

Sistem Rekomendasi Penawaran Produk Pada *Online Shop* Menggunakan *K-Means Clustering*

Farhan Naufal¹, Yulison Herry Chrisnanto², Ade Kania Ningsih³

^{1,2,3} Universitas Jenderal Achmad Yani Cimahi Jl. Terusan Sudirman, Cimahi 40513, Indonesia
e-mail: naufalfarhan256@gmail.com^{*1}, y.chrisnanto@gmail.com kania215@gmail.com³

INFORMASI ARTIKEL

Sejarah Artikel:

Diterima Redaksi: 01 Februari 2022

Revisi Akhir: 04 Juni 2022

Diterbitkan Online : 02 Juli 2022

Kata Kunci:

Online Shop, Sistem Rekomendasi,
Algoritma *K-Means Clustering*, *Sillhouette Coefficient*

Korespondensi:

Telepon / Hp : +6282113923299

E-mail : naufalfarhan256@gmail.com

A B S T R A K

Online Shop adalah salah satu fasilitas yang disajikan oleh internet, yang mampu mempermudah masyarakat dalam belanja tanpa harus bertatap muka dengan pelanggan, tanpa harus antri dan tawar menawar. Pertumbuhan ekonomi digital semakin besar persaingan bisnis juga akan semakin berat, akibatnya semakin banyak *online shop* tidak hanya menampilkan produk-produk tetapi juga perlu didukung oleh pemilihan produk yang tepat untuk menarik perhatian pelanggan. Terlalu banyaknya variasi produk yang ditawarkan secara *random* (acak) pada *online shop* membuat beberapa pelanggan kesulitan dalam menentukan produk yang akan dibeli. Berdasarkan permasalahan yang muncul maka penelitian mengenai Sistem Rekomendasi Penawaran Produk Pada *Online Shop* Menggunakan *K-Means Clustering* ini dilakukan. Sistem ini menggunakan algoritma *K-Means Clustering* serta dataset yang digunakan adalah data transaksi penjualan dari kurun waktu 1 tahun terakhir agar cakupannya tidak meluas dengan menggunakan data terbaru. Hasil dari penelitian ini ditemumakan bahwa ada 3 *cluster* yang memiliki karakteristik berbeda yaitu, *cluster 1* dengan karakteristik penjualan sedang dengan rentang umur pembeli 36-50 tahun, *cluster 2* dengan karakteristik penjualan terbanyak dengan rentang umur pembeli 18-26 tahun dan *cluster 3* dengan karakteristik penjualan rendah dengan rentang umur 27-35 tahun. Dari hasil *cluster* yang didapat bahwa produk yang direkomendasikan merupakan produk terpopuler dari setiap *cluster*-nya. Hasil perhitungan nilai *sillhouette coefficient* didapatkan *cluster* dengan jumlah 3 karena memiliki nilai paling mendekati $S_i = 1$ yaitu dengan nilai 0.5564393731451109.

1. PENDAHULUAN

Sistem rekomendasi merupakan suatu aplikasi untuk menyediakan dan merekomendasikan suatu item dalam membuat suatu keputusan yang diinginkan oleh pengguna [1]. Rekomendasi yang ditawarkan berdasarkan hasil observasi terhadap keadaan dan keinginan pengguna untuk menentukan item yang akan dipilih [2]. Sistem rekomendasi telah banyak digunakan oleh hampir sebagian besar area bisnis dimana konsumen perlu membuat suatu keputusan atau rekomendasi pilihan dari informasi yang disediakan [1]. Salah satu area bisnis yang menerapkan sistem rekomendasi yaitu jual beli *online* (*Online Shop*).

Online shopping atau yang sering disebut belanja *via online* sendiri adalah suatu proses pembelian barang atau jasa dari mereka yang menjual barang atau jasa melalui *internet* dimana antara penjual dan pembeli tidak pernah bertemu atau melakukan kontak secara fisik yang dimana barang yang diperjual belikan ditawarkan melalui *display* dengan gambar yang ada di suatu website atau toko maya. Setelahnya pembeli dapat memilih barang yang diinginkan untuk kemudian melakukan pembayaran kepada penjual melalui rekening bank yang bersangkutan. Setelah proses pembayaran di terima,

kewajiban penjual adalah mengirim barang pesanan pembeli ke alamat tujuan [3].

Di era globalisasi sekarang ini *online shop* semakin banyak diminati oleh masyarakat, terlebih lagi pada masa pandemi seperti ini yang dimana semua kegiatan dilakukan secara online dari mulai bekerja, belajar, hingga berbelanja. Menjadikan pertumbuhan pasar *online* terus meningkat, hal ini menjadi perhatian utama pelaku bisnis untuk mengembangkan teknologi guna mempermudah pelanggan dalam membeli produk yang mereka inginkan [4]. Pertumbuhan ekonomi digital semakin besar persaingan bisnis juga akan semakin berat, akibatnya semakin banyak *online shop* tidak hanya menampilkan produk-produk tetapi juga perlu didukung oleh pemilihan produk yang tepat untuk menarik perhatian pelanggan. Terlalu banyaknya variasi produk yang ditawarkan secara *random* (acak) pada *online shop* membuat beberapa pelanggan kesulitan dalam menentukan produk yang akan dibeli [4]. Oleh karena itu diperlukan suatu sistem yang dapat merekomendasikan produk berdasarkan yang diminati pelanggan untuk memudahkan dalam memilih produk yang akan dibeli. Berdasarkan permasalahan itulah penelitian mengenai Sistem Rekomendasi Penawaran Produk Pada *Online Shop* ini dilakukan.

2. METODE PENELITIAN

A. Pengumpulan Data

Pada tahap ini adalah pengumpulan data yang dilakukan untuk memperoleh informasi yang dibutuhkan dalam rangka mencapai tujuan penelitian. Tujuan dari langkah pengumpulan data dan teknik pengumpulan data ini adalah demi mendapatkan data yang valid, sehingga hasil dan kesimpulan penelitian pun tidak akan diragukan kebenarannya. Data yang dikumpulkan untuk penelitian ini adalah data transaksi penjualan pada 1 tahun terakhir hal ini dilakukan agar cakupannya tidak meluas dengan menggunakan data terbaru. Data diperoleh dari toko online Syakila Shop yang merupakan sebuah toko yang berada di daerah Bandung dengan menjual barang aksesoris seperti gelang, kalung dan yang paling utama adalah jam tangan. Data transaksi tersebut terdapat atribut no transaksi, tanggal, bulan, *customer*, umur *customer*, merek, *type*, harga jual, harga beli, margin, tanggal iklan, iklan, total iklan, total margin, dan pajak.

Tabel 1 . Contoh Data Transaksi

No	1	2
TGL	24-jun	28-jul
BULAN	JUNI	JULI
Customer	RAHMAWATI	SILMI KAFFA
Umur	45	36
MEREK	DW SET KALUNG	FOSSIL
TYPE	MK5799	ES3380
HARGA	679000	800000
BELI	365000	590000
MARGIN	314000	210000

B. Preprocessing Data

Pada tahap ini dilakukan untuk mempersiapkan data serta memproses data mentah sebelum dilakukannya proses *data mining*. Hal ini dilakukan agar data dapat terintegrasi dengan system yang akan digunakan pada saat proses penambangan data. Pada tahap *Pre-Processing Data* ini terdapat 3 tahapan yaitu *data cleaning*, *data selection*, dan *data transformation*.

1. Data Cleaning

Pada tahap ini ini dilakukan proses *cleaning* data untuk membersihkan data dengan cara melengkapi data yang kosong, menghapus data duplikat atau menghilangkan *noise* [5]. Selain itu pada tahap ini akan mengubah format data agar dataset dapat diintegrasikan dengan system aplikasi.

2. Data Selection

Pada tahap ini dilakukan setelah proses *cleaning* data selesai, yaitu dilakukannya pemilihan atribut agar data dapat di proses penambangan data maka perlu memilih atribut yang tepat sehingga terintegrasi dengan baik. Dari 15 atribut yang ada peneliti memilih 8 atribut yaitu no transaksi, bulan transaksi, *customer*, umur *customer*, merk, tipe, harga jual dan harga beli. Pemilihan atribut tersebut dilakukan karena hanya data tersebutlah yang diperlukan untuk proses data mining

seperti atribut no transaksi digunakan sebagai penomoran setiap transaksi, atribut bulan transaksi diperlukan untuk mengetahui jumlah penjualan setiap bulanya, atribut *customer* digunakan untuk mengetahui nama *customer* disetiap transaksinya, atribut umur *customer* digunakan untuk mengetahui rentang umur berapa saja yang memiliki kecenderungan melakukan transaksi pada produk tertentu, merk dibutuhkan untuk mengetahui nama produk pada setiap transaksinya, atribut *type* dibutuhkan untuk melengkapi informasi dari atribut merk karena *type* merupakan *type* dari merk itu sendiri, atribut harga jual dan harga beli digunakan untuk mengetahui margin pada setiap transaksinya.

3. Data Transformation

Selanjutnya dilakukan transformasi data yang berasal yang telah di *cleaning* dan *selection* dilakukan proses transformasi sehingga data siap untuk disimpan di *warehouse* [5]

C. Data Mining

Data mining adalah proses pengekstrakan informasi dari jumlah kumpulan data yang besar dengan menggunakan algoritma dan teknik gambar dari statistik, mesin pembelajaran dan sistem manajemen *database* [11]. Data mining yang disebut juga dengan *Knowledge Discovery in Database (KDD)* adalah sebuah proses secara otomatis atas pencarian data di dalam sebuah memori yang amat besar dari data untuk mengetahui pola dengan menggunakan alat seperti klasifikasi, hubungan (*association*), atau pengelompokan (*clustering*).

D. Clustering

Clustering pada suatu data pada dasarnya adalah suatu tahapan untuk menggolongkan suatu himpunan data yang atribut kelas belum dideskripsikan, berdasarkan konsepnya prinsip *clustering* adalah untuk memaksimalkan dan meminimalkan kemiripan intra kelas. Misalnya, ada suatu himpunan objek, langkah pertama dapat di klasterisasi menjadi beberapa himpunan kelas kemudian menjadi sebuah himpunan beraturan sehingga dapat diturunkan berdasarkan klasifikasi tertentu. *Clustering* sering digunakan sebagai tahap awal dalam proses *data mining*, dengan hasil *clustering* yang terbentuk akan menjadi input dari algoritma berikutnya yang digunakan [17].

Teknik *Clustering* adalah teknik yang paling banyak digunakan untuk prediksi masa depan [12]. Aplikasi ini dapat membantu *customer* dalam pengambilan keputusan untuk membeli suatu produk. Penelitian ini menggunakan analisis *cluster* untuk menyatukan suatu produk ke dalam kelompok sesuai dengan karakteristiknya. Data *clustering* merupakan bagian dari metode data mining yang mempunyai sifat tidak terawasi atau tidak memerlukan target *output (unsupervised)* yang digunakan untuk mengelompokkan sejumlah data besar ke dalam himpunan bagian yang disebut dengan *cluster* dimana setiap *cluster* adalah kumpulan objek data yang serupa

satu sama lain ditempatkan dalam *cluster* yang sama namun berbeda dengan objek dalam *cluster* lainnya [19]. Analisis *cluster* dapat dibagi menjadi teknik pengelompokan *hierarchical* (hirarki) dan teknik pengelompokan *non-hierarchical* (non-hirarki). Contoh dari teknik hirarki adalah single linkage, complete linkage, average linkage, median dan Ward. Sedangkan teknik non-hirarki meliputi *k-means*, *adaptif k-means*, *k-medoids*, *fuzzy c-means*, dan *fuzzy clustering* [14].

E. K-Means Clustering

K-Means Clustering adalah salah satu algoritma analisis klaster non hirarki. Analisis klaster merupakan salah satu alat untuk mengelompokkan data berdasarkan variabel atau *feature*. Metode ini mempartisi data ke dalam *cluster* sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu *cluster* yang sama dan data yang mempunyai karakteristik yang berbeda di kelompokkan ke dalam *cluster* yang lain. Algoritma ini juga merupakan metode penganalisaan data atau metode data mining yang melakukan proses pemodelan tanpa supervisi (*unsupervised*) dan merupakan salah satu metode yang melakukan pengelompokan data dengan sistem partisi [8].

Algoritma K-Means menggunakan proses secara berulang-ulang untuk mendapatkan basis data *cluster*. Dibutuhkan jumlah *cluster* awal yang diinginkan sebagai masukan dan menghasilkan jumlah *cluster* akhir sebagai output. Jika algoritma diperlukan untuk menghasilkan *cluster* K maka akan ada K awal dan K akhir. Metode K-Means akan memilih pola k sebagai titik awal centroid secara acak. Jumlah iterasi untuk mencapai *cluster* centroid akan dipengaruhi oleh calon *cluster* centroid awal secara random dimana jika posisi centroid baru tidak berubah. Nilai K yang dipilih menjadi pusat awal, akan dihitung dengan menggunakan rumus Euclidean Distance yaitu mencari jarak terdekat antara titik centroid dengan data/objek. Data yang memiliki jarak pendek atau terdekat dengan centroid akan membentuk sebuah *cluster* [9].

Berikut ini langkah-langkah yang terdapat pada algoritma K-Means:

1. Menentukan jumlah k, menentukan banyaknya *cluster* k dilakukan dengan beberapa pertimbangan seperti pertimbangan teoritis dan konseptual yang mungkin diusulkan untuk menentukan berapa banyak *cluster*.
2. Menentukan centroid awal, penentuan centroid awal dilakukan secara random/acak dari objek-objek yang tersedia sebanyak k *cluster*, kemudian untuk menghitung centroid *cluster* ke-i berikutnya.
3. Hitung jarak setiap objek ke masing-masing centroid dari masing-masing *cluster*. Untuk menghitung jarak antara objek dengan centroid dapat menggunakan rumus *euclidian distance* sebagai berikut :

$$d(x, y) = \sqrt{\sum_{k=1}^p \{x_i - y_i\}^2}$$

Keterangan:

Dij : Jarak objek antara objek i dan j

P : Dimensi data

Xi : Koordinat dari obyek i pada dimensi k

Yi : Koordinat dari obyek j pada dimensi

4. Mengelompokkan setiap data ke dalam *cluster* berdasarkan kedekatannya dengan *centroid* (jarak terkecil).
5. Memperbaharui nilai centroid, nilai centroid baru yang diperoleh dari rata-rata *cluster* yang bersangkutan. Untuk menentukan centroid yang baru digunakan rumus sebagai berikut :

$$\mu_k = \frac{1}{N_k} \sum_{j=1}^{N_k} x_j$$

Keterangan:

μ_k : Adalah titik tengah kelompok ke-I/k

Nk : Banyak data dalam *cluster* ke-I/k

I, K : Indeks dari *cluster*

J : Indeks dari variable (v1, v2, v3..)

Xj : Nilai data ke-j dalam ke-I/k

6. Setelah centroid baru ditemukan, lakukan lagi langkah untuk menghitung euclidean distance dengan centroid yang baru secara berulang sampai penentuan *cluster* menjadi tetap sehingga didapat *cluster* tetapnya.
7. Terakhir setelah didapat *cluster* tetapnya dihitung rata-rata setiap variabel hal digunakan untuk membandingkan kualitas suatu kelompok dengan kelompok, rumus untuk menghitung rata-rata variabel sebagai berikut :

$$x = \frac{1}{n} \sum_{i=1}^n x_i$$

Keterangan:

X = rata-rata hitung

Xi = nilai sample ke-I

N = jumlah sample

3. PERANCANGAN SISTEM

Pada proses pembangunan perangkat lunak yang dilakukan dalam pembangunan sistem rekomendasi penawaran produk ini menggunakan beberapa tahapan yang dilakukan dimulai dari analisa perancangan sistem, implementasi, dan pengujian perangkat lunak.

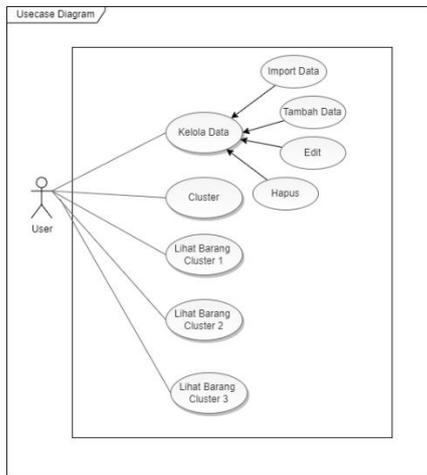
A. Analisa dan Perancangan Sistem

Pada tahapan ini menjelaskan tentang bagaimana penelitian ini dilakukan. Setiap tahapan penelitian akan

didasari pada literatur dan teori-teori oleh penelitian terdahulu yang digunakan sebagai pemahaman pendukung dan acuan dalam penelitian ini. Dengan begitu, hal tersebut akan menunjukkan bahwa setiap tahapan memiliki pegangan yang kuat.

1. Use case diagram

Use case diagram digunakan untuk memodelkan perilaku aktor yang berinteraksi dengan sistem pada aplikasi yang dibangun pada penelitian ini. Aktor merupakan user yang akan menggunakan aplikasi tersebut, sementara use case merupakan fungsi yang disediakan pada aplikasi tersebut [10]. Pada Gambar 3. dibawah merupakan use case diagram pada penelitian ini.



Gambar 1 Use Case Diagram

B. Implementasi Sistem

Setelah dilakukan analisa perancangan sistem tahapan berikutnya adalah tahapan implementasi perangkat lunak dimana tahapan ini yang mengimplementasikan hasil dari tahapan analisa kebutuhan perangkat lunak. Pada pengimplementasiannya menggunakan bahasa pemrograman python menggunakan framework flask untuk membuat keseluruhan website, dan penggunaan MySQL sebagai database. Berikut hasil dari pengimplementasian kebutuhan perangkat lunak.

a. Implementasi Halaman Dashboard

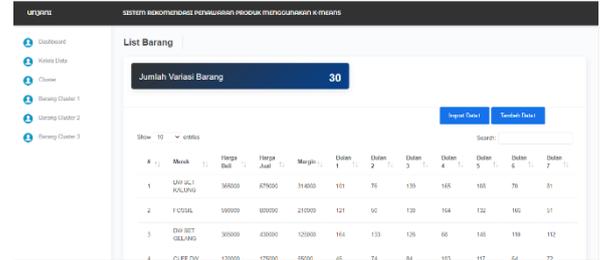
Halaman Dashboard merupakan halaman awal pada sistem ini yang terdapat informasi mengenai jumlah variasi data, jumlah data transaksi, serta jumlah hasil dari cluster.



Gambar 2 Halaman Dashboard

b. Implementasi Kelola Data

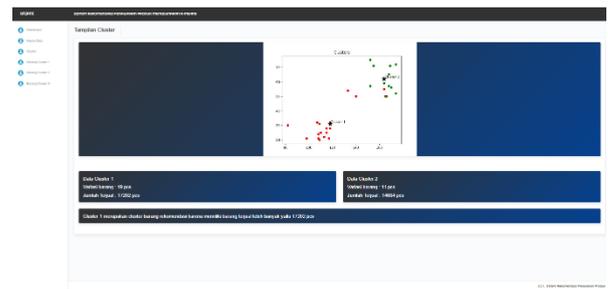
Halaman Kelola Data merupakan halaman yang digunakan untuk mengelola data produk pada sistem seperti menambah data, import data, hapus data dan edit data.



Gambar 3 Halaman Kelola Data

c. Implementasi Halaman Cluster

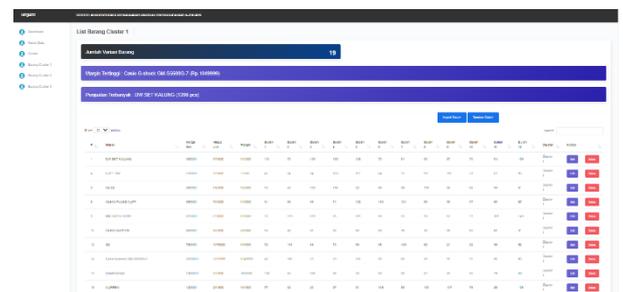
Halaman Kelola Cluster merupakan halaman yang menampilkan hasil dari proses data mining menggunakan K-Means Clustering.



Gambar 4 Cluster

d. Implementasi Halaman Barang Cluster

Halaman Barang Cluster merupakan halaman yang digunakan untuk menampilkan produk-produk yang sudah dikelompokkan berdasarkan cluster-nya.



Gambar 5 Barang Cluster

C. Implementasi K-Means Clustering

Menentukan Centroid Awal

Pada proses menentukan centroid awal dilakukan secara acak pada data sampel yang digunakan. Contoh penentuan centroid awal dapat dilihat pada Tabel 1.

Tabel 2 Menentukan Centroid Awal

Centroid	Merk	B1	B2	B3	B4	B5	...	B12
1	FOSSI L ME305 3	2	2	2	4	2		3
2	Casio G- shock GM- S5600 G-7	5	4	3	2	2		3
3	Origina l ZENO N B23- MEN	1	3	1	2	1		1

Pengelompokan Data

Proses pengelompokan data ke dalam suatu cluster dapat dilakukan dengan cara menghitung jarak terdekat menggunakan rumus euclidian distance dari suatu data ke sebuah titik centroid pada setiap kelompok yang ada. Berikut contoh perhitungan dari rumus jarak Euclidian Distance dapat dilihat pada Tabel 2.

Tabel 3 Perhitungan Euclidian Distance

Data ke-1	Proses	C1	C2	C3
	Euclidian Distance			
Data 1,	$\sqrt{(0-2)^2}$	4,795832		
Centroid	$\sqrt{(1-2)^2}$			
1	$\sqrt{(3-2)^2}$			
	$\sqrt{(2-4)^2}$			
	$\sqrt{(3-2)^2}$			
	$\sqrt{(1-3)^2}$			
	$\sqrt{(0-1)^2}$			
	$\sqrt{(1-2)^2}$			
	$\sqrt{(1-2)^2}$			
	$\sqrt{(3-3)^2}$			
	$\sqrt{(2-1)^2}$			
	$\sqrt{(1-3)^2}$			
Data 1,	$\sqrt{(0-5)^2}$	8,831761		
Centroid	$\sqrt{(1-4)^2}$			
2	$\sqrt{(3-3)^2}$			
	$\sqrt{(2-2)^2}$			
	$\sqrt{(3-2)^2}$			
	$\sqrt{(1-1)^2}$			
	$\sqrt{(0-4)^2}$			
	$\sqrt{(1-4)^2}$			
	$\sqrt{(1-4)^2}$			
	$\sqrt{(3-4)^2}$			
	$\sqrt{(2-4)^2}$			
	$\sqrt{(1-3)^2}$			
Data 1,	$\sqrt{(0-1)^2}$	4,690416		
Centroid	$\sqrt{(1-3)^2}$			
3	$\sqrt{(3-1)^2}$			

$\sqrt{(2-2)^2}$
$\sqrt{(3-1)^2}$
$\sqrt{(1-2)^2}$
$\sqrt{(0-1)^2}$
$\sqrt{(1-3)^2}$
$\sqrt{(1-0)^2}$
$\sqrt{(3-2)^2}$
$\sqrt{(2-1)^2}$
$\sqrt{(1-1)^2}$
...

Penentuan Cluster Yang Pertama

Dari perhitungan cluster euclidian distance didapat jarak terdekat antara objek dengan centroidnya, setelah itu ditentukan cluster. Dapat dilihat untuk produk dw set kalung termasuk ke cluster 3 karena dari ketiga centroid yang paling kecil jaraknya adalah C3. Contoh hasil penentuan cluster yang pertama dapat dilihat pada Tabel 3.

Tabel 4 Penentuan Cluster Pertama

No	Objek	Cluster (K)
1	DW SET KALUNG	3
2	FOSSIL ME3053	1
3	DW SET GELANG	1
4	CUFF DW	1
5	GUES	3
6	OLIVIA PLUSS CUFF	3
7	COACH	3
8	MICHAEL KORS	1
9	EDIFICIE	1
10	Daniel Wellington 3214	2
..		

Penentuan Centroid Baru

Penentuan centroid ini dilakukan agar menyesuaikan data dengan centroid yang sebenarnya. Dilakukan penentuan centroid sampai tidak ada perubahan centroid.

Tabel 5 Penentuan Centroid Baru

Centroid	Penentuan Centroid Baru
1	
Januari	$= \left(\frac{331}{143} \right)$ $= 2,314685$
Februari	$= \left(\frac{300}{138} \right)$ $= 2,173913$
Maret	$= \left(\frac{333}{144} \right)$ $= 2,3125$

April	$= \left(\frac{360}{145}\right)$
	$= 2,482759$
....
Desember	$= \left(\frac{362}{142}\right)$
	$= 2,510345$

Setelah centroid baru ditemukan, lakukan lagi langkah untuk menghitung euclidean distance dengan centroid yang baru secara berulang sampai penentuan *cluster* menjadi tetap sehingga didapat *cluster* tetapnya. Contoh hasil *cluster* tetap nya dapat dilihat pada Tabel 5.

Tabel 6 Hasil Cluster

No	Objek	Cluster (K)
1	DW SET KALUNG	3
2	FOSSIL ME3053	1
3	DW SET GELANG	3
4	CUFF DW	3
5	GUES	3
6	OLIVIA PLUSS CUFF	3
7	COACH	3
8	MICHAEL KORS	1
9	EDIFICIE	1
10	Daniel Wellington 3214	2
..		

Mencari Rata-Rata Variabel

Nilai rata-rata adalah angka representasi atau biasa juga disebut sebagai mean dari suatu kelompok data yang mewakili data secara keseluruhan. Dapat digunakan untuk membandingkan kualitas suatu kelompok dengan kelompok lain ,Mewakili kondisi suatu kelompok dengan 1 angka dan Mendeskripsikan suatu kelompok dengan singkat.

Tabel 7 Perhitungan Mencari Rata-Rata

Perhitungan	Proses Perhitungan
Nilai Sample 1	$= \left(\frac{153}{63}\right)$
	$= 2,428571$
Nilai Sample 2	$= \left(\frac{156}{63}\right)$
	$= 2,47619$
Nilai Sample 3	$= \left(\frac{132}{63}\right)$
	$= 2,095238$
....	

Tabel 8 Rata-Rata Variabel

K	B1	B2	B3	B4	B5	B6	...	B12
1	2,428 571	2,476 19	2,0 952 38	2,177 419	2,12 698	3		2,682 54
2	3	2,944 444	3,4 259 26	2,962 963	3,18 518	2,8 333		3,111 111
3	1,919 54	1,851 429	1,8 901 73	1,976 744	1,99 425 3	1,8 971 43		1,722 892

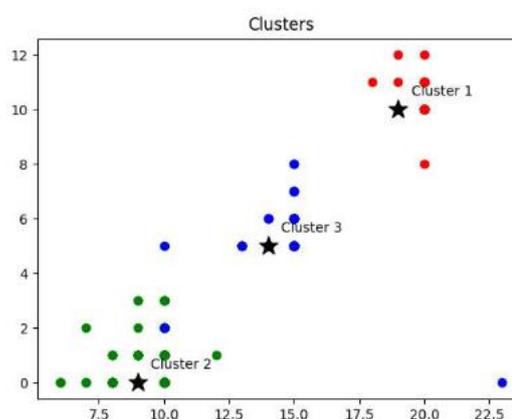
Berdasarkan perhitungan yang dihasilkan dari masing-masing variable, didapatkan karakteristik pada masing-masing *cluster*.berikut adalah penjelasan untuk setiap *cluster*.

- a. *Cluster* 1, memiliki karakteristik nilai rata-rata variabel penjualan bulan ke 1 sampai bulan ke 12 lebih tinggi dari *cluster* 3 tapi tidak lebih tinggi dari *cluster* 2. Dengan rentang umur pembeli 36-50 tahun.
- b. *Cluster* 2, memiliki karakteristik nilai rata-rata variabel penjualan bulan ke 1 sampai bulan ke 12 lebih tinggi dari *cluster* 1 dan 3. Dengan rentang umur pembeli 18-26 tahun.
- c. *Cluster* 3, memiliki karakteristik nilai rata-rata variabel penjualan bulan ke 1 sampai bulan ke 12 yang lebih rendah dari *cluster* 1 dan 2. Dengan rentang umur pembeli 27-35 tahun.

4. HASIL DAN PEMBAHASAN

A. Hasil

Pada tahap ini dilakukannya implementasi dari perancangan sistem yaitu menampilkan hasil *clustering* terhadap produk pada toko online syakila shop.



Gambar 6 Hasil Cluster

Dari hasil *cluster* diatas, dapat dilihat bahwa terdapat 3 *cluster* yang dihasilkan, dari ketiga *cluster* tersebut *cluster* 2 memiliki jumlah penjualan palik laku dengan rentang umur pembeli 18-26 tahun, *cluster* 1 memiliki jumlah penjualan sedang dengan rentang umur pembeli 36-50 tahun, *cluster* 3 memiliki jumlah

penjualan rendah dengan rentang umur pembeli 27-35 tahun.

B. Pengujian Cluster

Pengujian dilakukan terhadap semua data produk dengan memasukkan beberapa titik pusat secara acak kemudian dilakukan perhitungan nilai *silhouette coefficient* dari setiap titik pusat yang dimasukkan. Berikut adalah tabel proses pengujian terhadap data produk dengan memasukkan sebanyak 7 *cluster* yang berbeda dapat dilihat pada Tabel 8.

Tabel 9 Pengujian Cluster

Jumlah Cluster	Hasil Silhouette Coefficient
2	0.4815986035475924
3	0.5564393731451109
4	0.3099605480256625
5	0.2343006050013521
6	0.2292355977405526
7	0.22686121831886946
8	0.08479385867480424

Dari hasil pengujian 7 *cluster* yang dimasukan didapatkan *cluster* dengan jumlah 3 yang memiliki nilai *silhouette coefficient* paling mendekati dengan nilai $S_i = 1$ yaitu dengan nilai 0.5564393731451109. Nilai *silhouette* dapat dikatakan baik apabila bernilai positif karena titik pusat sudah berada di dalam *cluster* yang tepat sedangkan jika nilai *silhouette* dikatakan negatif menandakan terjadinya *overlapping* sehingga titik pusat berada diantara dua *cluster*.

5. KESIMPULAN

Penelitian ini menghasilkan perangkat lunak yang dapat mengelompokan data berdasarkan data transaksi penjualan pada sebuah *online shop*. Hasil akhir dari penelitian ini adalah mengelompokan sebuah data yang dapat digunakan sebagai bahan perekomendasi. Hasil dari data mining menggunakan *K-Means Clustering* menghasilkan 3 *cluster* yaitu, *cluster* 1 dengan karakteristik penjualan sedang dengan rentang umur pembeli 36-50 tahun, *cluster* 2 dengan karakteristik penjualan terbanyak dengan rentang umur pembeli 18-26 dan *cluster* 3 dengan karakteristik penjualan rendah dengan rentang umur 27-35. Dari hasil *cluster* yang didapat bahwa produk yang direkomendasikan merupakan produk terpopuler dari setiap *cluster*nya. Dari hasil 7 kali pengujian terhadap jumlah titik pusat, *cluster* yang berjumlah 3 memiliki nilai *silhouette coefficient* yang paling mendekati nilai $S_i=1$ yaitu dengan nilai 0.5564393731451109.

DAFTAR PUSTAKA

[1] B. T. W. Utomo and A. W. Anggriawan, "Sistem Rekomendasi Paket Wisata Se-Malang Raya Menggunakan Metode Hybrid Content Based

Dan Collaborative," *J. Ilm. Teknol. Inf. Asia*, vol. 9, no. 1, pp. 6–13, 2015.

- [2] W. A. Triyanto, "Association Rule Mining Untuk Penentuan Rekomendasi Promosi Produk," *J. SIMETRIS*, vol. Vol.5, no. No.2, pp. 121–126, 2014.
- [3] F. Ekonomika, D. A. N. Bisnis, and U. Diponegoro, "ANALISIS FAKTOR-FAKTOR YANG MEMPENGARUHI KEPUTUSAN PEMBELIAN DENGAN SISTEM PRE ORDER SECARA (Studi Kasus Pada Online Shop Chopper Jersey)," 2013.
- [4] C. S. Fatoni, E. Utami, and F. W. Wibowo, "Sistem Rekomendasi Produk Online Store Menggunakan Metode Apriori," *J. Inf. Politek. Indonusa Surakarta*, vol. 4, no. 2, pp. 20–27, 2018.
- [5] A. K. Ningsih and W. Witanti, "Sistem Rekomendasi Penjualan Menu Makanan di UMKM Kuliner Menggunakan Association Rule," *J. ICT Inf. Commun. Technol.*, vol. 19, no. 2, pp. 87–90, 2021, doi: 10.36054/jict-ikmi.v20i2.265.
- [6] A. Kurniawan, "Sistem Rekomendasi Produk Sepatu Dengan Menggunakan Menggunakan Metode Collaborative Filtering," *Semin. Nas. Teknol. Inf. dan Komun.*, vol. 2016, no. Sentika, pp. 610–614, 2016, [Online]. Available: <https://fti.uajy.ac.id/sentika/publikasi/makalah/2016/92.pdf>.
- [7] M. El Fikri, R. Ahmad, and R. Harahap, "Strategi Mengembangkan Kepuasan Pelanggan Online Shop Dalam Meningkatkan Penjualan (Studi Kasus Sabun Pyari)," *J. Manaj. Tools*, vol. 12, no. 1, pp. 87–105, 2020, doi: 10.1017/CBO9781107415324.004.
- [8] H. Sy, Rismayani, and A. Syam, "Data Mining Menggunakan Algoritma K-Means Pengelompokan Penyebaran Diare Di Kota Makassar," *SISITI Semin. Ilm. Sist. Inf. dan Teknol. Inf.*, vol. 8, no. 1, pp. 73–82, 2019.
- [9] I. M. A. W. Putra, G. Indrawan, and K. Y. E. Aryanto, "Sistem Rekomendasi Berdasarkan Data Transaksi Perpustakaan Daerah Tabanan dengan menggunakan *K-Means Clustering*," *J. Ilmu Komput. Indones.*, vol. 3, no. 1, pp. 18–22, 2018, [Online]. Available: <http://119.252.161.254/e-journal/index.php/jik/article/view/2749/1314>.
- [10] Y. Wautelet, "Using the RUP/UML business use case model for service development governance: A business and IT alignment based approach," *Proc. - 2020 IEEE 22nd Conf. Bus. Informatics, CBI 2020*, vol. 2, pp. 121–130, 2020, doi: 10.1109/CBI49978.2020.10069.
- [11] G. Abdurrahman, "Clustering Data Ujian Tengah Semester (UTS) Data Mining Menggunakan Algoritma K-Means," *Jurnal Sistem & Teknologi Informasi Indonesia (JUSTINDO)*, vol. 1, no. 2, pp. 71– 79, 2016.

- [12] M. H. I. Shovon and M. Haque, "Prediction of Student Academic Performance by an Application of *K-Means Clustering* Algorithm," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 2, no. 7, pp. 353–355, 2012.
- [13] L. Febriani, "Pengelompokan Mahasiswa Sistem Informasi Berdasarkan Tingkat Kompetensi Akademik Dengan Fuzzy K-Means," *Jurnal EKSIS*, vol. 5, no. 1, pp. 19–29, 2012.
- [14] F. A. Syam, "Implementasi Metode Klastering K-Means Untuk Mengelompokan Hasil Evaluasi Mahasiswa," *Jurnal Ilmu Komputer dan Bisnis*, vol. 8, no. 1, pp. 1857–1846, 2017
- [15] M. Ridwan, H. Suyono, and M. Sarosa, "Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier," *Jurnal EECCIS*, vol. 7, no. 1, pp.59–64, 2013.
- [16] A. Jananto, "Perbandingan Performansi Algoritma Nearest Neighbor dan SLIQ untuk Prediksi Kinerja Akademik Mahasiswa Baru," *Jurnal Teknologi Informasi DINAMIK*, vol. 15, no. 2, pp. 157–169, 2010.
- [17] N. Rohmawati, S. Defiyanti, and M. Jajuli, "Implementasi Algoritma K-Means Dalam Pengklasteran Mahasiswa Pelamar Beasiswa," *Jurnal Ilmiah Teknologi Informasi Terapan (JITTER)*, vol. 1, no. 2, pp. 62–68, 2015
- [18] F. N. R. F. J. Aziz, B. D. Setiawan, and I. Arwani, "Implementasi Algoritma K-Means untuk Klasterisasi Kinerja Akademik Mahasiswa," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 2, no. 6, pp. 2243–2251, 2018.
- [19] K. P. Chaudhari, R. A. Sharma, S. S. Jha, and R. J. Bari, "Student Performance Prediction System using Data Mining Approach," *IJARCCCE (International Journal of Advanced Research in Computer and Communication Engineering)*, vol. 6, no. 3, pp. 833–839, 2017
- [20] S. Sembiring, M. Zarlis, D. Hartama, R. S, and E. Wani, "Prediction Of Student Academic Performance By An Application Of DataMining Techniques," *International Conference on Management andArtificial Intelligence*, vol. 6, pp. 390–394, 2011.
- [21] A. Arief, " Rancang Bangun Sistem Rekomendasi Pariwisata *Mobile Advertising* Menggunakan Metode *Hybrid Filtering* Sebagai Pemberdayaan Masyarakat Usaha Kecil Menengah (UMKM) di Pulau Ternate", *Jurnal PROtek Vol. 03 No.1, Mei 2016*