

PENERAPAN ALGORITMA *K-MEANS CLUSTERING* UNTUK KLASTERISASI PENJUALAN *SMARTPHONE* PADA CARIN CELL

Atiq Lega¹⁾, Pitrasacha Adytia²⁾, Siti Lailiyah³⁾

^{1,2,3} Program Studi Teknik Informatika , STMIK Widya Cipta Dharma

^{1,2,3}JL. M. Yamin No.25, Samarinda75123

E-Mail: atiqlegamee@gmail.com¹⁾, pitra@gmail.com²⁾, lail.59a@wicida.ac.id³⁾

ABSTRAK

Setiap perusahaan dalam bidang perdagangan memiliki keinginan untuk mengembangkan usahanya dengan maksimal. Banyak perusahaan yang melakukan berbagai cara agar bisnis mereka selalu berkembang dengan melakukan analisis terhadap penjualan produk mereka dan bertujuan untuk dapat memperoleh laba dengan memanfaatkan sumber daya yang sudah dimiliki perusahaan. Salah satu strategi yang dapat digunakan ialah memanfaatkan data penjualan dengan melakukan pengelompokkan untuk melihat produk yang terjual lebih banyak dan sedikit terjual, agar dapat dilakukan evaluasi terkait penjualan produk *smartphone*.

Tujuan dari penelitian ini adalah untuk klusterisasi data penjualan *smartphone* pada Carin Cell dengan menggunakan algoritma *K-Means Clustering*. Dengan menggunakan tahapan Pengembangan Data. Data yang digunakan untuk klusterisasi terhitung dari Tahun 2022 sampai dengan Tahun 2023.

Hasil penelitian ini berupa model algoritma *K-Means Clustering* untuk mengelompokkan data penjualan pada Carin Cell dan mengetahui karakteristik dari cluster yang terbentuk. Model menghasilkan akurasi silhouette coefficient 78% dimana akurasi model tersebut memiliki struktur yang dihasilkan kuat. Penelitian menghasilkan 3 cluster, dengan interpretasi sebagai berikut: Cluster 0 ialah cluster dengan peminatan yang tinggi di kalangan kelompok harga standar sampai dengan mahal dan memiliki berbagai kelompok terjual yang telah dibuat, dengan peminatan pada memori 128GB hingga 512GB. Cluster 1 dan 2 merupakan cluster dengan peminatan pada kelompok harga di kalangan standar ke bawah/murah dan memiliki kelompok terjual sangat laris di setiap produk, dengan peminatan pada memori cenderung 128GB ke bawah. Setiap cluster memiliki minat yang tinggi terhadap tipe warna yang sama pada setiap pembelian *smartphone*, yaitu hitam, biru, dan abu-abu.

Kata Kunci: *Clustering*, Penjualan, *Smartphone*, Metode Penambangan Data, *K-Means*, *Data Mining*

1. PENDAHULUAN

Setiap perusahaan dalam bidang perdagangan memiliki keinginan untuk mengembangkan usahanya dengan maksimal agar tidak terjatuh dalam persaingan bisnis yang sangat ketat. Banyak perusahaan yang melakukan berbagai cara agar bisnis mereka selalu berkembang dengan melakukan analisis terhadap penjualan produk mereka dan bertujuan untuk dapat memperoleh laba dengan memanfaatkan sumber daya yang sudah dimiliki perusahaan. Perkembangan teknologi seperti *smartphone* dari tahun ke tahun semakin meningkat di Indonesia. Hal ini tidak lepas menjamurnya berbagai varian *smartphone*. Salah satu strategi yang dapat digunakan ialah memanfaatkan data penjualan dengan melakukan pengelompokkan untuk melihat produk yang terjual lebih banyak dan sedikit terjual, agar dapat dilakukan evaluasi terkait penjualan produk *smartphone*.

Tabel 1.1 Device Shipments 2022-2023

Device Type	2022 Shipments	2022 Growth (%)	2023 Shipments	2023 Growth (%)
PC	287.159	-16.0%	267.676	-6.8%
Tablet	136.938	-12.0%	132.963	-2.9%

Mobile Phone	1.395.247	-11.0%	1.339.505	-4.0%
Total Device	1.819.344	-11.9%	1.740.143	-4.4%

Sumber: gartner.com (Januari 2023)

Dapat dilihat pada tabel di atas menunjukkan bahwa mobile phone masih menjadi pencapaian dengan pembelian device tertinggi dibanding PC dan Tablet, hal ini menunjukkan bahwa *smartphone* memiliki pangsa pasar yang lebih besar.

2. RUANG LINGKUP

2.1 Rumusan Masalah

Berdasarkan latar belakang, maka dapat diketahui rumusan masalah dalam penelitian ini ialah sebagai berikut:

1. Bagaimanakah penerapan *K-Means* terhadap data penjualan Carin Cell ?
2. Bagaimanakah hasil *clustering* dari data penjualan *smartphone* ?.

2.2 Batasan Masalah

Adapun batasan masalah dalam penelitian ini sebagai berikut:

1. Penelitian ini dilakukan dengan menggunakan dataset histori penjualan smartphone pada Carin Cell Tahun 2022 s/d 2023.
2. *Clustering* berdasarkan variabel atau atribut yang meliputi nama barang smarphone dan penjualan barang smartphone.
3. Algoritma *K-Means* di gunakan untuk pengelompokkan data penjualan smartphone.
4. Proses *clustering* menggunakan bahasa pemrograman python.
5. Pendekatan machine learning yang digunakan yaitu Unsupervised Learning

3. BAHAN DAN METODE

3.1 Penjualan

Menurut Soraya dan Wahyudi (2021), menyatakan bahwa penjualan adalah kegiatan transaksi yang sesungguhnya untuk mencapai volume penjualan yang dikehendaki, yang bertujuan untuk mencari pembeli, mempengaruhi serta memberikan petunjuk agar pembeli dapat menyesuaikan kebutuhannya dengan produksi yang ditawarkan serta mengadakan perjanjian mengenai harga yang menguntungkan untuk kedua belah pihak. Selain mendapatkan volume penjualan yang diharapkan bertambah setiap bulannya, perusahaan juga memiliki tujuan untuk mendapatkan laba dan menunjang pertumbuhan perusahaan.

3.2 Metode Penambangan Data

Menurut E. Arriyanti dan P. Adytia (2020), menyatakan tahapan dari metode Penambangan Data memiliki 5 tahapan, yaitu:

1. Mendefinisikan domain data
Mendefinisikan domain data adalah menyatakan bahwa data dari mana dan terdiri data apa saja. Menjelaskan data yang didapat secara terperinci.
2. Mengumpulkan data
Pada tahapan ini mengumpulkan data adalah menjelaskan bagaimana tahapan data pada poin 1 dikumpulkan atau langkah mendapatkan data tersebut.
3. Pra-pemrosesan data
Deteksi pencilaan dan fitur penskalaan-penyandian-pemilihan adalah setidaknya dua tugas umum prapemrosesan. Mendeteksi dan pada akhirnya menghilangkan outlier, dan mengembangkan metode pemodelan yang kuat adalah dua strategi untuk menangani outlier. Direkomendasikan untuk menskalakan dan membawa data ke fitur yang sama untuk analisis lebih lanjut.
4. Memperbaiki model
Implementasi dari proses ini adalah berdasarkan jenis data mining. Memilih model terbaik adalah tugas tambahan.
5. Menginterpretasikan hasil
Metode penggalian data modern diharapkan dapat memberikan hasil yang sangat akurat dengan menggunakan model dimensi tinggi. Masalah penafsiran model-model ini dianggap sebagai tugas yang terpisah dengan teknik-teknik khusus untuk memvalidasi hasilnya. Seorang

pengguna tidak menginginkan ratusan halaman hasil numerik.

3.3 Clustering

Clustering merupakan sebuah proses dalam pembentukan kelompok-kelompok data yang berasal dari himpunan data yang belum diketahui kelas-kelasnya dan proses menentukan sebuah data termasuk dalam kelas-kelas tersebut (Putra, 2021). *Clustering* adalah salah satu teknik dari algoritma machine learning yaitu unsupervised learning. Menurut Muningsih dan Kiswati (2015) *Clustering* mengacu pada pengelompokkan data, observasi atau kasus berdasar kemiripan objek yang diteliti. Sebuah cluster adalah suatu kumpulan data yang mirip dengan lainnya atau ketidakmiripan data pada kelompok lain. Menurut Pertiwi,dkk (2022) *clustering* adalah metode pengelompokkan sebuah data (objek) yang mempunyai hubungan dan kemiripan antar data kedalam kelas-kelas yang sama. Kemiripan atau kesamaan objek diperoleh dengan menggambarkan kedekatan nilai atribut objek data yang disajikan sebagai titik dalam ruang multidimensi. Tujuan *clustering* adalah menggabungkan titik-titik data ke dalam kelompok yang homogen, artinya setiap anggota memiliki tingkat kemiripan yang tinggi dengan setiap anggota lainnya yang kemudian dikelompokkan ke dalam satu cluster.

3.4 K-Means

Menurut Wulandari, dkk (2022), metode *K-Means*, yang merupakan algoritma non-hierarkis pada umumnya. Setiap titik data ditetapkan ke pusat kluster terdekat, dan kemudian metode *K-means* memindahkan setiap pusat kluster ke titik data rata-rata dan akhir. *K-Means clustering* merupakan salah satu algoritma machine learning khususnya unsupervised learning yang populer digunakan, dalam penggunaan algoritma *K-Means* perlu diinisialisasikan nilai k (jumlah cluster) terlebih dahulu yang nantinya nilai k akan merujuk kepada jumlah centroid (titik pusat dari setiap *cluster*) yang kita butuhkan didataset. Analisis *Clustering* adalah teknik analisis multivariat yang dilakukan untuk mencari dan mengorganisasikan informasi tentang variabel, sehingga kelompok-kelompok secara kasar dapat didefinisikan sebagai metode kedekatan yang harus homogen secara internal, anggota mirip dengan anggota lain, dan berbeda secara eksternal ketika anggota ini tidak mirip mewujudkannya. Menurut Pertiwi, dkk (2022) menyatakan bahwa langkah-langkah dalam algoritma *K-Means* sebagai berikut.

1. Menentukan nilai k (jumlah *cluster*) pada *dataset*.
2. Tentukan nilai pusat (centroid). Penentuan nilai awal centroid dilakukan dengan cara acak dan nilai centroid ditentukan dari tahap iterasi dengan menggunakan rumus seperti dibawah ini:

$$V_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} X_{kj}$$

Keterangan:

1. V_{ij} = Centroid rata-rata *cluster* ke-i untuk variabel ke-j

2. N_i = Jumlah data yang menjadi anggota *cluster* ke- i
 3. i, k = Indeks dari *cluster*
 4. j = Indeks dari variabel
 5. X_{kj} = Nilai data ke- k variabel ke- j untuk *cluster* tersebut.
3. Melakukan perhitungan jarak disetiap titik centroid dengan titik objek menggunakan rumus jarak Euclidean (Euclidean Distance). Euclidean Distance adalah perhitungan jarak garis lurus dari dua titik dalam ruang Euclidean, dengan rumus seperti dibawah ini:

$$De = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2}$$
 Keterangan:
 1. De = Euclidean Distance
 2. i = Banyaknya objek
 3. (x, y) = Koordinat objek
 4. (s, t) = Koordinat centroid
 4. Mengelompokkan objek dengan memperhitungkan jarak minimum objek untuk menentukan setiap anggota *cluster*.
 5. Lakukan perulangan dari Langkah ke-2 sampai ke-4 hingga hasil nilai centroid sama dan anggota setiap *cluster* tidak berpindah ke *cluster* lain.

Menurut Hidayati, dkk (2021), menyatakan bahwa untuk menentukan kualitas hasil pengelompokan masing-masing perhitungan jarak, maka perlu dilakukan uji homogenitas. Pengujian dilakukan setelah mencapai konvergensi 0 di mana hasil pengelompokan terakhir sama dengan pengelompokan sebelumnya. Dengan kata lain, tidak ada data yang berpindah klaster. Pengujian dihitung menggunakan persamaan Silhouette coefficient. Langkah dalam menghitung Silhouette coefficient di mulai dengan mencari jarak rata-rata data ke- i dengan semua data di klaster yang sama, di sini kita asumsikan data ke- i berada di klaster A. Rumus dari $a(i)$ ditulis dalam persamaan sebagai berikut:

$$a(i) := \frac{1}{|A|-1} \sum_{j \in A, j \neq i} d(i, j)$$

Gambar 3.1 Persamaan $a(i)$

Dimana:

A = Banyaknya data di klaster A

Selanjutnya menghitung nilai $b(i)$, yang merupakan nilai minimum dari jarak rata-rata data ke- i dengan semua data di klaster berbeda. perhitungan jarak rata-rata data ke- i dengan semua data di klaster C ditulis sebagai berikut:

$$d(i, C) := \frac{1}{|C|} \sum_{j \in C} d(i, j)$$

Gambar 3.2 Persamaan $d(i, C)$

Dimana:

C = Banyaknya data di klaster C

Setelah menghitung $d(i, C)$ untuk semua klaster $C \neq A$, selanjutnya memilih nilai jarak paling minimum sebagai nilai $b(i)$.

$$b(i) := \min_{C \neq A} d(i, j)$$

Gambar 3.3 Persamaan $b(i)$

Jika klaster B memiliki nilai jarak minimum, maka $d(i, B) = b(i)$ yang disebut sebagai tetangga dari data ke-

i dan merupakan klaster terbaik kedua untuk data ke- i setelah klaster A. Setelah $a(i)$ dan $b(i)$ diketahui, maka proses terakhir menghitung Silhouette coefficient.

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i) - b(i)\}}$$

Gambar 3.4 Persamaan Silhouette Coefficient

Nilai $s(i)$ berada antara -1 dan 1, di mana setiap nilai diinterpretasi sebagai berikut:

1. $s(i) = 1$ = data ke- i digolongkan dengan baik (dalam A)
2. $s(i) = 0$ = data ke- i berada di tengah antara dua klaster (A dan B)
3. $s(i) = -1$ = data ke- i digolongkan dengan lemah (dekat ke klaster B daripada A)

Penafsiran silhouette coefficient ditunjukkan seperti pada tabel dibawah:

Tabel 3.1 Penafsiran Silhouette Coefficient

Silhouette Coefficient	Interpretasi
0.71 - 1.00	Struktur yang dihasilkan kuat
0.51 - 0.70	Struktur yang dihasilkan baik
0.26 - 0.50	Struktur yang dihasilkan lemah
≤ 0.25	Tidak terstruktur

3.5 Metode Penambahan Data

Adapun metode pengumpulan data yang dilakukan pada penelitian ini, yaitu:

3.5.1 Mendefinisikan domain data

Mendefinisikan domain data adalah menyatakan bahwa data dari mana dan terdiri data apa saja. Menjelaskan data yang didapat secara terperinci.

3.5.2 Mengumpulkan Data

Pada tahapan ini mengumpulkan data adalah menjelaskan bagaimana tahapan data pada poin 3.2.1 dikumpulkan atau langkah mendapatkan data tersebut.

3.5.3 Pra-pemrosesan Data

Deteksi pencilan dan fitur penskalaan-penyandian-pemilihan adalah setidaknya dua tugas umum prapemrosesan. Mendeteksi dan pada akhirnya menghilangkan outlier, dan mengembangkan metode pemodelan yang kuat adalah dua strategi untuk menangani outlier. Direkomendasikan untuk menskalakan dan membawa data ke fitur yang sama untuk analisis lebih lanjut.

3.5.4 Memperbaiki Model

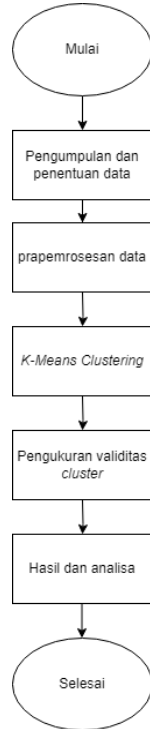
Implementasi dari proses ini adalah berdasarkan jenis data mining. Memilih model terbaik adalah tugas tambahan.

3.5.5 Menginterpretasikan Hasil

Metode penggalian data modern diharapkan dapat memberikan hasil yang sangat akurat dengan menggunakan model dimensi tinggi. Masalah penafsiran model-model ini dianggap sebagai tugas yang terpisah dengan teknik-teknik khusus untuk memvalidasi hasilnya. Seorang pengguna tidak menginginkan ratusan halaman hasil numerik.

3.6 Tahapan Penelitian

Penelitian *clustering* data penjualan smartphone pada Carin Cell menggunakan algoritma *K-Means* dilakukan melalui beberapa tahapan yang terstruktur sehingga proses penelitian dapat dilakukan dengan lebih sistematis. Adapun tahapan penelitian digambarkan sebagai berikut:



Gambar 3.5 Tahapan Penelitian

Berdasarkan Gambar 3.5, adapun penjabaran langkah-langkah tersebut sebagai berikut:

1. Pengumpulan dan penentuan data, pada tahap ini melakukan pengumpulan data pada tempat penelitian dan menentukan data yang cocok digunakan pada penelitian ini
2. Prampemrosesan data, pada tahap ini dilakukan proses pada data seperti pembersihan, penggabungan serta perubahan pada data agar data dapat dimasukkan ke dalam sebuah model *cluster*.
3. *K-Means Clustering*, pada tahap ini tahap implementasi algoritma *K-Means* terhadap data penjualan yang telah dilakukan prapemrosesan.
4. Pengukuran validitas *cluster*, pada tahap ini dilakukan pengukuran untuk melihat seberapa akurat model *cluster* yang didapat terhadap data penelitian.
5. Hasil dan analisa, pada tahap ini dilakukan penjabaran dan analisa terhadap *cluster* yang didapat menggunakan algoritma *K-Means*.

4. PEMBAHASAN

Hasil penelitian diuraikan berdasar model proses data science yaitu, Metode Penambangan Data dengan menggunakan algoritma *K-Means* dalam klasterisasi data penjualan smartphone pada Carin Cell tahun 2022 – 2023.

4.1 Mendefinisikan Domain Data

Pada tahap ini mendefinisikan domain data untuk menyatakan bahwa data yang didapat pada tempat penelitian. Data yang digunakan ialah data penjualan smartphone pada Carin Cell tahun 2022 sampai dengan 2023. Adapun gambaran data penjualan tahun 2022 yang didapat pada tempat penelitian seperti tabel berikut.

Tabel 4.1 Data Penjualan 2022

Tanggal	NamaPelanggan	Referensi	NamaBarang	IMEI	Merk	NamaGudang	Jumlah	Harga
01/01/2022	CASH		OPPO A15s 4/64 GB HITAM DINAMIS	867756053331095	OPPO	CARIN CELL WACHID HASYIM	1	2.200.000.00
01/01/2022	CASH		APPLE MACBOOK PRO 13" CHIP M1 CPU SCORE, GPU SCORE, SSD 512 GB SPACE GRAY (MYD921 D/A)	3609C03G84X7Q05F	APPLE	CARIN CELL 1	1	23.500.000.00
31/12/2022	CASH		REALME PAD MINI RAM 4/64 GB ABU ABU	220728862106000466	REALME	MINNE GADGET (ILYAS DAY ICHANK)	1	2.400.000.00
...

Pada Tabel 4.1 di atas memiliki ukuran 26.668 Baris dan 9 Kolom. Berdasarkan tabel 4.1, penjabaran setiap kolom seperti berikut:

1. Kolom “Tanggal” merupakan tanggal penjualan smartphone
2. Kolom “NamaPelanggan” merupakan metode pembelian pembeli
3. Kolom “Referensi” merupakan kolom tidak diketahui, karena tidak memiliki *value*
4. Kolom “NamaBarang” merupakan nama smartphone yang terjual
5. Kolom “IMEI” merupakan nomor seri dari smartphone yang terjual
6. Kolom “Merk” merupakan merk dari smartphone tersebut
7. Kolom “NamaGudang” merupakan tempat terjualnya smartphone
8. Kolom “Jumlah” merupakan kuantitas smartphone yang terjual
9. Kolom “Harga” merupakan harga dari sebuah smartphone

Selanjutnya adapun gambaran data penjualan tahun 2023 yang didapat pada tempat penelitian ialah seperti pada tabel berikut.

Tabel 4.2 Data Penjualan 2023

Tanggal	NamaPelanggan	NamaBarang	Merk	NamaGudang	Jumlah	Harga
01/01/2023	CASH	REALME NARZO 50iPRIME 3/32GB BIRU TUA	REALME	GUDANG UTA MA AHM AD DAH LAN	1	1.200.000.00
01/01/2023	CASH	VIVO Y22 6/128 GB STARLIT BLUE	VIVO	CARIN CELL1	1	3.000.000.00
...

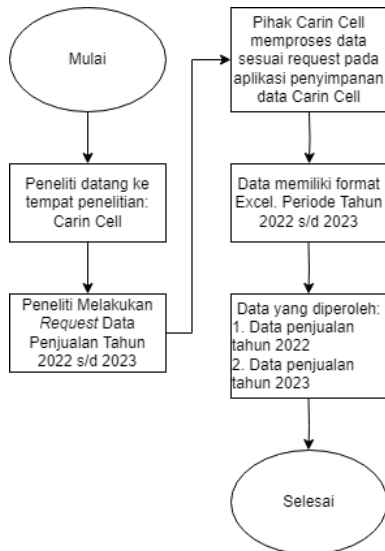
24/07/2023	CASH	VIVO Y02 3/32 GB ORC HID BLU E	VIVO	CARIN CELL LOAJANA N	1	1.400.00 0.00
------------	------	---	------	-------------------------------	---	------------------

Pada Tabel 4.2 di atas memiliki ukuran 13.274 Baris dan 7 Kolom. Berdasarkan tabel 4.2, penjabaran setiap kolom seperti berikut:

1. Kolom “Tanggal” merupakan tanggal penjualan *smartphone*
2. Kolom “NamaPelanggan” merupakan metode pembelian pembeli
3. Kolom “NamaBarang” merupakan nama *smartphone* yang terjual
4. Kolom “Merk” merupakan merk dari *smartphone* tersebut
5. Kolom “NamaGudang” merupakan tempat terjualnya *smartphone*
6. Kolom “Jumlah” merupakan kuantitas *smartphone* yang terjual
7. Kolom “Harga” merupakan harga dari sebuah *smartphone*

4.2 Mengumpulkan Data

Data penjualan yang digunakan dengan periode tahun 2022 sampai dengan tahun 2023 langsung didapatkan dari Carin Cell yang berada di Jalan Kyai Haji Ahmad Dahlan No.6, Sungai Pinang Luar, Kec. Samarinda Kota, Kota Samarinda, Kalimantan Timur 75117. Adapun tahapan pengambilan data sebagai berikut.



Gambar 4.1 Tahapan Pengambilan Data

Dari Gambar 4.1 menghasilkan 2 buah data yaitu data penjualan tahun 2022 dan tahun 2023, dimana data tahun 2022 ada pada Tabel 4.1 dan data tahun 2023 ada pada Tabel 4.2. Setelah mendapatkan data awal seperti pada Tabel 4.1 dan 4.2, maka dilakukan penggabungan data tahun 2022 dan 20223 serta pemilihan Data dan menambahkan kolom Id dengan tujuan untuk mempermudah mendapatkan informasi dan

meminimalkannya kehilangan informasi, dan dapat digunakan dengan baik terhadap proses data mining.

Penggabungan data tahun 2022 dan 20223 serta pemilihan Data dan menambahkan kolom Id. Adapun gambaran data yang telah dilakukan pada gambar diatas ialah seperti berikut.

Tabel 4.3 Data Penelitian

Id	Tanggal	NamaBarang	Merk	NamaGudang	Jumlah	Harga
1	2022-01-01	OPPO A15s 4/64 GB HITAM DINAMIS	OPPO	CARIN CELL WAC HID HASYIM	1	2200000
2	2022-01-01	APPLE MACBOOK PRO 13" CHIP M1, CPU 8CORE, GPU 8...	APPLE	CARIN CELL1	1	23500000
...
13724	2023-07-24	VIVO Y02 3/32 GB ORCHID BLUE	VIVO	CARIN CELL LOAJANAN	1	1400000

Tabel 4.3 memiliki ukuran 39.942 Baris dan 7 Kolom, dimana dari data tersebut akan dilakukan pra-pemrosesan, Maka langkah selanjutnya yang dilakukan setelah pemilihan data ialah pra-pemrosesan data. Berdasarkan tabel 4.3, maka penjabaran setiap kolom seperti berikut:

1. Kolom “Id” meruakan referensi dari index terhadap sebuah data
2. Kolom “Tanggal” merupakan tanggal penjualan *smartphone*
3. Kolom “NamaBarang” merupakan nama *smartphone* yang terjual
4. Kolom “Merk” merupakan merk dari *smartphone* tersebut
5. Kolom “NamaGudang” merupakan tempat terjualnya *smartphone*
6. Kolom “Jumlah” merupakan kuantitas *smartphone* yang terjual
7. Kolom “Harga” merupakan harga dari sebuah *smartphone*

4.3 Pra-pemrosesan Data

Pada tahap ini dilakukan pembersihan data bertujuan untuk menyiapkan data dengan cara menghapus dan memodifikasi data yang kurang relevan, duplikat, maupun yang tidak terformat agar data dapat dimasukkan ke dalam sebuah model *clustering*.

Tahap pertama pada preprocessing data ialah penghapusan nilai selain *smartphone* pada kolom

NamaBarang. Berikut adalah Tabel 4.4 data setelah dilakukan *pre-processing*.

Tabel 4.4 Data Setelah Pre-processing

Id	Tanggal	NamaBarang	Merk	NamaGudang	Jumlah	Harga
1	2022-01-01	OPPO A15s 4/64 GB HITAM DINAMIS	OPPO	CARIN CELL WACHID HASYIM	1	2200000
3	2022-01-01	OPPO A53 4/128 GB HITAM	OPPO	CARIN CELL 1	1	2700000
...
13724	2023-07-24	VIVO Y02 3/32 GB ORCHID BLUE	VIVO	CARIN CELL LOAJANAN	1	1400000

Pada Tabel 4.3 menunjukkan bahwa setelah dilakukan *pre-processing* adapun perubahan pada ukuran yaitu menjadi 37.626 baris.

```
1 # Melihat nama gudang
2 df['NamaGudang'].value_counts()
```

NamaGudang	
CARIN CELL 1	7563
CC5	5421
GUDANG UTAMA AHMAD DAHLAN	5341
M YAMIN	5185
CARIN CELL LOAJANAN	4170
CARIN CELL 3	3218
CARIN CELL WACHID HASYIM	3029
CARIN CELL LAMBUNTA	2987
CARIN CELL JAKARTA	587
GUDANG BESI D.JEMPANG	93
LDU , DEMOUNT , ADMIN ONLINE	19
GUDANG UTAMA	8
MINE GADGET (ILYAS DAN ICHANK)	3
RETURAN	2
Name: count, dtype: int64	

Gambar 4.2 Kolom Gudang

Pada Gambar 4.2 terlihat bahwa data tersebut masih mengandung penjualan smartphone pada semua gudang Carin Cell, maka dari itu dilakukan pembersihan data, yaitu dengan memilih gudang yang hanya berada ada Carin Cell Pusat di jalan Ahmad Dahlan. Berikut gambaran kode pemilahan nama gudang penjualan smartphone.

```
1 # Pengambilan data hanya di cabang utama ahmad dahlan
2 df = df.loc[df['NamaGudang'] == 'GUDANG UTAMA AHMAD DAHLAN']
3 df['NamaGudang'].value_counts()

NamaGudang
GUDANG UTAMA AHMAD DAHLAN    5341
Name: count, dtype: int64
```

Gambar 4.3 Hasil Kolom Gudang

Gambar 4.3 merupakan hasil dari pemilahan gudang. Maka dapat dilihat pada gambar di atas, bahwa data tersebut menunjukkan data penjualan pada Gudang Utama Ahmad Dahlan memiliki data 5.341 baris.data penjualan smartphone.

Adapun setelah dilakukan pemilihan data kolom "NamaGudang" yang hanya berada ada Carin Cell Pusat di jalan Ahmad Dahlan, berikut adalah tampilan tabel dari data tersebut.

Tabel 4.5 Data Setelah Pemilihan Gudang

Id	Tanggal	NamaBarang	Merk	NamaGudang	Jumlah	Harga
4	2022-01-01	NOKIA 105 PLUS DUAL SIM BLACK	NOKIA	GUDANG UTAMA AHMAD DAHLAN	1	2400000
6	2022-01-01	OPPO A53 4/128 GB HITAM	XIAOMI	GUDANG UTAMA AHMAD DAHLAN	1	6000000
...
13252	2023-07-24	VIVO Y02 3/32 GB ORCHID BLUE	REALME	GUDANG UTAMA AHMAD DAHLAN	1	2100000

Setelah melakukan eksplorasi, pada data yang didapat ada beberapa kesalahan pada value, maka dari itu dilakukan perubahan value seperti pada tabel berikut.

Table 4.6 Perubahan Value

Sebelum Perubahan "NamaBarang"	Setelah Perubahan "NamaBarang"
I PHONE	IPHONE
STRLGHT	STARLIGHT
STAR LIGHT	STARLIGHT

Langkah selanjutnya melakukan *feature engineering*, dimana melakukan penambahan kolom "Warna" berdasarkan kolom "NamaBarang" serta perubahan value pada kolom tersebut dengan tujuan untuk mendapatkan insight pada data penelitian. Adapun gambaran data yang didapat berdasarkan dilakukan *feature engineering* ialah sebagai berikut.

Tanggal	NamaBarang	Merk	NamaGudang	Jumlah	Harga	Warna
2022-01-01	NOKIA 105 PLUS DUAL SIM BLACK	NOKIA	GUDANG UTAMA AHMAD DAHLAN	1	2400000	BLACK
2022-01-01	XIAOMI MI 11T 5G 8/256 GB CELESTIAL BLUE	XIAOMI	GUDANG UTAMA AHMAD DAHLAN	1	6000000	BLUE
2022-01-01	XIAOMI POCO PHONE X3 PRO 8/256 GB FROST BLUE	XIAOMI	GUDANG UTAMA AHMAD DAHLAN	1	4100000	BLUE
2022-01-01	REALME C11 2021/4/64 GB BIRU DANAU	REALME	GUDANG UTAMA AHMAD DAHLAN	1	1700000	BIRU
2022-01-01	IPHONE 13 PRO MAX 128 GB SIERRA BLUE (MLL93PA)	IPHONE	GUDANG UTAMA AHMAD DAHLAN	1	21500000	BLUE

Gambar 4.4 Hasil Penambahan Kolom Warna

Setelah mendapatkan warna berdasarkan kolom "NamaBarang" maka dilakukan perubahan untuk menyesuaikan nama-nama warna. Untuk mempermudah dalam pembacaan warna maka perubahan dijabarkan pada tabel berikut.

Tabel 4.7 Penjabaran Kolom Warna

Warna Sebelum	Warna Sesudah	Warna Sebelum	Warna Sesudah	Warna Sebelum	Warna Sesudah
Silver	Perak	Snow	Biru	Midnight Grey	Abu
White	Putih	Cyan	Biru	Moonlight White	Putih
Yellow	Kuning	Bronze	Perunggu	Coral Orange	Jingga
Gold	Emas	Midnight	Biru	Ice Blue	Biru
Orange	Jingga	Sunshine	Jingga	Abu abu	Abu
Blue	Biru	Aqua	Biru	Midnight Green	Hijau
Brown	Coklat	Onyx	Abu	Graphite Black	Hitam

Purple	Ungu	Violet	Ungu	Malam	Hitam
Pink	Merah Muda	Hyperspace	Ungu	Ocean Green	Hijau
Gray/Gray	Abu	Graphite Gray/Gray	Abu	Black Black	Hitam
Black	Hitam	Ocean Blue	Biru	Coral	Jingga
Charcoal	Abu	Onyx Gray/Gray	Abu	Graphite	Abu
Laut	Biru	Moonlight Silver	Perak	Moonlight	Malam
Ocean	Biru	Moonlight Light	Putih	Blank	Tidak Diketahui
Ice	Biru	Black Red	Hitam		
Marigold	Emas	White Red	Putih		
Starlight	Putih	Midnight Black	Hitam		

Maka didapatkan gambaran dari perubahan value pada kolom "Warna" seperti pada gambar berikut.

```
1 df['Warna'].value_counts()

HITAM          1500
BIRU           1310
ABU            1009
HIJAU          515
PUTIH          354
PERAK          155
UNGU           147
EMAS           140
JINGGA         62
MERAH MUDA    41
PERUNGGU       36
MERAH         32
KUNING         29
TIDAK DIKETAHUI 11
Name: Warna, dtype: int64
```

Gambar 4.5 Hasil Kolom Warna

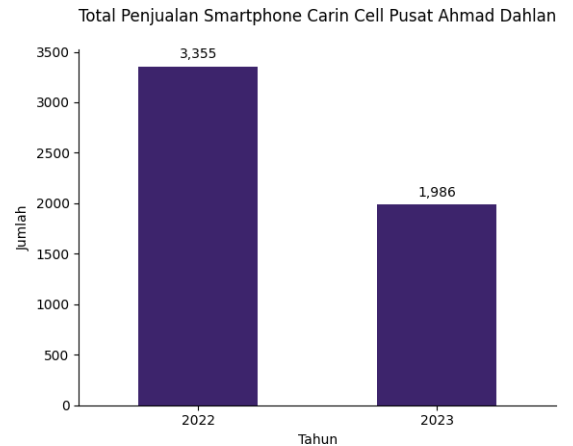
Selanjutnya melakukan pembersihan pada kolom "NamaBarang" dengan menghapus ukuran, warna dan tahun pada nama smartphone, dengan tujuan agar dapat lebih mudah melihat informasi terkait dengan kolom tersebut. Setelah melakukan pembersihan terhadap kolom "NamaBarang" dan data yang telah dibersihkan dinamai dengan nama kolom "Smartphone", adapun contoh gambaran output seperti pada gambar berikut.

```
3          NOKIA 105 PLUS DUAL SIM
5          XIAOMI MI 11T 5G
6          XIAOMI POCOPHONE X3 PRO
7          REALME C11
10         IPHONE 13 PRO MAX ( MLL93PA/A) TAM/IBOX
...
13213         ITEL S23 (S665L)
13214         NOKIA 150 new
13215         INFINIX HOT 30 NFC ( X6831 )
13250         XIAOMI REDMI NOTE 12 PRO 5G
13251         REALME C53
Name: Smartphone, Length: 5341, dtype: object
```

Gambar 4.6 Hasil Kolom NamaBarang

Setelah melakukan pembersihan data, maka langkah selanjutnya ialah melakukan eksplorasi data dengan memvisualisasikan agar mendapatkan insight

dari data tersebut. Visualisasi awal yang dilakukan ialah melihat visualisasi penjualan berdasarkan tahun.



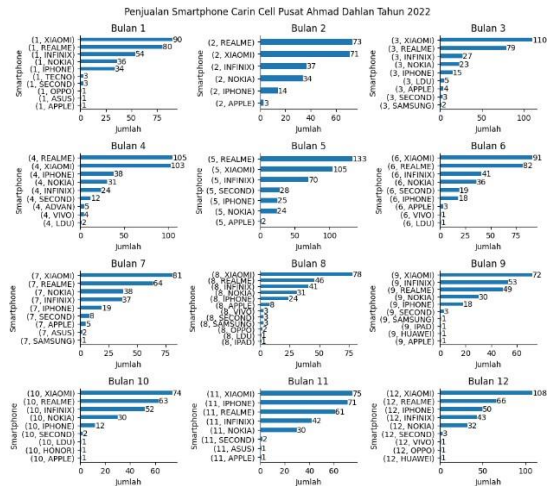
Gambar 4.7 Visualisasi Total Penjualan Smartphone Berdasarkan Tahun

Pada Gambar 4.6 bahwa penjualan smartphone pada tahun 2022 menyentuh hingga 3.355 penjualan smartphone dan pada tahun 2023 menyentuh penjualan smartphone hingga 1.986 penjualan. Selanjutnya melihat bagaimana penjualan per produk dalam tahun 2022 dan tahun 2023.



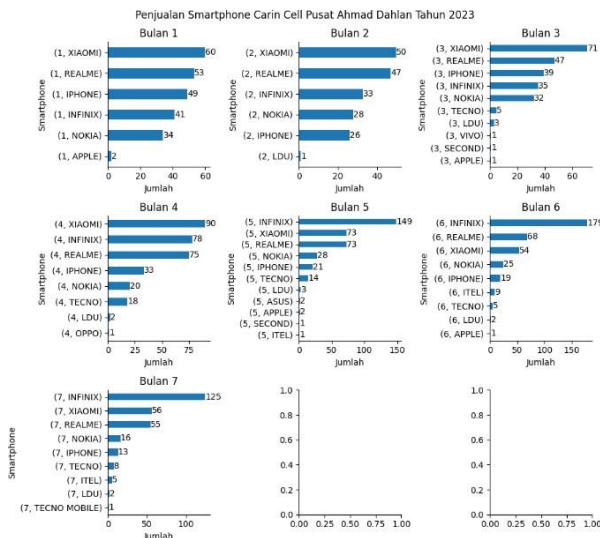
Gambar 4.8 Visualisasi Penjualan Smartphone Berdasarkan Merk

Gambar diatas menunjukkan bahwa penjualan smartphone pada Carin Cell Pusat pada tahun 2022 dengan merk Xiaomi menduduki penjualan terbanyak dimana terjual hingga 1058 smartphone, dan diikuti oleh realme terjual sebanyak 901 smartphone, dan ketiga oleh merk infinix yaitu sebanyak 521 smartphone yang terjual. Pada tahun 2023 Infinix menduduki peringkat pertama yaitu sebanyak 640 smartphone yang terjual, lalu diikuti oleh xiaomi dengan 454 smartphone terjual, dan yang ketiga diduduki oleh merk realme dengan penjualan sebanyak 418 smartphone. Adapun visualisasi untuk melihat penjualan perbulan di tahun 2022 dan 2023 ialah sebagai berikut.



Gambar 4.9 Visualisasi Penjualan Tahun 2022 Berdasarkan Merk

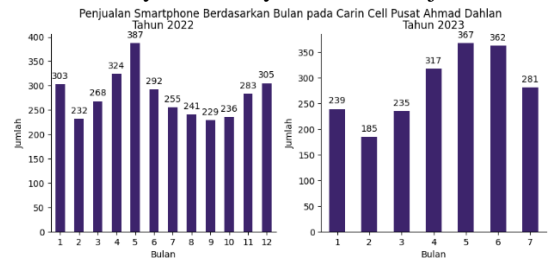
Gambar 4.9 diatas menunjukkan penjualan smartphone perbulan pada tahun 2022, dapat kita lihat bahwa penjualan smartphone dengan merk xiaomi dan realme merupakan penjualan yang selalu berada di peringkat teratas, pada bulan ke 1,3,6 hingga bulan 12 xiaomi menduduki peringkat pertama, dan merk realme menduduki peringkat pertama pada bulan ke 2, 4 dan 5. Penjualan tertinggi pada merk xiaomi ialah berada pada 113 kali terjual yaitu pada bulan ke 3, sedangkan pada merk realme berada pada 133 kali terjual pada bulan ke 5.



Gambar 4.10 Visualisasi Penjualan Tahun 2023 Berdasarkan Merk

Gambar 4.10 menunjukkan visualisasi penjualan smartphone perbulan pada tahun 2023. Pada tahun 2023 merk yang menduduki peringkat teratas ialah xiaomi dan infinix, dimana infinix memiliki peningkatan penjualan yang bisa mengalahkan penjualan smartphone merk realme di tahun 2023. Pada bulan ke 1 hingga ke 4 xiaomi menduduki peringkat teratas, dan penjualan terbanyak merk xiaomi berada pada bulan ke 4 yaitu sebanyak 90 kali terjual, sedangkan pada merk infinix menduduki peringkat pertama penjualan pada bulan ke 5 hingga ke 7, dan

penjualan terbanyak pada merk infinix ialah berada pada bulan ke 6 yaitu sebanyak 179 kali terjual.



Gambar 4.11 Visualisasi Total Penjualan Smartphone Berdasarkan Bulan dan Tahun

Pada Gambar 4.11 Diatas, bulan ke 4 dan ke 5 selalu memiliki penjualan yang tinggi, pada tahun 2022 menyentuh hingga 387 smartphone yang terjual begitu pun dengan tahun 2023 hampir menyentuh angka penjualan yang sama pada tahun sebelumnya di bulan ke 4, 5 dan 6, dengan penjualan terbanyak yaitu 367 smartphone yang terjual dibulan ke 5. Penjualan terendah berada pada bulan ke 2 di tahun 2023, begitu juga pada tahun 2022, bulan february merupakan penjualan yang sedikit di antara bulan yang lain.

Setelah eksplorasi data menggunakan visualisasi dari hasil pra-pemrosesan data dilakukan maka langkah selanjutnya melakukan transformasi data, yaitu dengan melakukan transformasi data terhadap kolom "Jumlah" dan "NamaBarang" untuk melihat seberapa banyak smartphone yang terjual, adapun gambaran data seperti pada gambar berikut.

```

1 df_intepret['Count'] = df_intepret.groupby(['Smartphone', 'Jumlah'])['Id'].transform('count')
2 df_intepret.drop_duplicates(subset=['Smartphone'], inplace=True)
3 df_intepret = df_intepret[['Id', 'Smartphone', 'Merk', 'Warna', 'Ukuran', 'Count']]
4 df_intepret.sort_values(ascending=False, by='Count', inplace=True)

1 df_intepret.rename(columns = {'Count': 'Terjual'}, inplace = True)
2 df_intepret

```

Id	Smartphone	Merk	Warna	Ukuran	Terjual	
3572	3573	IPHONE 11 TAM	IPHONE	UNGU	128	165
7115	7116	REALME C35	REALME	HIJAU	128	158
6761	6762	REALME C31	REALME	PERAK	64	150
5976	5977	XIAOMI REDMI NOTE 11	XIAOMI	BIRU	128	122
132	133	XIAOMI REDMI 10	XIAOMI	ABU	128	121
...
8736	8737	REALME 9 PRO FAJAR	REALME	HIJAU	128	1
8482	8483	INFINIX HOT 10 JADE	INFINIX	HITAM	64	1
22052	22053	SECOND IPHONE 13 PRO (EX INTER)	SECOND	EMAS	128	1
7981	7982	IPHONE 7 PLUS	IPHONE	EMAS	32	1
13213	13214	ITEL S23 8 PLUS/128 (S665L)	ITEL	HITAM	128	1

390 rows * 6 columns

Gambar 4.12 Transformation Columns Data

Adapun gambaran data dari Gambar 4.12 ialah sebagai berikut:

Tabel 4.8 Pengelompokkan Data

Id	Smartphone	Merk	Warna	Ukuran	Terjual
3573	IPHONE 11 TAM	IPHONE	UNGU	128	165
7116	REALME C35	REALME	HIJAU	128	158
...
13214	ITEL S23 8 PLUS/128 (S665L)	ITEL	HITAM	128	1

Tabel di atas menunjukkan data yang telah dilakukan pengelompokkan dan menghasilkan 390 baris.

Setelah data telah dikelompokkan maka langkah selanjutnya melakukan penambahan kolom dimana data transformasi dilakukan menggunakan fungsi `qcut` pada library `pandas` python untuk mendapatkan masing-masing kategori harga dan terjual, dimana fungsi tersebut melakukan pengelompokkan data berdasarkan input yang diberikan. Pada kolom kelompok harga dan kelompok terjual dilakukan 3 pengelompokkan, yaitu kelompok harga (mahal, standar, dan murah). Kelompok terjual (sangat laris, laris, kurang laris). Dengan fungsi `qcut` maka kategori tersebut akan terbuat secara otomatis berdasarkan distribusi data dan jumlah data yang kurang lebih sama di setiap kategori. Adapun langkah-langkah yang dilakukan oleh fungsi `qcut` ialah:

1. Mengurutkan data secara *ascending* (dimulai dari nilai terkecil ke terbesar)
2. Menentukan *quantile* berdasarkan jumlah yang diinput, pada transformasi yang dilakukan input = 3, maka data dibagi menjadi 3 *quantile*, yaitu:
 1. 0% - 33%, data meliputi 0% hingga 33% terendah, dimana batas bawah pada *quantile* ini ialah di 0% dan batas tertinggi ialah 33% terendah.
 2. 33% - 66%, data meliputi 33% hingga 66% terendah, dimana batas bawah pada *quantile* ini ialah di 33% terendah dan batas tertinggi ialah 66% terendah.
 3. 66% - 100%, data meliputi 66% hingga 100%, dimana batas bawah pada *quantile* ini ialah di 66% terendah dan batas tertinggi ialah nilai maksimum dari *quantile*.

Berikut gambaran data setelah dilakukan penambahan kolom.

```
Kelompok Harga
Mahal = Rp. 5950000 - Rp. 32999000
Standar = Rp. 2500000 - Rp. 5800000
Murah = Rp. 0 - Rp. 2400000

Kelompok Terjual
Sangat Laris 8 - 165
Laris 3 - 7
Kurang Laris 1 - 2
```

Gambar 4.13 Dataset After Transformation

Ketika sudah mendapatkan kolom terjual dan harga, maka langkah selanjutnya merubah data kategori yang berbentuk kategorikal menjadi numerik dengan tujuan agar data tersebut dapat dimasukkan ke dalam sebuah model. Berikut gambaran data setelah dilakukan konversi nilai kategori menjadi numerik.

Tabel 4.9 Konversi Kategori Ke Numerik

Kelompok Terjual	Hasil Konversi
Sangat Laris	1
Laris	2

Kurang Laris	3
--------------	---

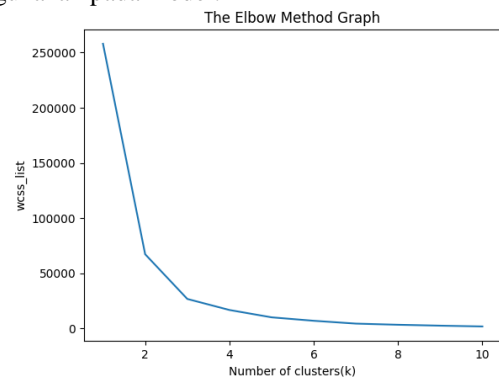
Tabel 4.10 Hasil Konversi dan transformasi

Id	Smartphone	Merak	Warna	Ukuran	Terjual	Harga	Kelompok Terjual	Kelompok Harga	LE_Kategori
3573	IPHONE 11 TAM	IPHONE	UNGU	128	165	11500000	Sangat Laris	Mahal	1
7116	REALM E C35	REALME	HIAU	128	158	2400000	Sangat Laris	Murah	1
...
13214	ITEL S238 PLU S8/128 (S665L)	ITEL	HITAM	128	1	1550000	Kurang Laris	Murah	3

Tabel di atas merupakan tabel yang telah ditambahkan kolom kategori Kelompok Terjual, Kelompok Harga dan LE_Kategori (Merupakan hasil konversi kolom Kelompok Terjual) dimana Tabel 4.9 tersebut memiliki 390 baris, yang mana data tersebut yang akan dimasukkan ke dalam sebuah model *K-Means Clustering*.

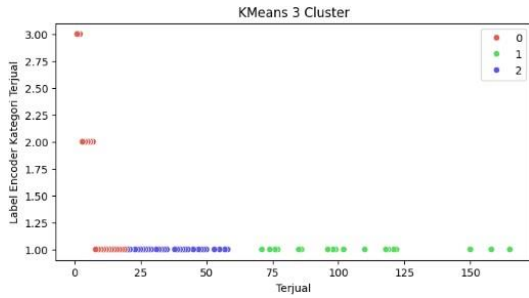
4.4 Memperbaiki Model

Pada tahap ini dilakukan pembuatan model *clustering* terhadap data yang sudah dilakukan preprocessing dan transformation data. Langkah pertama untuk membuat sebuah model *K-Means clustering* ialah menentukan variabel yang akan digunakan pada model. Variabel/kolom yang digunakan pada model ini ialah "Terjual" dan "LE_Kategori". Langkah selanjutnya mencari inertia/elbow untuk mencari klaster yang tepat digunakan pada model.



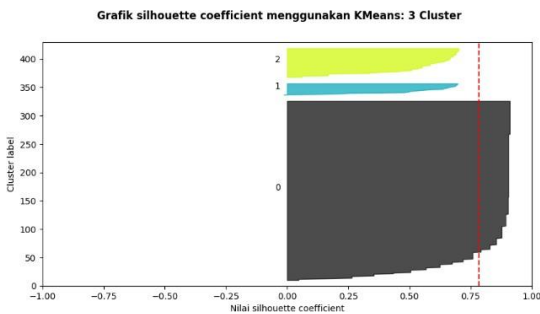
Gambar 4.14 Elbow

Pada gambar diatas metode elbow menunjukkan klaster yang tepat digunakan pada data penelitian ialah klaster = 3. Setelah mendapatkan klaster yang tepat maka langkah selanjutnya melakukan pelatihan model terhadap data yang sudah disiapkan, adapun hasil yang didapatkan ialah sebagai berikut.



Gambar 4.15 K-Means Scatter Plot

Selanjutnya melihat apakah model tersebut sudah cukup untuk dikatakan model yang kuat, maka dilakukan pengujian menggunakan silhouete coefficient. Adapun grafik silhouete coefficient seperti berikut.



Gambar 4.16 Visualisasi Silhouette Coefficient

Pada gambar diatas didapatkan bahwa hasil model K-Means clustering dengan cluster = 3 didapatkan akurasi silhouete coefficient sebesar 0.78 atau 78%. Maka dengan itu model tersebut dapat dikatakan struktur yang dihasilkan kuat, kategori intepretasi menyesuaikan dengan kategori pada Tabel 2.5. Dimana:

1. 0.71 – 1.00 = Struktur yang dihasilkan kuat
2. 0.51 – 0.70 = Struktur yang dihasilkan baik
3. 0.26 – 0.50 = Struktur yang dihasilkan lemah
4. ≤ 25 = Tidak terstruktur

Adapun gambaran data yang telah didapatkan hasil clustering menggunakan algoritma K-Means clustering ialah sebagai berikut:

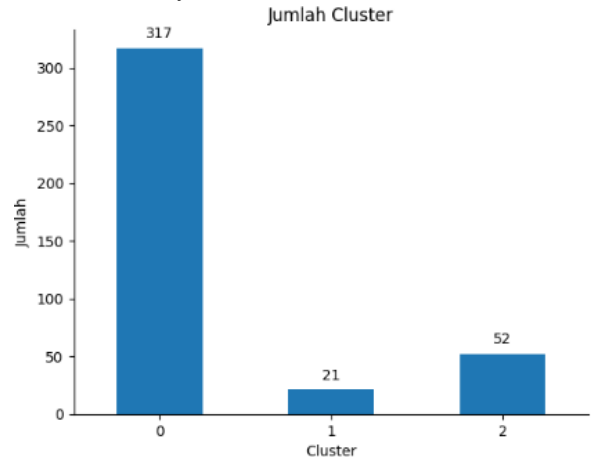
Tabel 4.11 Data setelah modeling K-Means Clustering

ID	Smartphone	Merk	Warna	Ukuran	Terjual	Harga	Kelompok_Terjual	Kelompok_Harga	LE_Kategori	Cluster
3573	IPHONE 11 TAM	IPHONE	UNGU	128	165	1150000	Sangat Laris	Mahal	1	1
7116	REALME C35	REALME	HIJAU	128	158	2400000	Sangat Laris	Murah	1	1
...
13214	ITEL S238 PLU SR/128 (S665L)	ITEL	HITAM	128	1	1550000	Kurang Laris	Murah	3	0

4.5 Intepretasi Hasil

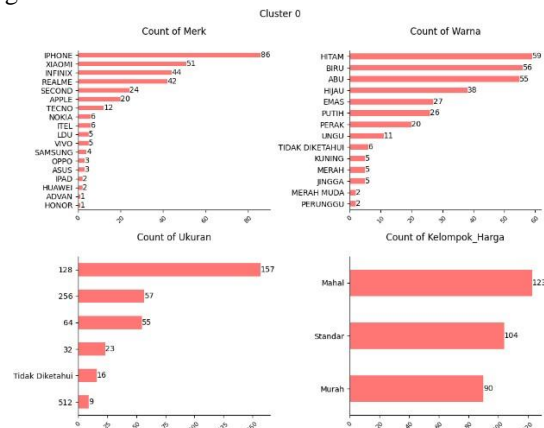
Penelitian ini menggunakan data penjualan smartphone pada Carin Cell Pusat Ahmad Dahlan. Adapun data yang didapat setelah melalui tahap data preprocessing ialah 390 smartphone dengan

memiliki atribut “Smartphone, Terjual, Kelompok Harga” dimana variabel yang digunakan untuk melakukan cluster terhadap algoritma K-Means clustering ialah variabel “Terjual” dan “Kelompok Terjual”, dan didapatkan cluster dengan struktur yang dihasilkan kuat yaitu 3 cluster.



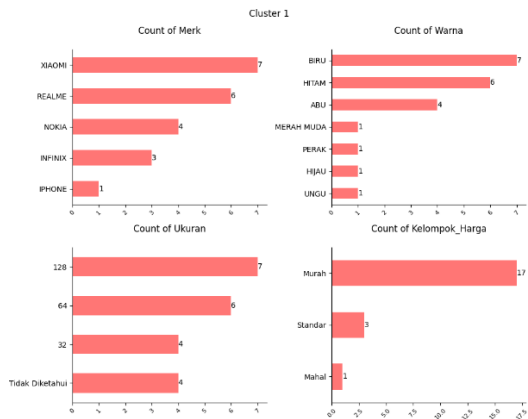
Gambar 4.17 Visualisasi Jumlah Cluster

Berdasarkan Gambar 4.17 diatas terdapat 3 cluster, dimana cluster 0 memiliki 317 produk smartphone, cluster 1 memiliki 21 produk smartphone, cluster 2 memiliki 52 produk smartphone. Cluster 0 merupakan cluster dengan frekuensi terbanyak diantara cluster yang lain.



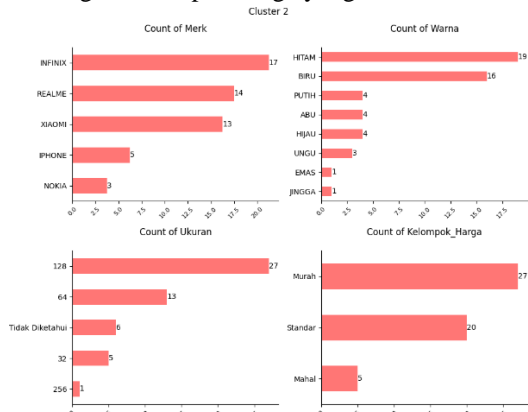
Gambar 4.18 Visualisasi Cluster 0

Berdasarkan Gambar 4.18 diatas, cluster 0 merupakan cluster dengan karakteristik penjualan dengan merk iphone, xiaomi, infinix, dan realme dengan terjual lebih dari sama dengan 30 penjualan serta memiliki warna penjualan smartphone pada hitam, biru dan abu dengan masing-masing memiliki penjualan lebih dari sama dengan 40 penjualan dengan ukuran memori terbanyak pada 128GB yaitu 157 penjualan, pada cluster 0 nampak memiliki minat terhadap smartphone dengan ukuran memori 512GB dengan sebanyak 9 penjualan. Cluster 0 merupakan cluster dengan karakteristik kelompok harga yang mahal.



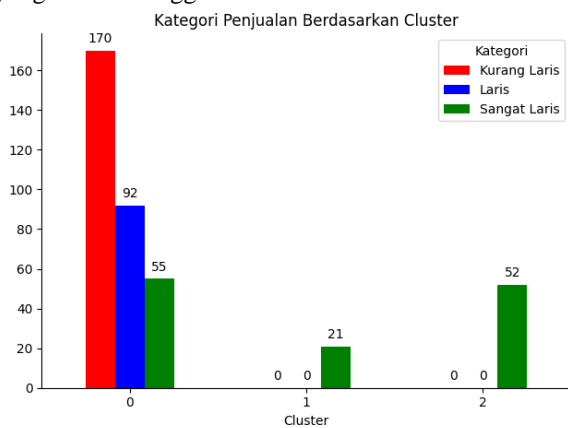
Gambar 4.19 Visualisasi Cluster 1

Berdasarkan Gambar 4.19 diatas *cluster 1* merupakan *cluster* yang memiliki karakteristik pembeli yang cenderung menyukai warna biru hitam dan abu-abu, dengan ukuran yang diminati terbanyak yaitu 128GB. Merk smartphone terjual terbanyak pada *cluster 1* ialah merk xiaomi. *Cluster 1* merupakan *cluster* dengan kelompok harga yang murah.



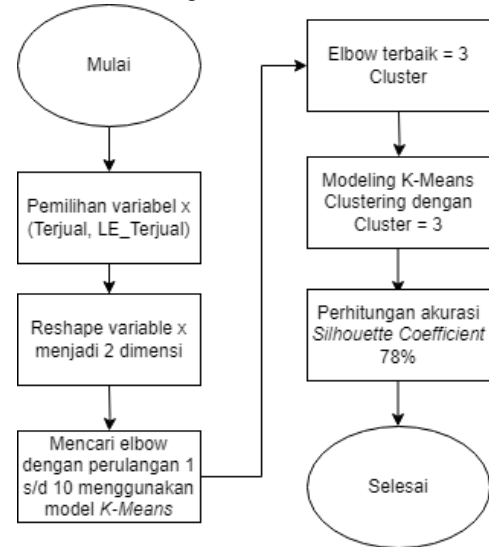
Gambar 4.20 Visualisasi Cluster 2

Pada *cluster 2* karakteristik dari *cluster*, yaitu memiliki penjualan lebih dari sama dengan sepuluh smartphone dengan ukuran terbanyak pada 128GB dengan 64GB dan cenderung memilih warna hitam dan warna biru. Merek terlaris pada *cluster* ini ialah infinix. *Cluster 2* merupakan *cluster* dengan kelompok harga yang standar hingga ke murah.



Gambar 4.21 Visualisasi Cluster Berdasarkan Kategori

Terdapat 3 karakteristik dalam kategori penjualan pada *cluster 0*, yaitu kurang laris, laris, dan sangat laris. *Cluster 1* merupakan *cluster* dengan karakteristik kategori penjualan sangat laris. *Cluster 2* juga merupakan *cluster* dengan karakteristik kategori penjualan yang sangat laris. Adapun *flowchart* dari penggunaan model *K-Means Clustering* yang digunakan ialah sebagai berikut:



Gambar 4.22 Proses Modeling K-Means Clustering

5. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan maka dapat diambil beberapa kesimpulan, Penerapan algoritma *K-means clustering* dengan variabel “Penjualan” dan “Kelompok Terjual” menggunakan $K = 3$ mendapatkan hasil silhouette coefficient 78% . Dengan interpretasi sebagai berikut:

- Cluster 0* merupakan *cluster* dengan peminatan pada kelompok harga dikalangan standar sampai dengan mahal dan memiliki berbagai macam kelompok terjual. Dengan peminatan pada memori 128GB hingga 512GB, sedangkan *cluster* yang lain tidak memiliki penjualan ukuran memori pada 512GB.
- Cluster 1* dan *2* merupakan *cluster* dengan peminatan pada kelompok harga dikalangan standar ke bawah/murah dan memiliki kelompok terjual sangat laris di setiap produk. Peminatan memori pada *cluster* ini yaitu cenderung pada 128gb kebawah
- Setiap *cluster* memiliki minat yang tinggi terhadap tipe warna yang sama pada setiap pembelian smartphone, yaitu hitam, biru dan abu-abu

6. SARAN

Berdasarkan kesimpulan yang telah dipaparkan sebelumnya, maka adapun saran terhadap penelitian selanjutnya sebagai berikut:

- Pengembangan penelitian yang akan datang perlu dilakukan dengan dataset yang relatif besar agar

tingkat keakuratan analisis mendapatkan *insight* yang luas.

2. Menambahkan atribut/variabel terkait data pembeli agar mendapatkan *insight* yang lebih luas dalam analisis *clustering*.
3. Analisa lebih lanjut terhadap faktor-faktor yang mempengaruhi penjualan *smartphone*.

7. DAFTAR PUSTAKA

- Hasibuan, F. P. (2021). Penerapan K-Means pada Pengelompokan Penjualan Produk Smartphone. *Jurnal Sains Teknologi dan Sistem Informasi*, 15-20.
- Hidayati, R., Zubair, A., Pratama, A. H., & Indand, L. (2021). Analisis Silhouette Coefficient pada 6 Perhitungan Jarak K-Means Clustering. *Techno.COM*, 186-197.
- Kurniawan, & Antoni, D. (2020). Visualisasi Data Penduduk Dalam Membangun E-government Berbasis Sistem Informasi Geografis (GIS). *Jurnal SISFOKOM (Sistem Informasi dan Komputer)*, 310-316.
- 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*.
- Muningsih, E., & Kiswati, S. (2015). Penerapan Metode K-Means Untuk Clustering Produk Online Shop Dalam Penentuan Stok Barang. *Jurnal Bianglala Informatika*, 10-17.
- Nurhayati, Busman, & Iswara, R. P. (2019). Pengembangan Algoritma Unsupervised Learning Technique Pada Big Data Analysis Di Media Sosial Sebagai Media Promosi Online Bagi Masyarakat. *JURNAL TEKNIK INFORMATIKA*, 79-96.
- Pertiwi, T. A., Halim, F. R., Afni, N., Fahrezi, I. R., & Pramana, D. (2022). Pemetaan Wilayah Da'wah Di Provinsi Riau Menggunakan Algoritma K-Means. *SENTIMAS: Seminar Nasional Penelitian dan Pengabdian Masyarakat*, 80-86.
- Putra, T. D. (2020). Analisis Keranjang Belanja dengan Algoritma Apriori Klasik pada Data Mining. *Jurnal Kajian Ilmiah (JKI)*, 59-66.
- Putra, Y. D., Sudarma, M., & Swarmadika, I. B. (2021). Clustering History Data Penjualan Menggunakan Algoritma K-means. *Ilmiah Teknologi Elektro*, 195-202.
- Soraya, A., & Wahyudi, A. D. (2021). Rancang Bangun Aplikasi Penjualan Dimsum Berbasis Web (Studi Kasus: Kedai Dimsum Soraya). *Jurnal Teknologi dan Sistem Informasi (JTSI)*, 43-48.
- Winarta, A., & Kurniawan, W. J. (2021). Optimasi Cluster K-means Menggunakan Metode Elbow Pada Data Pengguna Narkoba Dengan

Pemrograman Python. *Jurnal Teknik Informatika Kaputama (JTIK)*, 113-119.

- Wulandari, N., Farida, I. N., & Mahdiyah, U. (2022). Implementasi Metode K-Means Clustering Dalam Pengadaan Barang Di Toko N-Case. *Seminar Nasional Inovasi Teknologi*, 308-313.

