

Klasterisasi Tingkat Penjualan Produk Menggunakan Metode K-Means Untuk Penerapan Konsep Down-Selling

Studi Kasus pada ARTCH Indonesia

Fersyan Putra Samudra, Tacbir Hendro Pudjiantoro, Irma Santikarama
Jurusan Informatika, Fakultas Sains dan Infromatika
Universitas Jenderal Achmad Yani
Jl. Terusan Sudirman, Cimahi
fersyanra23@gmail.com

Abstrak— Pemanfaatan ilmu data mining telah dilakukan oleh beberapa perusahaan untuk mengolah data yang dimiliki untuk menghasilkan suatu informasi yang dapat digunakan untuk menentukan strategi yang akan dilakukan guna meningkatkan kinerja perusahaan, salah satunya di industri mode, seperti Artch Indonesia yang merupakan sebuah perusahaan yang bergerak di bidang fashion memiliki data penjualan yang besar setiap bulannya. Proses analisis terhadap data penjualan dengan jumlah yang banyak dan data yang beragam membutuhkan waktu yang cukup lama, sehingga informasi yang dihasilkan tidak dapat diperoleh dengan cepat dan pengetahuan yang mendalam tidak dapat ditemukan. Untuk itu penelitian ini bertujuan untuk melakukan teknik data mining klasterisasi dengan metode K-Means terhadap data penjualan dan menghasilkan informasi rekomendasi produk untuk penerapan konsep down-selling. Penelitian ini menghasilkan 5 klaster berdasarkan elbow method dan hasil pengujian silhouette score untuk jumlah klaster 5 adalah 0,538 dengan menghasilkan rekomendasi produk untuk diterapkan konsep down-selling sebanyak 36 macam produk.

Kata kunci — *Clustering; K-Means; Penjualan; Down-Selling.*

I. PENDAHULUAN

Pemanfaatan ilmu *data mining* telah dilakukan oleh beberapa perusahaan untuk mengolah data yang dimiliki untuk menghasilkan suatu informasi yang dapat digunakan untuk menentukan strategi yang akan dilakukan guna meningkatkan kinerja perusahaan, salah satunya di industri mode, *data mining* merupakan eksplorasi dan evaluasi informasi yang besar untuk menemukan pola dan aturan yang signifikan. Ilmu ini mengacu pada proses pencarian informasi tersembunyi dari sejumlah data yang besar melalui algoritma [1]. Ilmu ini berada dibawah bidang *data science* untuk mencari tahu fakta yang tersembunyi berdasarkan data historis [2]. Konsep data mining yang disebut juga *knowledge discovery in database* (KDD) merupakan aktivitas mengumpulkan dan menggunakan data historis untuk menemukan pola-pola atau hubungan dalam kumpulan data yang besar [3]. Salah satu teknik yang ada di dalam *data mining* yaitu klasterisasi, algoritma pengelompokkan

sejumlah data menjadi kelompok-kelompok data tertentu (*cluster*) [4]. Salah satu algoritma yang digunakan dalam klasterisasi adalah algoritma *K-Means Clustering* [5].

Dalam beberapa studi di bidang bisnis telah dilakukan *data mining* dengan teknik *clustering* untuk mengelompokkan data diantaranya Klasterisasi Pola Penjualan Pestisida Menggunakan Metode K-Means Clustering, hasil penelitian tersebut berhasil mengklusterkan produk pestisida menjadi 3 diantaranya produk yang sangat laku, laku dan tidak laku, dimana proses klasterisasi disini masih menggunakan tools yang sudah ada yaitu *RapidMiner* [6]. Kemudian pada studi dengan judul Penerapan Metode K-Means Untuk Clustering Produk Online Shop Dalam Penentuan Stok Barang, penelitian ini melakukan proses klasterisasi dengan menggunakan 3 variabel diantaranya jumlah transaksi, volume penjualan dan rata-rata penjualan yang menghasilkan 3 klaster yaitu produk yang sangat diminati, diminati dan kurang diminati [7].

Salah satu jenis usaha yang memiliki sejumlah data penjualan besar yang tersimpan adalah bidang usaha *fashion*, seperti Artch Indonesia yang merupakan sebuah perusahaan yang bergerak di bidang *fashion* yang berfokus pada penjualan tas, perusahaan ini berdiri sejak tahun 2010. Dalam proses penjualan tas terdapat sejumlah data besar yang tersimpan yang dapat dimanfaatkan untuk meningkatkan volume penjualan produk dan meningkatkan pemasukan perusahaan [8]. Proses analisis terhadap data penjualan dengan jumlah yang banyak dan data yang beragam membutuhkan waktu yang cukup lama, sehingga informasi yang dihasilkan tidak dapat diperoleh dengan cepat dan pengetahuan yang mendalam tidak dapat ditemukan.

Terikait dengan proses pengelompokkan data tersebut, digunakan suatu teknik untuk melakukan *clustering*. Proses *clustering* ini dapat dilakukan dengan menggunakan *K-Means Clustering*. Algoritma ini melakukan pengelompokkan set data menjadi sebuah klaster yang sudah ditentukan yang bertujuan untuk membentuk kelompok data yang terpisah yang memiliki kemiripan [9]. Hasil klasterisasi tingkat penjualan produk ini kemudian digunakan sebagai acuan untuk dapat diterapkan suatu konsep penjualan yang disebut

dengan *down-selling* dimana konsep ini mengelola produk-produk yang kurang diminati dan tidak terjual yang menyebabkan pendapatan perusahaan tidak optimal, salah satunya dengan cara memberi diskon pada produk yang kurang laku tersebut dan melakukan penawaran terhadap produk yang lebih murah [10]. Pemanfaatan konsep *down-selling* disini mengarahkan pelanggan untuk membeli produk yang memiliki nilai jual yang sesuai dengan kemampuan dan kebutuhan pelanggan untuk meningkatkan volume penjualan pada perusahaan. Teknik *down-selling* jarang diterapkan namun apabila dimanfaatkan dapat berpengaruh kepada loyalitas pelanggan dimana sebelumnya pernah membeli barang yang ditawarkan dari konsep *down-selling* ini.

Dari pembahasan tersebut maka dilakukan penelitian ini untuk dapat membantu perusahaan dalam memperoleh informasi mengenai produk berdasarkan tingkat penjualannya menggunakan teknik *data mining* metode *K-Means Clustering*. Hasil pengelompokan tersebut akan dijadikan acuan dalam penerapan strategi *down-selling*.

II. METODE

A. Pengumpulan data

Data penelitian diperoleh dari pengambilan data langsung ke instansi terkait yaitu ARTCH Indonesia yang diperlukan untuk mendukung penelitian yaitu data penjualan produk dari bulan Januari 2020 hingga Juni 2020 dengan jumlah 6174. Data transaksi penjualan ini berisi data penjualan jenis dan jumlah produk selama periode waktu tersebut.

B. Data Pre-processing

Pre-processing data merupakan sekumpulan data operasional yang perlu dilakukan seleksi sebelum tahap penggalian informasi dalam *Knowledge Discovery Database* (KDD) dimulai. Proses *cleaning* data dilakukan untuk menghilangkan data yang tidak konsisten, atau menghapus atribut yang tidak diperlukan [11]. Kegiatan yang dilakukan didalamnya meliputi:

- a. Pemilihan data (*Data Selection*), didefinisikan sebagai proses di mana data yang relevan dengan analisis diputuskan dan diambil dari pengumpulan data.
- b. *Cleaning*, proses menghilangkan data pengganggu (*noise*), *outlier*, inkonsistensi data dan perbaikan *Missing Value* untuk data yang akan ditransformasikan.

C. Transformasi Data

Transformasi Data didefinisikan sebagai proses mengubah data menjadi bentuk yang sesuai yang diperlukan oleh prosedur penambangan juga mengubah menjadi bentuk data numerik apabila diperlukan. Pada fase ini juga melakukan normalisasi atau proses mengubah data kedalam bentuk yang paling tepat/cocok. Salah satu teknik data *scaling* yang dapat digunakan yaitu *min-max normalization technique* [12]. Perhitungan teknik *min-max* dapat dilihat pada Persamaan 1.

$$v'_i = \frac{v_i - \min_A}{\max_A - \min_A} (\text{new_max}_A - \text{new_min}_A) + \text{new_min}_A \quad (1)$$

Keterangan:

v'_i : Nilai data yang baru dari hasil normalisasi min-max.

v_i : Nilai data yang akan dilakukan normalisasi.

\max_A : Nilai maksimum data.

\min_A : Nilai minimum data.

new_max_A : Nilai maksimum yang diharapkan dari proses normalisasi.

new_min_A : Nilai minimum yang diharapkan dari proses normalisasi.

Dilakukannya proses reparasi data diatas dilakukan agar data mentah yang telah dikumpulkan dapat diolah sesuai dengan kriteria yang dibutuhkan. Masalah kualitas data adalah salah satu kendala untuk menggunakan data secara efektif karena data yang kotor dapat menyebabkan kesalahan keputusan. Ini dapat memberikan berbagai layanan untuk organisasi dan hanya dengan data berkualitas tinggi, mereka dapat mencapainya layanan teratas dalam organisasi [13].

D. Clustering dengan metode K-means Clustering

Teknik *Clustering* dalam data mining memiliki arti sekelompok item yang mirip satu sama lain dalam grup dan tidak mirip dengan objek yang termasuk dalam kluster lain yang dikenal dengan istilah *similarity* [14]. Penggunaan *clustering* ini dapat mengklusterisasikan daerah yang padat dan meneumkan keterkaitan antara atribut yang ada. Kluster adalah sekumpulan data dimana objek didalam satu kluster tersebut memiliki kemiripan [9].

Algoritma *K-means* merupakan algoritma pengelompokan yang melakukan partisi terhadap set data menjadi sejumlah kluster yang sudah ditentukan di awal. Tujuannya adalah untuk membentuk kelompok yang terpisah dari sejumlah titik data [15]. *K-means* adalah pendekatan *iterative* yang menghitung nilai sentroid sebelum setiap iterasi. *K-means* adalah algoritma pengelompokan berbasis jarak yang membagi data menjadi sejumlah kluster dalam atribut numerik.

1. Tentukan jumlah kluster K dan jumlah iterasi maksimum.
2. Tentukan data acak yang menjadi centroid.
3. Hubungkan semua data observasi ke kluster terdekat dengan pengukuran jarak *Euclidean*. Rumus *Euclidean Distance* dapat dilihat pada Persamaan 2.

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (2)$$

4. Realokasi data ke masing-masing kelompok berdasarkan perbandingan jarak terdekat antar data dengan *centroid* masing-masing kelompok.
5. Menentukan nilai *centroid* yang baru dengan cara menghitung rata-rata dari kluster yang bersangkutan menggunakan Persamaan 3.

$$T(i) = \frac{x_1 + x_2 + x_3 + \dots + x_n}{\Sigma_n} \quad (3)$$

Keterangan:

$T(i)$ = Centroid data

x_1 = Data ke i pada atribut data ke k

Σ_n = Jumlah data

6. Lakukan perulangan dari langkah 3 - 5 hingga anggota tiap kluster tidak ada yang berubah.

Penentuan jumlah kluster dilakukan dengan menggunakan *Elbow Method*. Metode *Elbow* digunakan didalam teknik klusterisasi untuk menentukan jumlah kluster (k) terbaik. Penerapannya dengan mengukur jarak rata-rata kuadrat dari semua titik dalam sebuah kluster ke kluster *centroid* disebut dengan *Within Cluster Sum of Squares* (WCSS). Kemudian mengulangi proses ini untuk semua titik di kluster, lalu menjumlahkan nilai untuk kluster dan membaginya dengan jumlah poin. Selanjutnya menghitung rata-rata di semua kluster [16].

WCSS mengukur variabilitas pengamatan dalam setiap cluster. Secara umum, kluster yang memiliki jumlah kuadrat kecil lebih kompak daripada kluster yang memiliki jumlah kuadrat besar. Kluster yang memiliki nilai lebih tinggi menunjukkan variabilitas pengamatan yang lebih besar di dalam kluster.

Berikut ini tahapan algoritma metode Elbow dalam menentukan nilai K yang optimal [16]:

1. Menentukan total k;
2. Inisiasi perhitungan kluster dimulai dari nilai k=1 hingga total k;
3. Menghitung nilai kuadrat dari setiap jarak titik ke *centroid* nya.
4. Menjumlahkan hasil kuadrat semua titik.
5. Membandingkan nilai WCSS setiap jumlah k yang membentuk paling siku.

E. Interpretasi

Dengan data transaksi penjualan pada toko ARTCH yang akan dikelompokkan berdasarkan kemiripan setiap data. Pengelompokkan tersebut berdasarkan atribut yang dipilih saat pengumpulan dan preparasi data yang kemudian atribut tersebut akan diolah dengan menggunakan algoritma *K-means* yang akan menghasilkan kluster dan *centroid*. Maka dari *clustering* menggunakan metode *k-means* dapat menghasilkan pengetahuan yang menghasilkan visualisasi dan bentuk tabel untuk merepresentasikan hasil penggalan pengetahuan yang selanjutnya dapat diterapkan konsep penjualan *down-selling* oleh pemilik informasi ini.

Konsep *Down-selling* menawarkan produk yang diturunkan kepada pelanggan yang kesulitan menyelesaikan transaksinya. Sebagai contoh lain juga apabila ada pelanggan yang membatalkan pembelian, pelanggan ditawarkan produk yang sedang diskon agar pelanggan tidak pergi dengan tangan

kosong [17]. Lalu penerapan ini juga dapat meningkatkan nilai jual karena dapat terpusat terhadap pelanggan yang dapat meningkatkan kepuasan pelanggan dan loyalitas.

F. Silhouette Score

Silhouette Analysis digunakan untuk mempelajari dan memahami jarak pemisahan antara kluster yang dihasilkan. Analisis ini digunakan untuk mengukur seberapa dekat setiap objek dalam satu kluster dekat dengan objek lain di kluster lain [18]. Semakin dekat siluet objek ke nilai 1, semakin tinggi derajat kepemilikan objek dalam cluster. Sedangkan nilai yang mendekati -1 (nilai negatif) menunjukkan bahwa suatu objek kemungkinan berada pada kluster yang salah, karena kemiripan objek dengan objek lain pada kluster yang sama rendah [19]. Dalam penelitian ini penghitungan silhouette digunakan untuk menguji kualitas jumlah kluster yang sudah dihitung.

Untuk setiap poin data, perhitungan siluet terdapat tiga langkah:

1. Untuk setiap objek i dalam cluster, hitung jarak rata-rata dari objek ke seluruh objek dalam clusternya sendiri. Sebut jarak rata-rata antara objek i ke semua anggota kluster sebagai a_i .
2. Untuk setiap objek i dan kluster lain yang bukan kluster tempat saya berada, hitung jarak rata-rata dari objek ke seluruh objek di kluster lain yang paling dekat dengan kluster tempat saya berada (kluster tetangga). Panggil jarak rata-rata ke semua objek di kluster tetangga sebagai b_i .
3. Untuk setiap objek i, *silhouette coefficient* diperoleh dari Persamaan 4.

$$\frac{(b_i - a_i)}{\max(a_i, b_i)} \quad (4)$$

III. HASIL DAN DISKUSI

A. Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah data dari Perusahaan Artch Indonesia yang berisi riwayat transaksi dengan kurun waktu dari bulan Januari 2020 sampai dengan Juni 2020 yang terdiri dari 6174 *record*.

B. Data Pre-Processing

Sebelum tahap penggalan data selanjutnya perlu dilakukan data *pre-processing* untuk mengubah data transaksi yang sudah didapatkan agar diolah menjadi data yang siap untuk digunakan dalam *data mining*. Kegiatan yang akan dilakukan didalamnya yaitu *Data Selection* dan *Data Cleaning*.

1) Data Selection

Data yang dikumpulkan perlu dipilih dan dipisahkan menjadi kumpulan yang bermakna berdasarkan kebutuhan penggalan data.

TABEL 1. DATA SELECTION

No	Nama Produk	Harga Awal	Jumlah
1	ARTCH Outdoor Casual Sporty Snapback Trucker Baseball Five Panel	Rp 100.000	1
2	ARTCH Outdoor Casual Sporty Snapback Trucker Baseball Five Panel	Rp 100.000	1
5	ARTCH (tosca kuning) Sackbag tas seeut backpack ransel	Rp 50.000	1
4	ARTCH (tosca kuning) Sackbag tas seeut backpack ransel	Rp 50.000	1
...
6174	Tas Serut Artch [navy - maroon] sack bag drawsting bag gym sack	Rp 50.000	1

Seleksi data digunakan dalam data mining karena mereka menjadi dasar mempengaruhi jenis model data apa yang dibentuk. Contoh sampel data yang telah melalui tahap seleksi data dapat dilihat pada Tabel 1.

Pemilihan atribut yang digunakan dapat mendukung penggalian data yang sesuai dengan tujuan informasi yang akan didapatkan. Pemilihan atribut yang dibutuhkan adalah Nama Produk, Jumlah, Harga Awal. Atribut yang diambil mengacu pada penelitian sebelumnya yang relevan dengan penelitian ini yang melakukan klusterisasi produk menggunakan K-Means [6][7].

2) Data Cleaning

Tahapan selanjutnya melakukan pembersihan data yang didalamnya meliputi menghilangkan pencatatan data yang salah, outlier dan missing value untuk data yang akan ditransformasikan. Tujuan tahap ini agar menjaga kualitas data saat dilakukan proses data mining selanjutnya. Contoh data yang telah melalui tahap Data Cleaning dapat dilihat pada Tabel 2.

TABEL 2. DATA CLEANING

No	Nama Produk	Harga Awal	Jumlah
1	ARTCH Outdoor Casual Sporty Snapback Trucker Baseball Five Panel	Rp 100.000	1
2	ARTCH Outdoor Casual Sporty Snapback Trucker Baseball Five Panel	Rp 100.000	1
3	ARTCH (tosca kuning) Sackbag tas seeut backpack ransel	Rp 50.000	1
4	ARTCH (tosca kuning) Sackbag tas seeut backpack ransel	Rp 50.000	1
...
4969	Tas Serut Artch [navy - maroon] sack bag drawsting bag gym sack	Rp 50.000	1

Hasil data yang sudah dilakukan proses cleaning didapatkan jumlah record data dengan jumlah data sebanyak 4969.

C. Data Transformation

Tahap ini merupakan proses mengubah data transaksi menjadi bentuk yang sesuai dengan keperluan data mining. Selanjutnya melakukan normalisasi nilai setiap data dengan

menggunakan teknik data scaling yaitu min-max normalization.

Data yang telah melalui tahap transformasi ini dapat dilihat pada Tabel 3 dan Tabel 4.

TABEL 3. DATA TRANSFORMATION

No	Nama Produk	Harga Awal	Jumlah
1	ARTCH Outdoor Casual Sporty Snapback Trucker Baseball Five Panel	100000	9
2	ARTCH (tosca kuning) Sackbag tas seeut backpack ransel	50000	31
3	ARTCH [Apana -hitam abu] Hand Bag Mix tas vape bag slingbag tas selempang	120000	2
4	ARTCH [Steckbag - Magenta] slingbag tas selempang waterproof	120000	4
...
210	Tas Serut Artch [navy - maroon] sack bag drawsting bag gym sack	50000	9

Dari data Tabel 3 dilakukan transformasi nilai Harga Awal dan Jumlah menggunakan Min-Max Normalization dengan menggunakan rumus 1, hasil salah satu perhitungan dapat dilihat pada Persamaan 5.

$$\frac{100000-10000}{295000-10000}(1-0)+0=0,316 \tag{5}$$

Salah satu contoh perhitungan dilakukan adalah data produk ARTCH Outdoor Casual Sporty Snapback Trucker Baseball Five Panel dengan nilai Harga Awal 100000 yang di normalisasi menjadi 0,316.

TABEL 4. DATA TRANSFORMATION (NORMALISASI)

No	Nama Produk	Harga Awal	Jumlah
1	ARTCH Outdoor Casual Sporty Snapback Trucker Baseball Five Panel	0,316	0,038
2	ARTCH (tosca kuning) Sackbag tas seeut backpack ransel	0,140	0,143
3	ARTCH [Apana -hitam abu] Hand Bag Mix tas vape bag slingbag tas selempang	0,386	0,005
4	ARTCH [Steckbag - Magenta] slingbag tas selempang waterproof	0,386	0,014
...
210	Tas Serut Artch [navy - maroon] sack bag drawsting bag gym sack	0,140	0,038

Setiap record data dilakukan proses normalisasi menggunakan min-max normalization dengan skala 0 hingga 1.

D. Clustering dengan metode K-Means

Tahap selanjutnya melakukan penerapan teknik data mining yang telah dibahas sebelumnya, teknik yang digunakan yaitu teknik klusterisasi dengan implementasi algoritma K-Means terhadap dataset yang sebelumnya sudah melewati tahap preproses dan transformasi. Berikut merupakan tahap proses klusterisasi yang dilakukan:

1) Menentukan K Optimal

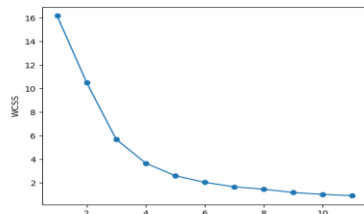
Untuk meningkatkan akurasi kluster akan digunakan suatu metode penentuan jumlah kluster (k) yang disebut dengan

Elbow Method, total k maksimal yang ditentukan adalah 11. Metode *Elbow* ini akan menentukan kualitas jumlah kluster terhadap dataset dengan mengukur jarak rata-rata kuadrat dari semua titik dalam sebuah kluster ke kluster *centroid* disebut dengan *Within Cluster Sum of Squares (WCSS)*. Nilai WCSS setiap kluster dapat dilihat pada Tabel 5.

TABEL 5. NILAI WCSS

Kluster	WCSS
1	16,18434
2	10,51558
3	5,704751
4	3,677146
5	2,608488
6	2,046677
7	1,664136
8	1,464443
9	1,217386
10	1,038481
11	0,854535

Sementara itu, visualisasi *elbow* diberikan oleh Gambar 1. Dapat dilihat berdasarkan metode *elbow* penentuan jumlah kluster yang optimal untuk dataset terkait adalah antara 4 hingga 6, maka dari itu dapat diambil jumlah kluster (k) yang optimal adalah 5.



Gambar 1. Grafik Elbow method

2) Perhitungan kluster dengan algoritma K-Means

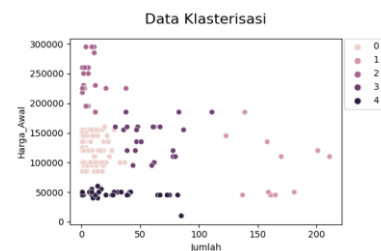
Jumlah kluster dalam algoritma K-Means dapat dipilih tanpa ketentuan namun dengan menerapkan metode *elbow* dapat meningkatkan kualitas dan akurasi klusterisasi. Setelah jumlah k ditentukan sebanyak 5 akan dilakukan perhitungan dengan metode k-means berdasarkan langkah-langkah pada tinjauan pustaka sebelumnya. Perhitungan menghasilkan setiap data terkelompok berdasarkan kluster yang dapat dilihat pada Tabel 6.

TABEL 6. HASIL KLUSTERISASI

Nama Produk	Harga Awal	Jumlah	Kluster
ARTCH Outdoor Casual Sporty Snapback Trucker Baseball Five Panel	100000	9	0
ARTCH (tosca kuning) Sackbag tas seout backpack ransel	50000	31	4
ARTCH [Steckbag - Magenta] slingbag tas selempang waterproof	120000	4	0
ARTCH [loreng abu tua] tas serut tas karung gymsack sackbag	45000	65	4
ARTCH [steckbag motif bunga] tas selempang slingbag bandung	120000	46	3

...
Tas Serut Artch [navy - maroon] sack bag drawstring bag gym sack	50000	9	4

Visualisasi hasil K-Means clustering dideskripsikan oleh Gambar 2 yang merujuk pada Tabel 6.



Gambar 2. Visualisasi klusterisasi

E. Interpretasi

Tahap selanjutnya interpretasi merupakan penerjemahan pola pola yang didapat dari hasil *data mining* agar mempermudah pemahaman terhadap data yang dilakukan oleh pengamat.

Setelah semua data terkluster akan dihitung bobot rata rata setiap kluster nya untuk menemukan makna dari setiap kluster dan selanjutnya menentukan kluster mana yang memiliki bobot terendah, maka akan didapatkan kluster yang setiap data didalamnya merupakan data produk yang memiliki harga yang rendah dan jumlah penjualan yang rendah, hal ini dilakukan menyesuaikan tujuan penelitian untuk menemukan kelompok produk yang cocok untuk diterapkan konsep *down-selling*. Hasil pengamatan kluster dapat dilihat pada Tabel 7.

TABEL 7. BOBOT KLASTER

Kluster	Harga Awal	Jumlah	Bobot
0	120695,6522	8,173913043	1,49858
1	94545,45455	164,0909091	2,502339
2	240500	7,076923077	0,594987
3	141818,1818	58,77272727	2,817878
4	46111,11111	26,33333333	0,506767

Setelah mendapatkan rata rata setiap kluster akan ditentukan kluster yang memiliki nilai bobot terendah dengan mengalikan rata-rata Harga dan Jumlah, pada data diatas disimpulkan bahwa kluster 4 merupakan kelompok data dengan harga dan jumlah penjualan rendah dengan nilai bobot 0,506767. Hasil rekomendasi produk terdapat pada Tabel 8.

TABEL 8. REKOMENDASI PRODUK

Nama Produk	Harga Awal	Jumlah
ARTCH [hijau fuji] tas serut sackbag gymsack ransel backpack	45000	2
ARTCH [lime / hijau stabilo] Tas Serut sackbag ransel backpack stringbag	45000	21
ARTCH [loreng coklat] tas serut sackbag tas karung waterproof	45000	2
ARTCH [lunar - coklat] Jackets lunar running	60000	14
ARTCH [merah] Tas Serut String Bag sackbag Drawstring Bag artch	45000	9
ARTCH [merah kopi] tas serut sackbag gymsack ransel	50000	7

Nama_Produk	Harga_Awal	Jumlah
ARTCH [mix mocca - pink] tas serut gymsack sabbag tas karung	50000	7
...
ARTCH [hijau fuji] tas serut sabbag gymsack ransel backpack	45000	2

Didapatkan hasil proses klasterisasi dimana anggota dalam klaster ini merupakan kelompok produk dengan tingkat penjualan yang rendah dan harga yang rendah.

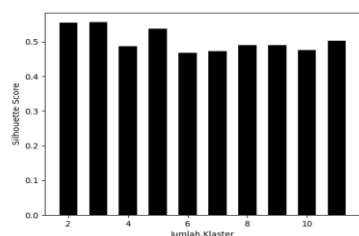
F. Pengujian dengan Silhouette Score

Perhitungan K-Means yang telah dilakukan di evaluasi dengan menggunakan *Silhouette Score* sebagai perbandingan kualitas jumlah klaster (k) yang digunakan dalam satu dataset. Pada dataset yang digunakan dalam penelitian ini jumlah k ditentukan sebanyak 5 berdasarkan hasil penerapan metode *elbow*. Pada pengujian ini menghitung dan melihat hasil *Silhouette Score* terhadap dataset. Hasil perhitungan *silhouette score* dapat dilihat pada Tabel 9.

TABEL 9. *SILHOUETTE SCORE*

Jumlah Klaster	Silhouette Score
2	0,554835691
3	0,556522064
4	0,486773759
5	0,537667434
6	0,46831974
7	0,472614921
8	0,49012164
9	0,490957787
10	0,476946679
11	0,503783568

Dari Tabel 8 dapat dilihat bahwa jumlah klaster dengan nilai *silhouette* yang tinggi ada pada klaster dengan jumlah 2, 3 dan 5. Sesuai dengan penerapan pada k-means dengan jumlah k yang digunakan dalam penelitian ini adalah 5 dengan skor siluet 0,538. Menjadikan klasterisasi yang dilakukan sudah cukup optimal. Pengujian ini dilakukan menyesuaikan dengan kasus yang dilakukan pada penelitian ini, untuk setiap pengujian akan berbeda beda hasilnya bergantung pada klaster dan data yang diuji.



Gambar 3. Visualisasi Silhouette Score

IV. KESIMPULAN

Dalam penelitian Klasterisasi Tingkat Penjualan Produk Menggunakan Metode *K-Means* Untuk Penerapan Konsep *Down-Selling* Di Artch Indonesia menghasilkan daftar rekomendasi produk yang dapat diterapkan konsep *down-selling*, didasari dari hasil implementasi *data mining* dengan

menggunakan Metode Klasterisasi *K-Means*. Hasil klasterisasi dengan jumlah klaster ditentukan sebanyak 5 berdasarkan *elbow method* dan hasil pengujian *silhouette score* untuk jumlah klaster 5 adalah 0,538, menunjukkan kelompok klaster 4 sebagai klaster dengan bobot terendah dan menghasilkan rekomendasi produk sebanyak 36 item. Dengan demikian tujuan yang diharapkan dari penelitian ini telah tercapai yaitu menghasilkan rekomendasi kelompok produk yang sesuai untuk penerapan konsep *down-selling*.

DAFTAR PUSTAKA

- [1] J. Liu *et al.*, "Data Mining and Information Retrieval in the 21st century: A bibliographic review," *Comput. Sci. Rev.*, vol. 34, 2019, doi: 10.1016/j.cosrev.2019.100193.
- [2] H. Li, Y. J. Wu, and Y. Chen, "Time is money: Dynamic-model-based time series data-mining for correlation analysis of commodity sales," *J. Comput. Appl. Math.*, vol. 370, p. 112659, 2020, doi: 10.1016/j.cam.2019.112659.
- [3] F., F. T. Kesuma, and S. P. Tamba, "Penerapan Data Mining Untuk Menentukan Penjualan Sparepart Toyota Dengan Metode K-Means Clustering," *J. Sist. Inf. dan Ilmu Komput. Prima (JUSIKOM PRIMA)*, vol. 2, no. 2, pp. 67–72, 2020, doi: 10.34012/jusikom.v2i2.376.
- [4] V. Y. Kiselev, T. S. Andrews, and M. Hemberg, "Challenges in unsupervised clustering of single-cell RNA-seq data," *Nat. Rev. Genet.*, vol. 20, no. 5, pp. 273–282, 2019, doi: 10.1038/s41576-018-0088-9.
- [5] F. Febrianti, M. Hafiyusholeh, and A. H. Asyhar, "Perbandingan Pengklusteran Data Iris Menggunakan Metode K-Means Dan Fuzzy C-Means," *J. Mat. "MANTIK"*, vol. 2, no. 1, p. 7, 2016, doi: 10.15642/mantik.2016.2.1.7-13.
- [6] S. A. Rahmah, "Klasterisasi Pola Penjualan Pestisida Menggunakan Metode K-Means Clustering (Studi Kasus Di Toko Juanda Tani Kecamatan Hutabayu Raja)," vol. 1, no. 1, pp. 1–5, 2020.
- [7] E. Muningsih and S. Kiswati, "Penerapan Metode K-Means untuk Clustering Produk Online Shop dalam Penentuan Stok Barang," *J. Bianglala Inform.*, vol. 3, no. 1, pp. 10–17, 2015.
- [8] C. Italina and Fakhrruzzi, "Pengaruh Strategi Pemasaran Terhadap Peningkatan Volume Penjualan Pakaian Pada Toko Grosir Kadafi Collection Di Kota Sigli Kabupaten Pidie," *J. Sains Ris.*, vol. 9, no. September 2019, pp. 61–67.
- [9] A. Saxena *et al.*, "A review of clustering techniques and developments," *Neurocomputing*, vol. 267, pp. 664–681, 2017, doi: 10.1016/j.neucom.2017.06.053.
- [10] B. Daniawan and A. Hermawan, "Sales Information System of Mobile Marketing at PT . World Innovative Telecommunication Using Up Selling , Cross Selling and Down Selling Methods," vol. 1, pp. 1–5, 2020, [Online]. Available: <https://jurnal.ubd.ac.id/index.php/te/article/view/411>.
- [11] C. A. Sugianto, A. H. Rahayu, and A. Gusman, "Algoritma K-Means untuk Pengelompokan Penyakit Pasien pada Puskesmas Cigugur Tengah," *J. Inf. Technol.*, vol. 2, no. 2, pp. 39–44, 2020, doi: 10.47292/joint.v2i2.30.
- [12] S. G. K. Patro and K. K. Sahu, "Normalization: A Preprocessing Stage," *Iarjset*, pp. 20–22, 2015, doi: 10.17148/iarjset.2015.2305.
- [13] F. Ridzuan and W. M. N. Wan Zainon, "A review on data cleansing methods for big data," *Procedia Comput. Sci.*, vol. 161, pp. 731–738, 2019, doi: 10.1016/j.procs.2019.11.177.
- [14] A. Dutt and M. A. Ismail, "Logical Review on Educational Data Mining," *Int. J. Comput. Commun. Netw.*, vol. 9, no. 3, pp. 39–42, 2020, doi: 10.30534/ijccn/2020/01932019.
- [15] G. Gan and M. K. P. Ng, "K-Means Clustering With Outlier Removal," *Pattern Recognit. Lett.*, vol. 90, pp. 8–14, 2017, doi: 10.1016/j.patrec.2017.03.008.
- [16] A. T. Rahman, Wiranto, and A. Rini, "Coal Trade Data Clustering Using K-Means (Case Study Pt. Global Bangkit Utama)," *ITSMART J.*

Teknol. dan Inf., vol. 6, no. 1, pp. 24–31, 2017, [Online]. Available: <https://jurnal.uns.ac.id/itsmart/article/download/11296/11108>.

- [17] S. M. Shugan and J. Xie, “Advance-selling as a competitive marketing tool,” *Int. J. Res. Mark.*, vol. 22, no. 3, pp. 351–373, 2005, doi: 10.1016/j.ijresmar.2004.11.004.
- [18] G. Ogbuabor and U. F. N., “Clustering Algorithm for a Healthcare Dataset Using Silhouette Score Value,” *Int. J. Comput. Sci. Inf. Technol.*, vol. 10, no. 2, pp. 27–37, 2018, doi: 10.5121/ijcsit.2018.10203.
- [19] S. N. Kane, A. Mishra, and A. K. Dutta, “Preface: International Conference on Recent Trends in Physics (ICRTP 2016),” *J. Phys. Conf. Ser.*, vol. 755, no. 1, 2016, doi: 10.1088/1742-6596/755/1/011001.