

## Penerapan *K-Means Clustering* Untuk Mengelompokkan Data Transaksi Penjualan (Studi Kasus pada Wijaya Hijab)

Lanjar Widyawati<sup>1</sup>, Veronica Lusiana<sup>2</sup>

Teknik Informatika, Universitas Stikubank Semarang, Semarang, Indonesia  
 \*e-mail Corresponding Author: lanjarwidya19@gmail.com

### Abstract

*UKM Wijaya Hijab Outfit is a shop engaged in Muslim fashion. The problem that occurs in this shop is the difficulty in supplying product stock that is in high demand, quite attractive, and less desirable by consumers. To prevent product shortages and excess stock, a system is needed to classify products using the K-Means Clustering Algorithm into 3 groups, namely highly desirable, moderately desirable, and less desirable. The data used in this study are sales transaction data for 12 months, out of 38 products processed using RapidMiner Software, there is 1 product that is a member of cluster 1 with the most interested category, 2 products of cluster member 2 categories of sufficient interest, and 35 products of cluster 3 members who are less fans. The expected objective of this research is to determine the stock of products so that excess and shortage of product stock does not occur. So that the inventory of goods can be controlled and can help improve stock management improvements and sales strategies at the store.*

**Keywords:** *Clustering; K-Means Algorithm; Stock Management; Sales Transaction Data*

### Abstrak

UKM Wijaya Hijab Outfit merupakan sebuah toko yang bergerak dibidang fashion muslim. Permasalahan yang terjadi pada toko ini yaitu kesulitan dalam persediaan stok produk yang sangat diminati, cukup diminati, dan kurang diminati oleh konsumen. Untuk mencegah kekurangan dan kelebihan stok produk maka dibutuhkan sistem untuk mengelompokkan produk menggunakan Algoritma *K-Means Clustering* menjadi 3 kelompok yaitu sangat diminati, cukup diminati, dan kurang diminati. Data yang digunakan pada penelitian ini yaitu data transaksi penjualan selama 12 Bulan dari 38 produk yang diolah menggunakan *Software Rapid Miner* terdapat 1 produk yang menjadi anggota *cluster* 1 dengan kategori peminat terbanyak, 2 produk anggota *cluster* 2 kategori cukup peminatnya, dan 35 produk anggota *cluster* 3 yang kurang peminatnya. Tujuan yang diharapkan dari penelitian ini adalah untuk menentukan penyetokan produk agar tidak terjadi kelebihan dan kekurangan stok produk. Sehingga persediaan stok barang dapat terkontrol dan dapat membantu memperbaiki peningkatan manajemen stok dan strategi penjualan pada toko.

**Kata kunci:** *Clustering; Algoritma K-Means; Manajemen Stok; Data Transaksi Penjualan*

### 1. Pendahuluan

Pada masa globalisasi perkembangan teknologi yang pesat merupakan salah satu aspek yang dapat dimanfaatkan untuk mencapai kemudahan, termasuk arus teknologi informasi. Teknologi informasi menghasilkan sejumlah besar data dalam beberapa cara, salah satunya adalah untuk tujuan berbisnis. Dalam persaingan berbisnis saat ini, kita dituntut untuk selalu mengembangkan bisnis agar bisa bertahan dengan persaingan, dalam persaingan berbisnis pengusaha dituntut untuk menemukan cara yang dapat meningkatkan pemasaran atau penjualan pada usahanya. Salah satunya yaitu dengan memanfaatkan data transaksi penjualan [1].

Hal ini tidak dimanfaatkan oleh salah satu UKM yang bergerak pada bidang fashion yaitu UKM Wijaya Hijab Outfit. Permasalahan yang terjadi pada UKM Wijaya Hijab Outfit yaitu tidak adanya manajemen persediaan yang baik sehingga menyulitkan dalam pengelolaan persediaan produk. Data yang tersedia seharusnya dapat digunakan sebagai sistem keputusan

untuk strategi bisnis, solusi bisnis dan dukungan infrastruktur di bidang teknis namun manajemen persediaan yang buruk akan menyulitkan hal ini sehingga dengan adanya penelitian ini diharapkan mampu membantu pihak UKM Wijaya Hijab Outfit dalam mengelompokkan data-data penjualan yang dimiliki untuk mempermudah mengelola persediaan produk melalui metode algoritma K-Means Clustering.

Data mining merupakan istilah yang digunakan untuk menemukan pengetahuan tersembunyi dalam basis data dengan proses semi-otomatis yang menggunakan statistik, matematika, kecerdasan buatan, dan teknik pembelajaran mesin untuk mengekstraksi dan mengidentifikasi pengetahuan dan informasi potensial yang berguna yang disimpan dalam basis data besar.[2] Teknik data mining biasanya terbagi dalam beberapa kategori yaitu deskripsi, estimasi, prediksi, klasifikasi, klastering dan asosiasi.[3] Pada penelitian yang dilakukan oleh Muhammad Ali Khofi Lutfi dan Agung Nilogiri (2019) dengan judul penelitian Implementasi Algoritma *K-Means Clustering* Untuk Pengelompokan Minat Konsumen Pada Produk *Online Shop* menghasilkan kesimpulan bahwa Dengan menerapkan algoritma k-means clustering, pengelompokan kepentingan konsumen dibagi menjadi 3 cluster. Produk dengan tingkat minat konsumen yang rendah masuk dalam klaster 1, produk dengan tingkat minat konsumen sedang masuk dalam klaster 2, dan produk dengan tingkat minat konsumen tinggi masuk dalam klaster 3.

Mengacu pada latar belakang di atas, maka penelitian ini menerapkan metode algoritma *K-Means Clustering*, dalam mengelompokkan data transaksi penjualan menjadi 3 (tiga) bagian yaitu sangat diminati, cukup diminati dan kurang diminati, oleh karena itu dengan pembagian seperti ini diharapkan bisa memudahkan Usaha Kecil dan Menengah Wijaya Hijab Outfit dalam menyusun strategi penyetokan barang dan mengelompokkan barang yang terjual dengan kriteria sangat diminati, maka menyetok barang dalam jumlah banyak, dengan kriteria cukup diminati, maka menyetok barang dengan jumlah sedang dan kriteria kurang diminati, maka menyetok barang dengan jumlah sedikit, serta dapat membantu memperbaiki peningkatan manajemen stok dan strategi penjualan pada toko.

## 2. Tinjauan Pustaka

Penelitian ini dilakukan karena masih adanya hasil yang kontradiktif antara penelitian-penelitian sebelumnya, sehingga diharapkan penelitian ini mampu melengkapi hasil penelitian sebelumnya dan mendapatkan hasil penelitian yang baik.

Pada penelitian yang dilakukan oleh Afiasari, Suarna, dan Rahaningsih (2020). "Implementasi Data Mining Transaksi Penjualan Menggunakan Algoritma *Clustering* dengan Metode *K-Means*" menggunakan metode *K-Means Clustering* untuk mengelompokkan produk terlaris. pengolahan data pada RapidMiner menghasilkan hasil produk terlaris yang dapat memperoleh informasi persediaan barang serta menunjukkan bahwa Toko Bill Lights dapat melakukan strategi penjualan dan produksi ulang berdasarkan barang yang laku terjual [5].

Penelitian yang dilakukan oleh Wahyudi dan Wibowo (2022). "Implementasi Metode *K-Means Clustering* Untuk Pengelompokan Data nStok Produk Toko Online Perdagangan Kaos menggunakan Algoritme K-Means" pada penelitian ini sistem berbasis web yang menggunakan metode *K-Means Clustering* dapat membantu mengelompokkan stok produk dengan kategori terlaris serta dapat menyelesaikan masalah klasifikasi pada toko [6].

Pada penelitian Avivah, dkk (2020). "Penerapan Algoritma K-Means untuk Mengelompokkan Penjualan Mebel, karena terjadi permasalahan sulitnya mengetahui data transaksi penjualan pada toko mebel penelitian ini menggunakan Algoritma *K-Means Clustering* untuk mengelompokkan barang. Dengan itu dapat membantu toko dalam menentukan produk yang diminati dan membantu memperlancar penjualan pada toko [7].

Pada penelitian yang dilakukan oleh Nawawi, dkk (2021). "Implementasi Algoritma *K-Means Clustering* Menggunakan Orange Untuk Penentuan Produk Busana Muslim Terlaris", pemilik toko menggunakan metode *K-Means Clustering* untuk mengelompokkan barang terlaris dengan bantuan software orange, pada toko menentukan produk terlaris yang dapat membantu menemukan solusi dalam mengoptimalkan persediaan barang serta meningkatkan promosi penjualan pada toko. [8]

Pada penelitian yang dilakukan oleh Normah, Nurajizah, dkk (2021). "Penerapan Data Mining Metode *K-Means Clustering* Untuk Analisa Penjualan Pada Toko Fashion Hijab Banten", melakukan penelitian untuk mengelompokkan data stok baju menggunakan metode Algoritma

*K-Measn Clustering* untuk mengembangkan strategi pemasaran pada toko dan mendapatkan informasi terkait produk yang sangat laris, laris serta kurang laris.[9]

Berdasarkan dari penelitian sebelumnya yang masih terdapat hasil yang berbeda-beda, maka penelitian ini melakukan pengujian kembali menggunakan metode *K-Means clustering* dimana data dikelompokkan menjadi 3 *cluster* yaitu sangat diminati, cukup diminati, dan kurang diminati. Selain itu objek yang digunakan dalam penelitian ini merupakan UKM yang bergerak di bidang fashion dengan produk yang beragam. Dengan menggunakan objek yang berbeda dari penelitian sebelumnya, diharapkan penelitian ini dapat memberikan hasil yang baik dan dapat membantu UKM Wijaya Hijab untuk memperbaiki manajemen persediaan produknya dan mempermudah dalam penentuan keputusan strategi bisnis dan memantau data penjualan.

### 3. Metodologi

#### 3.1. Metode Pengembangan Sistem

Model *waterfall* terdapat beberapa fase yang berurutan yaitu *requirements definition, sistem and software design, implimentation & unit testing, integration & testing, dan operation & maintenance*. Berikut tahapan model *waterfall* sebagai berikut (Nia Kusuma Wardhani dkk, 2018) [10].

##### 1. Requirement Definition

Tahapan *requirement* difokuskan untuk kebutuhan *software* yang dimana dapat mengetahui tentang program yang dibuat. Tahapan ini bertujuan untuk mempersiapkan dan juga menganalisis apa yang dibutuhkan oleh sistem.

##### 2. Sistem and Software Design

Tahapan *Design* dilakukan sebelum proses coding dimulai. Tujuan desain ialah untuk memberikan gambaran tentang yang harus dikerjakan dan bagaimana tampilan dari sebuah system yang diinginkan.

##### 3. Implementation and Unit Testing

Fase *implementation* tahapan pembuatan sistem pertama kali dikembangkan di program kecil yang disebut unit, yang terintegrasi dalam tahap selanjutnya. Setiap unit dikembangkan dan diuji untuk fungsionalitas yang disebut sebagai unit testing.

##### 4. Integration & Testing

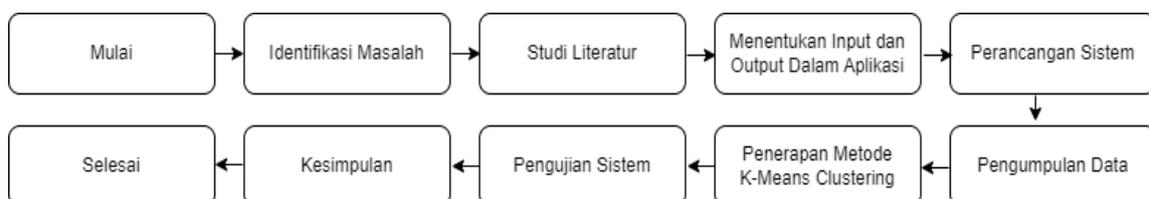
Tahap ini dilakukan pengujian perangkat lunak seperti pengujian menentukan hasil akurasi yang diperoleh menggunakan metode *kmeans* dan dilakukan untuk mengetahui *software* sudah sesuai desain yang diinginkan dan apakah masih ada kesalahan atau tidak.

##### 5. Operation & Maintenance

Fase *maintenance* merupakan tahapan dilakukan suatu perbaikan dan pemeliharaan disini *software* yang sudah jadi akan dijalankan atau dioperasikan oleh penggunanya. Pada tahap ini juga mencakup perbaikan apabila ditemukan suatu kesalahan error. Didalam proses pemeliharaan juga termasuk dalam memperbaiki kesalahan yang tidak ditemukan sebelumnya.

#### 3.2. Tahapan Penelitian

Alur penelitian terdapat sebuah gambaran umum tentang tahapan dari alur penelitian tersebut. Adapun rangkaian tahapan alur penelitian yang akan dilakukan dalam penelitian ini berikut alur penelitian yang bisa dilihat pada Gambar 1.



Gambar 1 Tahapan Penelitian

#### 3.3. K-Means Clustering

Algoritma *K-Means* adalah *algoritma non-hirarki* yang diturunkan dari metode pengelompokan data. Algoritma *K-Means* dimulai dari pembentukan partisi cluster di awal,

kemudian memperbaiki partisi cluster secara iteratif hingga partisi *cluster* tidak berubah secara signifikan.[11]

Algoritma *K-Means* adalah metode *non-hierarchical clustering* data yang mengelompokkan data ke dalam satu atau lebih cluster atau kelompok. Data dengan karakteristik yang sama dikelompokkan dan data dengan karakteristik berbeda dikelompokkan dengan cluster atau kelompok lain sehingga data dalam satu *cluster* atau kelompok memiliki variasi yang kecil. Pada tahap ini pusat cluster dipilih secara acak dari sekumpulan populasi data yang ada dan masing-masing dari komponen populasi akan ditandai pada pusat cluster. Posisi pusat cluster akan di hitung kembali sampai semua komponen data digolongkan kedalam tiap-tiap pusat cluster dan terakhir akan terbentuk posisi pusat *cluster* yang baru [12].

Proses Algoritma *K-Means*, langkah pertama yaitu menentukan jumlah cluster yang ingin dibentuk, pada penelitian ini terdapat 3 (tiga) *cluster*. Kemudian menentukan titik pusat pada setiap *cluster*, titik pusat awal ditentukan secara acak pada tabel 3. dibawah ini yaitu titik pusat awal *cluster*. Langkah kedua yaitu menghitung dan mengelompokkan ulang nilai centroid pada tahapan ini, sistem ini akan menghitung jarak dari titik pusat cluster ke setiap titik objek. Kemudian mengelompokkan sesuai dengan jarak terpendek antara data dan objek tersebut, pada jarak ini ditentukan dalam 1 (satu) kelompok yang paling dekat dengan pusat cluster. Langkah ketiga yaitu iterasi pusat *K-Means*, pada tahap ini dilakukan pengulangan sistem dan proses perhitungan dengan menggunakan titik pusat cluster (centroid) yang baru. Proses ini akan dilakukan terus menerus hingga kondisi menghentikan proses *K-Means Clustering* terpenuhi, yaitu hasil clustering dan nilai centroid yang terbentuk tidak berubah dari hasil perhitungan sebelumnya. Rumus perhitungan jarak terdekat menggunakan *Euclidean* dapat dilihat dibawah ini [13]:

$$D(X_2 - X_1) = \|X_2 - X_1\|_2 = \sqrt{\sum_{i=1}^p |X_2 - X_1|^2} \dots\dots\dots (1)$$

$$a_{ij} \begin{cases} 1 & d = \min \{D(X_i - C_1)\} \\ 0 & \text{Lainnya} \end{cases}$$

Keterangan:

X=Pusat Cluster

Y= Data

$a_{ij}$  :nilai keanggotaan titik  $X_i$  ke centroid  $c_1$

$d$  :jarak terpendek dari data  $X_i$  ke  $k$  (kelompok) setelah perbandingan

$c_1$  :centroid ke-1

## 4. Hasil dan Pembahasan

### 4.1. Pengolahan Data

Data yang telah diseleksi, diperoleh data transaksi penjualan dari UKM Wijaya Hijab Outfit pada Juni 2021 – Mei 2022. Tabel 1 merupakan data yang digunakan dalam penelitian ini.

Tabel 1. Data Transaksi Penjualan Bulan Juni 2021 – Mei 2022

No.	Nama Barang	Jumlah Penjualan	Rata - rata Penjualan perbulan
1	Square Pollycotton	7500	625
2	Scrunchie	2400	200
3	Salwa Bergo Instan	4300	358,3
4	Masker Kain Wijaya	924	77
5	Zara Dress	900	75
6	Salwa Bergo Kids	602	50,2
7	Mukena Army	500	41,7
8	Linen Basic Outer	0	0
9	Square Cornskin	176	14,7
10	Alila Dress	252	21

No.	Nama Barang	Jumlah Penjualan	Rata - rata Penjualan perbulan
11	Loli Top	450	37,5
12	Paris Premium S4	72	6
13	Inner Tanpa Lengan	300	25
14	Kinara Dress	1302	108,5
15	Jenna Blouse	620	51,7
16	Nai Dress	320	26,7
17	Flowy Dress	120	10
18	Nadira Dress	150	12,5
19	Zoya Midi Dress	400	33,3
20	Square Rubi S4	280	13,3
21	Pashmina Ceruty Babydoll	900	75
22	Aruna Top	600	50
23	Nudie bag	78	6,5
24	Square Voal Motif	180	15
25	Basic Oversize Shirt	200	16,7
26	Dina Kulot Bubble	228	19
27	Najwa Khimar	56	4,7
28	Ameena Bergo	420	35
29	Flower Dress	212	17,7
30	Yumi Skirt	109	9,08
31	Homey Dress	93	7,75
32	Maharani Mukena	330	27,5
33	Crop Cardi Rajut	0	0
34	Pashmina Inner	250	20,8
35	Nummy Top	400	33,3
36	Aura Set	190	15,8
37	Glossy Skirt	120	10
38	Leopard Bag	0	0

Langkah – langkah perhitungan K-Means

- 1) Langkah pertama yaitu menentukan jumlah *cluster* yang ingin dibentuk, pada penelitian ini terdapat 3 (tiga) *cluster*. Kemudian menentukan titik pusat pada setiap *cluster*, titik pusat awal ditentukan secara acak [14] pada tabel 4.2. dibawah ini yaitu titik pusat awal *cluster*. Data kemudian dimasukkan pada *Software Rapid Miner* dengan memasukkan 3 cluster yang sudah ditentukan, kemudian menerapkan metode *K-Means Clustering*. Ditentukan *cluster* menjadi 3 (tiga) yaitu C1, C2, C3. Gunakan *Rapidminer* untuk menemukan cluster. Kita harus mendapatkan nilai dari setiap cluster sesuai dengan atributnya. Dengan demikian, nilai C1, C2, dan C3 dapat dilihat pada tabel 2.

**Tabel 4.2 Titik Pusat Awal Cluster**

Centroid (C1)	Jumlah Penjualan	Rata-rata Penjualan Perbulan
C1	7500	635
C2	4300	358,3
C3	900	75

- 2) Langkah kedua yaitu menghitung dan mengelompokkan ulang nilai centroid pada tahapan ini, sistem ini akan menghitung jarak dari titik pusat cluster ke setiap titik objek. Kemudian mengelompokkan sesuai dengan jarak terdekat antara data dan objek tersebut, pada jarak ini ditentukan dalam 1 (satu) kelompok yang paling dekat dengan pusat cluster.
- 3) Langkah ketiga yaitu iterasi pusat *K-Means*, pada tahap ini dilakukan pengulangan sistem dan proses perhitungan dengan menggunakan titik pusat cluster (centroid) yang baru. Proses ini akan dilakukan terus menerus hingga kondisi menghentikan proses *K-Means Clustering* terpenuhi, yaitu hasil clustering dan nilai centroid yang terbentuk tidak berubah dari hasil perhitungan sebelumnya. Pada tabel 4.3 dibawah ini adalah hasil clustering pada iterasi akhir.

Tabel 3. Hasil *Clustering* Pada Tahap Akhir

C1	C2	C3	Jarak Terdekat	Kelompok
0	4164,386176	7189,600942	0	C1
5117,677696	953,2915202	2071,92326	953,2915	C2
3211,094656	953,2915202	3978,506288	953,2915	C2
6598,793829	2434,407653	590,8071746	590,8072	C3
6622,877018	2458,490842	566,7239883	566,724	C3
6921,907182	2757,521007	267,6939398	267,6939	C3
7024,260736	2859,874561	165,3405008	165,3405	C3
7525,996612	3361,610436	336,3957935	336,3958	C3
7349,38379	3184,997614	159,7831642	159,7832	C3
7273,123126	3108,73695	83,52266512	83,52267	C3
7074,436815	2910,050639	115,1644666	115,1645	C3
7453,747044	3289,360868	264,1462583	264,1463	C3
7224,956747	3060,570571	35,35693523	35,35694	C3
6219,4836	2055,097424	970,1173775	970,1174	C3
6903,84479	2739,458615	285,75632	285,7563	C3
7204,884655	3040,498479	15,28696807	15,28697	C3
7405,580666	3241,19449	215,9799136	215,9799	C3
7375,47668	3211,090504	185,8759569	185,876	C3
7124,612894	2960,226718	64,98851856	64,98852	C3
7245,866193	3081,489286	57,09130533	57,09131	C3
6622,877018	2458,490842	566,7239883	566,724	C3
6923,916883	2759,530707	265,6842033	265,6842	C3
7447,726247	3283,340071	258,1254645	258,1255	C3
7345,372693	3180,986517	155,7720118	155,772	C3
7325,300601	3160,914425	135,7000302	135,7	C3
7297,206315	3132,820139	107,6057477	107,6057	C3
7469,799736	3305,41356	280,1989764	280,199	C3
7104,540801	2940,154626	85,06060233	85,0606	C3
7313,259006	3148,87283	123,6584713	123,6585	C3
7416,619071	3252,232895	227,0183057	227,0183	C3
7432,674254	3268,288078	243,0734809	243,0735	C3
7194,852761	3030,466585	5,259388636	5,259389	C3

C1	C2	C3	Jarak Terdekat	Kelompok
7525,996612	3361,610436	336,3957935	336,3958	C3
7275,132826	3110,74665	85,53225149	85,53225	C3
7124,612894	2960,226718	64,98851856	64,98852	C3
7335,340799	3170,954623	145,7400749	145,7401	C3
7405,580666	3241,19449	215,9799136	215,9799	C3
7525,996612	3361,610436	336,3957935	336,3958	C3

Pada iterasi ke-3 perhitungan dihentikan, karena perhitungan pada iterasi 2 dan iterasi 3 dihasilkan sama atau tidak berubah dan dihasilkan pada tabel 4. dibawah ini.

Tabel 4. Hasil Akhir *Iterasi K-Means Clustering*

Cluster (C)	Jumlah Produk
Cluster 1	1 Produk
Cluster 2	2 Produk
Cluster 3	35 Produk

Hasil akhir perhitungan iterasi ke 3 (tiga):

C1 adalah barang yang sangat diminati yaitu:

1. Square Pollycotton

C2 adalah barang yang cukup diminati yaitu:

1. Scrunchie
2. Salwa Bergo Instan

C3 adalah barang yang kurang diminati yaitu:

- |                        |                          |                      |
|------------------------|--------------------------|----------------------|
| 1. Masker Kain Wijaya  | 13. Nai Dress            | 25. Ameena Bergo     |
| 2. Zara Dress          | 14. Flowy Dress          | 26. Flower Dress     |
| 3. Salwa Bergo Kids    | 15. Nadira Dress         | 27. Yumi Skirt       |
| 4. Mukena Army         | 16. Zoya Midi Dress      | 28. Homey Dress      |
| 5. Linen Basic Outer   | 17. Square Rubi S4       | 29. Maharani Mukena  |
| 6. Square Cornskin     | 18. Pashmina Ceruty      | 30. Crop Cardi Rajut |
| 7. Alila Dres          | 19. Aruna Top            | 31. Pashmina Inner   |
| 8. Loli Top            | 20. Nudie Bag            | 32. Nummy Top        |
| 9. Paris Premium S4    | 21. Square Voal Motif    | 33. Aura Set         |
| 10. Inner Tanpa Lengan | 22. Basic Oversize Shirt | 34. Glossy Skirt     |
| 11. Kinara Dress       | 23. Dina Kulot Bubble    | 35. sLeopard Bag     |
| 12. Jenne Blouse       | 24. Najwa Khimar         |                      |

#### 4.2. Pengujian RapidMiner

*Rapidminer* adalah *software* atau perangkat lunak untuk pengolahan data. *Rapidminer* menggunakan prinsip dan algoritme data mining, dikombinasikan dengan statistik, kecerdasan buatan, dan metode basis data, untuk mengekstraksi pola dari kumpulan data besar. *RapidMiner* memungkinkan pengguna untuk dengan mudah menggunakan operator untuk penghitungan data besar. Operator ini digunakan untuk memodifikasi data. Data sudah terhubung ke node pada operator, selanjutnya kita tinggal menghubungkannya ke node result untuk melihat hasilnya. Hasil yang ditampilkan oleh *Rapidminer* juga dapat divisualisasikan dengan grafik [15].

Data dianalisis menggunakan teknik *K-Means Clustering Rapidminer* untuk mengoptimalkan hasil pengelompokan data. Setelah dilakukan perhitungan manual menggunakan Algoritma *K-Means* selanjutnya dilakukan perhitungan menggunakan *Software Rapidminer*, hasil yang diperoleh yaitu sama, dapat disimpulkan bahwa hasil perhitungan

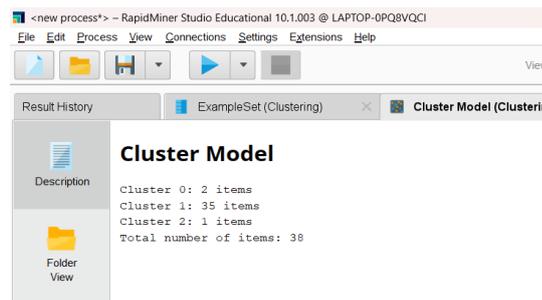
sudah benar. Pada gambar 2 dibawah ini adalah hasil pengolahan Data Transaksi Penjualan menggunakan Software RapidMiner.

Row No.	Nama Barang	cluster	Jumlah Penj...	Rata - rata P...
1	Square Polyc...	cluster_2	7500	625
2	Scrunchie	cluster_0	2400	200
3	Salwa B ergo l...	cluster_0	4300	358
4	Masker Kain ...	cluster_1	924	77
5	Zara Dress	cluster_1	900	75
6	Salwa B ergo ...	cluster_1	602	50
7	Mukena Army	cluster_1	500	42
8	Linen Basic O...	cluster_1	0	0
9	Square Corns...	cluster_1	176	15
10	Allia Dress	cluster_1	252	21
11	Loli Top	cluster_1	450	38
12	Paris Premiu...	cluster_1	72	6
13	Inner Tanpa Le...	cluster_1	300	25
14	Kinara Dress	cluster_1	1302	109
15	Jenna Blouse	cluster_1	620	52
16	Nai Dress	cluster_1	320	27
17	Flowy Dress	cluster_1	120	10

ExampleSet (38 examples, 2 special attributes, 2 regular attributes)

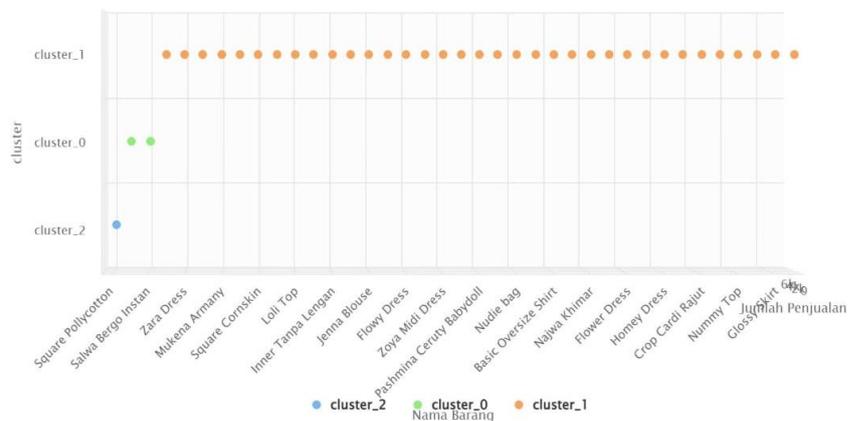
Gambar 2. Hasil Pengolahan Data dengan *Rapid Miner*

Pada gambar 2. diatas menunjukkan hasil perhitungan dengan RapidMiner yaitu saat input data dan klik menu run pada RapidMiner akan memunculkan hasil percluster pada setiap data.



Gambar 3. Hasil Clustering

Gambar 3 menunjukkan bahwa hasil saat data diklik menu *run* di RapidMiner didapat hasil *cluster 0* yang menunjukkan produk cukup diminati terdiri dari 2 items, *cluster1* menunjukkan produk kurang diminati ada 35 items, dan *cluster3* sebagai produk sangat diminati terdiri dari 1 items.



Gambar 4. Grafik Contoh Set *K-Means* dengan *RapidMiner*

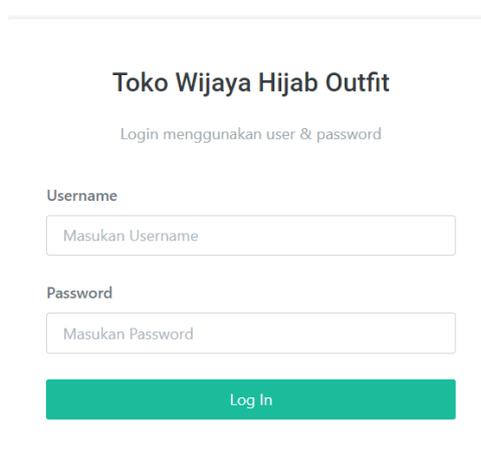
Pengolahan data transaksi penjualan menggunakan RapidMiner dapat dilakukan dengan mudah. Hasil yang dapat dilihat pada gambar 4 di atas menunjukkan grafik *Example Set K-Means Clustering* hasil perhitungan. Dari total 38 produk terdapat 1 produk C1, 2 produk C2, dan 35 produk C3.

#### 4.3. Hasil Program

Hasil program ini merupakan implementasi dari perencanaan sistem pada penelitian ini, berikut merupakan hasil implementasi perencanaan:

##### a. Halaman *Login*

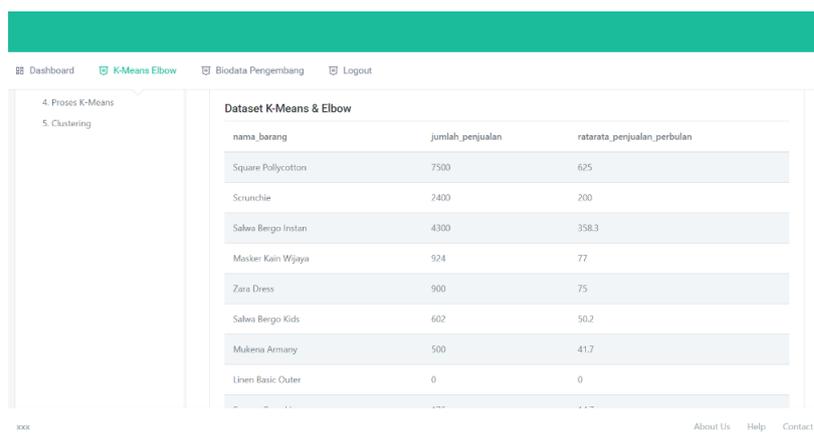
Halaman login dilihat pada gambar 5. pada halaman ini admin harus memasukkan username dan password agar dapat masuk sistem.



Gambar 5. Halaman Login

##### b. Halaman Data Produk

Setelah melakukan login terdapat halaman data produk, halaman ini digunakan untuk menginput data dengan cara manual yaitu choose file excel input semua data. Data yang telah diinput menggunakan file excel kemudian akan muncul ke halaman data produk pada gambar 6.

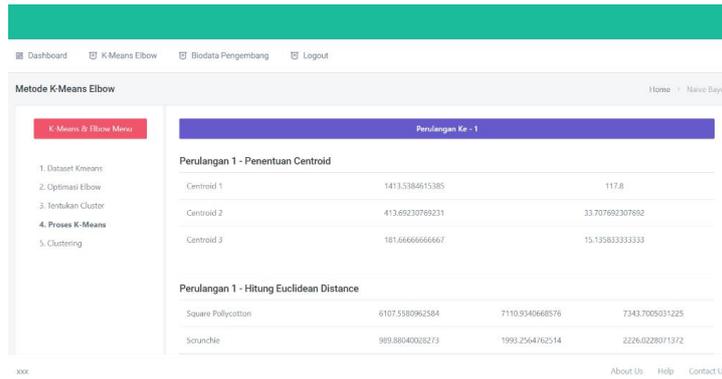


nama barang	jumlah penjualan	rata-rata penjualan perbulan
Square Polycotton	7500	625
Scrunchie	2400	200
Salwa Bergo Instan	4300	358.3
Mesker Kain Wjaya	924	77
Zara Dress	900	75
Salwa Bergo Kids	602	50.2
Mukena Armany	500	41.7
Linen Basic Outer	0	0

Gambar 6. Halaman Data Produk

##### c. Halaman Proses *K-Means*

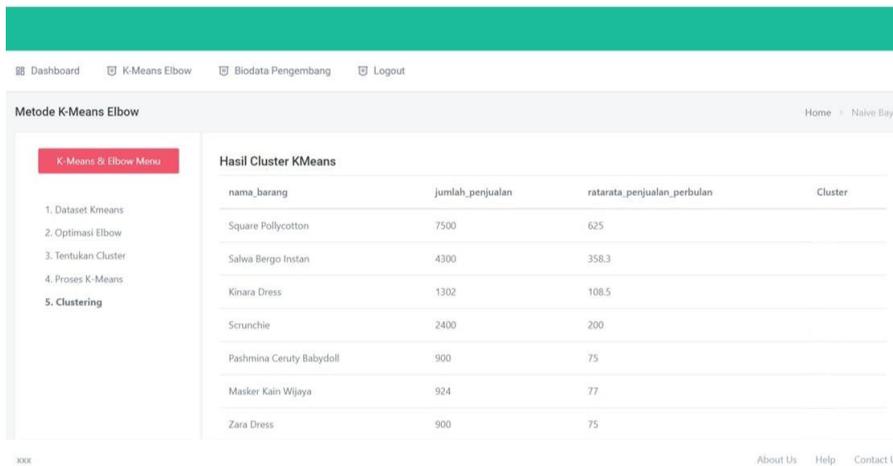
Pada proses *K-Means* data yang sudah diinput dan disimpan akan diproses setelah penentuan cluster dan akan menghasilkan perhitungan yang ada pada gambar 7.



Gambar 7. Halaman Proses K-Means

d. Halaman Clustering

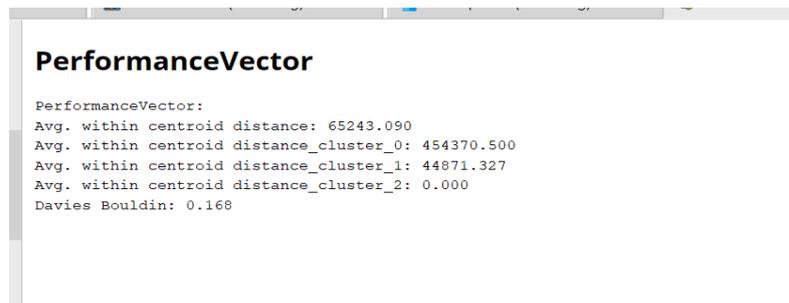
Halaman clustering digunakan untuk menampilkan hasil dari pengelompokkan data transaksi penjualan.



Gambar 8. Halaman Clustering

4.5. Pengujian Validasi

Validasi adalah proses membandingkan rata-rata pusat kluster awal dengan pusat kluster akhir. Pada saat verifikasi, data dihitung berdasarkan nilai K terdekat. Tujuan perhitungan adalah menghitung jumlah titik dengan label yang sama pada data latih. [16] Bias adalah faktor yang mempengaruhi yang mengarah pada hasil penelitian yang salah atau tidak akurat. Untuk mengurangi bias yang disebabkan oleh sampel tertentu, seluruh proses pelatihan dan pengujian akan diulangi beberapa kali dengan sampel yang berbeda. Hasil pengujian validasi ditunjukkan pada gambar 9.



Gambar 9. Pengujian Validasi Dengan RapidMiner

Pada pengujian validasi dihasilkan jumlah cluster yang paling optimal menggunakan metode *Davies Bouldin Index* (DBI) dengan hasil yang didapat yaitu *cluster* 0 mendapat hasil 454370.500, *cluster*-1 mendapat hasil 44871.327, dan *cluster*-2 mendapat hasil 0.000. Jika hasil pengujian mendapatkan hasil yang mendekati nilai Nol (0) maka semakin baik nilai cluster yang didapat. Dalam pengujian pada penelitian ini menggunakan metode DBI dihasilkan 3 *cluster* yang paling optimal dengan nilai DBI 0.168. Hasil dari temuan ini sama dengan yang dihasilkan dengan penelitian Puji Subekti, dkk.[17] Pada temuannya mereka mendapatkan hasil yang sama yaitu menemukan *cluster* yang paling optimal 3 cluster dengan nilai DBI 0.416. Berbeda dengan temuan Maulana, dkk yang menemukan *cluster* yang paling optimal yaitu 2 cluster dengan nilai DBI 0,837.[18]

## 5. Simpulan

Berdasarkan hasil penelitian tentang Algoritma *K-Means Clustering* Untuk Mengelompokkan Data Transaksi Penjualan (Studi Kasus: UKM Wijaya Hijab Outfit), dapat disimpulkan bahwa data transaksi penjualan ini sangat berguna untuk menjadi sebuah informasi bagi toko. Dengan penggunaan metode *K-Means Clustering* pada penelitian ini memberikan informasi kepada pemilik toko terkait pengelompokan produk yang sangat diminati, cukup diminati, dan kurang diminati. Data yang digunakan dalam penelitian ini yaitu data transaksi penjualan selama 12 Bulan dari 38 produk yang di dapat dan diteliti terdapat 1 produk yang menjadi anggota *cluster* 1 atau dapat diartikan sebagai produk yang memiliki peminat terbanyak, 2 produk anggota *cluster* 2 sebagai produk yang cukup peminatnya, dan 35 produk anggota *cluster* 3 yang memiliki peminat kurang. Penelitian ini dapat digunakan pemilik toko untuk menentukan penyetokan produk agar tidak terjadi kelebihan dan kekurangan stok produk. Sehingga persediaan stok barang dapat terkontrol dan dapat membantu memperbaiki peningkatan manajemen stok dan strategi penjualan pada toko. dan sistem *clustering* berbasis web dapat dikembangkan untuk membantu pemilik toko dalam memberikan informasi terkait pengelompokan produk.

## Daftar Referensi

- [1] M. H. Siregar, "Klasterisasi Penjualan Alat - alat Bangunan Menggunakan Metode K-Means," *Jurnal Teknologi Dan Open Source*, Vol. 1 No. 2, pp. 83-91, Desember 2018,
- [2] Y. Mardi, "Data Mining: Klasifikasi Menggunakan Algoritma C4.5," *Jurnal Edik Informatika Penelitian Bidang Komputer Sains dan Pendidikan Informatika*, vol. 2, no. 2, pp. 213-219, 2017.
- [3] D. R. Hermanto Wihono, "Prediksi Calon Pendorong Darah Potensial Dengan Algoritma Naïve Bayes, K-Nearest Neighbors dan Decision Tree C4.5," *JURIKOM (Jurnal Riset Komputer)*, vol. 7, no. 1, pp. 7-14, 2020.
- [4] A. K. Jeky Andreanus, "Sejarah, Teori Dasar dan Penerapan Reinforcement Learning :Sebuah Tinjauan Pustaka," *Jurnal Telematika*, vol. 12, pp. 113-118, 2017.
- [5] N. S. N. R. Nur Afiasari, "Implementasi Data Mining Transaksi Penjualan Menggunakan," *Jurnal SAINTEKOM*, vol. 13, no. 1, pp. 100-110, 2023.
- [6] A. W. Sukri Illahi Wahyudi, "Implementasi Metode K-Means Clustering Untuk Pengelompokan Data Stok Produk Toko Online Perdagangan Kaos," In *Prosiding Seminar Nasional Mahasiswa Fakultas Teknologi Informasi (SENAFTI)*, Vol. 1, No. 1, pp. 470-478). 2022.
- [7] A. F. N. S. Mutiara Rita Avivah, "Penerapan Algoritma K-Means untuk Mengelompokkan Penjualan Mebel," *Scientific Student Journal for Information, Technology and Science*, Vol. 1 No: 2, pp. 61-65, 2020.
- [8] M. T. A.A. Nia Kusuma Wardhani, "Sistem Informasi Manajemen Sumber Daya Manusia Berbasis Web (Studi Kasus : PT. Klik Teknologi Indonesia)," *Jurnal TECHNO Nusa Mandiri*, Vol. 15, No, 2, pp. 145-152 2018
- [9] V. K. Xindong Wu, *The top ten algorithms in data mining*, Tredemark: CRC Press, 2009.
- [10] F. Yunita, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Pada Penerimaan Mahasiswa Baru (Studi Kasus: Universitas Islam Indragiri)," *Sistemasi: Jurnal Sistem Informasi*, Vol. 7, No. 3, 238 – 249, 2018.

- 
- [11] Suprayogi, "Data Mining Clustering," *arXiv*, vol. 53, pp. 1689-1699, 2021
- [12] D. Novianti, "Implementasi Algoritma Naive Bayes Pada Data Set Hepatitis Menggunakan RapidMiner," *KOPERTIP: Scientific Journal of Informatics Management and Computer*, vol. XXI, pp. 49-54, 2019.
- [13] F. S. A. E. M Syukron Nawawi, "Implementasi Algoritma K-Means Clustering Menggunakan Orange Untuk Penentuan Produk Busana Muslim Terlaris," In *Prosiding Seminar Nasional Teknologi Informasi dan Komunikasi (SENATIK)*, Vol. 4, No. 1, pp. 789-797, 2021.
- [14] S. N. A. S. Normah, "Penerapan Data Mining Metode K-Means Clustering Untuk Analisa Penjualan Pada Toko Fashion Hijab Banten," *Jurnal Khatulistiwa Informatika*, Volume 7, No. 2, pp. 158-163, 2021.
- [15] I. S. D. E. I. Deny Franata Pasaribu and H. S. T. Suhada, "Memanfaatkan Algoritma K-Means Dalam Memetakan Potensi Hasil Produksi Kelapa Sawit PTPN IV Marihat," *BIOS: Jurnal Teknologi Informasi Dan Rekayasa Komputer*, VOL 2 NO 1 pp. 11-20, 2021.
- [16] A. P. W. U. P. Nurroby Wahyu Saputra and Prandanda Anugrah, "Pengembangan Bahan Ajar Data Mining Menggunakan Four-D Model dalam Kerangka Kerja CDIO," *Belantika Pendidikan*, Vol. 3, no.2, pp.48-58, 2020.
- [17] T. D. A. M. I. Puji Subekti, "Sistem Penentuan Konsentrasi Jurusan Bagi Mahasiswa Informatika Menggunakan Metode K-Means Di Institut Asia Malang," *Jurnal Manajemen Informatika*, Vol. 12 No. 1, pp. 25-39, 2022.
- [18] I. W. G. A. B. Muhammad Farid Iqbal Al-Rizki, "Prediksi Penyebaran Penyakit TBC dengan Metode K-Means Clustering Menggunakan Aplikasi RapidMiner," *JTERA (Jurnal Teknologi Rekayasa)*, Vol. 5, No. 1 pp. 1-10-, 2020.