

**PENERAPAN DATA MINING DENGAN ALGORITMA KMEANS  
TERHADAP DATA TRANSAKSI PENJUALAN  
(Studi Kasus: CV. Sogan Batik Rejodani)**

**TUGAS AKHIR**

**Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana Strata-1  
Program Studi Teknik Industri - Fakultas Teknologi Industri  
Universitas Islam Indonesia**



Nama : Lisa Alfiyani  
No. Mahasiswa : 19522383

**PROGRAM STUDI TEKNIK INDUSTRI  
FAKULTAS TEKNOLOGI INDUSTRI  
UNIVERSITAS ISLAM INDONESIA  
YOGYAKARTA  
2023**

## PERNYATAAN KEASLIAN

Saya mengakui bahwa tugas akhir ini adalah hasil karya saya sendiri kecuali kutipan dan ringkasan yang seluruhnya sudah saya jelaskan sumbernya. Jika dikemudian hari ternyata terbukti pengakuan saya ini tidak benar dan melanggar peraturan yang sah maka saya bersedia ijazah yang telah saya terima ditarik kembali oleh Universitas Islam Indonesia.

Yogyakarta, 9 September 2023



  
(Lisa Alfiyani)  
NIM 19522383

**SURAT BUKTI PENELITIAN**

**SOGAN BATIK REJODANI**  
Jl. Palagan Tentara Pelajar km 10 Sariharjo  
Ngaglik, Sleman 55581, Yogyakarta, Indonesia

**SURAT KETERANGAN MAGANG**

No :014-PBK/B/ HRD SOGAN BATIK/VII/2023

Nama yang bersangkutan dibawah ini telah melakukan Magang di Sogan Batik Rejodani dan berjalan dengan baik tanpa ada kendala, baik dari pihak yang bersangkutan maupun dari pihak perusahaan.

Nama : Lisa Alfiyani  
NIM : 19522383  
Fakultas : Teknik Industri Universitas Islam Indonesia  
Waktu Magang : Maret 2023 – 31 Agustus 2023

Demikian surat ini kami keluarkan sebagai bukti keterangan resmi dari Sogan Batik Rejodani untuk pihak yang bersangkutan yang telah melakukan Magang di perusahaan kami agar dapat dimanfaatkan sebaik-baiknya dengan penuh bertanggung jawab.

Yogyakarta, 8 Agustus 2023

HRD Sogan Batik Rejodani

  
Sogan  
(Budi Santoso)  
BATIK REJODANI



**LEMBAR PENGESAHAN PEMBIMBING**  
**PENERAPAN DATA MINING DENGAN ALGORITMA KMEANS TERHADAP**  
**DATA TRANSAKSI PENJUALAN**  
**(Studi Kasus: CV. Sogan Batik Rejodani)**



**Yogyakarta, 15 September 2023**

**Dosen Pembimbing**

**Danang Setiawan, S.T., M.T.**

**LEMBAR PENGESAHAN DOSEN PENGUJI**  
**PENERAPAN DATA MINING DENGAN ALGORITMA KMEANS TERHADAP**  
**DATA TRANSAKSI PENJUALAN (Studi Kasus: CV. Sogan Batik Rejodani)**

**TUGAS AKHIR**

**Disusun Oleh :**

**Nama : Lisa Alfiyani**

**No. Mahasiswa : 19522383**

Telah dipertahankan di depan sidang penguji sebagai salah satu syarat untuk memperoleh gelar Sarjana Strata-1 Teknik Industri Fakultas Teknologi Industri

**Universitas Islam Indonesia**

**Yogyakarta, 20 September 2023**

**Tim Penguji**

Danang Setiawan, S.T., M.T.

Ketua

Ir. Wahyudhi Sutrisno, S.T., M.M., M.T.

Anggota I

Taufik Abdurrahman, S.E.

Anggota II

**Mengetahui,**

**Ketua Program Studi Teknik Industri Program Sarjana**

**Fakultas Teknologi Industri**

**Universitas Islam Indonesia**



**Ir. Muhammad Rizki Nur Hafid, S.T., M.Sc., Ph.D., IPM**

**NIP 015220101**

## **HALAMAN PERSEMBAHAN**

Penulis mempersembahkan karya tulis ini untuk kedua orang tua penulis terkasih yang selalu senantiasa memberikan doa dan dukungan secara moril dan material. Serta untuk Bapak Danang Setiawan, S.T., M.T. yang telah membimbing hingga selesainya Tugas Akhir ini.

**MOTTO**

*"Jalani, nikmati, proses perjalanannya. Tidak perlu tergesa-gesa. Sabar, nanti pasti sampai"-Lisa Alfiyani*

*"Kalau ada kemauan pasti ada jalan, ada mimpi itu kita pasti bisa"- Feni Fitriyanti*

*"Kamu tidak bisa bergantung dengan orang lain terus-menerus, pada akhirnya yang bisa tolong kamu cuma diri kamu sendiri. Tapi jangan 100% bergantung sama diri sendiri juga, andalkan Tuhan."- Shania Gracia*

## KATA PENGANTAR

*Assalamu'alaikum Warahmatullahi Wabarakatuh.*

*Alhamdulillah Robbil Alamin* segala puji dan syukur bagi Allah SWT atas limpahan rahmat, hidayah dan nikmat-Nya, Shalawat serta salam senantiasa penulis kepada junjungan kita Nabi Muhammad SAW beserta keluarga, sahabat dan para pengikutnya hingga akhir zaman sehingga penulis dapat menyelesaikan Tugas Akhir yang berjudul "Penerapan Data Mining Dengan Algoritma *K-means* Terhadap Data Transaksi Penjualan" (Studi Kasus: CV. Sogan Batik Rejodani).

Tugas Akhir merupakan salah satu prasyarat untuk memperoleh gelar sarjana Strata Satu (S1) pada jurusan Teknik Industri, Fakultas Teknologi Industri, Universitas Islam Indonesia. Tugas Akhir ini bertujuan untuk menyelaraskan ilmu yang telah didapatkan selama penulis berada dibangku perkuliahan dengan realiti yang ada pada didunia nyata. Dalam penulisan Tugas Akhir ini, melalui proses yang panjang dan penuh halangan untuk mengerjakannya Alhamdulillah Allah SWT memudahkan jalan untuk dapat menyelesaikan penyusunan Tugas Akhir ini. Dengan segala kerendahan hati penulis mengucapkan terima kasih kepada:

1. Prof. Dr. Ir. Hari Purnomo., M.T., IPU., ASEAN, Eng. Selaku Dekan Fakultas Teknologi Industri, Universitas Islam Indonesia.
2. Bapak Ir. Muhammad Ridwan Andi Purnomo, S.T., M.Sc., Ph.D., IPM. Selaku Ketua Program Studi Teknik Industri Program Sarjana Fakultas Teknologi Industri, Universitas Islam Indonesia.
3. Bapak Danang Setiawan, S.T., M.T. Selaku Dosen Pembimbing yang selalu memberi masukan dan nasihat dalam penyusunan Tugas Akhir.
4. Bapak Taufik dan Ibu Iffa, selaku *Owner* Sogan Batik Rejodani yang telah memberikan kesempatan bagi penulis untuk melakukan penelitian di Sogan Batik Rejodani.
5. Bapak Suryanto dan Ibu Anisah, selaku orang tua terkasih yang telah memberikan doa dan dukungannya secara moril dan material sehingga penyusunan Tugas Akhir dapat terselesaikan dengan baik.
6. Teman-teman dekat penulis terkhususnya kepada Reza Novpri Andini, Ridwan Sidik Yahya, Azzam Farryal Haq, Hanif Maulana Albar, Mochamad Sahid Haikal, Annisa Dwiana Hamzah, Refki, dan Javier Bayu Nugroho yang selalu saling menguatkan dengan memberikan motivasi dan dukungan selama masa perkuliahan hingga akhirnya dapat menyelesaikan sampai penyusunan Tugas Akhir selesai. Senang sekali penulis dapat bertemu teman-teman semua, semoga pertemanan ini menjadi salah satu kenangan yang tidak terlupakan hingga tutupnya usia nanti.
7. Fransisca Saraswati Puspa Dewi dan Feni Fitriyanti yang telah memberikan motivasi dan rasa semangat untuk segera menyelesaikan Tugas Akhir ini dan terima kasih juga telah menemani penulis selama proses penyusunan Tugas Akhir hingga dapat terselesaikan dengan baik.
8. Teman-teman magang yaitu Haikal, Maulana, dan Surya yang telah menemani penulis selama magang hingga penyusunan Tugas Akhir ini dapat terselesaikan dengan baik. Terima kasih atas motivasi dan kegigihan teman-teman, senang sekali bisa mengenal teman-teman semua.

Berbagai kekurangan dan kesalahan mungkin pembaca temukan dalam penulisan Tugas Akhir ini. Maka dari itu, penulis sangat menerima segala kritikan dan saran yang membangun agar lebih baik lagi. Semoga apa yang menjadi kekurangan saat ini dapat

lebih baik lagi oleh peneliti selanjutnya. Karena kesempurnaan hanya milik Allah SWT, semoga karya ini senantiasa dapat memberikan manfaat. *Aamiin*.

*Wassalamu'alaikum Warahmatullahi Wabarakatuh.*

Sleman, 9 September 2023

A handwritten signature in black ink, appearing to be the initials 'Huf'.

Penulis

## ABSTRAK

CV. Sogan Batik Rejodani merupakan UMKM yang bergerak dalam bidang *fashion* yang memproduksi pakaian batik. Produk yang dihasilkan berupa *blouse* batik, kemeja batik, *dress* batik, *blazer* batik, gamis batik dan lain-lain. Setiap hari Sogan Batik melayani transaksi penjualan yang banyak, hal ini membuat data transaksi penjualan menumpuk, data sering kali diperlakukan hanya sebagai rekaman tanpa ada pengolahan. Sehingga, tidak memiliki nilai guna yang cukup baik untuk dimanfaatkan. Hal ini dapat menyebabkan kekecewaan pelanggan karena barang yang diinginkan sering habis dan Perusahaan dapat mengalami kerugian karena mengalami *overstock*. Penelitian ini bertujuan untuk mengetahui bagaimana data transaksi penjualan dimanfaatkan sehingga bisa mendapatkan strategi pemasaran yang cocok untuk Sogan Batik Rejodani dengan *cluster* variabel kode agen, dan daerah. Oleh karena itu, diperlukan proses pengolahan data besar dengan menggunakan suatu teknik data mining. Pada penelitian ini, teknik data mining yang digunakan yaitu metode *K-means Clustering* dengan bantuan pemrograman *python*. Tujuan pengelompokan data ini adalah untuk meningkatkan variasi antar kelompok dan mengurangi fungsi objektif yang dikumpulkan dalam suatu kelompok. Hasil yang didapatkan berupa visualisasi data yang menampilkan variabel-variabel yang digunakan. Dengan menerapkan data mining menggunakan Algoritma *K-means*, perusahaan dapat memanfaatkan jumlah data transaksi yang muncul pada setiap periodenya. Sehingga, perusahaan dapat melakukan analisis lebih mendalam dan luas lagi terhadap variabel-variabel yang ada.

**Kata Kunci:** *K-means*, Data Mining, *Python*, *Clustering*

## DAFTAR ISI

<b>PERNYATAAN KEASLIAN.....</b>	<b>ii</b>
<b>SURAT BUKTI PENELITIAN .....</b>	<b>iii</b>
<b>KATA PENGANTAR.....</b>	<b>viii</b>
<b>ABSTRAK .....</b>	<b>x</b>
<b>DAFTAR ISI .....</b>	<b>xi</b>
<b>DAFTAR TABEL .....</b>	<b>xiii</b>
<b>DAFTAR GAMBAR.....</b>	<b>xiv</b>
<b>BAB I PENDAHULUAN.....</b>	<b>1</b>
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	2
1.3 Tujuan Penelitian .....	3
1.4 Manfaat Penelitian .....	3
1.5 Batasan Penelitian .....	3
1.6 Sistematika Penulisan.....	3
<b>BAB II TINJAUAN PUSTAKA .....</b>	<b>5</b>
2.1 Kajian Induktif .....	5
2.1.1 Kajian Literatur .....	5
2.2 Kajian Deduktif.....	17
2.2.1 <i>Data Science</i> .....	17
2.2.2 <i>Data Mining</i> .....	18
2.2.3 <i>Clustering</i> .....	22
2.2.4 <i>Algoritma K-Means</i> .....	22
2.2.5 <i>Big Data</i> .....	24
2.2.6 <i>Python</i> .....	25
<b>BAB III METODE PENELITIAN .....</b>	<b>27</b>
3.1 Objek Penelitian .....	27
3.2 Metode Pengumpulan Data .....	27
3.3 Jenis Data .....	27
3.4 Alur Penelitian .....	28
<b>BAB IV PENGUMPULAN DAN PENGOLAHAN DATA.....</b>	<b>30</b>
4.1 Profil Perusahaan .....	30
4.2 Proses Bisnis .....	33
4.3 Pengolahan Data.....	35
4.3.1 <i>Data Selection</i> .....	35
4.3.2 <i>Preprocessing Data</i> .....	35
4.3.3 <i>Transformasi Data</i> .....	36

4.3.4	<i>Data Mining</i> .....	37
4.3.5	<i>Interpretation/Evaluation</i> .....	43
<b>BAB V</b>	<b>PEMBAHASAN</b> .....	<b>51</b>
5.1	Analisis Transaksi Penjualan .....	51
5.1.1	Variabel Kode Agen.....	51
5.1.2	Variabel Daerah .....	53
5.1.3	Variabel Kode Agen dan Daerah.....	54
5.1.4	Clustering Variabel Kode Agen.....	55
5.1.5	Clustering Variabel Daerah .....	56
5.2	Rekomendasi .....	57
<b>BAB VI</b>	<b>PENUTUP</b> .....	<b>58</b>
6.1	Kesimpulan .....	58
6.2	Saran.....	58
<b>DAFTAR PUSTAKA</b>	.....	<b>59</b>
<b>LAMPIRAN</b>	.....	<b>A</b>

**DAFTAR TABEL**

Table 2.1 Kajian Literatur.....	5
Table 4.2 Produk Sogan Batik Rejodani.....	31

## DAFTAR GAMBAR

Gambar 3.1 Alur Penelitian .....	28
Gambar 4.2 Proses Bisnis .....	34
Gambar 4.3 <i>Data Selection</i> .....	35
Gambar 4.4 <i>Preprocessing Data</i> .....	36
Gambar 4.5 <i>Syntax Pada Software Anaconda Prompt</i> .....	37
Gambar 4.6 <i>Pseudocode Untuk Mengimpor Fungsi</i> .....	37
Gambar 4.7 <i>Pseudocode Untuk Membaca File</i> .....	38
Gambar 4.8 <i>Pseudocode Mengidentifikasi Nilai Kosong</i> .....	38
Gambar 4.9 <i>Pseudocode Menghapus Nilai Yang Kosong</i> .....	38
Gambar 4.10 <i>Pseudocode Melihat Nilai Duplikat</i> .....	38
Gambar 4.11 <i>Pseudocode Mengecek Nilai Unik</i> .....	39
Gambar 4.12 <i>Pseudocode Menghapus Nilai Unik Yang Besar</i> .....	39
Gambar 4.13 <i>Univariate Analysis Kode agen</i> .....	39
Gambar 4.14 <i>Univariate Analysis Daerah</i> .....	39
Gambar 4.15 <i>Univariate Analysis Produk</i> .....	39
Gambar 4.16 <i>Bivariate Analysis Variabel kode agen dan daerah</i> .....	40
Gambar 4.17 <i>Bivariate Analysis Variabel kode agen dan produk</i> .....	40
Gambar 4.18 <i>Pseudocode Membaca File</i> .....	40
Gambar 4.19 <i>Pseudocode Ubah Tipe Data Menjadi Kategori</i> .....	40
Gambar 4.20 <i>Pseudocode LabelEncoder</i> .....	40
Gambar 4.21 <i>Pseudocode K-means Cluster</i> .....	41
Gambar 4.22 <i>Pseudocode Result</i> .....	41
Gambar 4.23 Model Kmeans .....	41
Gambar 4.24 <i>Pseudocode Kmeans</i> .....	41
Gambar 4.25 <i>Pseudocode Menamai Cluster</i> .....	42
Gambar 4.26 <i>Pseudocode Inverse Transform</i> .....	42
Gambar 4.27 <i>Pseudocode Histplot</i> .....	42
Gambar 4.28 <i>Pseudocode Plot Bivariate Analysis Kode_agen Result</i> .....	42
Gambar 4.29 <i>Pseudocode Plot Bivariate Analysis Daerah Result</i> .....	42
Gambar 4.30 <i>Plot Variabel Kode_agen Januari</i> .....	43
Gambar 4.31 <i>Plot Variabel Daerah Januari</i> .....	43
Gambar 4.32 <i>Plot Variabel Kode_agen &amp; Daerah Januari</i> .....	44
Gambar 4.33 <i>Plot Variabel Kode_agen Result Januari</i> .....	44
Gambar 4.34 <i>Plot Variabel Daerah Result Januari</i> .....	45
Gambar 4.35 <i>Plot Variabel Kode_agen Februari</i> .....	45
Gambar 4.36 <i>Plot Variabel Daerah Februari</i> .....	46
Gambar 4.37 <i>Plot Variabel Kode_agen &amp; Daerah Februari</i> .....	46
Gambar 4.38 <i>Plot Variabel Kode_agen Result Februari</i> .....	47
Gambar 4.39 <i>Plot Variabel Daerah Result Februari</i> .....	47
Gambar 4.40 <i>Plot Variabel Kode_agen Maret</i> .....	48
Gambar 4.41 <i>Plot Variabel Daerah Maret</i> .....	48
Gambar 4.42 <i>Plot Variabel Kode_agen &amp; Daerah Maret</i> .....	49
Gambar 4.43 <i>Plot Variabel Kode_agen Result Maret</i> .....	49
Gambar 4.44 <i>Plot Variabel Daerah Result Maret</i> .....	50

## **BAB I**

### **PENDAHULUAN**

#### **1.1 Latar Belakang**

Dalam dunia bisnis yang sangat bersaing saat ini, suatu perusahaan dituntut untuk terus mengembangkan bisnis untuk tetap kompetitif. Ada beberapa cara untuk mencapai hal tersebut, seperti meningkatkan kualitas produk, menambah jenis produk, dan mengurangi biaya operasional bisnis. Salah satu cara untuk mencapai hal ini adalah dengan menggunakan analisis data penjualan perusahaan (Siregar et al., n.d.). Dalam penelitian (Nurajizah & Salbinda, 2021), Data Mining adalah suatu proses penggalian data yang sangat penting dari data. Selain itu, proses ini juga menggali pola data dari berbagai jenis basis data, seperti data berorientasi objek, data transaksi, data *warehouse*, dan data relasional. Data Mining dapat membantu bisnis membuat keputusan cepat (Nurajizah, 2019). Pada penelitian (Nurajizah & Salbinda, 2021), Menentukan pola penjualan dengan melihat kecenderungan pembelian konsumen dapat membantu menentukan produk mana yang paling laris dan produk mana yang kurang laku terjual. Hal ini juga dapat membantu mengontrol stok persediaan dan membantu perusahaan merancang strategi pemasaran (Normah et al., 2020).

*K-Means* yaitu salah satu metode pengelompokan data non-hierarki yang ada ke dalam bentuk dua atau lebih kelompok. Data yang memiliki karakteristik yang sama dikelompokkan ke dalam kelompok yang sama, dan data dengan karakteristik yang berbeda dikelompokkan ke dalam kelompok yang berbeda. Tujuan pengelompokan data ini adalah untuk meningkatkan variasi antar kelompok dan mengurangi fungsi objektif yang dikumpulkan dalam suatu kelompok (Asroni et al., 2018). Sebuah perusahaan pastinya memiliki data transaksi penjualan. Namun, tidak semua perusahaan dapat memanfaatkan data penjualan tersebut salah satu yang menjadi alasan mengapa perusahaan tidak dapat mengolah data yang ada yaitu karena data transaksi penjualan yang setiap harinya semakin banyak tetapi sumber daya manusia dari perusahaan masih bingung untuk memanfaatkannya. Oleh karena itu, dengan melakukan analisis data

penjualan dengan menerapkan algoritma *K-Means Clustering*. Dari analisis data dapat membentuk *cluster-cluster* penjualan sehingga menjadi suatu informasi yang bermanfaat untuk perusahaan.

CV. Sogan Batik Rejodani merupakan UMKM yang bergerak dalam bidang fashion yang memproduksi pakaian batik. Produk yang dihasilkan berupa blus batik, kemeja batik, blazer batik, gamis batik dan lain-lain. Setiap hari Sogan Batik melayani transaksi penjualan yang banyak, salah satu masalah yang sering dihadapi yaitu ketersediaan kain sering kali tidak memadai kebutuhan pelanggan. Seperti, jika ada stok barang yang terlalu sedikit tetapi permintaannya tinggi, atau jika ada stok barang yang banyak tetapi permintaannya rendah. Hal ini dapat menyebabkan kekecewaan pelanggan karena barang yang diinginkan sering habis dan mengalami kerugian bagi perusahaan karena beberapa barang yang tidak laku dapat menumpuk. Setiap hari data transaksi penjualan menumpuk, data sering kali diperlakukan hanya sebagai rekaman tanpa ada pengolahan. Sehingga, tidak memiliki nilai guna yang cukup baik untuk dimanfaatkan.

Berdasarkan pada permasalahan tersebut, dalam menentukan perencanaan strategi pemasaran yang tepat dengan hasil *cluster* yang didapatkan, salah satunya yaitu dengan menerapkan data *mining*. Dengan memanfaatkan data transaksi penjualan sebagai salah satu cara untuk menganalisis data untuk menemukan pola dalam kumpulan data. Metode yang akan digunakan adalah *clustering k-means*, pengelompokan objek data dalam kelompok yang sama dan berbeda. Dengan menerapkan data *mining* menggunakan Algoritma *K-means*, perusahaan dapat memanfaatkan jumlah data transaksi yang muncul pada setiap periodenya. Sehingga, perusahaan dapat menentukan strategi pemasaran yang efektif dan efisien berdasarkan hasil *clustering*.

## 1.2 Rumusan Masalah

Pada penelitian ini, didapatkan rumusan masalah sebagai berikut:

1. Bagaimana hasil *clustering* dari variabel kode agen dan daerah?
2. Bagaimana hasil dari variabel kode agen dan daerah yang memiliki peminat yang lebih banyak?

### 1.3 Tujuan Penelitian

Pada penelitian ini terdapat beberapa tujuan penelitian, berikut adalah tujuan penelitian:

1. Mengetahui hasil *clustering* variabel kode agen dan daerah.
2. Untuk mengetahui kode agen dan daerah yang peminatnya banyak.

### 1.4 Manfaat Penelitian

Berikut adalah manfaat dari penelitian ini, diharapkan:

1. Bagi peneliti, lebih memahami dan mengetahui secara sistematis tentang penggunaan dari metode analisis *k-means clustering* pada data transaksi penjualan.
2. Bagi para pembaca, diharapkan dengan adanya penelitian ini menambah wawasan dan menambah informasi dalam memanfaatkan *data mining*.
3. Untuk CV. Sogan Batik Rejodani diharapkan dapat membantu sebagai acuan, maupun saran, dalam memanfaatkan ilmu *data mining* dengan mengelola Big Data pada perusahaan.

### 1.5 Batasan Penelitian

Berikut merupakan batasan penelitian yang dilakukan:

1. Data yang digunakan peneliti adalah data sekunder yang diperoleh dari data Transaksi Penjualan selama 3 bulan terakhir di Perusahaan.
2. Data yang akan diolah merupakan data transaksi penjualan pada bulan Januari hingga Maret 2023.
3. Metode analisis yang digunakan yaitu *k-means clustering*.
4. Data diolah menggunakan bantuan dari bahasa pemrograman *python*.

### 1.6 Sistematika Penulisan

Agar penulisan lebih terstruktur, berikut merupakan sistematika penulisan pada penelitian ini:

BAB I PENDAHULUAN

Bab ini menjelaskan tentang latar belakang dilakukannya penelitian, rumusan masalah, tujuan penelitian, batasan masalah, metode penelitian dan sistematika penulisan laporan tugas akhir.

## BAB II KAJIAN LITERATUR

Pada Bab ini, menjelaskan tentang semua landasan teori yang ada hubungannya dengan penelitian yang sedang dilakukan.

## BAB III METODE PENELITIAN

Pada Bab ini, menjelaskan mengenai metodologi penelitian yang berisi objek penelitian, gambaran umum perusahaan, dan alur penelitian.

## BAB IV PENGUMPULAN DAN PENGOLAHAN DATA

Pada Bab ini, pengumpulan data dilakukan dengan studi lapangan seperti melakukan wawancara dengan pemilik perusahaan dan orang-orang yang terlibat dengan perusahaan. Pengolahan data dilakukan dengan metode Algoritma *K-Means* dan alat bantu oleh *python*.

## BAB V PEMBAHASAN

Pada bab ini, menjelaskan hasil dari data yang sudah diolah pada bab sebelumnya.

## BAB VI KESIMPULAN DAN SARAN

Pada bab ini, menjelaskan tentang kesimpulan dan saran penelitian yaitu hasil akhir dari data transaksi penjualan sogan batik yang terbentuk dan juga saran untuk peneliti berikutnya.

## BAB II

### TINJAUAN PUSTAKA

#### 2.1 Kajian Induktif

##### 2.1.1 Kajian Literatur

Kajian literatur yaitu sebuah landasan teoritik yang berfungsi sebagai pedoman untuk sumber hipotesis yang digunakan dalam penelitian. Kajian literatur dilakukan dari berbagai sumber tertulis, seperti buku, makalah, jurnal, dan laporan penelitian. Pada penelitian ini, berikut merupakan literatur yang digunakan:

*Table 2.1. Kajian Literatur*

No	Judul	Metode	Tujuan	Hasil dan Kesimpulan
1.	Analisa Cluster Data Transaksi Penjualan Minimarket Selama Pandemi Covid-19 dengan Algoritma K-means. (Negara, 2021)	Algoritma k-means. Dengan mengelompokkan data berdasarkan kemiripan karakteristiknya.	Untuk menemukan kelompok produk dengan data penjualan terbanyak sehingga manajemen stok dapat terpenuhi dan meningkatkan transaksi penjualan.	Dengan menggunakan algoritma pada 278480 data transaksi, terbentuk tiga cluster data penjualan: cluster 2 memiliki penjualan tertinggi sebanyak 57 produk, cluster 1 memiliki penjualan sedang sebanyak 5 produk, dan cluster 0 memiliki penjualan rendah sebanyak 5 produk. Hasilnya menunjukkan bahwa model klusterisasi

No	Judul	Metode	Tujuan	Hasil dan Kesimpulan
				dengan <i>confusion matrix</i> memiliki akurasi 87%. Hasil ini akan membantu pemilik Minimarket Berkah Abadi membuat keputusan tentang bagaimana mengelola stok barang mereka.
2.	Implementasi Data Mining Penjualan Menggunakan <i>Algoritma Clustering</i> dengan Metode <i>K-Means</i> . (Afiasari et al., 2023)	<i>Algoritma K-Means Clustering</i> menggunakan bantuan <i>software RapidMiner</i> .	Untuk memecahkan masalah pemanfaatan informasi persediaan dengan mengelompokkan produk persediaan berdasarkan karakteristik produk dengan menggunakan teknik data mining.	Hasil dari pengelompokan tiga kelompok yang dilakukan menggunakan metode <i>K-Means</i> terdiri dari data set yang terdiri dari 160 item produk. 88 item yang dikategorikan laku terjual masuk ke <i>cluster</i> 0, 26 item yang dikategorikan tidak laku terjual masuk ke <i>cluster</i> 1, dan 46 item yang dikategorikan kurang

No	Judul	Metode	Tujuan	Hasil dan Kesimpulan
				laku terjual masuk ke <i>cluster 2</i> . Toko <i>Bill Lights</i> dapat menerapkan rencana penjualan dan produksi ulang yang berfokus pada barang yang laku.
3.	Pemodelan Segmentasi Transaksi Jual Beli Produk Menggunakan Pendekatan Model <i>K-Means dan Subtractive Clustering</i> Studi Kasus Survei Pada Beberapa Cabang Optik Retail. (Arvio et al., 2022)	Menggunakan <i>Algoritma K-Means dan Subtractive Clustering</i> . Data yang digunakan merupakan hasil survei penjualan produk kacamata di 3 cabang optik.	Untuk dapat melihat penjualan yang paling banyak diminati pelanggan, sehingga perusahaan dapat melakukan prediksi dan analisa untuk penjualan di tahun berikutnya yang berdampak pada modal yang harus dikeluarkan oleh perusahaan.	Didapat tiga kelompok dengan kriteria merk produk, jumlah penjualan, dan harga yang berbeda di setiap cabang. Namun, dalam perhitungan subtraktif mencari nilai maksimal potensial yang ada, sehingga satu kelompok yang dihasilkan menunjukkan produk kacamata yang paling laris di masing-masing cabang.
4.	Implementasi Metode <i>Data Mining K-Means Clustering</i>	Metode <i>K-Means Clustering</i> dan	Menerapkan metode <i>K-Means</i>	Penelitian ini berhasil mengelompokkan

No	Judul	Metode	Tujuan	Hasil dan Kesimpulan
	<i>Clustering</i> Terhadap Data Pembayaran Transaksi Menggunakan Bahasa Pemrograman <i>Python</i> Pada CV Digital Dimensi. (Manalu & Gunadi, 2022)	menampilkan hasilnya dalam bentuk visual menggunakan bahasa pemrograman <i>Python</i> dan pustaka <i>Scikit-Learn</i> .	<i>Clustering</i> untuk mengelompokkan data-data transaksi pembayaran pada CV Digital Dimensi dan menampilkan hasilnya dalam bentuk grafik visual dengan menggunakan bahasa pemrograman <i>Python</i> dan pustaka <i>Scikit-Learn</i> .	data transaksi pembayaran menjadi lima klaster. Klaster-klaster ini dapat digunakan sebagai referensi untuk menentukan strategi bisnis perusahaan. Setiap klaster memiliki tingkatan tertentu. Ada tingkat tinggi, tingkat sedang, dan tingkat rendah.
5.	Pendekatan <i>K-Means Cluster</i> Pada Pengelompokan Hotel Berbintang Di Kota Pontianak. (Dofiri, 2023)	Metode yang digunakan dalam penelitian ini yaitu <i>K-Means Cluster</i> dimana banyaknya <i>cluster</i> yang akan terbentuk sebanyak empat <i>cluster</i> .	Untuk mengelompokkan hotel-hotel berbintang yang ada di Kota Pontianak agar pihak manajemen hotel dapat meningkatkan kualitas dan	Hasil penelitian menunjukkan bahwa terbentuk empat <i>cluster</i> , masing-masing dengan beberapa hotel di Kota Pontianak. <i>Cluster</i> pertama mencakup hotel dengan bintang empat, <i>cluster</i> kedua

No	Judul	Metode	Tujuan	Hasil dan Kesimpulan
			kuantitas dari masing-masing hotel.	mencakup hotel dengan bintang 3, <i>cluster</i> ketiga mencakup hotel dengan bintang 2, dan <i>cluster</i> keempat mencakup hotel dengan bintang 1.
6.	Pengelompokan pembagian zakat dengan menggunakan metode <i>clustering k-means</i> . (Islami, 2023)	Menggunakan algoritma <i>k-means clustering</i> dan untuk pengujian menggunakan <i>Davies Bouldin Index</i> .		Dengan menggunakan metode <i>clustering k-means</i> , pembagian zakat dikelompokkan dengan menggunakan penghasilan, kondisi rumah, dan kepemilikan harta, dengan jumlah data 1133 dan $k=4$ . <i>Cluster</i> 0 dan 2 memiliki beberapa pemohon yang masuk ke dalam kelompok data yang layak menjadi penerima, sedangkan <i>cluster</i> 1 dan 3 memiliki semua pemohon yang masuk

No	Judul	Metode	Tujuan	Hasil dan Kesimpulan
7.	<i>Applications of data mining and machine learning framework in aquaculture and fisheries: A review.</i> (Gladju et al., 2022)	Menggunakan data mining dan machine learning	Implementasi data mining dan machine learning dalam domain akuakultur dan perikanan yang canggih.	ke dalam kelompok data yang layak menjadi penerima. Terlepas dari tantangan terkait volume data, kompleksitas, kualitas dan keamanan, dalam waktu dekat, penerapan sistem dan alat data cerdas tidak bisa dihindari dalam pengelolaan akuakultur dan perikanan karena memfasilitasi pengambilan keputusan yang proaktif dan berbasis pengetahuan yang dapat mencegah kecelakaan/kerugian besar di budidaya. Penggunaan <i>drone</i> dan robot otonom tak berawak (misalnya <i>SeaDrone</i> , <i>Shoal</i> ,

No	Judul	Metode	Tujuan	Hasil dan Kesimpulan
				<p><i>OpenROV, OceanOne, Deep Trekker, Aquabotix, Apium Swarm Robotics, Blueye Pioneer, dan PowerRay)</i> untuk eksplorasi/operasi bawah air dan integrasinya dengan sistem data merupakan hal yang menjanjikan bagi ekosistem. Pemantauan dan budidaya kandang intensif.</p>
8.	<p><i>Identifying key success factors for startups With sentiment analysis using text data mining.</i> (Asgari et al., 2022)</p>	<p><i>Text data mining</i> dengan menganalisis <i>random forest, support-vector machine (SVM)</i> dan <i>multilayer perceptron (MLP)</i> untuk menguji</p>	<p>Untuk mengidentifikasi faktor-faktor dasar keberhasilan, yang penting bagi <i>startup</i> menjadi model bisnis yang sukses dan menguntungkan</p>	<p>Metode MLP berdasarkan kriteria akurasi menawarkan akurasi lebih tinggi dibandingkan metode pengujian lainnya. Algoritma pemungutan suara dan penumpukan digunakan untuk</p>

No	Judul	Metode	Tujuan	Hasil dan Kesimpulan
		pelabelan data yang tidak berlabel.	dari waktu ke waktu.	meningkatkan keakuratan algoritma. Dengan menggunakan metode voting, akurasi hampir sama dengan hasil yang diperoleh dari MLP dan dengan metode <i>stacking</i> akurasi kurang dari ketiga metode tersebut. Dengan menggunakan kata <i>cloud</i> , diindikasikan bahwa <i>trigram</i> yang paling negatif adalah inovasi <i>startup</i> terkait perubahan iklim. Proses akselerasi <i>startup</i> , mendorong penyelesaian lebih cepat, memberikan produk terbaik di awal proyek, praktik manajemen yang buruk, dan fokus hanya pada properti

No	Judul	Metode	Tujuan	Hasil dan Kesimpulan
				dikelompokkan sebagai sentimen <i>negative</i> .
9.	<i>Customer Relationship Management Based on SPRINT Classification Algorithm under Data Mining Technology.</i> (Sun & Tan, 2022)	Data mining dengan algoritma klasifikasi SPRINT	Penerapan teknologi data mining untuk meningkatkan tingkat manajemen dan pengambilan keputusan perusahaan.	<i>Algoritma</i> klasifikasi SPRINT digunakan dalam klasifikasi pelanggan, yang meningkatkan akurasi klasifikasi pelanggan. FP-growth, algoritma <i>asosiation rule</i> diterapkan membuat sistem menjadi lebih praktis. Dengan ini, Perusahaan dapat mempertahankan pelanggan dan membuat toko mendapatkan keuntungan maksimal.
10.	<i>Data Mining in Internet of Things Systems: A Literature Review.</i> (Sarhan, 2022)	Algoritma data mining	Untuk membantu para manajer dan operator di Perusahaan makanan dalam menentukan	Berdasarkan penelitian ini, <i>deviation detection</i> telah sangat membantu untuk aplikasi <i>IoT</i> seperti

No	Judul	Metode	Tujuan	Hasil dan Kesimpulan
			tingkat risiko	<i>smart homes, smart</i>
			keamanan pangan	<i>agriculture, smart</i>
			yang akurat, dan	<i>traffic, and parking</i>
			pengambilan	<i>system</i> dan lain
			keputusan.	sebagainya.
11.	<i>Data Mining Methods: K-Means Clustering Algorithms.</i> (Annas, 2023)	Algoritma means	K- Untuk melakukan tinjauan terhadap varian algoritma <i>K-means clustering</i> .	Berbagai modifikasi yang tersedia dari desain dan implementasi algoritma <i>K-means</i> untuk meningkatkan kinerja dan kecepatan pengelompokannya. Studi saat ini, menemukan bahwa perbaikan mencakup semua aspek utama dari desain algoritma, proses, <i>output</i> , dan modifikasi konsep.
12.	<i>Implementation of Data Mining Using K-Means Clustering Method to Determine Sales Strategy In S&amp;R Baby Store.</i> (Wahyudi & Silfia, 2022)	Data mining menggunakan algoritma <i>K-means clustering</i> dengan metode CRISP-DM. Implementasi	Untuk mendapatkan strategi penjualan yang baik agar dapat meningkatkan laba penjualan.	Data transaksi penjualan dengan jumlah 4 atribut dan membentuk 4 <i>cluster</i> yang terdiri dari sangat laris, laris, cukup laris dan

No	Judul	Metode	Tujuan	Hasil dan Kesimpulan
		dengan <i>Rapidminer.</i>		kurang laris. <i>cluster</i> kedua dengan jumlah 944 produk, <i>cluster</i> ketiga dengan jumlah 2 produk, dan <i>cluster</i> keempat dengan jumlah 43 produk.
13.	<i>Customer clustering using K-means clustering: Supporting customer relationship management system.</i> (Rosianingsih et al., 2021)	<i>K-means Clustering</i>	Untuk meningkatkan penjualan dari setiap distributor perusahaan.	Dari sistem yang diusulkan memiliki fitur yang dapat membantu distributor untuk melakukan pemesanan secara mandiri, konfirmasi pembayaran, <i>retur</i> , <i>complain</i> , dan penukaran poin <i>reward</i> . Hasil dari sistem telah dilakukan uji penerimaan sistem dengan menggunakan kuesioner dengan 5 responden. Hasil uji menunjukkan tingkat kepuasan distributor terhadap sistem.

No	Judul	Metode	Tujuan	Hasil dan Kesimpulan
14.	<i>An Effective and Adaptable K-means clustering Algorithm for Big Data Cluster Analysis.</i> (Hu et al., 2023)	Algoritma	Untuk membandingkan dengan algoritma <i>k-means</i> , <i>XK-means</i> , <i>DDK-means</i> , dan <i>Canopyk-means</i> pada 10 set data <i>open source</i> .	Menunjukkan bahwa algoritma <i>LK-means</i> memiliki hasil pencarian yang lebih baik merata, sangat meningkatkan kemampuan pencarian global, kemampuan pemrosesan data yang besar, dan pusat <i>cluster</i> yang tidak merata dari algoritma-algoritma <i>means</i> . Artinya algoritma <i>LK-means</i> dapat secara efektif meningkatkan efek pengelompokan data dan memiliki kecepatan iterasi yang lebih baik.
15.	<i>Analysis K-means of Covid-19 cases in Bandar Lampung and South Lampung.</i> (Untoro et al., 2023)	Data mining dengan algoritma K-means	Untuk mengetahui pengelompokan zona yang terkena Covid-19.	Berdasarkan perhitungan dengan menggunakan metode <i>k-means</i> dibagi menjadi 4 <i>cluster</i> ,

No	Judul	Metode	Tujuan	Hasil dan Kesimpulan
				yaitu zona 0, zona 1, zona 2, zona 3. Zona 0, zona 1 terdiri dari 37 data zona 2 terdiri dari 25 data, dan zona 3 terdiri dari 3 data paling banyak yaitu 43 data.

## 2.2 Kajian Deduktif

### 2.2.1 Data Science

Menurut penelitian (Sembiring, 2023), *Data science* adalah ide tentang menyatukan statistik, analisis data, informatika, dan teknik yang berkaitan dengan pemahaman dan analisis data fenomenal (Hayashi, 1988). Ini menggunakan teori dan pendekatan dari matematika, statistik, ilmu komputer, ilmu informasi, dan pengetahuan domain. *Data science* tidak sama dengan *computer science* dan *information science*. Jim Gray, pemenang *Turing Award*, mendefinisikan ilmu data sebagai "paradigma keempat" ilmu pengetahuan-empiris, teoritis, komputasi, dan sekarang digerakkan oleh data. Dia juga menegaskan bahwa kemajuan teknologi informasi telah mengubah sains (Tolle et al., 2011).

Menurut (Tryana et al., 2021), *data science* adalah bidang interdisipliner yang berfokus mengumpulkan pengetahuan dari kumpulan data yang besar, biasanya disebut *big data*, dan menerapkan pengetahuan dan wawasan yang diperoleh dari data untuk memecahkan masalah dalam berbagai domain aplikasi. Bidang ini mencakup persiapan data untuk analisis, merumuskan masalah *data science*, menganalisis data, mengembangkan solusi berbasis data, dan menyajikan temuan untuk menginformasikan keputusan. Oleh karena itu, menggabungkan keterampilan dalam bidang komputer,

statistik, ilmu data, matematika, visualisasi informasi, integrasi data, desain grafis, sistem kompleks, komunikasi, dan bisnis. Ahli statistik Nthan Yau mengaitkan ilmu data dengan interaksi manusia-komputer, di mana orang harus dapat mengontrol dan menjelajahi data dengan mudah. *American Statistical Association* menetapkan tiga "komunitas profesional" utama pada tahun 2015: manajemen *database*, statistik, *machine learning*, dan sistem terdistribusi paralel.

Menurut (Tryana et al., 2021), data *science* adalah proses mendestilasi, mengekstraksi, atau menggali *insight* dari data. Tujuan akhir dari bidang ini adalah untuk menemukan *insight* dari data. Jumlah data yang diolah dapat sangat besar. *Insight* seperti berlian atau emas: mereka berharga meskipun sedikit atau berukuran kecil. *Insight* tersebut dapat berupa model data yang dibuat atau informasi penting yang akan membantu pengambilan keputusan. Untuk mendapatkan pemahaman yang berguna dari data, perlu ada rasa ingin tahu yang kuat dari individu atau organisasi. Ini adalah kebutuhan, karena ada masalah yang ingin diselesaikan dengan data. Setelah itu, seorang data *scientist* dapat melakukan banyak hal dengan memanfaatkan ilmu dan teknologi yang tepat untuk mendapatkan *insight*.

Menurut (Tryana et al., 2021), berpendapat bahwa data *science* bukan suatu bidang baru, tetapi hanya nama lain untuk ilmu statistik. Ahli lain berpendapat bahwa data *science* berbeda dari statistik karena berkonsentrasi pada masalah dan pendekatan khusus yang dihadapi dengan data digital. *Data science* menangani data kuantitatif dan kualitatif (seperti gambar), dan menekankan prediksi dan tindakan. Sebaliknya, Vasant Dhar menyatakan bahwa statistik menekankan data kuantitatif dan deskriptif.

### **2.2.2 Data Mining**

Menurut (Sharda et al., 2018), organisasi menggunakan berbagai metode data *mining* untuk memperoleh pemahaman yang lebih baik tentang pelanggan dan operasi mereka, serta untuk memecahkan masalah yang lebih kompleks. Data *mining* adalah langkah analisis dari proses "*Knowledge Discovery in Databases*", atau *KDD*. Selain langkah analisis dasar, data *mining* juga melibatkan aspek *database* dan manajemen data, *pre-processing* data, pertimbangan tentang model dan *inference considerations*, metrik ketertarikan, pertimbangan kompleksitas, *post-processing* struktur yang ditemukan,

visualisasi, dan *updating online*. Data mining, menurut (Turban et al., 2005) adalah proses menemukan pola, hubungan, dan kecenderungan penting dalam sekumpulan *big data* yang tersimpan di dalam penyimpanan. dilakukan dengan menggunakan teknik pengenalan pola seperti teknik matematika dan statistik, yang digunakan untuk menguraikan penemuan pengetahuan di dalam *database*.

Menurut (Suntoro, 2019), data *mining* adalah proses mengekstraksi informasi bermanfaat dari basis data yang sangat besar. Ini harus dilakukan untuk menjadi informasi baru dan membantu dalam pengambilan keputusan. Sedangkan menurut (Witten et al., 2016), data *mining* adalah proses mengumpulkan data dari berbagai sumber dan mengubahnya menjadi pola, informasi, atau pengetahuan penting. Tujuannya adalah untuk meningkatkan keuntungan, mengurangi biaya, atau bahkan keduanya. Menurut (Tryana et al., 2021). Analisis data *mining* adalah analisis semi-otomatis atau otomatis dari sejumlah *big data* untuk mengeksplorasi pola yang sebelumnya tidak diketahui. Metode analisis ini termasuk analisis kelompok data (*cluster analysis*), analisis catatan yang tidak biasa (deteksi *anomaly*), dan analisis pola dependensi (*association rule mining* dan *sequential pattern mining*). Pola-pola ini kemudian dapat dilihat sebagai semacam ringkasan input data, dan dapat digunakan dalam analisis lebih lanjut, seperti dalam *machine learning* dan analisis prediktif. Misalnya, langkah data *mining* dapat mengidentifikasi beberapa kelompok data, yang kemudian dapat digunakan oleh sistem pendukung keputusan untuk mendapatkan hasil prediksi yang lebih akurat. Pengumpulan data, persiapan data, dan interpretasi dan pelaporan hasil bukanlah bagian dari langkah data *mining*, tetapi termasuk sebagai langkah tambahan dalam proses *KDD* secara keseluruhan.

Perbedaan analisis data dan data *mining* yaitu data *mining* menggunakan model *machine learning* dan model statistik untuk menemukan pola tersembunyi dalam sejumlah data yang besar, sedangkan analisis data digunakan untuk menguji model dan hipotesis pada data set, terlepas dari jumlah datanya. hal Ini membedakan analisis data dari data *mining*. Istilah "*dredging*", "*fishing*", dan "*snooping*" mengacu pada penggunaan teknik data *mining* untuk mengambil sampel bagian dari kumpulan data yang lebih besar, yang mungkin terlalu kecil untuk membuat kesimpulan statistik yang dapat dipercaya

tentang validitas setiap pola yang ditemukan. Untuk menguji hipotesis baru dengan jumlah data yang lebih besar, teknik ini dapat digunakan (Tryana et al., 2021).

Istilah data *mining* dan *knowledge discovery in databases (KDD)* sering kali digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi di dalam suatu basis data yang besar. Sebenarnya kedua istilah tersebut memiliki konsep yang berbeda, tetapi berkaitan satu sama lain. Salah satu tahapan dalam keseluruhan proses *KDD* adalah *data mining*. Proses *KDD* secara garis besar dapat dijelaskan sebagai berikut (Tryana et al., 2021).

1. *Data Selection*

Sebelum tahap penggalian informasi dalam *KDD* dimulai, data yang dipilih dari sekumpulan data operasional harus dipilih dan disimpan dalam suatu berkas yang berbeda dari basis data operasional.

2. *Pre-processing/Cleaning*

Sebelum masuk ke tahap proses data *mining*, data yang menjadi fokus *KDD* harus dibersihkan. Proses pembersihan mencakup membuang duplikasi data, memeriksa data yang tidak konsisten, memperbaiki kesalahan, seperti kesalahan cetak (tipografi), dan juga melengkapi data sebelumnya dengan data atau informasi lain yang relevan dan diperlukan untuk *KDD*, seperti data atau informasi eksternal.

3. *Transformation Coding*

Proses koding dalam *KDD* adalah proses kreatif dan bergantung pada jenis atau pola informasi yang akan dicari dalam basis data. Ini juga merupakan proses transformasi pada data yang telah dipilih sehingga, dapat digunakan untuk proses data *mining*.

4. *Data Mining*

Data *mining* adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma yang digunakan dalam data *mining* sangat beragam, dan pemilihan teknik atau algoritma yang tepat sangat bergantung pada tujuan dan proses *KDD* secara keseluruhan.

### 5. *Interpretation/Evaluation*

Pola informasi yang dihasilkan dari proses data *mining* harus ditampilkan dalam bentuk yang mudah dipahami dan dipahami oleh pihak yang berkepentingan. Ini adalah tahap dalam proses *KDD* yang disebut interpretasi. Tahap ini mencakup memeriksa apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya.

Berdasarkan dari (Larose & Larose, 2014), data *mining* dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan yaitu:

#### 1. Deskripsi

Terkadang peneliti dan analis sederhana ingin mencoba menggambarkan pola dan kecenderungan dalam data estimasi-estimasi, yang hampir sama dengan data klasifikasi, kecuali variabel target estimasi lebih berorientasi numerik daripada kategori.

#### 2. Prediksi

Prediksi dan klasifikasi hampir mirip dengan estimasi dan klasifikasi, kecuali prediksi nilai dari hasil di masa mendatang.

#### 3. Klasifikasi

Dalam klasifikasi, terdapat target variabel kategori yang digunakan. Sebagai contoh, penggolongan pendapatan dapat dibagi menjadi kategori tinggi, sedang, dan rendah.

#### 4. Klaster

Klaster adalah pengelompokan *record*, pengamatan, atau perhatian dan membentuk kelas objek yang memiliki kemiripan. Klaster adalah kumpulan *record* yang serupa satu sama lain dan tidak serupa dengan *record* dalam klaster lainnya.

#### 5. Asosiasi

Menentukan fitur yang muncul dalam satu waktu adalah tugas asosiasi dalam data *mining*. Dalam dunia bisnis analisis keranjang belanja (*market basket analysis*) adalah istilah yang lebih umum digunakan.

### 2.2.3 *Clustering*

Menurut (Berkhin, 2006) *Clustering* adalah metode *KDD*, Pengelompokan sekumpulan objek disebut *clustering* atau analisis klaster. *Clustering* dilakukan sehingga objek dalam kelompok yang sama (disebut *cluster*) lebih mirip satu sama lain dalam beberapa hal dibandingkan dengan objek dalam kelompok lain (disebut *cluster*). *Clustering* membagi data ke dalam grup yang terdiri dari objek yang memiliki karakteristik yang sama. Ini mengelompokkan item data ke dalam sejumlah kecil grup, sehingga masing-masing grup memiliki persamaan penting (Garcia et al., 2009). *Clustering* digunakan dalam berbagai aplikasi data *mining*, termasuk eksplorasi data ilmu pengetahuan, akses ke informasi dan *text mining*, penggunaan basis data spasial, dan analisis *web* (Andayani, 2007).

*Clustering* adalah salah satu alat bantu data *mining* yang digunakan untuk mengelompokkan item ke dalam klaster-klaster. Klaster adalah sekelompok atau sekumpulan objek data yang sangat mirip satu sama lain dan sangat mirip dengan objek data dalam klaster yang berbeda. Klaster dapat terdiri dari satu atau lebih klaster, tergantung pada seberapa mirip objek-objek tersebut dengan satu sama lain dalam salah satu dari klaster tersebut. Salah satu tujuan utama metode klaster adalah untuk mengumpulkan sejumlah data atau objek ke dalam kelompok, sehingga setiap kelompok memiliki jumlah data yang seminimal mungkin (Santoso, 2007). Selain itu, analisis klaster membagi data dari himpunan ke dalam sejumlah kelompok yang memiliki kesamaan data yang lebih besar daripada kelompok lain (Jang et al., 2005). Kemampuan *clustering* ini dapat digunakan untuk mengetahui struktur data untuk aplikasi yang lebih luas seperti klasifikasi, pengolahan gambar, dan pengenalan pola.

### 2.2.4 *Algoritma K-Means*

*K-Means* adalah teknik pengelompokan data non hirarki (sekatan) yang bertujuan untuk membagi data yang sudah ada ke dalam dua atau lebih kelompok. Data dengan karakteristik yang sama dimasukkan ke dalam kelompok yang sama, dan data dengan karakteristik yang berbeda dimasukkan ke dalam kelompok yang berbeda. Menurut (Prasetyo, 2012) tujuan dari pengelompokan data ini adalah untuk mengurangi fungsi yang tidak penting yang ditetapkan selama proses pengelompokan. Proses pengelompokan umumnya bertujuan untuk meminimalkan variasi dalam suatu kelompok

dan memaksimalkan variasi antar kelompok. Metode algoritma *K-Means* menggabungkan data ke dalam kelompok yang berbeda, dengan karakteristik yang berbeda untuk setiap kelompok (Asroni et al., 2018). Proses *clustering* digambarkan sebagai berikut (Santoso, 2007)

- 1) Pilih jumlah kluster k.
- 2) Pusat kluster ini dapat diinisialisasi dengan berbagai cara. Tetapi biasanya terjadi secara acak.
- 3) Masukkan setiap objek atau data ke kluster terdekat. Pada titik ini, perlu menghitung jarak antara semua data ke setiap pusat kluster. Setiap data yang masuk dalam kluster tertentu ditentukan oleh jarak paling dekat diantara data kluster tersebut. Dapat menggunakan teori jarak euklidian, yang dirumuskan sebagai berikut:

$$d(x, y) = |x - y|^2 = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Keterangan:

$d(x, y)$  = jarak antara data pada titik x dan y

x = titik data pertama (pusat kluster)

y = titik data kedua (data dari N)

n = jumlah atribut data

- 4) Hitung kembali pusat kluster menggunakan keanggotaan kluster saat ini. Pusat kluster adalah rata-rata dari semua objek atau data yang termasuk dalam kluster tertentu. Juga dapat menggunakan median dari kluster tersebut jika diperlukan. Oleh karena itu, standar rata-rata atau *mean* bukan satu-satunya standar yang dapat digunakan.
- 5) Setelah itu, tugaskan kembali setiap objek dengan pusat kluster baru. Jika pusat kluster tidak berubah, proses *clustering* selesai. Jika tidak, kembali ke langkah (3) sampai pusat kluster tidak berubah lagi.

### 2.2.5 *Big Data*

Big data merupakan pengumpulan, pengelolaan, dan analisis data, untuk menghasilkan pengetahuan dan mengungkapkan pola tersembunyi. Big data dapat berasal dari apa pun yang dikumpulkan di sekitar kita (Sönmez et al., 2019). Di bidang big data, orang berbicara tentang cara menganalisis, mengekstrak, atau menangani kumpulan data yang sangat besar atau kompleks sehingga perangkat pemrosesan data yang sifatnya tradisional dapat ditangani.

Dipopulerkan oleh Jhon Mashey pada tahun 1990-an, istilah "*big data*" telah digunakan sejak saat itu. Dalam kebanyakan kasus, *big data* mencakup kumpulan data yang sangat besar yang tidak dapat ditangani oleh perangkat pemrosesan data konvensional. Perangkat pemrosesan data tradisional biasanya digunakan untuk menangkap, mengkurasi, mengelola, dan memproses data dalam waktu yang sangat singkat (Snijders et al., 2012). Filosofi *big data* berfokus pada data tidak terstruktur, tetapi juga mencakup data terstruktur, semi terstruktur, dan tidak terstruktur.

Saat ini, jumlah dan ukuran kumpulan data yang tersedia terus meningkat karena pengumpulan data dari berbagai perangkat seperti perangkat *internet of things*, sensor jarak jauh, perangkat seluler, kamera, sensor, mikrofon, dan jaringan sensor nirkabel, serta *software logs*, *readers radiofrequency identification (RFID)*, dan perangkat lainnya. *IBM* mengatakan pada tahun 2013 bahwa 2,5 exabites (2,5 kali 260 *bytes*) data dihasilkan setiap hari sejak tahun 1980-an, dan kapasitas teknologi untuk menyimpan data meningkat dua kali lipat setiap empat puluh bulan. *IDC* memperkirakan peningkatan eksponensial dalam volume data global, meningkat dari 4,4 *zettabytes* menjadi 44 *zettabytes* dari 2013 hingga 2020. Mereka juga memperkirakan bahwa pada tahun 2025 volume data global akan mencapai 163 *zettabytes* (Reinsel et al., 2017).

Dalam bidang teknologi informasi. Sejak 2015, *big data* menjadi alat penting dalam operasi bisnis karena membantu karyawan bekerja lebih efisien dan membuat pengumpulan dan distribusi teknologi informasi menjadi lebih mudah. *TI Operation Analytics (TIOA)* adalah istilah yang mengacu pada penggunaan big data untuk menyelesaikan masalah teknologi informasi dan pengumpulan data perusahaan (Solnik, 2016) Departemen teknologi informasi dapat memprediksi masalah yang mungkin terjadi

dan bergerak maju untuk memberikan solusi bahkan dalam manajemen sistem dengan menerapkan prinsip big data ke dalam konsep *machine intelligence dan deep computing*. Ini terjadi dengan menawarkan platform yang menyatukan data *silos* individual dan menghasilkan wawasan dari keseluruhan sistem, bukan dari kumpulan data yang terisolasi.

### 2.2.6 Python

Dari penelitian (Al Faruqi, 2021), menjelaskan bahwa *Python* adalah bahasa pemrograman interpretatif multiguna dengan filosofi perancangan yang berfokus pada tingkat keterbacaan kode. *Python* diklaim sebagai bahasa yang menggabungkan kapabilitas, kemampuan, dengan sintaksis kode yang sangat jelas, dan dilengkapi dengan fungsionalitas pustaka standar yang besar serta komprehensif. *Python* bisa dibilang bahasa pemrograman dengan tujuan umum yang dikembangkan secara khusus untuk membuat *source code* mudah dibaca. *Python* juga memiliki *library* yang lengkap sehingga memungkinkan *programmer* untuk membuat aplikasi yang mutakhir dengan menggunakan *source code* yang tampak sederhana (Perkovic, 2015). *Python* adalah bahasa pemrograman interpretatif dengan filosofi perancangan yang berpusat pada tingkat keterbacaan kode. Dengan sintaksis kode yang sangat jelas dan banyak fitur pustaka standar, Python dianggap sangat kuat (Khatib Sulaiman et al., 2023). *Input* dan *output* sangat penting untuk pemrograman, menurut (Syahrudin & Kurniawan, 2018) karena tanpanya program tidak dapat berjalan dan berinteraksi dengan pengguna. Untuk mempelajari bahasa pemrograman, penting untuk memahami strukturnya daripada menghafal kode karena setiap pembaruan atau pengembangan terkadang mengubah *syntax* atau kode.

*Python* adalah bahasa pemrograman yang umum digunakan untuk berbagai tujuan. Bahasa ini mirip dengan *C++*, *JavaScript*, *C#*, dan *Java Oracle*. Pada sekitar tahun 1980-an, Guido van Rossum mengambil nama *Python* dari serial televisi favoritnya, "*Monty Pythons Flying Circus*". Beberapa alasan mengapa *Python*, bahasa pemrograman menjadi populer:

- a. Fleksibel dan mudah digunakan, membuatnya mudah dipelajari.

- b. *Python* digunakan oleh komunitas *Data Science* karena memiliki standar pemrograman yang lebih tinggi dibandingkan bahasa pemrograman lain seperti *R*.
- c. *Python* juga cocok untuk pekerjaan naskah seperti *DevOps*, yang membutuhkan kemampuan pemrograman yang lebih tinggi daripada bahasa pemrograman tradisional.
- d. *Python* dapat dijalankan di hampir semua sistem operasi, terutama *Windows*, *MacOs*, dan *Linux*.
- e. Dengan modul jangkauannya yang luas, *python* dapat digunakan untuk mengembangkan fitur dasar bahasa pemrograman. *Python* adalah bahasa pemrograman yang gratis.

## **BAB III**

### **METODE PENELITIAN**

Metode penelitian dapat didefinisikan sebagai cara kerja yang digunakan selama proses penelitian, baik dalam pencarian data maupun pengungkapan fenomena yang ada (Zulkarnaen, 2020)

#### **3.1 Objek Penelitian**

Objek pada penelitian ini adalah data transaksi penjualan pada CV. Sogan Batik Rejodani yang beralamat di Jalan Palagan Tentara Pelajar KM 10 Dusun Rejodani RT 01 / RW 01, Tambak Rejo, Sariharjo, Kec. Ngaglik, Kabupaten Sleman, Daerah Istimewa Yogyakarta 5558. Data transaksi penjualan yang akan dianalisis yaitu pada tahun 2023.

#### **3.2 Metode Pengumpulan Data**

Dalam penelitian ini, data yang akan digunakan yaitu data historis, referensi jurnal, dan wawancara dengan pihak terkait. Berikut adalah beberapa metode yang akan digunakan untuk mengumpulkan data:

1. Data Historis

Penelitian ini akan menggunakan data historis dari penjualan Sogan Batik Rejodani pada Januari 2023.

2. Referensi Jurnal

Untuk mendukung pernyataan peneliti, jurnal penelitian sebelumnya yang menggunakan metode yang sama akan digunakan sebagai referensi untuk penelitian ini.

3. Wawancara

Melakukan tanya jawab dengan pemilik dan pihak-pihak yang terkait dengan Sogan Batik.

#### **3.3 Jenis Data**

Berikut merupakan jenis data pada penelitian ini:

1. Primer

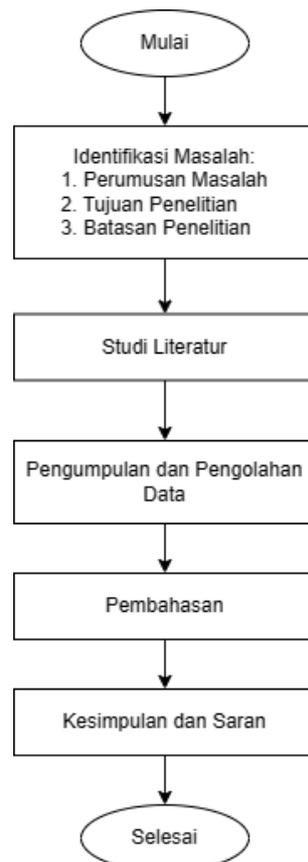
Data primer merupakan data yang informasinya diperoleh dari sumbernya langsung melalui observasi, wawancara, maupun kuesioner. Data primer yang dibutuhkan pada penelitian ini yaitu data transaksi penjualan pada bulan Januari 2023 dan wawancara dengan pihak terkait di Sogan Batik.

## 2. Sekunder

Data sekunder yang digunakan dalam penelitian ini adalah data atau informasi yang diperoleh dari sumber lain. Seperti penelitian sebelumnya dan referensi dalam jurnal yang mendukung diskusi tentang temuan penelitian ini.

### 3.4 Alur Penelitian

Pada penelitian ini, agar penelitian berjalan dengan baik maka dibuatlah alur penelitian. Di bawah ini merupakan alur dari penelitian yang dilakukan, sebagai berikut:



Gambar 3.1 Alur Penelitian

Berdasarkan Gambar 3.1, dapat dijelaskan sebagai berikut:

1. Pada tahapan identifikasi masalah, dilakukan dengan melakukan pengamatan langsung kondisi lapangan dan wawancara kepada *Owner* Sogan Batik Rejodani. Selanjutnya didapatkan rumusan masalah. Dari rumusan masalah tersebut diketahui tujuan akhir dari penelitian ini. Lalu, batasan dari penelitian ini yaitu data yang digunakan data Transaksi Penjualan selama 3 bulan terakhir di Perusahaan, menggunakan metode *K-means clustering*, dan dengan bantuan bahasa pemrograman *python*.
2. Studi literatur ini dilakukan untuk mendukung metodologi penelitian yang akan dilakukan. Dari hasil studi literatur, untuk mencapai tujuan penelitian ini peneliti menggunakan metode *Clustering K-Means* dengan memakai *software jupyter notebook* yang memakai bahasa pemrograman *Python*.
3. Pada proses pengumpulan data dilakukan dengan menggunakan data transaksi Sogan Batik pada tahun 2023 dan melakukan wawancara dengan pihak terkait di Sogan Batik. Sebagai informasi tambahan yang tidak diketahui oleh peneliti dan membantu untuk melakukannya pembahasan. Proses pengolahan data menggunakan *clustering K-Means* dengan bantuan alat Bahasa pemrograman *Python*.
4. Pada tahap pembahasan, setelah dilakukan pengumpulan dan pengolahan data selanjutnya hasil dari pengolahan data tersebut dijabarkan secara jelas.
5. Kesimpulan dan Saran, pada tahapan ini berisi tentang kesimpulan yang dapat diambil pada bagian pembahasan untuk menjawab permasalahan pada penelitian ini. Selain itu, saran berisi rekomendasi untuk penelitian lanjutan dan Sogan Batik yang dapat diberikan oleh peneliti.

## BAB IV

### PENGUMPULAN DAN PENGOLAHAN DATA

#### 4.1 Profil Perusahaan

CV. Sogan Batik Rejodani adalah perusahaan yang bergerak dalam industri pembuatan dan perdagangan batik dengan konsep modern dan target kalangan menengah ke atas. Perusahaan ini merupakan produsen batik cap asli dan batik tulis. Produk yang dihasilkan meliputi baju batik, kebaya batik, kemeja batik, bolero, *blazer*, *blouse*, dan lain-lain. Setiap produk batik yang dirancang memiliki cerita atau filosofi tersendiri. K.H Muhammad Darum bercita-cita untuk meningkatkan kesejahteraan masyarakat Rejodani, sehingga mendirikan CV. Sogan Batik Rejodani. Selanjutnya, keturunannya, Ibu Iffah M. Dewi, melanjutkan keinginan ini dengan memperbaiki dan menjelaskan jenis bisnis yang akan digunakannya. Awal sebelum CV. Sogan Batik Rejodani proses bisnis utamanya adalah pembuatan dan penjualan batik, Sogan Batik juga dikenal sebagai Sogan *Village*, yang menawarkan banyak wisata kuliner dan berbagai macam masakan tradisional. Namun, seiring berjalannya waktu bisnis itu tidak dapat bertahan lama. Pada akhirnya, CV Sogan Batik Rejodani menjadi fokus utamanya adalah membuat dan menjual batik. Batik yang dibuat melalui sistem *preorder* dan stok siap pakai.

Perusahaan ini berlokasi di Jalan Palagan Tentara Pelajar KM 10 Dusun Rejodani RT 01 / RW01, Tambak Rejo, Sariharjo, Kec. Ngaglik, Kabupaten Sleman, Daerah Istimewa Yogyakarta 5558. Sogan Batik Rejodani melakukan pemasaran, perdagangan dan distribusi dengan berbagai jenis produk, pemasaran yang dilakukan dengan menjual produk Baju koko, kemeja, *dress*, tunik, *blouse* dan *blazer*. Saat ini CV. Sogan Batik Rejodani memasarkan produknya secara *online* yang bekerjasama dengan beberapa *marketplace* dan untuk penjualan secara *offline* konsumen dapat menemukan produk sogan di beberapa *event fashion show* ataupun langsung datang ke *gallery*. Dalam pemasarannya Sogan Batik Rejodani mendistribusikan produk *ready stock* maupun *preorder* melalui *e-commerce* atau aktif secara *online* di Tiktok Shop, Instagram, dan Shopee. Sogan Batik Rejodani juga memperkenalkan produknya melalui *fashion show* seperti Indonesia *Modest Fashion Week* 2017, Jogja *Virtual Fashion Show* 2020, Jogja

*Fashion Week Trend Show 2019/2020, Jogja Fashion week 2021, dan Festival Batik 2022.* Sogan banyak sekali memproduksi produk batik, tabel di bawah ini merupakan beberapa produk yang dijual batik sogan, sebagai berikut:

Table 4.2 Produk Sogan Batik Rejodani

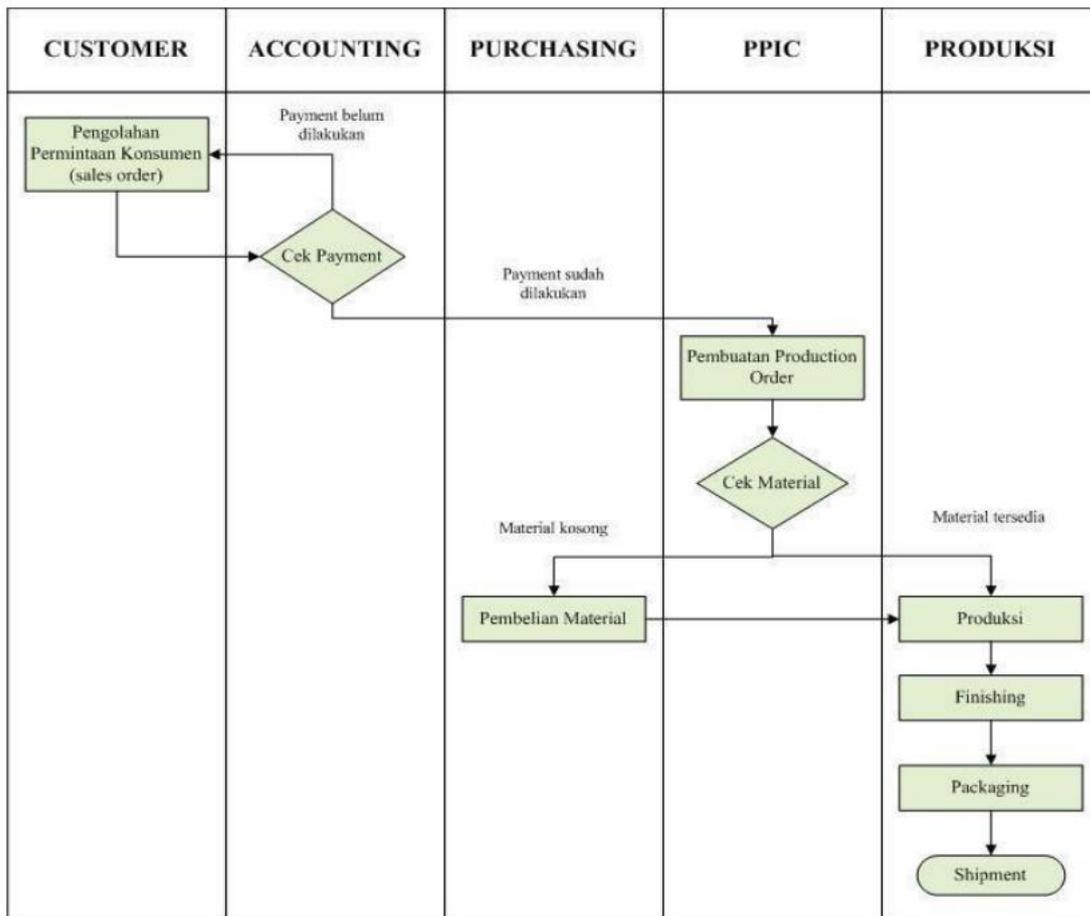
No	Nama Produk	Gambar	Filosofi
1.	Indang <i>Outer</i>		<p>Aksara Jawa 3; angka 3, Masyarakat Manggarai memiliki kebiasaan berkumpul dan mendengarkan petuah para tetua, hal ini sering disebut 3 “ng”, yaitu:</p> <ul style="list-style-type: none"> <li>• Undang artinya mengundang untuk berkumpul.</li> <li>• Indang artinya nasehat.</li> <li>• Ondang artinya selesai dan membiarkan pulang.</li> </ul> <p>Bagian belakang tertulis dalam aksara Jawa “dakwah”; nasehat kebaikan atau indang.</p>
2.	Syakira <i>Dress</i> Koleksi Wulan Sari		<p>Syakira, artinya syukur atau terima kasih. Terinspirasi dari syair: “<i>Wajabas syakru ‘alainaa, maa da’aa lillahi daa’i</i>” Artinya: "Wajiblah umatnya bersyukur atasnya, ketika seorang penyeru mengajak kepada Allah."Bagian depan</p>

No	Nama Produk	Gambar	Filosofi
3.	Purnama <i>Long Outer</i> Koleksi Wulan Sari		tertulis dengan aksara Jawa wulan sari atau bulan purnama. Terinspirasi dari syair “ <i>Thola'al badru 'alainaa, min Tsaniyyatil Wadaa</i> ”. Bagian belakang: Batik tulis yang menggambarkan sinar bulan purnama. Bagian depan: aksara Jawa bertuliskan Wulan Sari (Bulan Purnama).
4.	<i>Woelan Dress</i>		Terinspirasi dari syair kerinduan pada Sang Nabi SAW: “ <i>Thola'al badru 'alainaa</i> ”, Bagian depan: Batik tulis aksara Jawa Wulan Sari (Bulan yang Indah). Aplikasi ini bisa dilepas pasang dan berfungsi sebagai saku. Bagian atas dan belakang: aplikasi melingkar menggambarkan bulan.
5.	Dakwah <i>Blouse</i> Koleksi Al Mustafa Sogan Batik Rejodani Kombinasi Batik Tulis Dan Batik Cap Asli		Aksara jawa dibagian depan dibaca “dakwah”.

No	Nama Produk	Gambar	Filosofi
6.	Grapyak Dress Koleksi Dahayu Sister Brand		Terinspirasi dari sifat Grapyak dalam bahasa Jawa yang artinya ramah tamah. Memberi sapa dan salam kepada sesama perempuan, menghormati, bertutur kata baik tanpa memandang status sosial. <i>Ruffle dress</i> dengan motif garis dan kembang dahayu serta warna coklat dan merah simbol kehangatan dan keceriaan.
7.	Taat Dress		Terinspirasi dari syair: " <i>Ayyuhal mab'utsuu fiinaa, ji'ta bilamril mutho</i> ". Motif batik cap salur atau garis menggambarkan kebaikan yang ditaati. Motif batik tulis aksara Jawa: wulan sari (bulan purnama).

#### 4.2 Proses Bisnis

Berikut di bawah ini merupakan proses bisnis perusahaan CV. Sogan Batik Rejodani:



Gambar 4.2 Proses Bisnis

Dari gambar di atas menjelaskan bahwa proses bisnis yang ada pada CV. Sogan Batik Rejodani. Proses pertama adalah menerima permintaan dari pelanggan. Pelanggan dapat memesan CV Sogan Batik Rejodani melalui telepon, *WhatsApp*, *Shopee*, atau langsung ke *gallery* CV Sogan Batik Rejodani. Setelah itu, pihak layanan pelanggan menerima dan memasukkan *order* ke dalam data. Kemudian bagian akuntansi melakukan cek pembayaran. Jika pembayaran diterima, bagian akuntansi memberikan informasi kepada bagian PPIC untuk meneruskan *order*, dan jika pembayaran belum diterima bagian akuntansi memberikan informasi kembali ke bagian *customer service* untuk melakukan konfirmasi ulang kepada pelanggan. Dari bagian PPIC akan mengeluarkan *production order* yang berfungsi untuk melakukan perintah produksi. Setelah membuat *production order*, bagian PPIC mengecek ketersediaan bahan yang akan digunakan dalam proses

produksi. Jika bahan lengkap dan terpenuhi, bahan tersebut dapat langsung dikirim ke produksi. Jika tidak, bagian PPIC mengirimkan *order* ke bagian pembeli untuk membeli bahan yang kosong atau tidak lengkap. Setelah itu, bagian produksi melakukan proses pembuatan *order* hingga selesai dan kemudian dikirim ke bagian *finishing*. Di bagian ini, dilakukan proses setrika dan *packing* ke dalam *plastic wrap*. Dari bagian *finishing*, dikirim ke bagian *packaging* dan kemudian dikirim ke bagian pengiriman untuk dikirim ke konsumen.

### 4.3 Pengolahan Data

#### 4.3.1 Data Selection

Penelitian ini menggunakan riwayat data transaksi pada Sogan Batik Rejodani selama 3 bulan pada tahun 2023. Riwayat data transaksi dipisah perbulan berbentuk *Microsoft excel*. Berikut merupakan contoh riwayat data transaksi yang ada pada Sogan Batik Rejodani:

Kode Agen	Telp	Nama	Kota	Nama Produk	Qty	Jumlah Transak
LINDS	+62 813-1622-4333	Ibu Susi	Sleman	OKEKUNI TUNIK	1	568000
LSI	+62.0813.1582.5879	ari Apriana Rahmawatkarta Timur, Pasar Rel	KO L.PANJANG, SY/		4	1868500
LSI	+62.0813.2778.8488	Budi nuryaman	Ciamis, Lakhok	JUNG DRESS HITAM	1	516000
LSI	+62.0813.6312.4403	INDAH	LUBUK SIKARAH	TA DRESS, SYAKIRA	2	1550000
LSI	+62.65.8227.4176	Juwarni	GAN BATIK REJOD/	TAAT DRESS	1	656000
LSI	+62.6654.0082	lianti Mongkok HK +	NGAGLIK	AM 01, SABAR DRE	2	738000
LSI	+62.6844.3259	ss (Susan Ismanto HK	Yogyakarta, Klaton	JAMILA TUNIK, Hij	3	1251000
LSI	+62.9549.0929	i AAN Express Taiwai	GAN BATIK REJOD/IRA DRESS, FLO BU		5	1800000
LSI	+62.9609.8510	Express (BJE) Wina	DEPOK-SLM	KEMEJA L.PANJAN	4	2962000
LSI	+62.9851.7795	ress (Zaenah +852 98	Jakarta Timur, Ciracas	DRESS, Hijab Voal W	2	738000
SAG	6.20812E+12	IBU DEWI	Klaten, Klaten Selatan	TER, PURNAMA LC	2	1197000
SAG	6.20812E+12	IBU NIDA	Sogan Batik Rejodani	TAAT DRESS	1	820820
SAG	6.20813E+13	TUTI HAKIM	Metro, Metro Timur	DINA KEMEJA L PA	3	1290000
SAG	6.20813E+13	Nuning Fauziah Affian	Wonogiri	OELAN DRESS BLAC	1	350000
SAG	6.20813E+13	UHANIN ZAMRODA	Blitar, Selopuro	NA DRESS, ALYAA I	2	1365000
SAG	6.20821E+13	NINING	ogor, Bogor Barat - Kc	KEMEJA L.PANJAN	3	2055000
SAG	6.20856E+12	DITA PRETIASARI	Palembang, Plaju	K, JAMILA TUNIK, J	11	7616000
LINDS	6.20857E+13	Tambahan linds 827	Kendal, Gemuh	ESS, SYAKIRA KOK	2	1115000
LINDS	6.20857E+13	Chelsi	GAN BATIK REJOD/	AFIQA DRESS (NEW	1	580000
LINDS	6.20867E+13	supari	Kendal, Gemuh	OELAN DRESS BLAC	1	350000
WDDS	6.20879E+13	MIROTUL ADAWIYA	SUKODONO-SDA	JANG, AGHISNA DR	4	2488500
SAG	6.20882E+14	MBAK ILFIN	Sogan Batik Rejodani	TEING OUTER	1	860000
LINDS	6.20895E+13	Wiwik Larasati	Kendal, Gemuh	DRESS HITAM (NEV	5	2916750
ATK	6211	IIN ZACKY	Sogan Batik Rejodani	N DRESS, SYAKIRA I	2	1196500
NANA	+62811-913-816	Ibu Luci	akarta Selatan, Jagakars	K, PURNAMA KOK	8	2895000

Gambar 4.3 Data Selection

#### 4.3.2 Preprocessing Data

Dalam *preprocessing* data, peneliti melakukan reduksi data di mana data yang tidak valid seperti data transaksi yang gagal diproses karena pembatalan pemesanan. Berikut merupakan contoh hasil data setelah dilakukan *preprocessing* data:

kode agen	kota	nama produk
NANA	Kendal	Nekabail Dress, Nekabail Kemeja L Pendek
LSI	Jambi	INNER DRESS MANARA (APLIKASI)
LSI	Ponorogo	LAZUARDI TUNIK
LSI	Lampung	ZAMRUD MADINA KEMEJA L PANJANG, INNER HITAM LANGIT MADINA, INNER HITAM LANGIT MADINA
LSI	Yogyakarta	AGNI TUNIK
LSI	Yogyakarta	TATONG DRESS
an Batik Rejoc		HIJRA REV
SAG	Magelang	SYAKIRA DRESS, SYAKIRA KOKO L PANJANG
SAG	Blitar	JAMILA TUNIK, SHAHABI DRESS HITAM
SAG	an Batik Rejoc	INDANG OUTER, KEMBANG SEPATU KEMEJA L PENDEK, NOESA KOKO L PENDEK
SAG	Sukabumi	TOMING OUTER, OBI NEKA 01 (BUNDLING)
SAG	Pati	MAHBUBAH DRESS BROWN
SAG	an Batik Rejoc	NOESA DRESS, NEKABAIL KEMEJA L PANJANG
SAG	Yogyakarta	TIBA OUTER
LINDS	an Batik Rejoc	WAFIQA DRESS
SAG	Bogor	TITONG DRESS
ATK	an Batik Rejoc	WAFIQA DRESS, MAHBUBAH DRESS BROWN
LSI	an Batik Rejoc	ZAMRUD MADINA DRESS, OKEKUNI TUNIK 02 BROWN
MPL	an Batik Rejoc	WAFIQA DRESS (NEW) M, WAFIQA DRESS (NEW) L, TOMING OUTER
NANA	Semarang	MUNIRA DRESS (RAYON TWILL)

Gambar 4.4 *Preprocessing* Data

Setelah dilakukan *preprocessing* data pada setiap bulan pada tahun 2023, peneliti hanya memakai variabel-variabel kode agen, dan kota. Dalam data *mining* variabel tersebut dibutuhkan untuk mendefinisikan transaksi yang terbentuk serta barang apa saja yang termasuk dalam suatu transaksi. Pada penelitian (Arrafii, 2021), variabel *quantity* tidak digunakan karena jumlah setiap barang yang dibeli tidak berpengaruh, yang memengaruhi adalah apakah suatu barang dibeli atau tidak, dikarenakan ingin mencari tahu apa hubungan membeli suatu barang dan membeli beberapa barang lainnya. Jumlah transaksi yang terjadi di 3 bulan terakhir pada tahun 2023 sebanyak 583 transaksi dan setelah dilakukan reduksi data jumlah transaksi menjadi 547 transaksi.

Pada variabel kode agen yang di maksud yaitu penamaan kode agen berasal dari nama masing-masing *customer service* yang disingkat dan nama *channel* pembelian. Apabila seorang *customer* hendak melakukan pembelian suatu produk, *customer* dapat menghubungi bagian *customer service* yang akan dipilih. Dari perusahaan telah menyediakan beberapa kontak yang dapat dihubungi untuk melakukan transaksi penjualan. Selain itu, *customer* dapat melakukan pembelian melalui sosial media Tiktok, dan Shopee.

#### 4.3.3 *Transformasi Data*

Pada tahapan transformasi data, data yang sudah melalui proses *preprocessing* data akan diubah ke dalam bentuk biner menggunakan bahasa pemrograman *Python* supaya memudahkan peneliti akan mengolah data.

#### 4.3.4 Data Mining

1. Pertama-tama user sudah menginstall *software Jupyter Notebook*, setelah itu membuka program *anaconda prompt* dan menulis syntax “D:” untuk membaca *harddisk* D karena folder yang ingin dibuka terletak di *harddisk* D; “D:\>cd ”Syarat KP\_TA” maksud dari arti kata *cd* artinya *change directory* menuju folder yang di inginkan seperti folder *Python* yang diinginkan peneliti, dan folder *Python* berada di folder TA; “D: Syarat KP\_TA>jupyter notebook” memasukan *syntax jupyter notebook* untuk membuka *software jupyter notebook* yang akan terbuka pada *default browser* pengguna.

```
(base) C:\Users\lisaa>d:
(base) D:\>cd "Syarat KP_TA"
(base) D:\Syarat KP_TA>jupyter notebook
```

Gambar 4.5 *Syntax* Pada *Software Anaconda Prompt*

2. Setelah menjalankan *software jupyter notebook*, membuat file baru berformat *ipnyb* dan mengaktifkan fungsi-fungsi yang terdapat di dalam *Python* seperti mengaktifkkn fungsi, *pandas*, *numpy*, *seaborn*, dan *matplotlib* pada *Python*. Berikut penjelasan berupa *pseudocode*.

```
In [238]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

from sklearn.cluster import KMeans
from sklearn.preprocessing import OneHotEncoder, LabelEncoder, StandardScaler
```

Gambar 4.6 *Pseudocode* Untuk Mengimpor Fungsi

3. Setelah fungsi-fungsi sudah dinyalakan saatnya untuk mengimpor *file excel* ke dalam *software jupyter notebook* dengan cara: pastikan *file* yang diimpor sudah berada didalam folder yang sama dengan *file .ipnyb*, setelah itu menulis *syntax* “*baca = pd.read\_excel("nama file.xlsm")*”; “*baca*”. Kata “*baca =*” mengartikan definisi dari *syntax* yang akan dimasukkan; sedangkan “*pd.read\_excel*” suatu *syntax* untuk membaca *file excel*; “*nama file.xlsm*” adalah *file* yang akan diimpor ke *jupyter notebook* dan “*baca*” adalah suatu perintah untuk menjalankan pemrograman tersebut,

```
In [239]: df = pd.read_excel("./data transaksi januari.xlsx")
df.head()
```

Gambar 4.7 *Pseudocode* Untuk Membaca File

- Selanjutnya mengecek *missing values*, untuk mengidentifikasi nilai yang kosong atau tidak terbaca dalam data. Karena *missing value* bisa menyebabkan perubahan hasil analisis dan mengurangi akurasi model.

```
In [241]: df.isnull().sum()
```

Gambar 4.8 *Pseudocode* Mengidentifikasi Nilai Kosong

- Jika terdapat missing values atau nilai kosong pada data maka harus dihapus dengan kode berikut:

```
In [243]: df.dropna(inplace=True)
```

Gambar 4.9 *Pseudocode* Menghapus Nilai Yang Kosong

- Lalu mengecek nilai duplikat, fungsi *duplicated* ini untuk menemukan nilai yang berulang dalam sebuah *list*, *array*, atau *dataframe*. Terlihat dalam data ini tidak ada nilai duplikat yang artinya data aman.

```
In [242]: df.duplicated().sum()
Out[242]: 0
```

Gambar 4.10 *Pseudocode* Melihat Nilai Duplikat

- Selanjutnya *Exploratory Data Analysis* (EDA) dan *Data Preprocessing*, dalam hal ini yang dilakukan pertama kali yaitu mengecek nilai unik dari data, jumlah data dan *frequency* dari nilai modus dari data. Jika dilihat data *telp* dan nama memiliki nilai unik yang besar dan bisa dibilang nilai semua datanya adalah unik. Oleh karena itu, data *telp* dan nama harus dihapus karena data tersebut dapat membuat model menjadi kurang akurat.

```
In [244]: df.describe(exclude="number").T
```

```
Out[244]:
```

	count	unique	top	freq
kode_agen	110	10	NANA	54
telp	110	110	+62+886932514059	1
nama	110	110	Jumadi	1
daerah	110	20	Jawa Tengah	22
produk	110	77	WAFIQA DRESS (NEW)	8

Gambar 4.11 *Pseudocode* Mengecek Nilai Unik

```
In [245]: df.drop(["telp", "nama"], axis=1, inplace=True)
df.head()
```

Gambar 4.12 *Pseudocode* Menghapus Nilai Unik Yang Besar

8. Selanjutnya melakukan *Univariate Analysis* yaitu analisis yang melihat distribusi hanya satu variabel. Berikut untuk *variable* kode agen, daerah, dan produk:

```
In [246]: # Univariate Analysis of kode_agen
fig, ax = plt.subplots(figsize=(10, 3), layout="constrained")
sns.histplot(data=df["kode_agen"], ax=ax)
ax.set_title("Kode Agen")
plt.show()
```

Gambar 4.13 *Univariate Analysis* Kode agen

```
In [247]: # Univariate Analysis of daerah
fig, ax = plt.subplots(figsize=(30, 15), layout="constrained")
sns.histplot(data=df["daerah"], ax=ax)
ax.set_title("daerah")
plt.show()
```

Gambar 4.14 *Univariate Analysis* Daerah

```
In [248]: # Univariate Analysis of produk
fig, ax = plt.subplots(figsize=(30, 15), layout="constrained")
sns.histplot(data=df["produk"], ax=ax)
ax.set_title("Produk")
plt.show()
```

Gambar 4.15 *Univariate Analysis* Produk

9. Melakukan *Bivariate Analysis* yaitu analisis yang melihat hubungan antara dua variabel. Berikut *Bivariate Analysis* dari variabel kode agen dan daerah, kode agen dan produk.

```
In [249]: # Bivariate Analysis of kode agen vs daerah
fig, ax = plt.subplots(figsize=(12, 5), layout="constrained")
sns.countplot(data=df, x="daerah", hue="kode_agen")
labels = ax.get_xticklabels()
plt.setp(labels, rotation=45, horizontalalignment="right")
plt.legend(bbox_to_anchor=(1, 0.2))
plt.tight_layout()
plt.show()
```

Gambar 4.16 *Bivariate Analysis* Variabel kode agen dan daerah

```
In [250]: # Bivariate Analysis of kode_agen vs produk
fig, ax = plt.subplots(figsize=(12, 5), layout="constrained")
sns.countplot(data=df, x="produk", hue="kode_agen")
labels = ax.get_xticklabels()
plt.setp(labels, rotation=45, horizontalalignment="right")
plt.legend(bbox_to_anchor=(1, 0.2))
plt.tight_layout()
plt.show()
```

Gambar 4.17 *Bivariate Analysis* Variabel kode agen dan produk

10. Selanjutnya, membuat model *machine learning k-means*. Pertama, cek terlebih dahulu tipe datanya dan ubah tipe datanya seperti berikut:

```
In [251]: df.info()
```

Gambar 4.18 *Pseudocode* Membaca File

```
In [252]: df["kode_agen"] = df["kode_agen"].astype("category")
df["daerah"] = df["daerah"].astype("category")
df["produk"] = df["produk"].astype("category")
```

Gambar 4.19 *Pseudocode* Ubah Tipe Data Menjadi Kategori

11. Langkah selanjutnya melakukan *preprocessing* data dengan *LabelEncoder*. Mengubah data kategori menjadi data numerik/angka dengan *LabelEncoder*, agar membuat data lebih mudah dipahami oleh *machine learning*.

```
In [253]: one_dict = {}
for col in df.columns:
    one_dict[col] = LabelEncoder().fit(df[col])
df[col] = one_dict[col].transform(df[col])
df.head()
```

Gambar 4.20 *Pseudocode LabelEncoder*

12. Jika semua data telah menjadi data angka dan siap dimasukkan ke dalam model *machine learning*. Selanjutnya, membuat model *machine learning KMeans*

dengan 3 *cluster*. Menampilkan hasil dari pemodelan ke *pandas* dengan nama *result*. Berikut *Pseudocode*:

```
In [254]: km = KMeans(n_clusters=3)

In [255]: y_predict = km.fit_predict(df)
df["result"] = y_predict
df.head()
```

Gambar 4.21 *Pseudocode K-means Cluster*

13. Selanjutnya untuk melihat visualisasi dari model tersebut, yaitu:

```
In [256]: df1 = df[df.result == 0]
df2 = df[df.result == 1]
df3 = df[df.result == 2]

plt.scatter(df1.qty, df1.jumlah_transaksi, color="green")
plt.scatter(df2.qty, df2.jumlah_transaksi, color="red")
plt.scatter(df3.qty, df3.jumlah_transaksi, color="blue")

plt.xlabel("Kode agen")
plt.ylabel("daerah")
plt.show()
```

Gambar 4.22 *Pseudocode Result*

14. Lalu menghapus variabel "*qty*" dan variabel "*jumlah\_transaksi*" dan membuat ulang model *Kmeans*.

```
In [257]: df.drop(["qty", "jumlah_transaksi"], axis=1, inplace=True)

In [258]: df.head()
```

```
Out[258]:
```

	kode_agen	daerah	produk	result
0	4	5	38	2
1	2	3	16	1
2	2	6	20	1
3	2	10	76	0
4	2	19	1	1

Gambar 4.23 Model *Kmeans*

```
In [259]: kmeans = KMeans(n_clusters=3)
y_predict = kmeans.fit_predict(df)
df["result"] = y_predict
```

Gambar 4.24 *Pseudocode Kmeans*

15. Mengubah data *cluster* 0 menjadi Tinggi, *cluster* 1 menjadi Sedang, dan *cluster* 2 menjadi Rendah.

```
In [260]: choices = {0: "Tinggi", 1: "Sedang", 2: "Rendah"}
df["result"] = [choices[i] for i in df["result"]]
df
```

Gambar 4.25 *Pseudocode* Menamai *Cluster*

16. Kemudian mencoba membalikkan data yang telah ditransformasi ke bentuk semula.

```
In [261]: for i in df.columns[:-1]:
df[i] = one_dict[i].inverse_transform(df[i])

df
```

Gambar 4.26 *Pseudocode* Inverse Transform

17. Setelah itu visualisasikan data *cluster*, hal ini untuk melihat visualisasi berupa *histplot* dengan kode berikut:

```
In [262]: sns.histplot(df["result"])
```

Gambar 4.27 *Pseudocode* Histplot

18. Untuk melihat visualisasi terhadap kode agen hasil *clustering*, yaitu:

```
In [263]: # Bivariate Analysis of kode agen result
fig, ax = plt.subplots(figsize=(12, 5), layout="constrained")
sns.countplot(data=df, x="kode_agen", hue="result")
labels = ax.get_xticklabels()
plt.setp(labels, rotation=45, horizontalalignment="right")
plt.legend(bbox_to_anchor=(1, 0.2))
plt.tight_layout()
plt.show()
```

Gambar 4.28 *Pseudocode* Plot Bivariate Analysis Kode\_agen Result

19. Untuk melihat visualisasi terhadap daerah hasil *clustering*, yaitu:

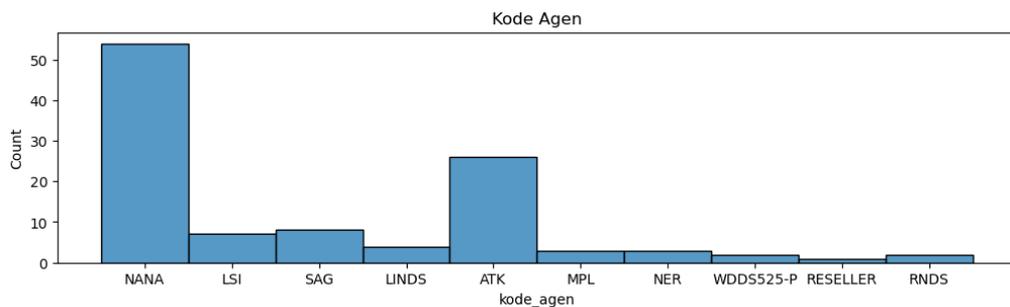
```
In [264]: # Bivariate Analysis of daerah vs result
fig, ax = plt.subplots(figsize=(12, 5), layout="constrained")
sns.countplot(data=df, x="daerah", hue="result")
labels = ax.get_xticklabels()
plt.setp(labels, rotation=45, horizontalalignment="right")
plt.legend(bbox_to_anchor=(1, 0.2))
plt.tight_layout()
plt.show()
```

Gambar 4.29 *Pseudocode* Plot Bivariate Analysis Daerah Result

### 4.3.5 Interpretation/Evaluation

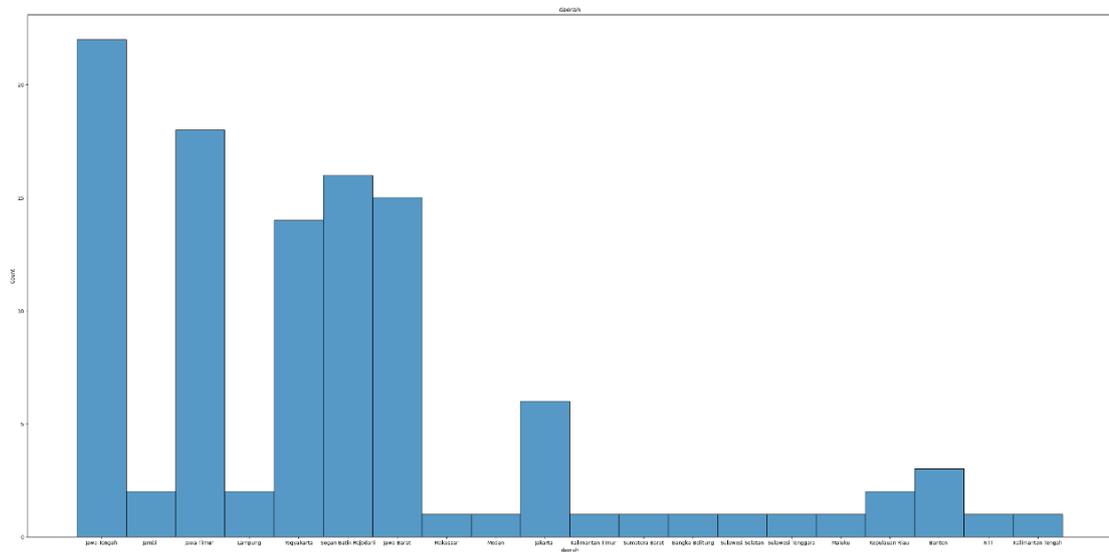
#### 1. Bulan Januari 2023

- Pada gambar di bawah ini merupakan hasil dari data yang telah diolah menggunakan pemrograman *python*, terlihat bahwa kode\_agen Nana adalah yang terbanyak sedangkan untuk kode\_agen Reseller yang paling sedikit.



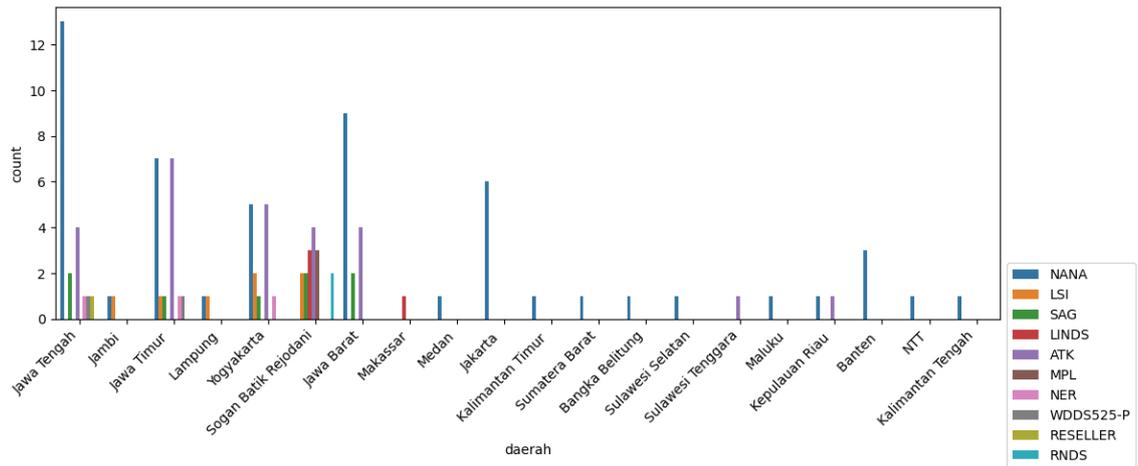
Gambar 4.30 Plot Variabel Kode\_agen Januari

- Gambar di bawah ini merupakan visualisasi dari *Univariate Analysis* dari variabel daerah. Data tersebut menunjukkan bahwa daerah Jawa Tengah dan Jawa Timur merupakan daerah yang terbanyak dalam melakukan transaksi pembelian.



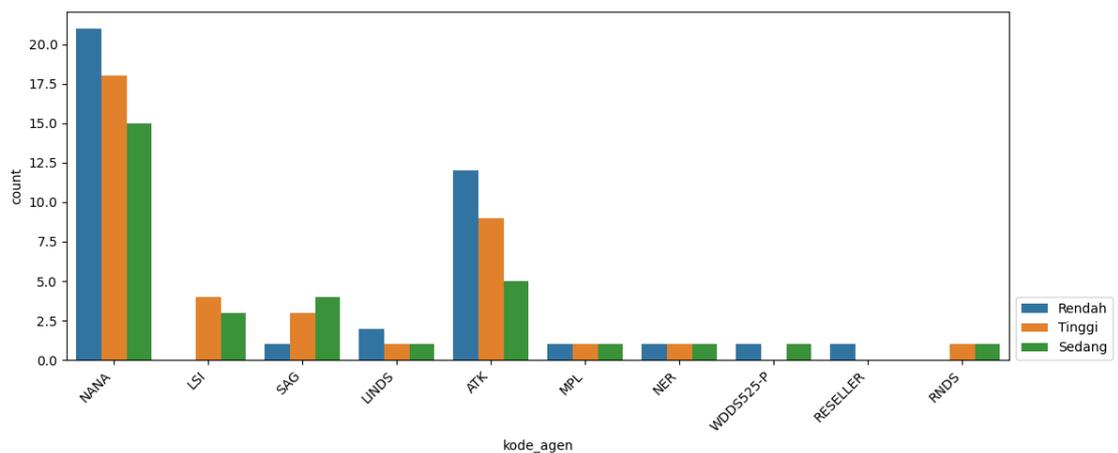
Gambar 4.31 Plot Variabel Daerah Januari

- Gambar di bawah ini merupakan visualisasi dari *Bivariate Analysis* untuk kategori kode\_agen dan daerah. Berdasarkan gambar tersebut, menunjukkan bahwa terdapat 10 agen yang dapat dihubungi untuk melakukan pembelian produk. Dari agen-agen tersebut banyaknya daerah yang memakai produk yaitu dengan kode\_agen Nana.



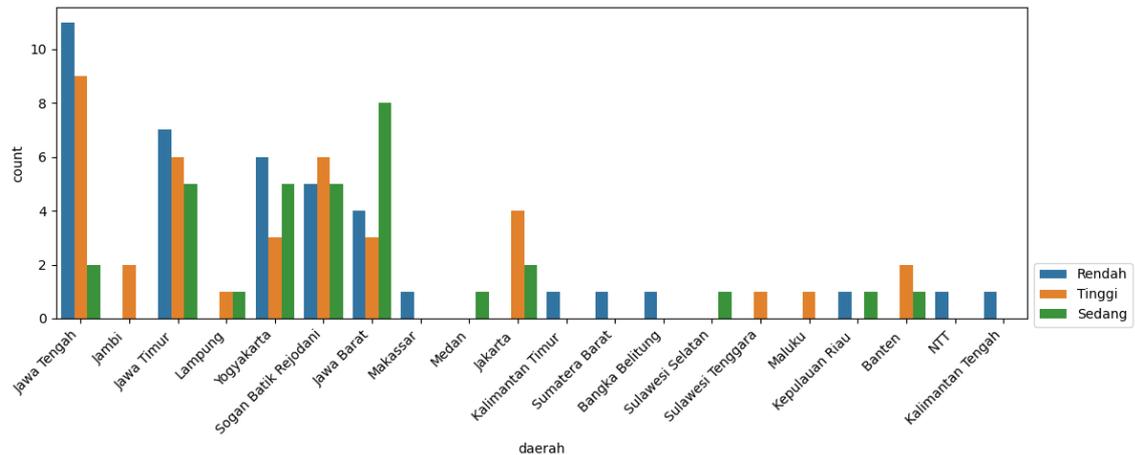
Gambar 4.32 *Plot Variabel Kode\_agen & Daerah Januari*

- Berdasarkan gambar di bawah ini, kode\_agen Nana terlihat yang banyak pemakainya dengan terbagi *cluster* rendah, tinggi, dan sedang. Menampilkan bahwa *cluster* rendah adalah yang terbanyak dibanding yang lainnya.



Gambar 4.33 *Plot Variabel Kode\_agen Result Januari*

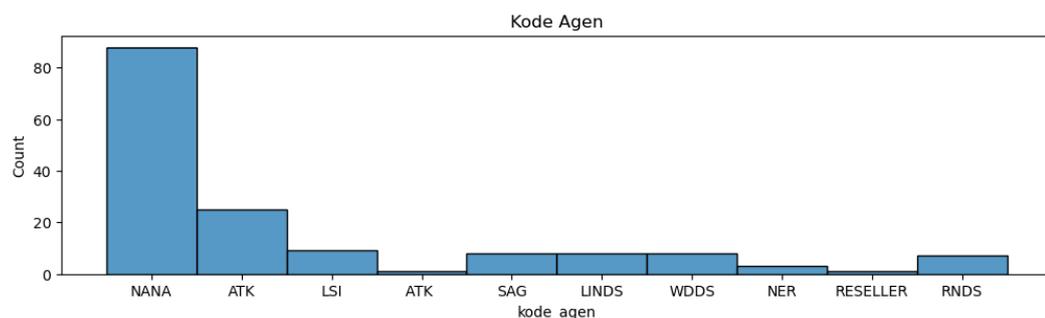
- Gambar ini menampilkan visualisasi hasil *cluster* dari variabel daerah, menunjukkan data terbanyak dalam transaksi pembelian adalah Jawa Tengah dengan kebanyakan tergolong *cluster* rendah dan *cluster* tinggi.



Gambar 4.34 *Plot Variabel Daerah Result Januari*

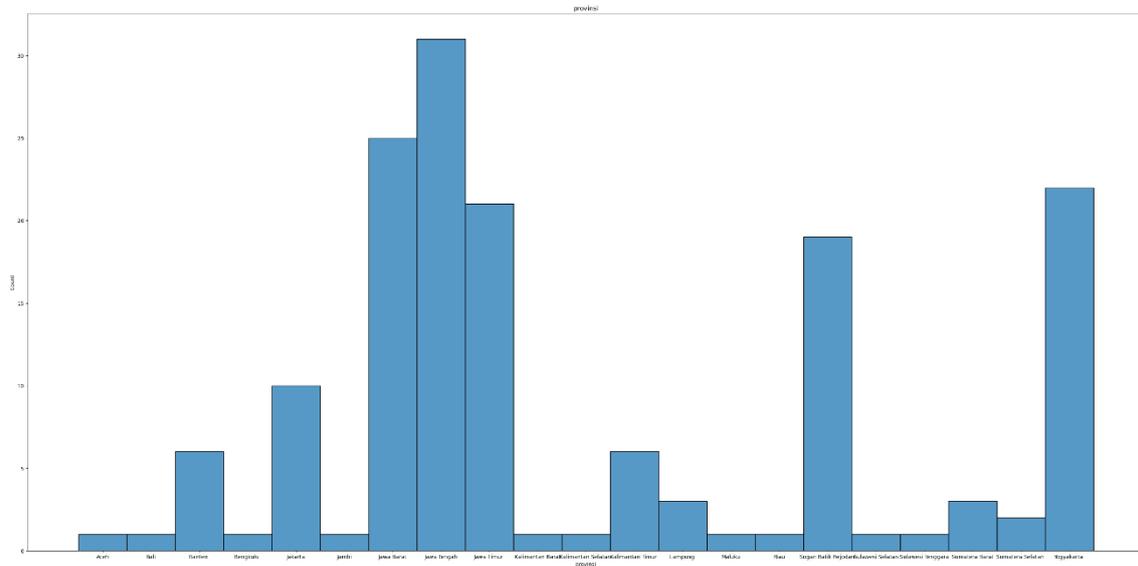
## 2. Bulan Februari 2023

- Pada gambar di bawah ini merupakan hasil dari data yang telah diolah menggunakan pemrograman *python*, terlihat bahwa kode\_agen Nana adalah yang terbanyak sedangkan untuk kode\_agen Reseller dan ATK yang paling sedikit.



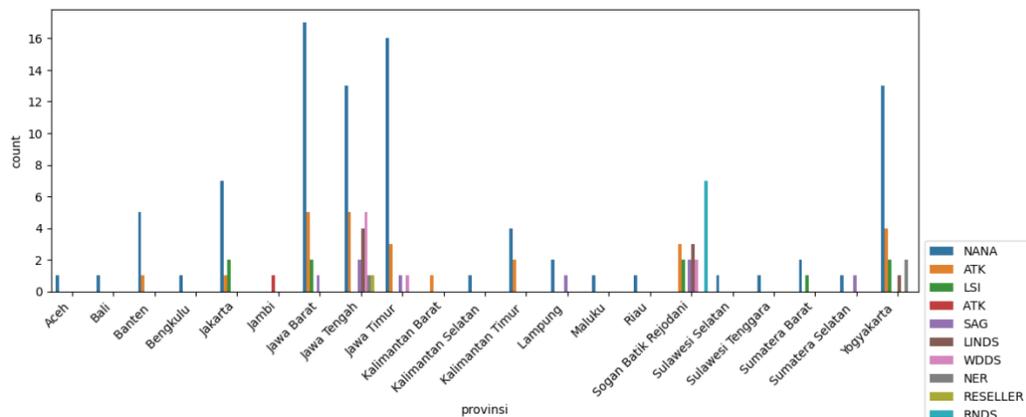
Gambar 4.35 *Plot Variabel Kode\_agen Februari*

- Gambar di bawah ini merupakan visualisasi dari *Univariate Analysis* dari variabel provinsi. Data tersebut menunjukkan bahwa daerah Jawa Tengah dan Jawa Barat merupakan daerah yang terbanyak dalam melakukan transaksi pembelian.



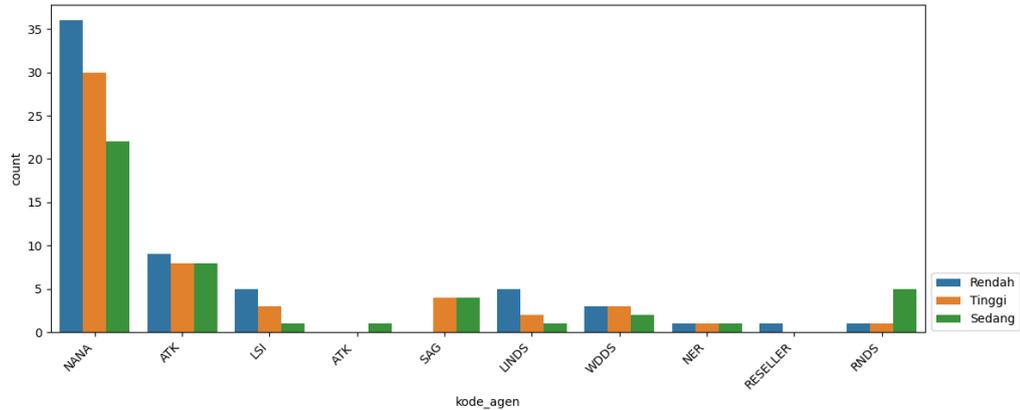
Gambar 4.36 *Plot* Variabel Daerah Februari

- Gambar di bawah ini merupakan visualisasi dari *Bivariate Analysis* untuk kategori kode\_agen dan daerah. Berdasarkan gambar tersebut, menunjukkan bahwa terdapat 10 agen yang dapat dihubungi untuk melakukan pembelian produk. Dari agen-agen tersebut banyaknya daerah yang memakai produk yaitu dengan kode\_agen Nana.



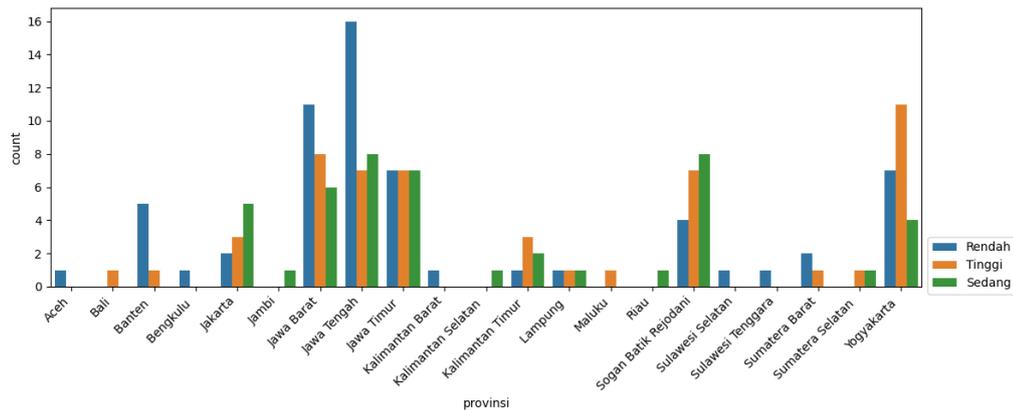
Gambar 4.37 *Plot* Variabel Kode\_agen & Daerah Februari

- Berdasarkan gambar di bawah ini, kode\_agen Nana terlihat yang banyak pemakainya dengan terbagi *cluster* rendah, tinggi, dan sedang. Menampilkan bahwa *cluster* rendah dan *cluster* tinggi adalah yang terbanyak dibanding yang lainnya.



Gambar 4.38 *Plot Variabel Kode\_agen Result Februari*

- Gambar ini menampilkan visualisasi hasil *cluster* dari variabel provinsi, menunjukkan data terbanyak dalam transaksi pembelian adalah Jawa Tengah dengan kebanyakan tergolong *cluster* rendah.

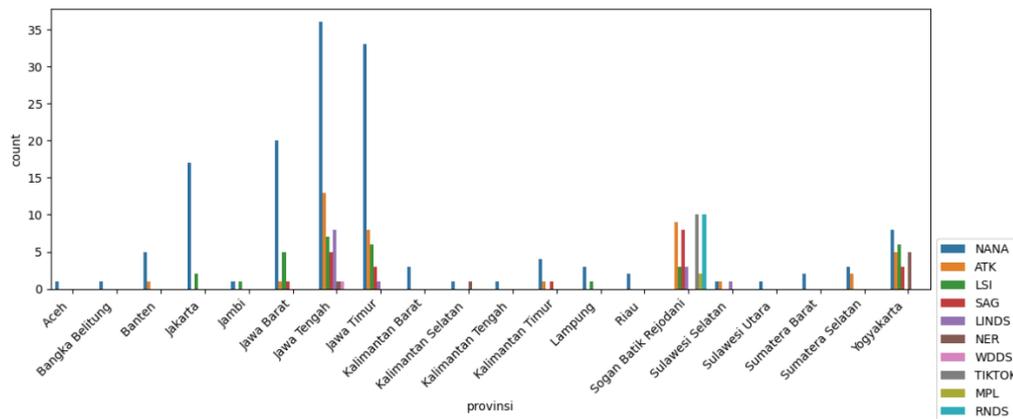


Gambar 4.39 *Plot Variabel Daerah Result Februari*

### 3. Bulan Maret 2023

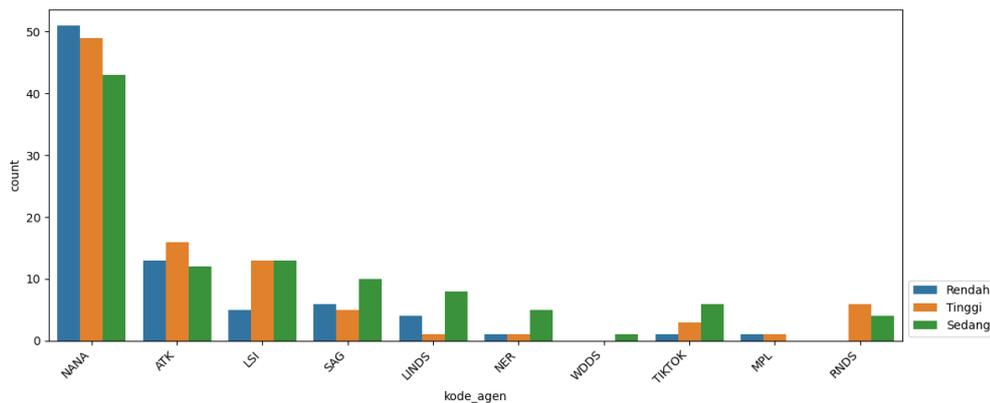
- Pada gambar di bawah ini merupakan hasil dari data yang telah diolah menggunakan pemrograman *python* terlihat bahwa kode\_agen Nana adalah yang terbanyak sedangkan untuk kode\_agen WDDS dan MPL yang paling sedikit.





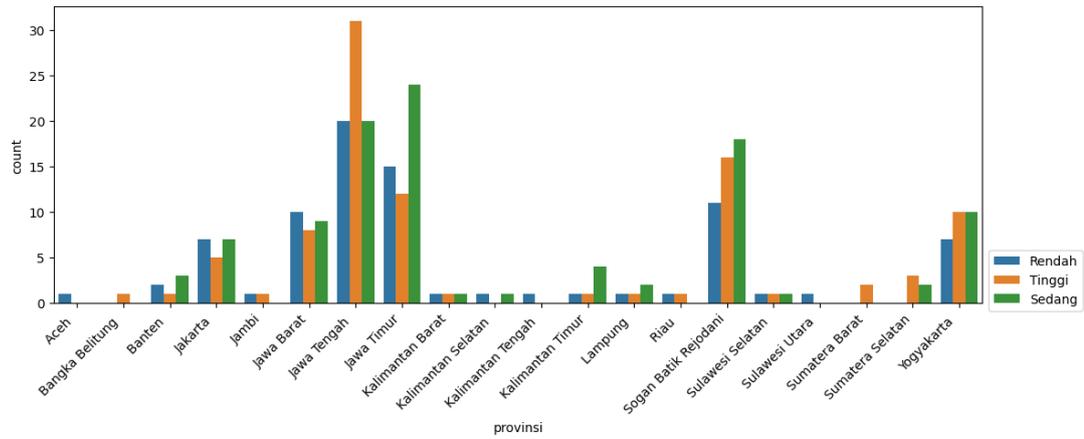
Gambar 4.42 *Plot* Variabel Kode\_agen & Daerah Maret

- Berdasarkan gambar di bawah ini, kode\_agen Nana terlihat yang banyak pemakainya dengan terbagi *cluster* rendah, tinggi, dan sedang. Menampilkan bahwa terdapat *cluster* tinggi, sedang, dan rendah yang terbanyak dibanding yang lainnya.



Gambar 4.43 *Plot* Variabel Kode\_agen *Result* Maret

- Gambar di bawah ini menampilkan visualisasi hasil *cluster* dari variabel provinsi, menunjukkan data terbanyak dalam transaksi pembelian adalah Jawa Tengah dengan kebanyakan tergolong *cluster* rendah dan Jawa Timur dengan kebanyakan tergolong *cluster* tinggi.



Gambar 4.44 *Plot* Variabel Daerah *Result* Maret

## **BAB V**

### **PEMBAHASAN**

#### **5.1 Analisis Transaksi Penjualan**

Dari pengolahan data diketahui terdapat 547 transaksi yang terjadi di Sogan Batik Rejodani pada bulan Januari, Februari, dan Maret di tahun 2023. Dari hasil pengolahan data yang ada, berikut ini peneliti akan menjelaskan hasil data yaitu:

##### **5.1.1 Variabel Kode Agen**

Bagian ini menjelaskan visualisasi *univariate analysis* pada variabel kode agen. Berdasarkan dari riwayat data transaksi Sogan Batik Rejodani selama bulan Januari terdapat 110 data transaksi pembelian. Diketahui bahwa kode agen NANA adalah yang terbanyak dalam melakukan transaksi pembelian, sedangkan untuk kode agen yang paling sedikit melakukan transaksi penjualan selama bulan Januari berada pada kode agen *RESELLER*. Transaksi penjualan terbanyak kedua ada pada agen ATK. Kemudian, kode agen SAG, LSI, LINDS, MPL, NER, WDDS, dan RNDS. Dari banyaknya kode agen tersebut agen NANA lebih mendominasi dibandingkan dengan agen lainnya. Maka aktivitas transaksi penjualan pada bulan Januari selain agen NANA masih terdapat agen yang penjualannya belum memenuhi target, terlihat dari banyaknya transaksi penjualan yang masuk pada agen-agen tersebut.

Pada bulan Februari bahwa kode agen NANA adalah yang terbanyak, sedangkan untuk kode agen *RESELLER* dan ATK yang paling sedikit. Hal ini menunjukkan bahwa *customer* lebih sering melakukan transaksi pembelian dengan menghubungi kode agen NANA. Dapat dilihat pada grafik bahwa kode agen *RESELLER* paling sedikit pada bulan Februari. Agen ATK berada pada posisi kedua terbanyak dilanjut oleh kode agen LSI, SAG, LINDS, WDDS, RNDS, dan terakhir NER. Dari banyaknya kode agen tersebut agen NANA masih lebih mendominasi dibandingkan dengan agen lainnya seperti pada bulan Januari. Aktivitas transaksi penjualan pada bulan Februari selain agen NANA masih terdapat agen yang penjualannya belum memenuhi target, terlihat dari banyaknya transaksi penjualan yang masuk pada agen-agen tersebut.

Pada bulan Maret kode NANA masih tertinggi seperti bulan Januari dan Februari. Sedangkan, untuk agen yang aktivitas transaksi penjualannya rendah yaitu WDDS, dan MPL. Kode agen tertinggi setelah NANA yaitu ada ATK, LSI, SAG, LINDS, TIKTOK, RNDS, dan NER. Dari banyaknya kode agen tersebut, agen NANA lebih mendominasi dibandingkan dengan agen lainnya. Maka, dalam kinerjanya telah sesuai dengan kemampuan dan keahlian dalam melayani *customer* dengan baik. Hal ini diukur melalui produktivitas kerja dan pencapaian target yang telah ditetapkan perusahaan. Dengan memberikan pelayanan yang baik, dapat memikat hati pelanggan. Sehingga, hal ini menjadi salah satu faktor yang mempengaruhi tingginya penjualan oleh agen NANA.

Target penjualan dapat terpenuhi disebabkan oleh faktor kualitas pelayanan yang baik. Menurut (Lovelock & Wirtz, 2014), kualitas pelayanan adalah suatu tolak ukur dari seberapa baik tingkat pelayanan yang diberikan sesuai harapan pelanggan. Kualitas layanan yang baik berarti sesuai dengan harapan pelanggan dan dilakukan secara konsisten. Pada penelitian (Riana & Nafiati, 2021), menjelaskan persepsi kualitas produk dan persepsi kualitas pelayanan berpengaruh positif terhadap tingkat penjualan UMKM. Menurut (Indrasena, 2022), Kualitas layanan dapat diukur melalui pemenuhan keinginan dan kebutuhan pelanggan serta ketepatan penyampaian untuk memenuhi harapan pelanggan. Kualitas layanan yang lebih baik akan membuat konsumen lebih sering membeli produk atau jasa. Maka, pentingnya suatu perusahaan memiliki karyawan yang berkompeten dibidangnya dan menghargai perusahaan, *customer*, serta rekan kerjanya. Berdasarkan penelitian (Suhendi et al., 2017), menjelaskan bahwa dalam meningkatkan target penjualan maka perlu adanya promosi agen dan bonus. Kedua faktor tersebut dapat memberikan motivasi agen dalam meningkatkan kinerjanya. Dengan diberikan promosi dan bonus untuk agen, karyawan pasti merasa nyaman, lebih dihargai, diperhatikan, dibutuhkan kemampuan kinerjanya, dengan demikian akan terjadi sebuah loyalitas yang tinggi pada perusahaan.

### 5.1.2 Variabel Daerah

Pada variabel daerah Jawa Tengah merupakan daerah yang terbanyak dalam melakukan transaksi pembelian kemudian dilanjut dengan Jawa Timur, Sogan Batik, Jawa Barat, dan Yogyakarta. Untuk penamaan pada daerah Sogan Batik artinya beberapa transaksi pembelian dilakukan di tempat tersebut atau *Gallery* sogan. Terdapat 10 provinsi (daerah) yang memiliki data transaksi pembelian terkecil yaitu Makasar, Medan, Kalimantan Timur, Sumatera Barat, Bangka Belitung, Sulawesi Selatan, Sulawesi Tenggara, Maluku, NTT, Kalimantan Tengah. Berdasarkan data, banyaknya peminat pada daerah-daerah tersebut masih cenderung sedikit dan jarang. Hal ini dapat disebabkan oleh faktor jarak antar luar pulau, karena terdapat beberapa *customer* dengan alamat pengirimannya di luar pulau. Sehingga, berakibat pada ongkos kirim yang lebih mahal dibandingkan dengan pengiriman di pulau Jawa.

Pada bulan Februari, menampilkan 20 daerah yang pernah melakukan transaksi pembelian produk sogan yaitu ada Aceh, Bali, Banten, Bengkulu, Jakarta, Jawa Barat, Jawa Tengah, Jawa Timur, Kalimantan Barat, Kalimantan Selatan, Kalimantan Timur, Lampung, Maluku, Riau, Sogan Batik Rejodani, Sulawesi Selatan, Sulawesi Tenggara, Sumatera Barat, Sumatera Selatan, dan Yogyakarta. Terlihat bahwa Jawa Tengah merupakan daerah yang terbanyak dalam melakukan transaksi pembelian lalu daerah banyak kedua dan ketiga yaitu Jawa Barat dan Yogyakarta. Hal ini terjadi karena sebagian besar *customer* berasal dari daerah tersebut. Berdasarkan data, banyaknya peminat pada daerah-daerah tersebut masih cenderung sedikit dan jarang. Hal ini dapat disebabkan oleh faktor jarak antar luar pulau, karena terdapat beberapa *customer* dengan alamat pengirimannya di luar pulau. Sehingga, berakibat pada ongkos kirim yang lebih mahal dibandingkan dengan pengiriman di pulau Jawa. Terlihat bahwa dalam penjualan daerah yang peminatnya lebih banyak yaitu *customer* yang berada di pulau Jawa.

Pada bulan Maret, terlihat 20 daerah yang pernah melakukan transaksi pembelian produk sogan yaitu ada Aceh, Bangka Belitung, Banten, Jakarta, Jambi, Jawa Barat, Jawa Tengah, Jawa Timur, Kalimantan Barat, Kalimantan Selatan, Kalimantan Tengah, Kalimantan Timur, Lampung, Riau, Sogan Batik Rejodani, Sulawesi Selatan, Sulawesi Utara, Sumatera Barat, Sumatera Selatan, dan Yogyakarta. Terlihat bahwa Jawa Tengah

merupakan daerah yang terbanyak dalam melakukan transaksi pembelian lalu daerah banyak kedua dan ketiga yaitu Jawa Timur dan Yogyakarta. Hal ini terjadi karena sebagian besar *customer* berasal dari daerah tersebut. Sedangkan untuk daerah yang paling rendah ada Aceh, Bangka Belitung, Kalimantan Tengah, Sulawesi Utara, Jambi, Kalimantan Selatan, Riau, Sumatera Barat, Kalimantan Barat, Sulawesi Selatan, Lampung, Sumatera Selatan, Banten, dan Kalimantan Timur. Artinya peminat produk sogan (*customer*) dari daerah-daerah tersebut cenderung lebih sedikit dibandingkan dengan daerah lainnya. Banyaknya peminat di daerah-daerah tersebut masih cenderung sedikit dan jarang. Hal ini dapat disebabkan oleh faktor jarak antar luar pulau, karena terdapat beberapa *customer* dengan alamat pengirimannya di luar pulau. Sehingga, berakibat pada ongkos kirim yang lebih mahal dibandingkan dengan pengiriman di pulau Jawa. Terlihat bahwa dalam penjualan daerah yang peminatnya lebih banyak yaitu *customer* yang berada di pulau Jawa.

Menurut (Indrasena, 2022), promosi berpengaruh terhadap target penjualan. Promosi penjualan adalah bagian penting dari strategi pemasaran karena dapat menarik respon pelanggan lebih cepat dan efektif karena memberi penjual nilai positif untuk menarik pelanggan dan meningkatkan penjualan. Hal ini didukung oleh penelitian (Syafira & Sukmawati, 2021), bahwa promosi berpengaruh secara signifikan terhadap keputusan pembelian.

### **5.1.3 Variabel Kode Agen dan Daerah**

Pada bagian ini menjelaskan visualisasi *bivariate analysis* pada variabel kode agen dan daerah. Di bulan Januari, terdapat 10 kode agen yang ada pada Sogan Batik yaitu NANA, LSI, SAGA, LINDS, ATK, MPL, NER, WDDS, RESELLER, dan RINDS. Kode agen ini merupakan nama dari *customer service* sogan batik. Dari sepuluh kode agen tersebut berdasarkan visualisasi yang didapatkan kategori kode agen dan daerah, *customer service* NANA yang paling tinggi dalam melakukan transaksi pembelian. Maka, *customer* lebih sering menghubungi kode agen NANA untuk melakukan order produk. Pada daerah Jawa Tengah menampilkan bahwa kode NANA memiliki grafik yang tinggi, bahwa transaksi pembelian pada daerah tersebut didominasi oleh kode NANA. Daerah yang didominasi oleh kode NANA selain Jawa Tengah yaitu Jawa Barat, Jawa Timur, dan Jakarta. Serta

terlihat juga bahwa kode ATK di daerah Jawa Timur cukup banyak. Faktor yang menyebabkan agen NANA selalu mendapatkan penjualan tertinggi dibandingkan agen lain yaitu agen NANA dapat memberikan pelayanan yang baik sehingga *customer* di daerah tersebut lebih memilih agen NANA.

Pada Bulan Februari kategori kode agen dan daerah menunjukkan bahwa kode agen yang lebih sering muncul pada 20 daerah yaitu kode agen NANA dan yang paling tinggi ada berada pada daerah Jawa Barat, Jawa Timur, Jawa Tengah, dan Yogyakarta. Adanya aktivitas transaksi pembelian lebih sering dilakukan oleh kode agen NANA pada daerah-daerah tersebut. Sebaliknya, begitu juga dengan kode agen yang jarang muncul pada setiap daerah yang melakukan transaksi pembelian. Jika kode agen jarang muncul di daerah tersebut artinya tidak ada atau sedikitnya aktivitas transaksi penjualan yang dilakukan pada kode agen tersebut.

Pada bulan Maret diketahui kode NANA lebih mendominasi di 20 daerah dibandingkan dengan kode lainnya. Daerah tertinggi ada pada Jawa Tengah dan Jawa Timur dan terdapat beberapa kode agen juga ada pada daerah tersebut namun, tidak terlalu tinggi seperti kode agen NANA. Artinya pada daerah tersebut kode NANA lebih sering melakukan aktivitas transaksi penjualan, begitu juga dengan kode agen lainnya. Sebaliknya, apabila kode agen jarang muncul pada daerah-daerah tersebut maka agen tersebut jarang melakukan transaksi penjualan pada daerah tersebut. Penjualan lebih banyak di daerah tersebut dengan kode agen NANA, faktor yang mempengaruhi penjualan NANA karena memberikan pelayanan yang baik, seperti dalam memberikan informasi terkait produk baru kepada *customer*.

#### **5.1.4 Clustering Variabel Kode Agen**

Pada bulan Januari variabel kode\_agen terbagi menjadi 3 *cluster*, yaitu *cluster* 0 atau rendah, *cluster* 1 atau tinggi, dan *cluster* 2 atau sedang. Terlihat bahwa kode NANA yang paling banyak pemakainya dengan *cluster* 0, *cluster* 1 dan *cluster* 2. Yang termasuk dalam *cluster* rendah adalah yang terbanyak dibanding yang lainnya. Artinya *customer* lebih sering melakukan transaksi pembelian melalui agen tersebut. Terdapat beberapa agen yang hanya memiliki dua *cluster* yaitu kode agen LSI, WDDS, dan RNDS serta satu *cluster* yaitu kode agen RESELLER dengan masing-masing tinggi grafiknya cukup

rendah. Hal ini terjadi karena *customer* tidak banyak membeli produk melalui kode agen tersebut. Faktor mengapa kode agen NANA dalam *cluster* 0 (rendah) cukup tinggi, hal tersebut dikarenakan banyaknya *customer* melakukan pembelian produk tidak lebih dari 3 item.

Pada bulan Februari dari banyaknya kode agen pada perusahaan tersebut menunjukkan bahwa kode agen NANA termasuk *cluster* 0, *cluster* 1, dan *cluster* 2 yaitu rendah, tinggi, dan sedang. Pada tampilan terlihat bahwa *cluster* 0 lebih banyak dibanding *cluster* lainnya. Maka, banyaknya *customer* lebih sering melakukan pembelian tidak lebih dari 3 *items* pada kode agen tersebut.

Pada bulan Maret diketahui bahwa kode agen NANA, ATK, LSI, SAG, LINDS, NER, dan TIKTOK termasuk ke dalam *cluster* 0, *cluster* 1, dan *cluster* 2. Sedangkan WDDS termasuk *cluster* 1, MPL termasuk *cluster* 0 dan *cluster* 2, dan RNDS termasuk *cluster* 1 dan 2. Dari banyaknya kode agen, menunjukkan bahwa kode agen NANA adalah yang terbanyak dalam *cluster* 0 dibandingkan dengan *cluster* lainnya. Maka, banyaknya *customer* lebih sering melakukan pembelian tidak lebih dari 3 item pada kode agen tersebut.

### **5.1.5 Clustering Variabel Daerah**

Pada bulan Januari, hasil *cluster* dari variabel daerah berdasarkan produk yang dibeli menunjukkan bahwa data terbanyak dalam transaksi pembelian adalah Jawa Tengah dengan kebanyakan tergolong *cluster* rendah dan *cluster* tinggi. Maka, pada daerah tersebut peminat untuk membeli produk soğan cukup banyak karena terbagi menjadi tiga *cluster* rendah, tinggi, dan sedang. Namun, yang paling tinggi adalah *cluster* 0 atau rendah. Pada daerah Jambi hanya memiliki satu *cluster* yaitu *cluster* 1 atau tinggi, artinya tingkat banyaknya produk yang dibeli *customer* cukup banyak namun tidak banyak *customer* dari daerah Jambi yang melakukan transaksi pembelian produk soğan. Begitu juga dengan daerah-daerah lain yang hanya memiliki satu *cluster*. Faktor mengapa daerah tersebut dalam *cluster* 0 (rendah) cukup tinggi, hal tersebut dikarenakan banyaknya *customer* melakukan pembelian produk tidak lebih dari 3 item.

Pada bulan Februari diketahui hasil *cluster* menampilkan bahwa variabel daerah Jawa Tengah termasuk ke dalam *cluster* 0, *cluster* 1, dan *cluster* 2. Terlihat bahwa *cluster*

0 lebih banyak dibandingkan dengan cluster lainnya. Artinya banyaknya *customer* lebih sering melakukan pembelian tidak lebih dari 3 *items* pada daerah tersebut.

Pada bulan Maret diketahui bahwa daerah Aceh, Kalimantan Tengah, dan Sulawesi Utara termasuk *cluster* 2. Daerah Bangka Belitung dan Sumatera Barat termasuk *cluster* 0. Daerah Banten, Jakarta, Jawa Barat, Jawa Tengah, Jawa Timur, Kalimantan Barat, Kalimantan Timur, Lampung, Sogan Batik, Sulawesi Selatan, dan Yogyakarta termasuk *cluster* 0, *cluster* 1, dan *cluster* 2. Daerah Jambi dan Riau termasuk *cluster* 1 dan *cluster* 2. Daerah Kalimantan Selatan termasuk *cluster* 0 dan *cluster* 1. Dari 20 daerah tersebut Jawa Tengah adalah daerah yang tingkat *cluster* 0 lebih banyak dibandingkan dengan *cluster* lainnya. Maka, banyaknya *customer* lebih sering melakukan pembelian tidak lebih dari 3 *items* pada daerah tersebut. Sehingga, banyaknya *cluster* 1 terdapat pada daerah Jawa Timur. Oleh sebab itu, banyaknya *customer* lebih sering melakukan pembelian lebih dari 5 *items* pada daerah tersebut.

## 5.2 Rekomendasi

Berdasarkan pembahasan di atas, masih terdapat kode agen dan daerah yang penjualannya masih rendah. Maka dari itu, berikut merupakan rekomendasi yang peneliti bisa berikan, yaitu:

1. Perusahaan dapat melakukan analisis lebih luas terhadap daerah yang sudah di klusterkan. Selain variabel kode agen dan daerah, variabel lainnya masih bisa dimanfaatkan lagi untuk mendapatkan sebuah *insight*. Yaitu variabel produk, tanggal transaksi pembelian, dan lain sebagainya.
2. Perusahaan dapat memberikan promo gratis ongkir untuk daerah yang berada di luar pulau Jawa. Karena terdapat daerah luar pulau Jawa yang melakukan pemesanan produk. Namun, beberapa daerah peminatnya cenderung sedikit. Oleh karena itu, ongkos kirim bisa menjadi faktor dalam keputusan pembelian konsumen.

## **BAB VI**

### **PENUTUP**

#### **6.1 Kesimpulan**

Berdasarkan hasil pembahasan di atas, didapatkan kesimpulan sebagai berikut ini.

1. Berdasarkan data bulan Januari 2023 daerah Jawa Tengah adalah daerah yang terbanyak dalam transaksi pembelian dengan kebanyakan data tergolong *cluster* 0 (rendah) dan *cluster* 1 (tinggi). Pada bulan Februari dan Maret, daerah Jawa Tengah masih terlihat sebagai daerah yang terbanyak dalam transaksi pembelian dengan data tergolong dalam *cluster* 0 (rendah) lebih banyak dibandingkan dengan *cluster* lainnya. Artinya banyaknya *customer* lebih sering melakukan pembelian tidak lebih dari 3 *items* pada daerah tersebut.
2. Dari 3 bulan terakhir, agen NANA lebih mendominasi dibandingkan dengan agen lainnya. Pada variabel daerah, Jawa Tengah lebih mendominasi dibandingkan dengan daerah yang berada di luar pulau jawa.

#### **6.2 Saran**

Berikut merupakan saran dari peneliti yang bisa diberikan, yaitu:

1. Pada penelitian selanjutnya, dapat menggunakan jumlah transaksi yang lebih besar dengan periode waktu yang lebih lama. Sehingga, nilai data transaksi penjualan yang dihasilkan menghasilkan tingkat akurasi yang lebih tinggi.
2. Data dari penelitian ini, dapat dilakukan analisis lebih luas pada variabel-variabel lainnya yaitu variabel produk, tanggal transaksi pembelian, dan lain sebagainya.
3. Dapat menggunakan teknik, algoritma, dan aplikasi *data mining* yang berbeda sehingga dapat menghasilkan informasi yang variatif.

## DAFTAR PUSTAKA

- Afiasari, N., Suarna, N., & Rahaningsi, N. (2023). Implementasi Data Mining Transaksi Penjualan Menggunakan Algoritma Clustering dengan Metode K-Means. *Jurnal SAINTEKOM*, 13(1), 100–110. <https://doi.org/10.33020/saintekom.v13i1.402>
- Al Faruqi, M. (2021). Sistem Pemetaan Posisi Objek Kendaraan Menggunakan Pengolahan Citra Pada Area 360°. (*Doctoral Dissertation, Universitas Komputer Indonesia*).
- Andayani, S. (2007). Pembentukan cluster dalam Knowledge Discovery in Database dengan Algoritma K-Means. In *Seminar Nasional Matematika Dan Pendidikan Matematika*, 70–75.
- Annas. (2023). Data Mining Methods: K-Means Clustering Algorithms. *International Journal of Cyber and IT Service Management (IJCITSM)*, 3(1), 40–47.
- Arrafii, F. R. (2021). *ANALISIS POLA PERILAKU KONSUMEN MENGGUNAKAN METODE MARKET BASKET ANALYSIS (MBA) PADA TOKO SAYUR KELUARGA (TOSAGA) DI MASA PANDEMI COVID-19*.
- Arvio, Y., Riwurohi, J. E., Erno, ;, & Dewantara, K. (2022). Pemodelan Segmentasi Transaksi Jual Beli Produk Menggunakan Pendekatan Model K-Means dan Subtractive Clustering Studi Kasus Survey Pada Beberapa Cabang Optik Retail. *Prosiding Seminar Nasional Energi*, 3.
- Asgari, T., Daneshvar, A., Chobar, A. P., Ebrahimi, M., & Abrahamyan, S. (2022). Identifying key success factors for startups With sentiment analysis using text data mining. *International Journal of Engineering Business Management*, 14, 1–18. <https://doi.org/10.1177/18479790221131612>
- Asroni, A., Fitri, H., & Prasetyo, E. (2018). Penerapan Metode Clustering dengan Algoritma K-Means pada Pengelompokan Data Calon Mahasiswa Baru di Universitas Muhammadiyah Yogyakarta (Studi Kasus: Fakultas Kedokteran dan Ilmu Kesehatan, dan Fakultas Ilmu Sosial dan Ilmu Politik). *Semesta Teknika*, 21(1). <https://doi.org/10.18196/st.211211>
- Berkhin, P. (2006). A survey of clustering data mining techniques. In *Grouping multidimensional data: Recent advances in clustering. Heidelberg: Springer Berlin Heidelberg*, 25–71.
- Dofiri, D. S. (2023). Pendekatan K-means Cluster Pada Pengelompokan Hotel Berbintang Di Kota Pontianak. *Kabilah: Journal of Social Community*, 8(1), 412–419.

- Garcia, M. H., Ullman, J. D., & Widom, J. (2009). *DATABASE SYSTEMS The Complete Book Second Edition*.
- Gladju, J., Kamalam, B. S., & Kanagaraj, A. (2022). Applications of data mining and machine learning framework in aquaculture and fisheries: A review. In *Smart Agricultural Technology* (Vol. 2). Elsevier B.V. <https://doi.org/10.1016/j.atech.2022.100061>
- Hayashi, C. (1988). What is data science? Fundamental concepts and a heuristic example. In *Data Science, Classification, and Related Methods: Proceedings of the Fifth Conference of the International Federation of Classification Societies (IFCS-96)*, 40–51.
- Hu, H., Liu, J., Zhang, X., & Fang, M. (2023). An Effective and Adaptable K-means Algorithm for Big Data Cluster Analysis. *Pattern Recognition*, 139.
- Indrasena, R. W. (2022). Pengaruh Kualitas Pelayanan, Harga, Dan Promosi Terhadap Keputusan Pembelian Pada Shopee Food. *Jurnal Ilmu Dan Riset Manajemen (JIRM)*.
- Islami, A. A. (2023). Pengelompokan pembagian zakat dengan menggunakan metode clustering k-means. *Jurnal CoSciTech (Computer Science and Information Technology)*, 154–163.
- Jang, J. S. R., Sun, C. T., & Mizutani, E. (2005). Neuro-Fuzzy and Soft Computing-A Computational Approach to Learning and Machine Intelligence [Book Review]. *IEEE Transactions on Automatic Control*, 42(10), 1482–1484. <https://doi.org/10.1109/tac.1997.633847>
- Khatib Sulaiman, J., Setiyawati, D., & Cahyono, N. (2023). Analisa Sentimen Pengguna Sosial Media Twitter Terhadap Perokok di Indonesia. *Indonesian Journal of Computer Science Attribution*.
- Larose, D. T., & Larose, C. D. (2014). Discovering knowledge in data: an introduction to data mining. *John Wiley & Sons*, 4.
- Lovelock, C., & Wirtz, J. (2014). *Services Marketing: People, Technology, Strategy*. Pearson Education, Inc. New Jersey.
- Manalu, D. A., & Gunadi, G. (2022). IMPLEMENTASI METODE DATA MINING K-MEANS CLUSTERING TERHADAP DATA PEMBAYARAN TRANSAKSI MENGGUNAKAN BAHASA PEMROGRAMAN PYTHON PADA CV DIGITAL DIMENSI. *Infotech: Journal of Technology Information*, 8(1), 43–54. <https://doi.org/10.37365/jti.v8i1.131>

- Negara, I. S. (2021). Analisa Cluster Data Transaksi Penjualan Minimarket Selama Pandemi Covid-19 dengan Algoritma K-means. *JOINTECS (Journal of Information Technology and Computer Science)*, 153–163.
- Normah, Rifai, B., & Sari Pita. (2020). Analisa Transaksi Penjualan Obat menggunakan Algoritma Apriori. *Jurnal Informatika Dan Komputer*, 22(2).
- Nurajizah, S. (2019). Analisa Transaksi Penjualan Obat menggunakan Algoritma Apriori. *Jurnal INOVTEK Polbeng- Seri Informatika* , 4(1).
- Nurajizah, S., & Salbinda, A. (2021). Penerapan Data Mining Metode K-Means Clustering Untuk Analisa Penjualan Pada Toko Fashion Hijab Banten. *Jurnal Teknik Komputer AMIK BSI*, 7(2). <https://doi.org/10.31294/jtk.v4i2>
- Perkovic, L. (2015). *Introduction to computing using python: An application development focus* (Second edition). John Wiley & Sons.
- Prasetyo, E. (2012). Data Mining: Konsep dan Aplikasi menggunakan MATLAB. *Data Mining*.
- Reinsel, D., Gantz, J., & Rydning, J. (2017). *Data Age 2025: The Evolution of Data to Life-Critical Don't Focus on Big Data; Focus on the Data That's Big*. [www.idc.com](http://www.idc.com)
- Riana, R. I., & Nafiati, L. (2021). Pengaruh Persepsi Etika Bisnis Islam, Persepsi Kualitas Produk Dan Persepsi Kualitas Pelayanan Terhadap Tingkat Penjualan UMKM Kota Yogyakarta. *Jurnal REKSA: Rekayasa Keuangan, Syariah, Dan Audit*, 08(01), 59–66.
- Rosianingsih, S., Setiawan, A., & Boentarman, M. (2021). Customer clustering using K-means clustering: Supporting customer relationship management system. *The 6th Engineering Science and Technology International Conference (ESTIC 2021)*, 2691(1).
- Santoso, B. (2007). *Data mining : Teknik pemanfaatan data untuk keperluan bisnis*. Graha Ilmu .
- Sarhan, A. M. (2022). Data Mining in Internet of Things Systems: A Literature Review. In *Journal of Engineering Research* (Vol. 6, Issue 5).
- Sharda, R., Delen, D., & Turban, E. (2018). Business intelligence, analytics, and data science: a managerial perspective. *Business Intelligence*.
- Siregar, M. H. (2018). KLASERISASI PENJUALAN ALAT-ALAT BANGUNAN MENGGUNAKAN METODE K-MEANS (STUDI KASUS DI TOKO ADI BANGUNAN). *Jurnal Teknologi Dan Open Source*, 1(2), 83–91.

- Snijders, Matzat, Reips, & U-d. (2012). “Big Data” : big gaps of knowledge in the field of internet science Citation for. [www.tue.nl/taverne](http://www.tue.nl/taverne)
- Sönmez, F., Perdahçı, Z. N., & Aydin, M. N. (2019). Big Data Analytics and Models. *Handbook of Research on Big Data Clustering and Machine Learning*, 10–33.
- Suhendi, D., Bonus, P., Pencapaian, A., Penjualan, T., & Jiwa, J. A. (2017). Graduate Program Universitas Galuh Master of Manajemen Studies Program PENGARUH PROMOSI DAN BONUS AGEN TERHADAP PENCAPAIAN TARGET PENJUALAN JASA ASURANSI Jiwa PT. BRINGIN Jiwa SEJAHTERA. 1, 113–122. <http://jurnal.unigal.ac.id/index.php/managementreview>
- Sun, Y., & Tan, X. (2022). Customer Relationship Management Based on SPRINT Classification Algorithm under Data Mining Technology. *Computational Intelligence and Neuroscience*, 2022. <https://doi.org/10.1155/2022/6170335>
- Suntoro, J. (2019). *Data mining: Algoritma dan Implementasi dengan Pemrograman PHP*. PT Elex Media Komputindo.
- Syafira, B., & Sukmawati, N. (2021). ANALISIS PERSEPSI HARGA, PROMOSI, KUALITAS LAYANAN, DAN KEMUDAHAN PENGGUNAAN TERHADAP KEPUTUSAN PEMBELIAN PRODUK PADA MARKETPLACE SHOPEE.
- Syahrudin, A. N., & Kurniawan, T. (2018). *Input dan Output Pada Bahasa Pemrograman Python*.
- Tolle, K. M., Tansley, D. S. W., & Hey, A. J. G. (2011). The fourth Paradigm: Data-intensive scientific discovery. *Proceedings of the IEEE*, 99(8), 1334–1337. <https://doi.org/10.1109/JPROC.2011.2155130>
- Tryana, M., Chalis, S., & Hasibuan, F. (2021). *DATA SCIENCE Strategi UMKM Dalam Pengambilan Keputusan*.
- Turban, E., Aronson, J. E., & Liang, T.-P. (2005). *Decision Support System and Intelligent Systems*.
- Untoro, M. C., Wisnubroto, M. S., Nurdyanti, Y., & Farida, D. R. (2023). Analysis K-means of Covid-19 cases in Bandar Lampung and South Lampung. *In AIP Conference Proceedings*.
- Wahyudi, T., & Silfia, T. (2022). IMPLEMENTATION OF DATA MINING USING K-MEANS CLUSTERING METHOD TO DETERMINE SALES STRATEGY IN S&R BABY STORE. *In Journal of Applied Engineering and Technological Science* (Vol. 4, Issue 1).
- Witten, I. H., Frank, E., & Hall, M. A. (2016). *Data Mining Practical Machine Learning Tools and Techniques Third Edition*.

## LAMPIRAN

Lampiran 1 data transaksi penjualan yang telah di pindahkan dari *website* perusahaan ke dalam bentuk *excel* yang akan diolah menggunakan *Python*.

Kode_Agen	Telp	Nama	Kota	Nama Produk	Qty	Jumlah
LINDS	+62 813-1622-4333	Ibu Susi	Sleman	OKEKUNI TUNIK	1	
LSI	+62.0813.1582.5879.	Sari Apriana Rahmawati	Jakarta Timur, Pasar Rebo	SYAKIRA DRESS, SYAKIRA KOKO L.PANJANG, S	1, 1, 1, 1	
LSI	+62.0813.2778.8488.	Budi turyaman	Cianis, Lakkok	BERNAUNG DRESS HITAM (NEW)	1	
LSI	+62.0813.6312.4403.	INDAH	LUBUK SIKARAH	SYAKIRA DRESS, SYAKIRA DRESS	1, 1	
LSI	+62.65.8227.4176.	Juwani	SOGAN BATIK REJODANI	TAAT DRESS	1	
LSI	+62.6654.0082.	CCE (Markin Yulianti Mongkok HK +852 6654 00 NGAGLIK	Yogyakarta, Kraton	ITIBIS TUNIC HITAM 01, SABAR DRESS BRICK (N	1, 1	
LSI	+62.6844.3259.	Indo Express (Susan Isnanto HK 68443259)	Yogyakarta, Kraton	OKEKUNI TUNIK, JAMILA TUNIK, Hijab Voal Wu	1, 1, 1, 1	
LSI	+62.9549.0929.	Yunus Dwi Lestari AAN Express Taiwai HK +8529	SOGAN BATIK REJODANI	SYAKIRA DRESS, SYAKIRA DRESS, SYAKIRA DR	1, 1, 1, 1, 1	
LSI	+62.9609.8510.	Berkah Jaya Express (BJE) Wina (96098510)	DEPOK-SLM	SHAHABI DRESS HITAM (NEW), ILMISHAHABI K	1, 1, 1, 1	
LSI	+62.9851.7795.	LN Express (Zamah +852 98517795)	Jakarta Timur, Ciracas	TAAT DRESS, Hijab Voal Wulan Sari	1, 1	
SAG	6.20812E+12	IBU DEWI	Klaten, Klaten Selatan	INDANG OUTER, PURNAMA LONGOUTER	1, 1	
SAG	6.20812E+12	IBU NIDA	Sogan Batik Rejodani	TAAT DRESS	1	
SAG	6.2081272585505	TUHI HAKIM	Metro, Metro Timur	LANGOT MADINA ABAYA, ZAMRUD MADINA KE	1, 1, 1	
SAG	6.20813E+13	Nuning Fauziah Affiani	Wonorejo	WOELAN DRESS BLACK	1	
SAG	6.20813E+13	YUHANIN ZAMRODAH	Blitar, Selopuro	AGHISNA DRESS, ALYAA DRESS	1, 1	
SAG	6.20821E+13	NINING	Bogor, Bogor Barat - Kota	TAAT DRESS, TAAT KEMEJA L.PANJANG, SYAKI	1, 1, 1, 1	
SAG	6.20856E+12	DITA PRETHASARI	Palemang, Plaju	JAMILA TUNIK, JAMILA TUNIK, JAMILA TUNIK, J	11	
LINDS	6.20857E+13	Tambahan linds 827	Kendal, Gemuh	SYAKIRA DRESS, SYAKIRA KOKO L.PANJANG	1, 1	
LINDS	6.20857E+13	Chehsi	SOGAN BATIK REJODANI	WAFIQA DRESS (NEW)	1	
LINDS	6.20867E+13	supari	Kendal, Gemuh	WOELAN DRESS BLACK	1	
WDSD	6.20879E+13	DMROTUL ADAWIYAH	SUKODONO-SDA	TAAT DRESS, TAAT KEMEJA L.PANJANG, AGHIS	4	
SAG	6.20882E+14	MBAK ILFEN	Sogan Batik Rejodani	TEING OUTER	1	
LINDS	6.20895E+13	Wiwik Larasati	Kendal, Gemuh	SHAHABI DRESS HITAM (NEW), SHAHABI DRES	3, 1, 1	
ATK	6211	IN ZACKY	Sogan Batik Rejodani	BULAN DRESS, SYAKIRA DRESS	1, 1	
NANA	+62811-913-816	Ibu Luci	Jakarta Selatan, Jagakarsa	LULANA TUNIK, LULANA TUNIK, LULANA TUNI	3, 2, 3	
NANA	6.28112E+11	Irviana Finanda	Bandung Barat, Lembang	TAAT DRESS, SYAKIRA KOKO L.PANJANG, Hijab	1, 1, 1	
NANA	+62811236438	Luthfi N	Bandung, Ciendo	SYAKIRA DRESS, Bucked Hat Wulan Sari	1, 1	
NANA	6.28112E+11	Titi Lestari	Deltamas, Deltamas Selatan	TAAT DRESS, Bucked Hat Wulan Sari	1, 1	

Lampiran 2 *Import file* data berbentuk *excel* ke dalam pemrograman untuk dilakukan pengolahan data.

```

In [238]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

from sklearn.cluster import KMeans
from sklearn.preprocessing import OneHotEncoder, LabelEncoder, StandardScaler

In [239]: df = pd.read_excel("./data_transaksi_januari.xlsx")
df.head()

Out[239]:
   kode_agen  telp          nama  daerah          produk  qty  jumlah_transaksi
0      NANA  +62+886932514059      Jumadi  Jawa Tengah  Nekabali Dress, Nekabali Kemeja L Pendek  2.0  1137500
1      LSI   +62.0812.7260.7723.      Nikmah  Jambi          INNER DRESS MANARA (APLIKASI)  1.0  370000
2      LSI   +62.0822.2560.4514.      Sunardi / Turat  Jawa Timur  LAZUARDI TUNIK  1.0  455000
3      LSI   Godong          Nurul Khasanah  Lampung  ZAMRUD MADINA KEMEJA L PANJANG, INNER HITAM LA...  3.0  1061000
4      LSI   +62.5583.5912.      JPBAI (Kartika CWB HK +852 55835912)  Yogyakarta  AGNI TUNIK  1.0  445000

In [240]: df.shape # data with row 110 and 7 columns
Out[240]: (111, 7)

In [241]: df.isnull().sum()
Out[241]:
kode_agen    1
telp         0
nama         1
daerah       1
produk       1

```