidan'
'Wedang Uuwuh komplit instan Gamafood | Teh herbal antioksidan | tea detox']
[0 1 2 3 4 5 6 7 8]

['KAB. BADUNG' 'KAB. BANJARNEGARA' 'KAB. BANTUL' 'KAB. BOGOR'
'KAB. CILACAP' 'KAB. KEBUMEN' 'KAB. KULON PROGO' 'KOTA JAKARTA SELATAN']
[0 1 2 3 4 5 6 7]

['BALI' 'DI YOGYAKARTA' 'DKI JAKARTA' 'JAWA BARAT' 'JAWA TENGAH']
[0 1 2 3 4]

['anggananda_4' 'chessyofficial1720_' 'floretta.ratna' 'gp066111'
'heryanifatmah' 'kamell.officialsr' 'pzw380dp1x' 'verbysaja'
'yulijaspatriana']
[0 1 2 3 4 5 6 7 8]

['L' 'P']
[0 1]

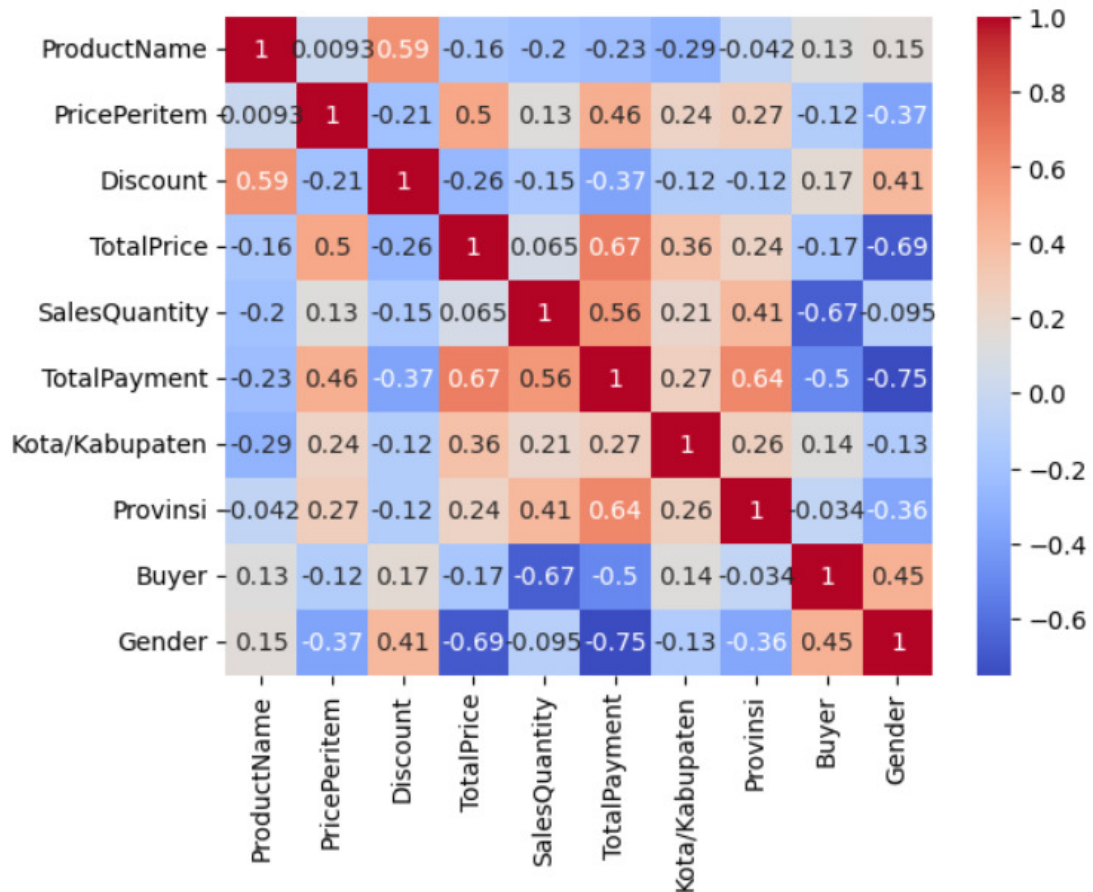
```

Gambar 4. 3 Proses *Encoding*

### 4.3 Pemilihan Variabel

Data yang sudah didapatkan selanjutnya akan dipilih variabel apa saja yang akan digunakan untuk proses *clustering* melalui proses pemilihan variabel yang mempengaruhi tingkat penjualan dan dibuktikan melalui proses pengecekan variabel

yang saling berkorelasi dengan menggunakan metode korelasi *Product Moment* (*Pearson*). Berdasarkan faktor yang mempengaruhi tingkat penjualan dan hasil korelasi setiap variabel yang telah dilakukan akan dipilih 10 variabel dari 16 variabel pada data transaksi penjualan. Variabel yang dimaksud adalah *ProductName*, *PricePeritem*, *Discount*, *TotalPrice*, *SalesQuantity*, *TotalPayment*, *Kota/Kabupaten*, *Provinsi*, *Buyer*, dan *Gender*. Adapun hasil korelasi antar variabel dapat dilihat pada Gambar 4.4.



Gambar 4. 4 Korelasi Setiap Variabel

Pada Gambar 4.4 dapat diketahui bahwa matriks korelasi tersebut menunjukkan koefisien korelasi antar variabel. Jika nilai  $r$  mendekati 1 atau -1 menunjukkan hubungan yang kuat antar kedua variabel, sedangkan jika nilai  $r$  mendekati 0 maka menunjukkan hubungan kedua variabel lemah. Pada gambar diagonal utama memiliki nilai 1 yang menunjukkan bahwa setiap variabel memiliki korelasi yang kuat pada dirinya sendiri. Dan sama halnya korelasi antara variabel *TotalPrice* dan *TotalPayment* memiliki nilai  $r=0.67$  yang menunjukkan kedua variabel tersebut memiliki korelasi yang kuat. Sedangkan korelasi antara variabel *ProductName* dan *PricePeritem* memiliki nilai 0.0093 yang menunjukkan hubungan antara kedua variabel sangat lemah. Adapun rekapitulasi korelasi setiap variabel dapat dilihat pada Tabel 4.3.

Tabel 4. 3 Matriks Korelasi Setiap Variabel

No	Variabel	Korelasi		
		Lemah	Moderat	Kuat
1	ProductName    PricePeritem	0.0093		
2	ProductName    Discount		0.59	
3	ProductName    Buyer	0.13		
4	ProductName    Gender	0.15		
5	TotalPrice    TotalPayment			0.67
6	TotalPayment    Provinsi		0.64	
7	PricePeritem    TotalPrice		0.5	

No	Variabel	Korelasi		
		Lemah	Moderat	Kuat
8	ProductName    TotalSales		0.46	
9	Discount    ProductName		0.59	
10	Discount    Gender		0.41	
11	SalesQuantity    TotalPayment		0.56	
12	SalesQuantity    Provinsi		0.41	
13	TotalPrice    Kota/Kabupaten		0.36	
14	TotalPrice    Provinsi	0.24		
15	TotalPayment    Gender			-0.75
16	SalesQuantity    Buyer			-0.67
17	TotalPrice    Gender			-0.69

Berdasarkan rekapitulasi korelasi matriks pada Tabel 4.3 didapatkan 5 variabel yang memiliki korelasi yang kuat yakni 2 variabel dengan korelasi positif yaitu korelasi antara variabel *TotalPrice* dengan *TotalPayment* serta 3 variabel lainnya dengan korelasi negatif. Dan juga menunjukkan adanya hubungan positif yang moderat antar variabel.

#### 4.4 Tahapan Klasterisasi

##### 4.4.1 Load Dataset pada Google Spreadsheet

Pada penelitian ini *Google Spreadsheet* sebagai database yang berfungsi untuk menyimpan data yang telah di *preprocessing* dan hasil *clustering*. Untuk melakukan

*load* data ke *Google Spreadsheet* dapat melalui *Google Drive*. Dataset yang telah diolah di *Google Collabs* akan disimpan di *Google Drive* dengan format CSV (*Comma Separated Values*). Adapun *script* yang digunakan untuk menyimpan dataset ke *Google Drive* dapat dilihat pada Gambar 4.5.

```
from google.colab import drive
drive.mount('/content/drive')
df.to_csv('/content/drive/MyDrive/Preprocessing Data/dataset.csv', index=False)
```

Mounted at /content/drive

Gambar 4. 5 Simpan Dataset ke Google Drive

#### 4.4.2 Singkronisasi Google Spreadsheets

Singkronisasi antara RStudio dengan *Google Spreadsheet* dilakukan untuk *import dataset* ke lingkungan kerja R. Singkronisasi dilakukan dengan menggunakan *package googlesheet4* dan *function read.csv* atau *read.table* untuk membaca *dataset*. Proses singkronisasi dapat dilihat pada Gambar 4.6.

```
> df <- read_sheet('1FP629z37Re7M-r4jTopPbpexDnnQIiknrJvwKK_dm2Q')
The googlesheets4 package is requesting access to your Google account.
Enter '1' to start a new auth process or select a pre-authorized account.
1: send me to the browser for a new auth process.
2: annisajafar14@gmail.com
3: hairatunnisa14@gmail.com
Selection: 2
Auto-refreshing stale OAuth token.
✓ Reading from dataset.
✓ Range dataset.
>
```

Gambar 4. 6 Proses Autentikasi

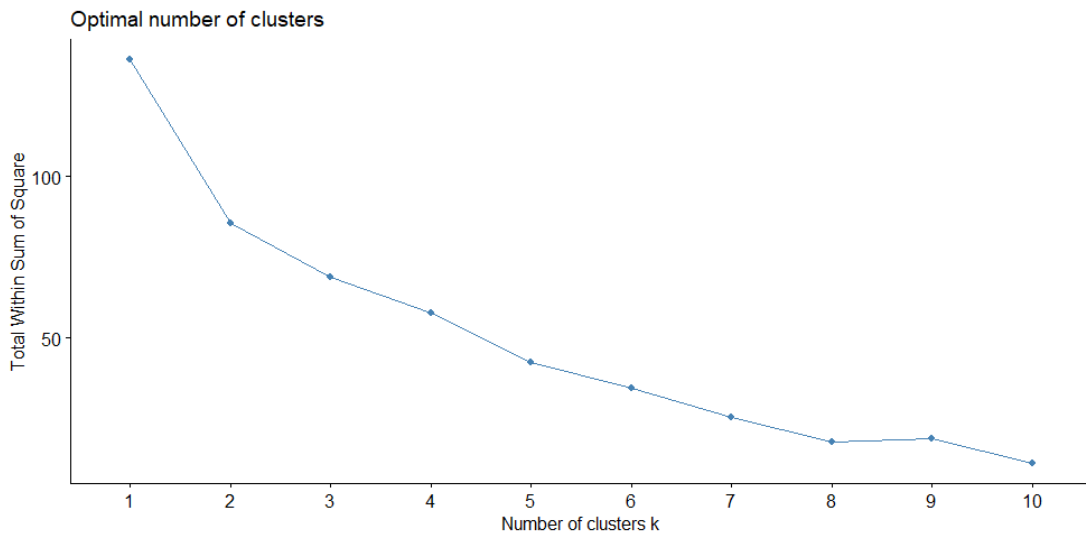
Pada Gambar 4.6 menunjukkan proses *load* data dari *Spreadsheets* dengan nama dataset “dataset” berada di *sheet* “dataset”. Tampilan dari *dataset* yang telah di *load* ke Rstudio dapat dilihat pada Gambar 4.7.

	OrderNumber	Resi	ProductName	PricePeritem	Discount	TotalPrice	SalesQuantity	TotalPayment	Kota.Kabupaten	
1	230405BMVY7PA2	10007007454520		6	39.0	5.85	33.15	8	53.050	2
2	230405BMVY7PA2	10007007454520		2	3.0	2.10	18.90	8	53.050	2
3	230416ADGHGBUD	10007084785874		4	30.0	0.00	150.00	6	233.900	4
4	230416ADGHGBUD	10007084785874		3	83.0	0.00	83.00	6	233.900	4
5	230417B2AN2J2Q	10007085267146		6	39.0	5.85	33.15	12	160.550	5
6	230417B2AN2J2Q	10007085267146		2	3.0	3.00	27.00	12	160.550	5
7	230417B2AN2J2Q	10007085267146		3	83.0	0.00	83.00	12	160.550	5
8	230501JDKHW6DB	004203427561		0	50.0	0.00	100.00	2	102.000	6

Gambar 4. 7 Tampilan Dataset di RStudio

#### 4.4.3 Penentuan Jumlah Cluster

Sebelum melakukan proses *clustering* maka terlebih dahulu menentukan jumlah *cluster* optimal. Untuk menentukan jumlah *cluster* optimal menggunakan metode *elbow*. Dalam penentuan jumlah *cluster* optimal menggunakan metode *elbow* yaitu dengan melihat grafik yang membentuk siku. Grafik pada Gambar 4.8 menunjukkan penurunan nilai yang signifikan dari *cluster* 1 ke *cluster* 2 dan nilai distorsinya (nilai *Sum of Squared Errors*) sudah mulai stabil. Dengan demikian jumlah *cluster* yang digunakan adalah 2 *cluster* dan 3 *cluster* untuk perbandingan. Metode *elbow* ini dicari menggunakan bahasa pemrograman R pada *tools* RStudio. Adapun *output* yang dihasilkan dapat dilihat pada Gambar 4.8.



Gambar 4. 8 Grafik Metode Elbow

#### 4.4.4 Proses Clustering

Data yang sudah diolah akan di *cluster* menggunakan algoritma K-Means dan dijalankan pada tools RStudio. Jumlah *cluster* yang digunakan adalah 2 dan 3 *cluster*. Sebelum melakukan *clustering* terlebih dahulu membuat data frame baru yang berisi variabel *ProductName*, *PricePeritem*, *Discount*, *TotalPrice*, *SalesQuantity*, *TotalPayment*, Kota.Kabupaten, Provinsi, *Buyer*, dan *Gender*. 10 kolom yang dipilih memiliki indeks 3 hingga 16 dalam data frame. Tindakan ini memungkinkan untuk mengambil *subset* data yang hanya berisi kolom-kolom tertentu sesuai dengan indeks yang telah ditentukan. Indeks 3 ditujukan untuk variabel *ProductName*, indeks 4 (*PricePeritem*), indeks 5 (*Discount*), indeks 6 (*TotalPrice*), indeks 7 (*SalesQuantity*), indeks 8 (*TotalPayment*), indeks 9 (Kota.Kabupaten), indeks 10 (Provinsi), indeks 15 (*Buyer*), dan 16 adalah indeks variabel *Gender*. *Subset* data dapat dilihat pada Gambar 4.9.

```
> data <- df[, c(3, 4, 5, 6, 7, 8, 9, 11)]
```

Gambar 4. 9 *Select Variable* untuk *Clustering*

Selanjutnya melakukan *clustering* pada data frame yang telah dibuat, masing-masing proses *clustering* dilakukan sebagai berikut.

a. *Clustering* dengan 2 cluster

Gambar 4.10 menunjukkan *script* yang digunakan untuk proses *clustering* dengan parameter 2 cluster dan *random set* atau *nstart* sebanyak 25 yang artinya proses klusterisasi akan dilakukan sebanyak 25 kali secara acak dan seluruh hasil perhitungan akan dimasukkan ke dalam variabel “*clustering*”.

```
> clustering <- kmeans(data, centers=2, nstart=25)
```

Gambar 4. 10 Proses *Clustering* dengan KMeans

Setelah proses pada Gambar 4.10 dilakukan didapatkan hasil yang dapat dilihat pada Gambar 4.11. Proses *clustering* yang dilakukan mendapatkan akurasi 62.2% atau kualitas *clustering* cukup baik.

```
> clustering
K-means clustering with 2 clusters of sizes 13, 5

Cluster means:
  ProductName PricePeritem Discount TotalPrice SalesQuantity TotalPayment Kota.Kabupaten Provinsi
1  4.769231    30.30769 3.957692   39.42692     3.230769     62.39154     2.615385 1.846154
2  3.600000    47.60000 1.770000   75.23000     9.600000    189.89000     4.600000 4.000000

  Buyer Gender
1 4.153846    1.0
2 0.600000    0.6

Clustering vector:
[1] 1 1 2 2 2 2 2 1 1 1 1 1 1 1 1 1 1

within cluster sum of squares by cluster:
[1] 18094.71 21227.18
   (between_SS / total_SS =  62.2 %)

Available components:

[1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss" "betweenss"
[7] "size"         "iter"         "ifault"
> |
```

Gambar 4. 11 Hasil *Clustering* dengan 2 Cluster



Gambar 4.11 menunjukkan dua *cluster* berhasil dibentuk. *Cluster* pertama terdiri dari 13 data dan *cluster* kedua terdiri dari 5 data. Untuk informasi mengenai masing-masing *centroid* dapat ditemukan pada bagian "*Cluster Means*". *Clustering vector* menunjukkan label yang mewakili setiap data dalam *cluster*, seperti pada Gambar 4.11 terlihat bahwa data yang berada di urutan pertama dimasukkan ke dalam *cluster* 1 dan data yang berada pada urutan ketiga dimasukkan ke dalam *cluster* 2.

b. *Clustering* dengan 3 *cluster*

Pada tahap ini dilakukan *clustering* dengan 3 *cluster* untuk membandingkan hasil *clustering* yang telah dilakukan sebelumnya. Proses *clustering* untuk 3 *cluster* dapat dilihat pada Gambar 4.12.

```
> clustering <- kmeans(data, centers=3, nstart=25)
```

Gambar 4. 12 Proses *Clustering* KMeans dengan 3 *Cluster*

Pada Gambar 4.12 menunjukkan hasil *clustering* dengan 3 *cluster* memiliki persebaran unit data dengan rincian *cluster* 1 memiliki 12 anggota, *cluster* 2 memiliki 2 anggota, dan *cluster* 3 memiliki 4 anggota. Dari 3 *cluster* yang dihasilkan secara keseluruhan nilai antar anggota dalam satu *cluster* (*within cluster sum of squares*) *cluster* 2 (3649.500) dan *cluster* 3 (9922.161) lebih rendah dari nilai *within cluster sum of squares* *cluster* 1 (11934.971) yang mengindikasikan adanya kepadatan yang baik dan homogenitas dalam kelompok tersebut. Dengan nilai *between\_SS* (*between sum of squares*) atau nilai antar *cluster* dengan *cluster*

lainnya dibagi nilai total\_SS (*total sum squares*) sebesar 76.5% atau akurasi tersebut menggambarkan *cluster* yang terbentuk memiliki perbedaan yang cukup signifikan antara satu sama lain yang dapat merepresentasikan seberapa berbeda nilai antar *cluster* dengan *cluster* lainnya. Hasil *clustering* dapat dilihat pada Gambar 4.13.

```
> clustering
K-means clustering with 3 clusters of sizes 12, 2, 4

Cluster means:
  ProductName PricePeritem Discount TotalPrice SalesQuantity TotalPayment Kota.Kabupaten Provinsi
1  5.166667    28.66667    4.2875    34.37917     3.333333     59.09083     2.333333 1.916667
2  3.500000    56.50000    0.0000   116.50000     6.000000    233.90000     4.000000 4.000000
3  2.750000    43.75000    2.2125    60.78750     9.500000    145.91250     5.250000 3.250000
  Buyer Gender
1 3.916667    1
2 0.000000    0
3 2.500000    1

Clustering vector:
[1] 1 1 2 2 3 3 3 3 1 1 1 1 1 1 1 1 1 1 1 1

within cluster sum of squares by cluster:
[1] 11934.971 3649.500 9922.161
 (between_SS / total_SS = 75.5 %)

Available components:
[1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss" "betweenss"
[7] "size"         "iter"         "ifault"
> |
```

Gambar 4. 13 Hasil *Clustering* dengan 3 Cluster

Berdasarkan Tabel 4.4 dapat dilihat bahwa hasil *clustering* dengan jumlah 3 *cluster* lebih baik dibandingkan dengan hasil *clustering* dengan 2 *cluster*.

Tabel 4. 4 Perbandingan Hasil *Clustering*

Jumlah Cluster	Between_SS / Total_SS
2	0.622
3	0.755

#### 4.4.5 Validasi Hasil *Clustering*

Proses validasi dilakukan untuk melihat kualitas atau kekuatan *cluster* dengan menggunakan perhitungan *Sum of Squares Error* (SSE). *Script* yang digunakan untuk mencari nilai SSE setiap *cluster* dapat dilihat pada Gambar 4.14.

```
> sse <- sum(clustering$withinss)
```

Gambar 4. 14 Mencari Nilai SSE Setiap *Cluster*

Berdasarkan tabel 4.5 nilai SSE dari masing-masing *cluster* yang mengalami penurunan paling besar adalah K=2 atau semakin rendah nilai SSE yang dihasilkan maka akan semakin bagus *cluster* yang terbentuk. Namun, akurasi *cluster* yang terbentuk kurang baik seperti yang ditampilkan pada Gambar 4.11 yang menunjukkan akurasi *clustering* dengan jumlah K=2 adalah 62.2%. Maka dari itu untuk meningkatkan akurasi *cluster* dan nilai SSE masih berada di rentang kecil seperti pada Tabel 4.5 yang diberi warna biru yaitu dengan meningkatkan nilai K menjadi 3 dengan akurasi sebesar 75.5% seperti yang ditampilkan pada Gambar 4.13 dan nilai SSE.

Tabel 4. 5 Hasil Validasi *Clustering*

Jumlah K	Nilai SSE	Selisih
1	136	0
2	85.50009	50.49991
3	66.35388	19.14621
4	53.68341	12.67047
5	42.25248	11.43093
6	32.6887	9.56378
7	25.47643	7.212267

Jumlah K	Nilai SSE	Selisih
8	17.71171	7.764723
9	13.21444	4.497268

#### 4.4.6 Simpan Dataset

Dataset hasil *clustering* yang dilakukan di RStudio akan disimpan di *database Google Spreadsheets* yang diberi nama “Hasil\_Klastering” dengan menggunakan *script* pada Gambar 4.15.

```
> sheet_name <- "Hasil_Klastering"
> write_sheet(final_data, sheet = sheet_name)
✓ Creating new sheet: fightable-nabarlek.
```

Gambar 4. 15 Upload Dataset Hasil Clustering

Untuk mempermudah integrasi antara RStudio dan *Spreadsheets* terdapat pilihan untuk mengintegrasikan hasil *clustering* langsung ke *Spreadsheets*. Langkah awalnya adalah dengan membuat sebuah *sheet* baru yang akan digunakan untuk menampung hasil *clustering*. Setelah itu, data hasil *clustering* dapat dengan mudah ditambahkan ke dalam lembar kerja yang telah disiapkan sebelumnya. Dengan menjalankan *script* pada Gambar 4.15 proses tersebut akan secara otomatis membuat lembar kerja baru dengan nama yang dihasilkan secara otomatis.

#### 4.5 Integrasi Google Spreadsheets dan App Script

Ketika terjadi penambahan data pada dataset, *App Script* akan otomatis menginisiasi proses *clustering* untuk mengidentifikasi pola dan kelompok baru. Proses ini akan dipicu oleh fungsi *trigger* yang telah ditetapkan untuk merespon setiap inputan

data yang masuk. Hasil *clustering* ini kemudian akan diintegrasikan secara langsung ke dalam dashboard visualisasi yang telah dibuat menggunakan *Looker Studio*.

Pada *App Script*, fungsi *doPost* digunakan untuk memproses file yang diupload ke direktori *Google Drive*. Fungsi ini menggunakan ID yang telah ditetapkan untuk direktori tujuan kemudian menambahkan data yang diupload ke dalam database *Spreadsheets*. Selain itu, fungsi *doPost* juga digunakan untuk melakukan proses *clustering* yang diterapkan pada data yang telah diupload. Setelah proses *clustering* berlangsung variabel *message* pada *interface* akan menampilkan pesan konfirmasi bahwa file berhasil diupload dan telah di *clustering* yang dapat dilihat pada Gambar 4.16.

```
function doPost(e) {
  Logger.log(JSON.stringify(e));

  var destination_id = '117VBj_UXJANT5kBcCbDBjZbHc0ttZS7N';
  var destination = DriveApp.getFolderById(destination_id);

  var data = Utilities.base64Decode(e.parameter.fileData);
  var blob = Utilities.newBlob(data, e.parameter.mimeType, e.parameter.fileName);
  var file = destination.createFile(blob);

  var fileId = file.getId();
  addDataToSheet(fileId);

  performClustering();

  var htmlOutput = HtmlService.createTemplateFromFile('Upload');
  htmlOutput.message = 'File berhasil diupload dan diklaster!!!';
  return htmlOutput.evaluate();
}
```

Gambar 4. 16 Fungsi *doPost*

Untuk memperbaharui data pada *Spreadsheets* digunakan fungsi *addDataToSheet* dengan menetapkan URL dari *Spreasheets* yang dituju. Fungsi ini

akan membuka *sheet* lembar kerja dan mencari *sheet* berdasarkan nama yang ditetapkan. Jika *sheet* tersebut belum ada, *sheet* baru akan dibuat dengan *row* yang mencakup kolom “*ProductName*”, “*PricePeritem*”, “*Discount*”, “*TotalPrice*”, “*SalesQuantity*”, “*TotalPayment*”, “*Kota.Kabupaten*”, “*Provinsi*”, “*Buyer*”, dan “*Gender*”, “*File URL*”, dan “*Timestamp*”. URL dari direktori tempat menyimpan file yang diupload diambil menggunakan fungsi *getFileUrl*. Adapun *script* yang digunakan dapat dilihat pada Gambar 4.17.

```
function addDataToSheet(fileId) {
  var url = 'https://docs.google.com/spreadsheets/d/14-5B_0k2_wNiK9wLo3K14bqteVzfjdsGGTk_MUEmTsk/edit#gid=1978881507'; // URL OF GOOGLE SHEET
  var ss = SpreadsheetApp.openByUrl(url);
  var recordsSheet = ss.getSheetByName("DataFile");

  if (!recordsSheet) {
    recordsSheet = ss.insertSheet("DataFile"); // Create the sheet if it doesn't exist
    recordsSheet.appendRow(["ProductName", "ProductCategory", "PricePeritem", "SalesQuantity", "File URL", "Timestamp"]); // Header row
  }

  var username = Session.getActiveUser().getEmail();
  var timestamp = new Date();
  var fileUrl = getFileUrl(fileId);
  var data = getDataFromFile(fileId);

  data.forEach(function(row) {
    recordsSheet.appendRow([row.ProductName, row.ProductCategory, row.PricePeritem, row.SalesQuantity, fileUrl, timestamp]);
  });
}
```

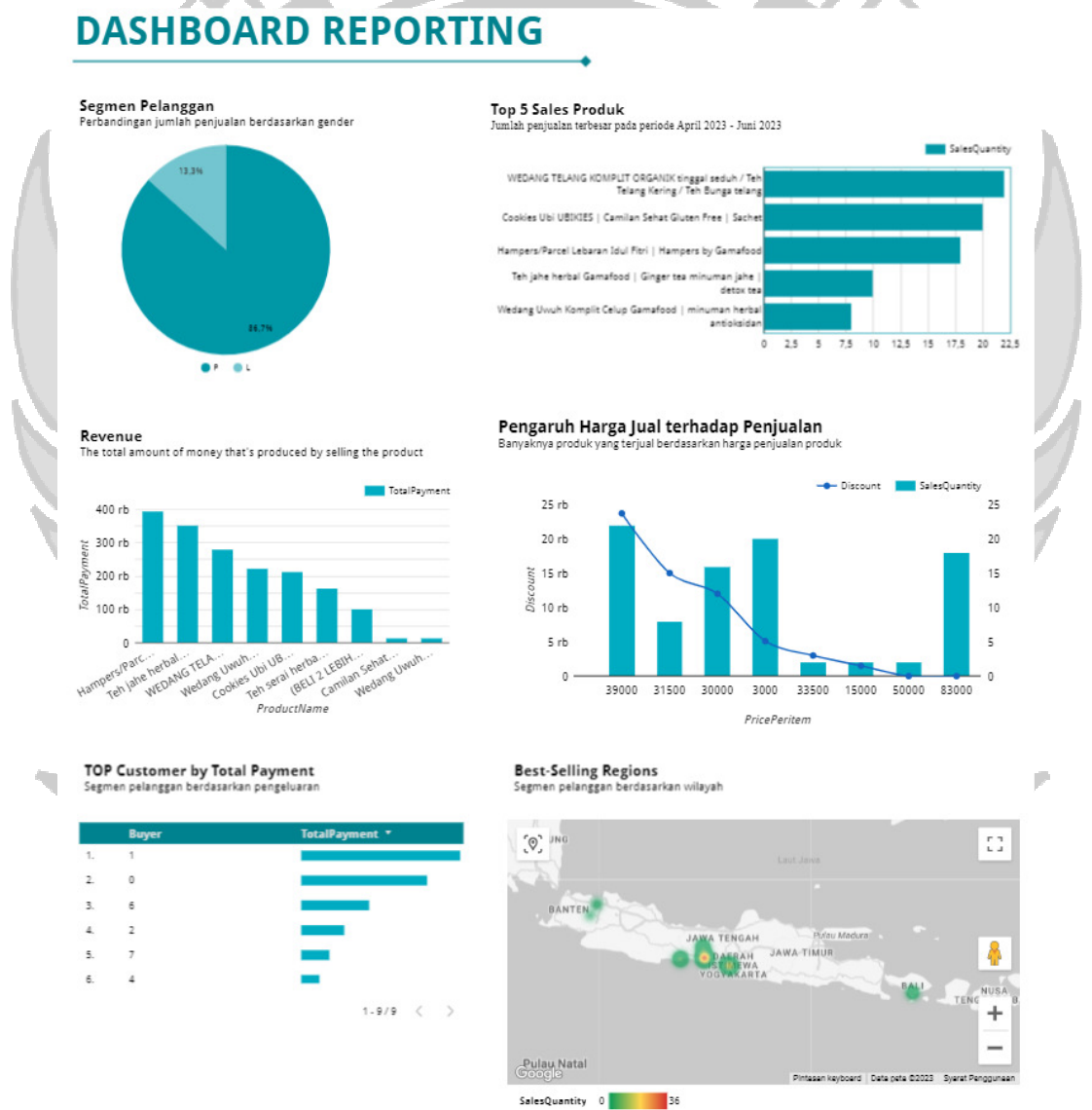
Gambar 4. 17 *Function addDataToSheet*

#### 4.6 Visualisasi

Hasil visualisasi dibuat sebagai sarana untuk membantu tim pemasaran dalam melihat rangkuman hasil *clustering* yang telah dilakukan. Visualisasi ini dibuat dalam bentuk *dashboard* menggunakan alat visualisasi data, yaitu *Looker Studio*. Dashboard visualisasi tersebut kemudian akan diintegrasikan ke dalam *Google Sites* agar

memungkinkan para karyawan dengan mudah mengakses dan melakukan *clustering* jika ada penambahan data baru.

Hasil visualisasi yang diberikan terdiri dari enam *chart* dengan masing-masing kategori *chart*, yaitu *pie chart*, *bar chart*, dan *map chart*. Dashboard atau *chart* yang digunakan pada penelitian ini dapat dilihat pada Gambar 4.18.



Gambar 4. 18 *Dashboard Reporting Customer Segmentation*

a. *Pie Chart*

*Pie chart* digunakan untuk membandingkan segmen pelanggan berupa persentase perbandingan jumlah penjualan berdasarkan *gender*. Dapat diketahui *gender* yang paling banyak melakukan pembelian atau *repeat order* yaitu *gender* perempuan sebesar 86,7% dan *gender* laki-laki hanya 13.3%.

b. *Bar Chart*

Dalam visualisasi yang dilakukan terdapat tiga *bar chart* yang digunakan dengan masing-masing *chart* terdiri dari satu *bar* dan satu metrik. Untuk *bar chart* “*Top 5 Sales Produk*” merupakan lima top produk dengan penjualan terbanyak selama periode April 2023 – Juni 2023 atau dapat disimpulkan produk yang paling diminati pelanggan dan adanya *repeat order* pelanggan yaitu produk herbal Wedang Telang. Dengan mengetahui produk yang paling diminati oleh setiap segmen, perusahaan dapat mengalokasikan stok dengan lebih efektif dan memastikan distribusi yang tepat untuk memenuhi permintaan pelanggan. Sedangkan diagram kombinasi menunjukkan pengaruh harga jual dan pemberian diskon terhadap jumlah penjualan produk. Grafik tabel “*Top Customer by Total Payment*” menampilkan segmen pelanggan dengan pengeluaran terbanyak berdasarkan pembelian produk dan *repeat order* yang dilakukan.

c. *Map Chart*

*Map chart* merupakan bentuk visualisasi geografis yang digunakan untuk menampilkan wilayah (provinsi dan kabupaten/kota) *customers* berdasarkan jumlah pembelian produk. Setiap kabupaten/kota memiliki warna sesuai dengan



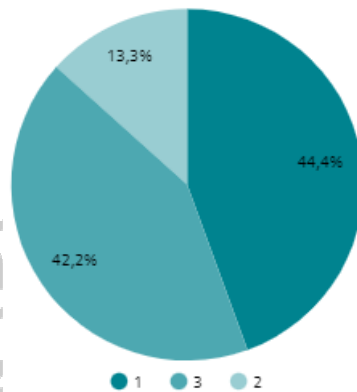
persebaran wilayah atau alamat *customers*. Pada Gambar 4.18 menunjukkan warna merah merupakan warna dengan nilai penjualan produk yang paling tinggi yang berada di Kabupaten Kebumen Jawa Tengah. Untuk warna kuning menunjukkan wilayah dengan tingkat penjualan menengah. Sedangkan warna hijau menunjukkan daerah dengan penjualan produk minim atau rendah.

#### 4.7 Analisis Hasil *Clustering*

Dari hasil visualisasi yang dilakukan dapat ditarik beberapa *insights* yang dapat dijadikan acuan dalam penentuan strategi pemasaran.

##### a. Persentase *cluster*

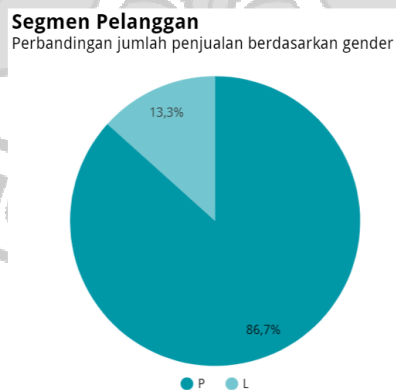
Berdasarkan hasil klasterisasi dari 10 variabel terdapat 3 *cluster* yang terbentuk dan secara umum hasil *cluster* produk dikelompokkan menjadi 3 kategori berdasarkan tingkat penjualannya yaitu rendah, sedang, dan tinggi. Dari grafik persebaran *cluster* pada Gambar 4.19 diperoleh sebesar 44.4% dari total dataset yang tergabung ke dalam *cluster* 1, 13.3% yang tergabung ke dalam *cluster* 2, dan 42.2% dari total dataset yang tergabung ke dalam *cluster* 3. *Cluster* 1 merupakan data penjualan produk yang tergolong tinggi. Untuk data yang tergabung ke dalam *cluster* 3 tergolong jenis produk dengan penjualan sedang. Sedangkan data yang tergabung ke dalam *cluster* 2 tergolong rendah dengan selisih 28.9% dari penjualan produk di *cluster* 3. Dan ketiga *cluster* yang terbentuk didapatkan 5 produk teratas dengan penjualan tertinggi pada periode April 2023 – Juni 2023 yang dapat dilihat pada Gambar 4.18.



Gambar 4. 19 *Persentase Cluster*

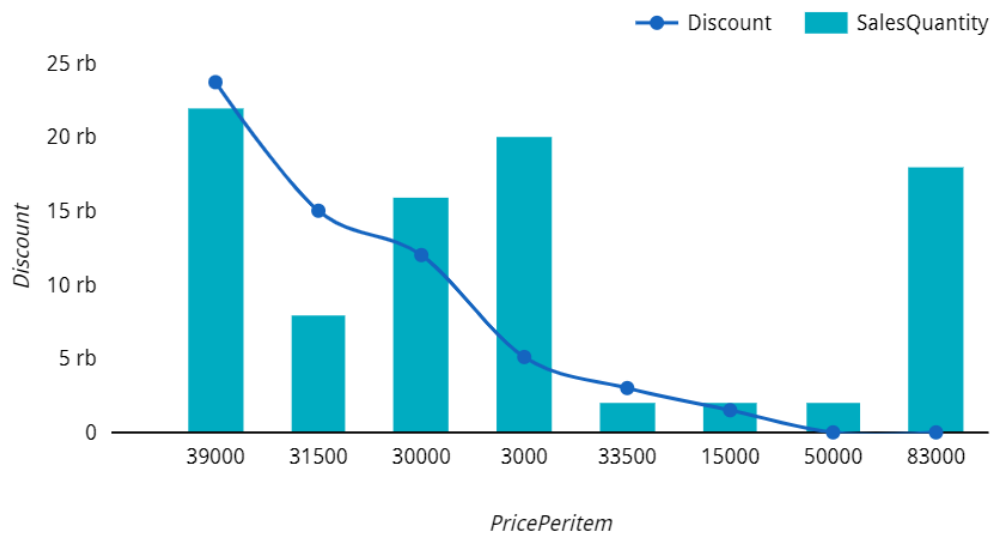
b. Segmen pasar

Data pada grafik *pie chart* pada Gambar 4.20 menunjukkan adanya pengaruh yang cukup kuat dari perbedaan *gender* terhadap minat beli konsumen dengan produk herbal. Terlihat selisih tingkat penjualan sekitar 73,4% dan signifikan yang menunjukkan pembelian produk dan *repeat order* terbanyak dilakukan oleh *customer* dengan *gender* perempuan. Berdasarkan segmentasi tersebut maka dapat dilakukan *event* promosi atau pendekatan ke segmen *gender* laki-laki terkait pola hidup sehat dengan produk herbal.



Gambar 4. 20 Grafik Perbandingan Penjualan Berdasarkan *Gender*

Pada Gambar 4.21 diagram kombinasi bertumpuk yang ditampilkan memberikan informasi adanya hubungan antara penetapan harga jual produk dan pemberian diskon dengan keputusan pelanggan dalam membeli produk. Semakin besar diskon yang diberikan penjualan produk pun ikut meningkat.



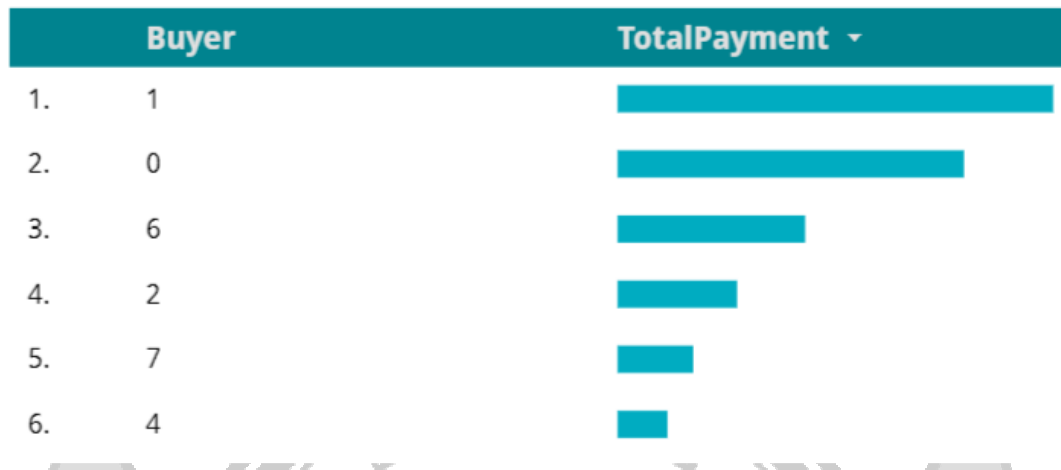
Gambar 4. 21 Grafik Perbandingan Harga Jual dan Pemberian diskon

Untuk persebaran data pada *map chart* didapatkan pola baru terkait jangkauan penjualan berdasarkan geografis. Pada Gambar 4.22 menunjukkan terdapat pembelian produk yang dilakukan di luar pulau Jawa yang artinya ada peluang untuk memperluas segmen pasar.



Gambar 4. 22 Segmen Pelanggan Berdasarkan Geografis

Berdasarkan grafik *bar (table column)* pada Gambar 4.23 segmentasi pelanggan berdasarkan total pembelian yang dilakukan dapat memberikan *insight* dalam mengidentifikasi pola pembelian pelanggan dan dapat memetakan tingkatan pelanggan. Dengan pemahaman yang lebih mendalam tentang perilaku pembelian pelanggan, perusahaan dapat mengambil langkah-langkah strategis untuk meningkatkan retensi pelanggan, meningkatkan nilai transaksi rata-rata, dan mengoptimalkan promosi pemasaran. Selain itu, perusahaan dapat mengembangkan penawaran khusus yang sesuai dengan karakteristik setiap segmen. Dan dengan menggunakan data dari tingkatan pelanggan perusahaan dapat menawarkan insentif khusus kepada pelanggan dengan pembelian lebih tinggi, seperti diskon eksklusif, penawaran bundel, atau hadiah loyalitas. Hal ini akan membantu mendorong pelanggan untuk meningkatkan nilai belanja.



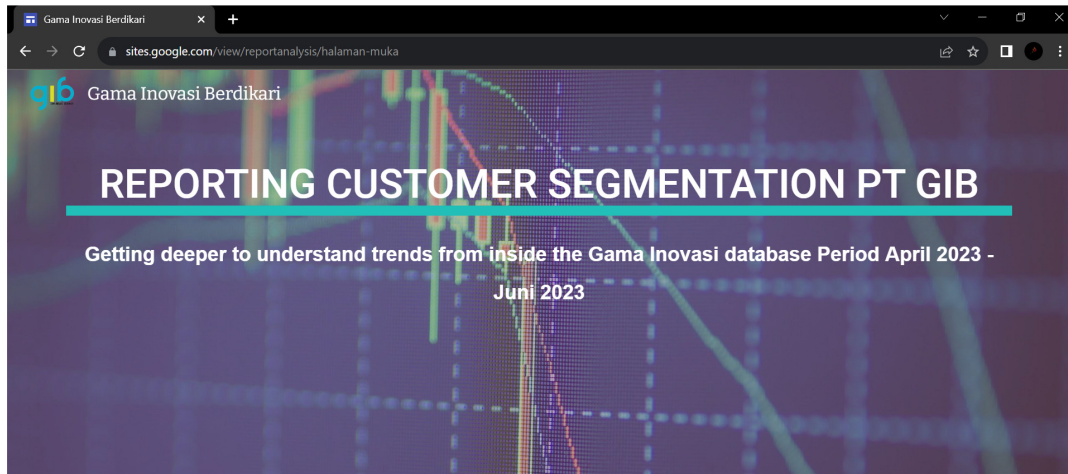
Gambar 4. 23 Grafik *Customer by Total Payment*

#### 4.8 Implementasi Sistem

Untuk menampilkan dashboard hasil klusterisasi dan input data baru dilakukan dengan membuat *interface* berbasis website menggunakan *Google Sites* yang bersifat publik. Sistem ini ditujukan sebagai wadah untuk mempermudah dalam melakukan pengelompokan data baru dan menampilkan informasi dari hasil pengelompokan dalam bentuk grafik dalam menentukan strategi bisnis dan produktivitas.

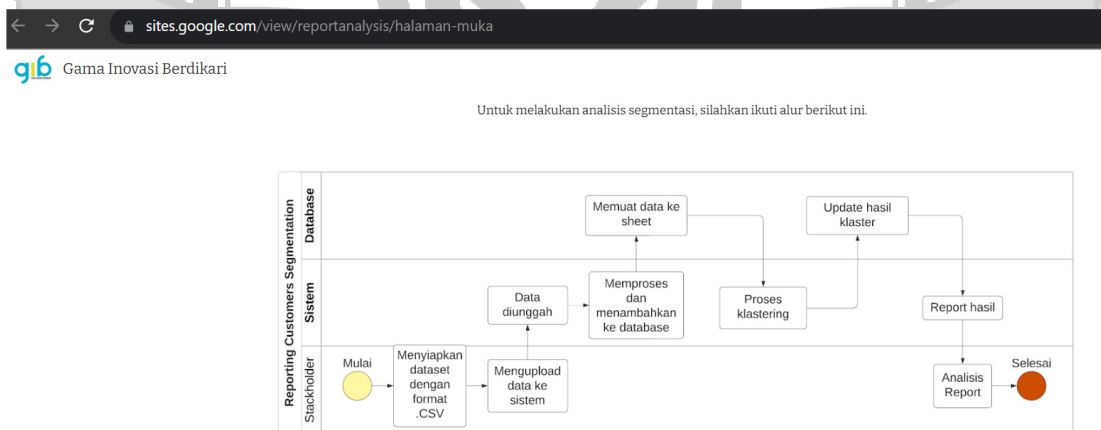
Dalam pembuatan *interface* ada beberapa fitur *Sites* yang digunakan, yaitu fitur integrasi dashboard visualisasi (*Looker Studio*) dan integrasi *App Script* untuk proses klusterisasi data baru. Gambar 4.24 merupakan tampilan *Sites* yang telah diintegrasikan dengan *Looker Studio* dan *App Script*.

*Web interface* yang dibuat terdiri dari 5 bagian, yaitu *header*, diagram alur penggunaan sistem, bagian *upload* dan klusterisasi data tambahan, dan bagian *reporting* hasil *clustering*. Gambar 4.24 adalah tampilan yang menunjukkan *header* dari sistem yang telah dibuat.



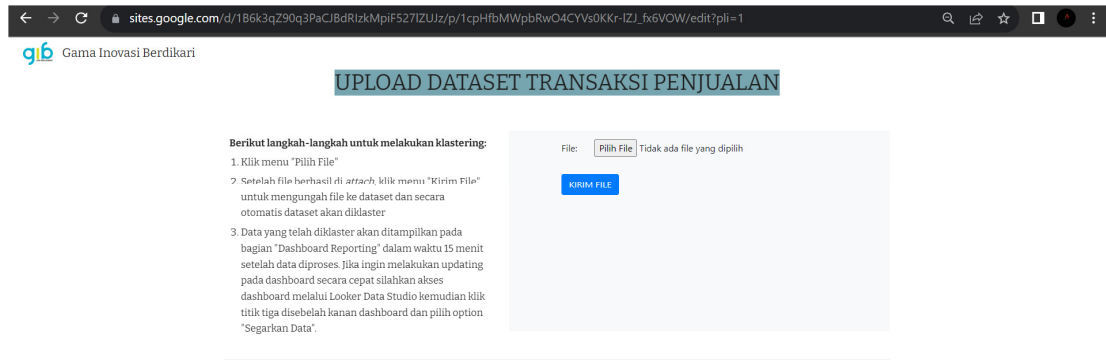
Gambar 4. 24 Tampilan Bagian *Header* Pada Website

Gambar 4.25 menunjukkan diagram alur untuk membantu *user* dalam menggunakan sistem.



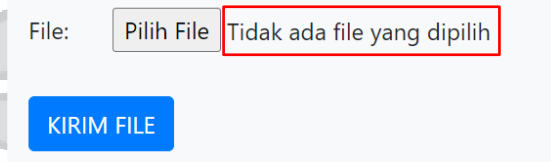
Gambar 4. 25 Tampilan Diagram Alur Penggunaan Sistem

Untuk tampilan upload dan klasterisasi data dapat dilihat pada Gambar 4.26.



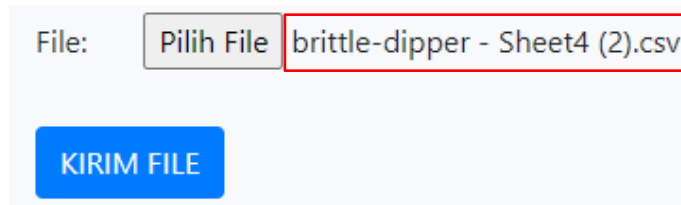
Gambar 4. 26 Tampilan Menu Upload dan Klastering

Jika terdapat data baru yang akan di *cluster*, user dapat melakukan *choose* file dengan format *.CSV (Comma Separated Values)* dengan memperhatikan parameter yang telah ditetapkan. Untuk melakukan *upload* file ke *database Spreadsheets* dapat mengklik *button* "KIRIM FILE". Setelah *dataset* berhasil ditambahkan ke *database* proses klasterisasi dengan algoritma *KMeans* akan secara otomatis diinisiasi sesuai dengan konfigurasi *trigger* yang telah ditetapkan. Tampilan sebelum file atau *dataset* dipilih dapat dilihat Gambar 4.27.



Gambar 4. 27 Tampilan Kondisi Sebelum Pilih File

Untuk tampilan *interface* ketika data diupload dari lokal dan berhasil di *attachment* dapat dilihat pada Gambar 4.28 yang diberi kotak merah.



Gambar 4. 28 Upload File

Setelah data berhasil di *attach* seperti pada Gambar 4.28, selanjutnya klik menu “KIRIM FILE” untuk mengunggah file ke dataset *Spreadsheets*. Ketika data berhasil ditambahkan ke dataset secara otomatis data tersebut akan di *cluster*. Tampilan *interface* setelah data berhasil diunggah dan di *cluster* dapat dilihat pada Gambar 4.29.



Gambar 4. 29 Tambahkan File ke Dataset dan Klasterisasi

Pada *Spreadsheets* terdiri dari dua sheet, yaitu sheet untuk hasil data yang di *cluster* dan sheet untuk data *discretization* atau data yang ditransformasikan ke dalam bentuk kategorikal yang akan diintegrasikan ke *Looker Studio*. Sheet “Hasil\_Klastering” data *discretization* dapat dilihat pada Gambar 4.30.



	A	B	C	D	E	F	G	H	I	J	K
1	ProductName	PricePerItem	Discount	TotalPrice	SalesQuantity	TotalPayment	Kota	Provinsi	Gender	Buyer	Cluster
2	WEDANG TELANG KOMPLIT 39	5.85		33.15	8	53.05	KAB. KEBUMEN	DI YOGYAKARTA	P		1
3	Cookies Ubi UBIKIES   Camil 3	44.928		45.187	8	53.05	KAB. KEBUMEN	DI YOGYAKARTA	P		2
4	Teh jahe herbal Gamafood   30	0		150	6	233.9	KOTA JAKARTA SE	JAWA TENGAH	L		1
5	Hampers/Parcel Lebaran Idu. 83	0		83	6	233.9	KOTA JAKARTA SE	JAWA TENGAH	L		1
6	WEDANG TELANG KOMPLIT 39	5.85		33.15	12	160.55	KAB. BOGOR	JAWA TENGAH	P		1
7	Cookies Ubi UBIKIES   Camil 3	3		27	12	160.55	KAB. BOGOR	JAWA TENGAH	P		1
8	Hampers/Parcel Lebaran Idu. 83	0		83	12	160.55	KAB. BOGOR	JAWA TENGAH	P		1
9	(BELI 2 LEBIH HEMAT) WED/ 50	0		100	2	102	KAB. BANJARNEG	DI YOGYAKARTA	P		1
10	Wedang Uwuh Komplit Celu 45.077	6		57	2	58	KAB. BADUNG	DKI JAKARTA	P		1
11	WEDANG TELANG KOMPLIT 39	12		66	2	67.5	KAB. KULON PR	JAWA BARAT	P		1
12	Wedang Uwuh Komplit Celu 45.077	3		45.074	3	84.5	KAB. CILACAP	JAWA TENGAH	P		2
13	Teh jahe herbal Gamafood   30	3		27	3	84.5	KAB. CILACAP	JAWA TENGAH	P		1
14	Teh serai herbal Gamafood   30	3		27	3	84.5	KAB. CILACAP	JAWA TENGAH	P		1
15	Wedang Uwuh komplit insta 33.5	3		45.076	2	14.995	KAB. KEBUMEN	DI YOGYAKARTA	P		2
16	Camilan Sehat   Keripik KAS 15	45.047		45.059	2	14.995	KAB. KEBUMEN	DI YOGYAKARTA	P		2
17	Teh jahe herbal Gamafood   30	3		27	1	33.57	KAB. BADUNG	DKI JAKARTA	P		1
18	Wedang Uwuh Komplit Celu 45.077	6		57	3	80.215	KAB. BANTUL	BALI	P		1
19	Teh serai herbal Gamafood   30	3		27	3	80.215	KAB. BANTUL	BALI	P		1
20											
21											
22											
23											
24											
25											
26											

Gambar 4. 30 Hasil *Clustering* (Data *Discretization*)

Data yang telah di *cluster* akan dikembalikan ke dataset *Spreadsheets* dengan nama sheet “*Update\_Klastering*” seperti yang ditampilkan pada Gambar 4.31. Gambar yang diberi kotak hijau adalah data yang baru ditambahkan dan gambar yang diberi kotak merah adalah *update* dari hasil *clustering* berdasarkan data baru yang ditambahkan.

	A	B	C	D	E	F	G	H	I	J	K
1	ProductName	PricePerItem	Discount	TotalPrice	SalesQuantity	TotalPayment	Kota	Provinsi	Buyer	Gender	Cluster
2	6	39	5.85	33.15	8	53.05	2	1	1	1	1
3	2	3	2.1	18.9	8	53.05	2	1	1	1	2
4	4	30	0	150	6	233.9	4	4	0	0	1
5	3	83	0	83	6	233.9	4	4	0	0	1
6	6	39	5.85	33.15	12	160.55	5	4	1	1	1
7	2	3	3	27	12	160.55	5	4	1	1	1
8	3	83	0	83	12	160.55	5	4	1	1	1
9	0	50	0	100	2	102	6	1	7	1	1
10	7	31.5	6	57	2	58	7	2	8	1	1
11	6	39	12	66	2	67.5	3	3	4	1	1
12	7	31.5	3	28.5	3	84.5	1	4	6	1	2
13	4	30	3	27	3	84.5	1	4	6	1	1
14	5	30	3	27	3	84.5	1	4	6	1	1
15	8	33.5	3	30.5	2	14.995	2	1	3	1	2
16	1	15	1.5	13.5	2	14.995	2	1	3	1	2
17	4	30	3	27	1	33.57	7	2	5	1	1
18	7	31.5	6	57	3	80.215	0	0	2	1	1
19	5	30	3	27	3	80.215	0	0	2	1	1
20	1	39.000	5.850	33.150	8	53.050	1	1	1	1	1
21	2	3.000	2.100	18.900	8	53.050	1	1	1	1	1
22	3	30.000	0	150.000	6	233.900	0	2	0	0	1
23	1	83.000	0	83.000	6	233.900	0	2	0	0	1
24	2	39.000	5.850	33.150	12	160.550	2	2	1	1	1

Gambar 4. 31 Hasil *Clustering* Setelah Penambahan Data

Pada Gambar 4.31 menunjukkan sebelum adanya penambahan data, data pada baris ketiga masuk ke dalam *cluster* 1. Namun, setelah data baru ditambahkan data tersebut berpindah ke *cluster* 2 yang dapat dilihat pada Gambar 4.30. Begitupun dengan data lainnya yang juga di *cluster* berdasarkan penambahan data baru.

	A	B	C	D	E	F	G	H	I	J	K
1	ProductName	PricePerItem	Discount	TotalPrice	SalesQuantity	TotalPayment	Kota	Provinsi	Buyer	Gender	Cluster
2	6	39	5.85	33.15	8	53.05	2	1	1	1	1
3	2	3	2.1	18.9	8	53.05	2	1	1	1	1
4	4	30	0	150	6	233.9	4	4	0	0	2
5	3	83	0	83	6	233.9	4	4	0	0	2
6	6	39	5.85	33.15	12	160.55	5	4	1	1	3
7	2	3	3	27	12	160.55	5	4	1	1	3
8	3	83	0	83	12	160.55	5	4	1	1	3
9	0	50	0	100	2	102	6	1	7	1	3
10	7	31.5	6	57	2	58	7	2	8	1	1
11	6	39	12	66	2	67.5	3	3	4	1	1
12	7	31.5	3	28.5	3	84.5	1	4	6	1	1
13	4	30	3	27	3	84.5	1	4	6	1	1
14	5	30	3	27	3	84.5	1	4	6	1	1
15	8	33.5	3	30.5	2	14.995	2	1	3	1	1
16	1	15	1.5	13.5	2	14.995	2	1	3	1	1
17	4	30	3	27	1	33.57	7	2	5	1	1
18	7	31.5	6	57	3	80.215	0	0	2	1	1
19	5	30	3	27	3	80.215	0	0	2	1	1
20											
21											
22											
23											
24											
25											
26											

Gambar 4. 32 Hasil *Clustering* Sebelum Penambahan Data

Data hasil *clustering* tersebut kemudian akan ditampilkan pada dashboard *reporting* dalam waktu 15 menit. Namun, jika *user* ingin menampilkan visualisasi dengan cepat yaitu dengan melakukan *refresh data* dengan mengakses *Looker Studio* di bagian bawah *Dashboard Reporting* yang dapat dilihat pada Gambar 4.33.



Gambar 4. 33 Tampilan *Dashboard Reporting*

#### 4.9 Hasil Pengujian *System Usability Scale (SUS)*

Sistem yang sudah dibuat diuji tingkat kegunaannya oleh 1 orang responden, yaitu Kepala Pemasaran PT GIB. Responden mengeksplorasi dashboard yang sudah dibuat. Kemudian, responden mengisikan kuesioner dengan pertanyaan yang sudah disediakan. Hasil dari kuesioner pengujian ini dapat dilihat pada Lampiran 4.

Hasil perhitungan dari pengujian *usability* terhadap sistem pada kuesioner yang diambil dari 1 responden yang diisi oleh Kepala Pemasaran PT GIB sebagai berikut:

$$\begin{aligned} \text{Skor SUS} &= ((R1 - 1) + (5 - R2) + (R3 - 1) + (5 - R4) + (R5 - 1) + \\ &\quad (5 - R6) + (R7 - 1) + (5 - R8) + (R9 - 1) + (5 - R10)) * 2.5) \\ &= ((4 - 1) + (5 - 2) + (4 - 1) + (5 - 2) + (4 - 1) + (5 - 3) + \\ &\quad (4 - 1) + (5 - 2) + (5 - 1) + (5 - 3)) \times 2.5 \\ &= (3 + 3 + 3 + 3 + 3 + 2 + 3 + 2 + 3 + 3 + 4 + 2) \times 2.5 \\ &= 29 \times 2.5 \\ &= 72.5 \end{aligned}$$

Dari hasil perhitungan dengan metode SUS memberikan skor sebesar 72.5. Dengan skor tersebut, tingkat *usability* sistem dalam *percentile ranks* sebesar 0.725% dan termasuk kategori “Baik” dengan *grade scale* B.

## BAB V KESIMPULAN DAN SARAN

### 5.1 Kesimpulan

Berdasarkan hasil pengujian pada penelitian ini, dapat disimpulkan bahwa:

1. Sistem *Business Intelligence* berhasil dibuat menggunakan algoritma K-Means yang dapat melakukan *clustering* secara otomatis terhadap data transaksi penjualan PT Gama Inovasi Berdikari dan telah berjalan sesuai fungsi dari setiap fitur berdasarkan kondisi data inputan baru.
2. Hasil dari pengujian *System Usability Scale* menghasilkan skor 72,5 yang termasuk ke dalam *adjective rating* “Good” yang berarti sistem *Business Intelligence* dapat melakukan *cluster* pelanggan dan memvisualisasikan hasil *cluster* tersebut dengan baik.

### 5.2 Saran

Hal-hal yang dapat dijadikan pertimbangan untuk penelitian selanjutnya, yaitu:

1. Dapat menggunakan metode lain dalam melakukan *clustering* multidimensi untuk melakukan segmentasi pelanggan atau dapat dikombinasikan dengan K-Means *Clustering*.
2. Dapat menggunakan metode lain untuk melakukan otomatisasi *clustering*.

## DAFTAR PUSTAKA

- Ahmed, S. R. A., Al Barazanchi, I., Jaaz, Z. A., & Abdulshaheed, H. R. (2019). Clustering algorithms subjected to K-mean and gaussian mixture model on multidimensional data set. *Periodicals of Engineering and Natural Sciences*, 7(2), 448–457.
- Aprilyani, R., Budianto, A., & Herlina, N. (2020). Pengaruh Karakteristik Produk dan Kebijakan Harga Terhadap Minat Beli Konsumen1. Aprilyani R, Budianto A, Herlina N. Pengaruh Karakteristik Produk dan Kebijakan Harga Terhadap Minat Beli Konsumen. *Bus Manag Entrep*. 2020;2(2):131–46. *Business Management and Entrepreneurship*, 2(2), 131–146.
- Buana, W., & Sari, B. N. (2022). Analisis User Interface Meningkatkan Pengalaman Pengguna Menggunakan Usability Testing pada Aplikasi Android Course. *DoubleClick: Journal of Computer and Information Technology*, 5(2), 91–97.
- Cordeiro, M. C. R., Martinez, J.-M., & Peña-Luque, S. (2021). Automatic water detection from multidimensional hierarchical clustering for Sentinel-2 images and a comparison with Level 2A processors. *Remote Sensing of Environment*, 253, 112209. <https://doi.org/10.1016/j.rse.2020.112209>
- Ekasetya, V. A., & Jananto, A. (2020). Klusterisasi Optimal Dengan Elbow Method Untuk Pengelompokan Data Kecelakaan Lalu Lintas Di Kota Semarang. *Jurnal Dinamika Informatika*, 12(1), 20–28. <https://doi.org/10.35315/informatika.v12i1.8159>
- Fernando, D. (2018). Visualisasi Data Menggunakan Google Data Studio. *Jurnal Pengabdian Masyarakat*, 2(1), 71–77. <https://ejournal.lppmunsera.org/index.php/snartisi/article/view/808>
- Gautam, R., Vanga, S., Ariese, F., & Umopathy, S. (2015). Review of multidimensional data processing approaches for Raman and infrared

spectroscopy. *EPJ Techniques and Instrumentation*, 2.  
<https://doi.org/10.1140/epjti/s40485-015-0018-6>

- Jubaidah, S., & Zulkarnain, M. R. (2020). Penggunaan google sites pada pembelajaran matematika materi pola bilangan SMP kelas VIII SMPN 1 Astambul. *Lentera: Jurnal Ilmiah Kependidikan*, 15(2), 68–73.
- Jumadi, J., Yupianti, Y., & Sartika, D. (2021). Pengolahan Citra Digital Untuk Identifikasi Objek Menggunakan Metode Hierarchical Agglomerative Clustering. *JST (Jurnal Sains Dan Teknologi)*, 10(2), 148–156.
- Karami, A., & Widharto, Y. (2023). Perancangan Business Intelligence dan Segmentasi Pelanggan Menggunakan K Means Clustering Berdasarkan RFM Model. *Industrial Engineering Online Journal*, 12(1).
- Kiat, A. B. H. (2020). *Penerapan Metode K-Means dengan Metode Elbow untuk Segmentasi Pelanggan Menggunakan Model RFM (Recency, Frequency, Monetary)*. Universitas Muhammadiyah Malang.
- Merliana, N. P. E., & Santoso, A. J. (2015). *Analisa Penentuan Jumlah Cluster Terbaik pada Metode K-Means Clustering*.
- Mirantika, N., Syamfithriani, T. S., & Trisudarmo, R. (2023). *Implementasi Algoritma K-Medoids Clustering Untuk Menentukan Segmentasi Pelanggan*. 17, 196–204.
- Mishra, P., Biancolillo, A., Roger, J. M., Marini, F., & Rutledge, D. N. (2020). New data preprocessing trends based on ensemble of multiple preprocessing techniques. *TrAC Trends in Analytical Chemistry*, 132, 116045.  
<https://doi.org/https://doi.org/10.1016/j.trac.2020.116045>
- Murpratiwi, S. I., Agung Indrawan, I. G., & Aranta, A. (2021). Analisis Pemilihan Cluster Optimal Dalam Segmentasi Pelanggan Toko Retail. *Jurnal Pendidikan*

*Teknologi Dan Kejuruan*, 18(2), 152. <https://doi.org/10.23887/jptk-undiksha.v18i2.37426>

Nisa, N., Firdaus, D., & Aprilia, R. (2023). Implementasi Business Intelligence Untuk Menganalisis Jumlah Guru SD SMP SMA SMK Di Jawa Barat. *Simpatik: Jurnal Sistem Informasi Dan Informatika*, 3(1), 11–16.

Oktavian, A. R., & Rumaisa, F. (2022). Penerapan Business Intelligence Terhadap Data Penjualan UMKM (Foodendez) Menggunakan Metode Algoritma Apriori Dalam Menentukan Segmentasi Pasar. *Jurnal Media Informatika Budidarma*, 6(3), 1740–1745.

Paembonan, S., & Abduh, H. (2021). Penerapan Metode Silhouette Coefficient untuk Evaluasi Clustering Obat. *Pena Teknik: Jurnal Ilmiah Ilmu-Ilmu Teknik*, 6(2), 48. [https://doi.org/10.51557/pt\\_jiit.v6i2.659](https://doi.org/10.51557/pt_jiit.v6i2.659)

Pal, D., & Vanijja, V. (2020). Perceived usability evaluation of Microsoft Teams as an online learning platform during COVID-19 using system usability scale and technology acceptance model in India. *Children and Youth Services Review*, 119, 105535. <https://doi.org/10.1016/j.childyouth.2020.105535>

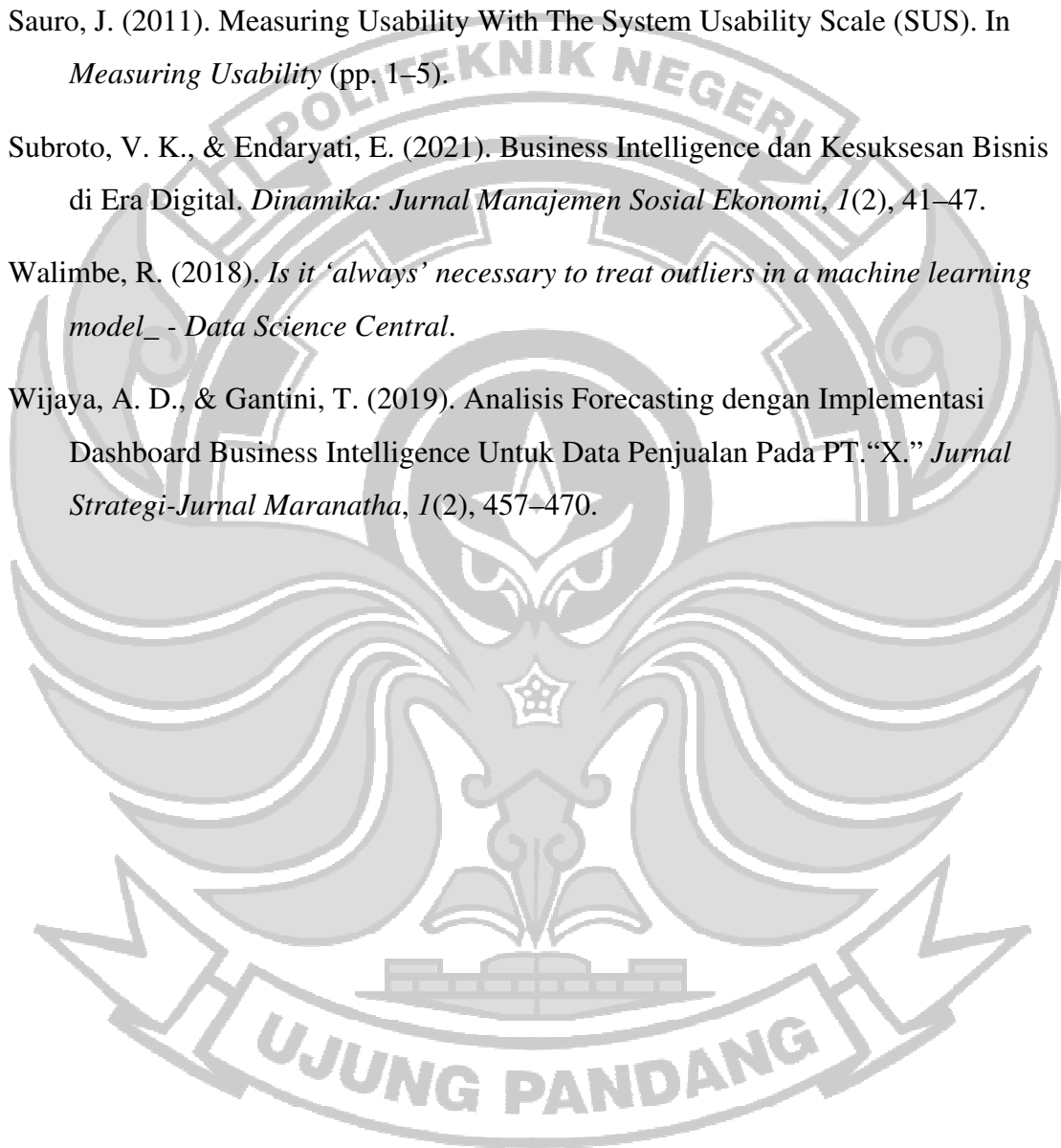
Priharsari, D. (2020). Pengolahan SEM Covariance-Based dengan Modul Lavaan pada Penelitian Sistem Informasi. *Jurnal Sistem Informasi, Teknologi Informasi, Dan Edukasi Sistem Informasi*, 1(1).

Purba, D., & Purba, M. (2022). Aplikasi Analisis Korelasi dan Regresi menggunakan Pearson Product Moment dan Simple Linear Regression. *Citra Sains Teknologi*, 1(2), 97–103.

Rahman, A. (2020). *Rancang Bangun Sistem Informasi Penjadwalan Menggunakan Google Apps Script Pada Politeknik Kampar*. Universitas Islam Negeri Sultan Syarif Kasim Riau.



- Rumahorbo, A. C., & Sekarwati, K. A. (2020). Penerapan Data Mining Dengan Menggunakan Algoritma C4. 5 Pada Klasifikasi Fasilitas Kesehatan Provinsi Di Indonesia: Array. *Jurnal Ilmiah Komputasi*, 19(1), 27–38.
- Sauro, J. (2011). Measuring Usability With The System Usability Scale (SUS). In *Measuring Usability* (pp. 1–5).
- Subroto, V. K., & Endaryati, E. (2021). Business Intelligence dan Kesuksesan Bisnis di Era Digital. *Dinamika: Jurnal Manajemen Sosial Ekonomi*, 1(2), 41–47.
- Walimbe, R. (2018). *Is it 'always' necessary to treat outliers in a machine learning model\_ - Data Science Central*.
- Wijaya, A. D., & Gantini, T. (2019). Analisis Forecasting dengan Implementasi Dashboard Business Intelligence Untuk Data Penjualan Pada PT.“X.” *Jurnal Strategi-Jurnal Maranatha*, 1(2), 457–470.



## LAMPIRAN

### Lampiran 1. Surat Izin Penelitian



**KEMENTERIAN PENDIDIKAN, KEBUDAYAAN,  
RISET, DAN TEKNOLOGI**

**POLITEKNIK NEGERI UJUNG PANDANG**

Jalan Perintis Kemerdekaan Km. 10 Tamalanrea, Makassar 90245  
Telepon: (0411) 585365, 585367, 585368; Faksimili: (0411) 586043  
Laman: [www.poliupg.ac.id](http://www.poliupg.ac.id) / E-Mail: pnup@poliupg.ac

Nomor : 0177/PL10.5.6/PP.02.10/2023

22 Februari 2023

H a l : **Permohonan Izin Penelitian**

**Yth. Wakil Direktur Bidang Akademik, Kemahasiswaan dan Alumni**  
Politeknik Negeri Ujung Pandang  
Makassar

Sehubungan dengan adanya pelaksanaan Tugas Akhir bagi mahasiswa Jurusan Teknik Elektro **Program Studi D4 Teknik Komputer dan Jaringan Angkatan 2019**, maka kami mohon kepada Bapak untuk menerbitkan Surat Permohonan Izin untuk melakukan Penelitian pada :

**PT Gama Inovasi Berdikari**

Universitas Gadjah Mada, Blk. K, Jl. Asem Kranji No.7, Sendowo, Sinduadi, Kec. Mlati,  
Kabupaten Sleman,  
Daerah Istimewa Yogyakarta 55284

kepada mahasiswa kami :

N a m a : Hairatunnisa  
Stambuk : 42519034  
W a k t u : 22 Februari s.d 17 April 2023

Demikian surat pengantar ini, atas perhatian dan kerjasama yang baik diucapkan terimakasih.

Ketua Jurusan Teknik Elektro PNUP,



**Ahmad Rizal Sultan, S.T., M.T., Ph.D.**  
**NIP 197609212000031001**

## Lampiran 2. Query Lengkap Proses Clustering

```
library(cluster)
library(factoextra)
library(ggplot2)
library(googleheets4)

# Load dataset
df <- read_sheet('1FP629z37Re7M-r4jTopPbpexDnnQIiknrJvwKK_dm2Q')

# select variables for clustering
data <- data[, c(3, 4, 5, 6, 7, 8, 9, 10, 15, 16)]
data

# Menentukan jumlah kluster optimal
fviz_nbclust(data, kmeans, method = "wss")

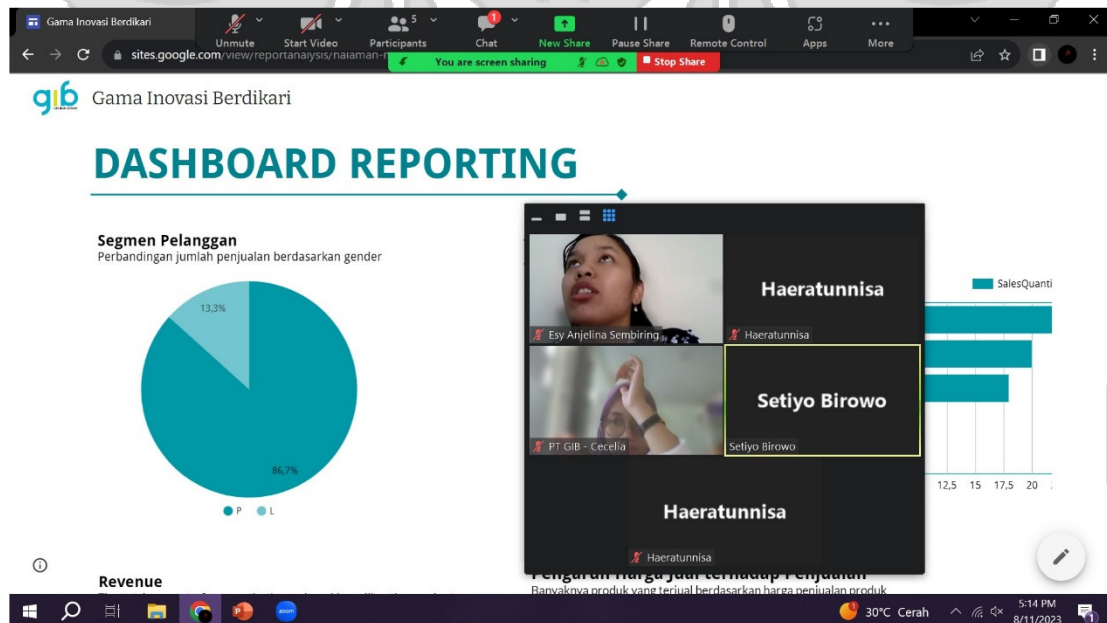
# Proses clustering
clustering <- kmeans(data, centers=3, nstart=25)
clustering

final_data <- cbind(data, cluster = clustering$cluster)

# Ubah final_data menjadi dataframe
final_data <- as.data.frame(final_data)

# upload data to googlesheet
sheet_name <- "Hasil_klustering"
write_sheet(hasil_klustering, sheet = sheet_name)
```

## Lampiran 3. Pengujian Sistem oleh Stakeholder PT GIB



Lampiran 4. Lembar Kuesioner Pengujian SUS (*System Usability Scale*)

LEMBAR KUESIONER *USABILITY*

Jawablah pertanyaan berikut dengan memberi satu tanda centang (✓) pada setiap pertanyaan pada kolom jawaban yang tersedia.

<b>Keterangan:</b>		
STS : Sangat Tidak Setuju	TS : Tidak Setuju	RG : Ragu-ragu
ST : Setuju	SS : Sangat Setuju	

	STS	TS	RG	ST	SS
1. Saya pikir akan menggunakan sistem ini lagi	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
	1	2	3	4	5
2. Saya merasa sistem ini rumit untuk digunakan.	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	1	2	3	4	5
3. Saya pikir sistem ini mudah digunakan.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
	1	2	3	4	5
4. Saya membutuhkan bantuan dari orang lain atau teknisi dalam menggunakan sistem ini.	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	1	2	3	4	5
5. Saya pikir fitur-fitur sistem ini berjalan dengan semestinya.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
	1	2	3	4	5
6. Saya pikir ada terlalu banyak ketidaksesuaian di dalam sistem ini.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	1	2	3	4	5
7. Saya merasa orang lain akan memahami cara menggunakan sistem ini dengan cepat.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
	1	2	3	4	5
8. Saya merasa sistem ini rumit untuk digunakan.	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	1	2	3	4	5
9. Saya merasa tidak ada hambatan dalam menggunakan sistem ini.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
	1	2	3	4	5
10. Saya perlu membiasakan diri terlebih dahulu sebelum menggunakan sistem ini.	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	1	2	3	4	5

Yogyakarta, 15 Agustus 2023

Cecelia Dwi W

## Lampiran 5. Hasil Wawancara Penerapan Sistem

### DAFTAR PERTANYAAN TERKAIT METODE PERANCANGAN STRATEGI MELALUI PENERAPAN *BUSINESS INTELLIGENCE* DAN VISUALISASI HASIL *CLUSTERING* PADA DATA PELANGGAN

1. Metode apa yang biasa digunakan oleh perusahaan untuk membuat strategi pemasaran yang dapat meningkatkan profit perusahaan dan jumlah pelanggan?

Jawaban: Dalam meningkatkan profit dan menambah jumlah pelanggan, perusahaan menggunakan metode promosi untuk dua saluran penjualan, yaitu metode promosi online seperti Giveaway, Bundling Produk, Diskon, dan sebagainya, serta metode promosi offline seperti Event dan penawaran langsung kepada mitra dengan komunikasi tatap muka maupun komunikasi telemarketing.

2. Apakah perusahaan terkait pernah menggunakan segmentasi pelanggan untuk membantu dalam pembuatan strategi pemasaran?

Jawaban: Belum, untuk saat ini segmentasi pelanggan dipilih secara acak.

3. Apakah perusahaan pernah menggunakan *Business Intelligence* dan metode *Clustering* untuk menentukan segmentasi pelanggan dalam membuat strategi pemasaran dan membaca tren bisnis?

Jawaban: Belum, maka kami harap hasil *Business Intelligence* dan metode *Clustering* yang dilakukan oleh peneliti dapat bermanfaat bagi kami dan dapat dijadikan acuan dalam mengembangkan bisnis dan menyesuaikan target pasar.

4. Apakah metode *clustering* dapat membantu perusahaan dalam membuat strategi pemasaran?

Jawaban: Ya.



UJUNG PANDANG

5. Apakah kategori klaster multidimensi (*Price Peritem, Total Sales, Sales Quantity*) memiliki peranan dalam penentuan strategi pemasaran?

Jawaban: Ya, namun ketiga hal tersebut hanya bagian kecil yang masih perlu menambah beberapa faktor pertimbangan lainnya dalam menentukan strategi pemasaran.

6. Menurut Anda, apakah visualisasi dari hasil kluster dapat membantu perusahaan dalam menentukan strategi pemasaran yang tepat serta dapat membaca tren bisnis?

Jawaban: Ya, kami sangat membutuhkan visualisasi dari hasil kluster tersebut untuk membantu pembuatan strategi pemasaran yang tepat.

7. Apa kritik dan saran Anda terkait penerapan *Business Intelligence* dan metode *Clustering* terhadap segmentasi pelanggan dalam membuat strategi bisnis?

Jawaban: Pada saat pengujian, hasil analisisnya masih belum menjawab segmentasi pelanggan yang tepat sasaran untuk melakukan pemasaran produk. Masih perlu waktu dalam pengolahan data yang lebih rinci dan pengembangan fitur sistemnya agar bisa menentukan segmentasi pasar yang cocok untuk produk ini.

Yogyakarta, 08 Agustus 2023



Cecelia Dwi W