

# Penerapan Metode *K-Means* dalam Penjualan Produk Souq.Com

Fadli Amin<sup>1</sup>, Dini Sri Anggraeni<sup>2</sup>, Qurrotul Aini<sup>3</sup>

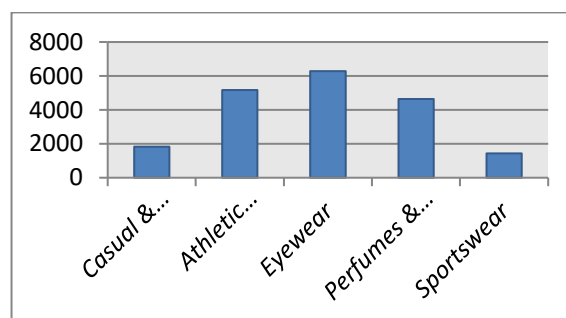
**Abstrak**—Banyak produk yang dijual oleh perusahaan Souq.com. Perusahaan ini mengalami penumpukan di beberapa produknya karena terlalu banyak menyediakan produk yang kurang diminati oleh pelanggan, sehingga banyak produk yang tidak terjual. Perusahaan perlu melakukan pengelompokan untuk setiap produk untuk mengetahui barang mana yang paling diminati dan kurang diminati oleh pelanggan. Penelitian ini bertujuan menentukan tiga kelompok *cluster* dengan kemiripan produknya agar dijadikan rekomendasi bagi manajemen perusahaan dalam merencanakan stok barang. Pada penelitian ditentukan 3 *cluster* dengan *cluster* 1 merupakan produk yang paling laris, *cluster* 2 produk yang laris dan yang *cluster* 3 produk yang kurang laris. Penelitian ini mengambil 5 kategori produk dengan masing-masing kategori ada 10 jenis produk. Hasil perhitungan dengan menggunakan Ms. Excel menunjukkan bahwa *cluster* 1 mempunyai 18 anggota, untuk *cluster* 2 mempunyai 6 anggota dan *cluster* 3 mempunyai 26 anggota. Dari perhitungan dengan Rapidminer didapat bahwa anggota pada *cluster* 1 sebanyak 22 anggota, *cluster* 2 sebanyak 5 anggota dan *cluster* 3 sebanyak 23 anggota. Evaluasi *cluster* dengan DBI didapatkan hasil yang cukup baik dengan nilai 0,431. Pengukuran akurasi, *recall* dan *precision* hasil perhitungan Ms.Excel mendapatkan nilai masing-masing 62%, 67% dan 59%. Untuk perhitungan menggunakan Rapidminer mendapatkan nilai akurasi 64%, *recall* 81% dan *precision* 88%. Hasil perbandingan *clustering* membuktikan perhitungan menggunakan Rapidminer mendapatkan nilai yang *accuracy*, *recall* dan *precision* yang lebih tinggi.

**Kata Kunci**— Penjualan, Clustering, Metode *K-Means*, Souq

## I. PENDAHULUAN

Souq.com merupakan salah satu *platform e-commerce* yang dimiliki oleh Amazon. Souq.com sekarang merupakan anak perusahaan dari Amazon, dan bertindak sebagai tangan Amazon untuk wilayah Timur Tengah. Souq.com menjual berbagai macam pakaian, tas, alat-alat elektronik, sepatu, aksesoris, dan produk-produk terkait lainnya untuk pria, wanita sampai anak-anak. Perusahaan ini awalnya merupakan situs lelang *online* yang kemudian beralih ke pasar *e-commerce* layaknya Amazon pada tahun 2011, sedangkan Souq.com sendiri berdiri pada tahun 2005.

Souq.com merupakan *platform e-commerce* terbesar di Timur tengah, tetapi meskipun begitu dengan banyaknya *e-commerce* yang semakin banyak dan menjual produk yang sama Souq.com memerlukan rekomendasi barang mana yang perlu mereka stok dengan jumlah banyak dan sedikit, dengan mengambil data dari penjualan produk sebelumnya. Masalah yang sering dihadapi yaitu terlalu banyak menyediakan produk yang kurang diminati oleh pelanggan sehingga banyak produk yang menumpuk dan tidak terjual. Pada akhirnya mereka harus menjual produk tersebut dengan memberikan harga promosi agar tidak mengalami kerugian yang besar.



Gambar 1. Penjualan Barang Souq.com

Penelitian sejenis dilakukan pada Toko swalayan MM.TIKA [1] yang menggunakan metode *k-means* dan membuat dua *clustering* antara barang laku dan tidak laku. Hasil penelitian tersebut mengungkapkan bahwa produk yang laku adalah makanan dan minuman, sedangkan produk yang tidak laku yaitu kosmetik. Hasil dari penelitian [2] didapat pengelompokan hingga iterasi ke-5, dimana titik pusat tidak lagi berubah dan tidak ada data yang berpindah antar *cluster*. Hasil pengelompokan dibagi menjadi 3 *cluster*. *Cluster* pertama merupakan barang paling laris ( $C_1$ ) sebanyak 8 barang, barang yang laris ( $C_2$ ) sebanyak 26 barang dan yang terakhir barang kurang laris ( $C_3$ ) sebanyak 16 barang.

Pada penelitian [3] Penerapan metode *clustering* bisa menentukan pembelian dari stok barang yang memang dibutuhkan dengan segera, dari riset yang dilakukan didapatkan hasil bahwa kelompok barang yang laris adalah 10 item sehingga pembelian stok barang diutamakan pada 10 item tersebut. Hasil dari penelitian [4] didapatkan kesimpulan

<sup>2</sup>D. S. Anggraeni, mahasiswa Prodi Sistem Informasi UIN Syarif Hidayatullah Jakarta (e-mail: [dini.srianggraeni17@mhs.uinjkt.ac.id](mailto:dini.srianggraeni17@mhs.uinjkt.ac.id))

<sup>3</sup>Q. Aini, Prodi Sistem Informasi UIN Syarif Hidayatullah Jakarta (e-mail: [qurrotul.aini@uinjkt.ac.id](mailto:qurrotul.aini@uinjkt.ac.id))

bahwa sekolah yang asal mulanya SMA (Sekolah Menengah Pertama) maka pada umumnya jurusan yang diambil ialah Sistem Informasi dan jika sekolah asalnya adalah SMK maka jurusan yang umumnya diambil adalah Teknik Informatika.

Tujuan dari penelitian ini adalah untuk mengetahui *cluster* dari produk-produk yang dijual di Souq.com, sehingga data yang didapatkan dapat dijadikan sebagai rekomendasi bagi manajemen dalam merencanakan stok produk agar toko tidak mengecewakan pelanggan karena barang yang ingin dibeli tidak tersedia. Dari latar belakang yang telah dijelaskan sebelumnya, maka dapat dirumuskan permasalahan pada penelitian ini adalah bagaimana mengimplementasikan metode *k-means* dalam *clustering* produk penjualan pada Souq.com dengan *cluster* produk tersebut kedalam tiga kelompok yaitu yang paling laris, cukup laris dan kurang laris.

## II. KAJIAN PUSTAKA

### A. Data Mining

*Data mining* atau pengembangan data bisa diartikan sebagai metode pemilihan, pencarian dan pemodelan dari data yang jumlahnya besar untuk menciptakan pola atau kecenderungan yang umumnya tidak dirasakan kehadirannya [6]. Data yang dibuat sangat bermacam-macam, hampir disemua sektor, seperti data di sektor pemerintahan, sektor pendidikan, sektor keuangan, dan pada sektor lainnya lagi. Data yang bermacam-macam ini kemudian dikembangkan sehingga membagikan pemahaman yang aktual. *Data mining* merupakan pengamatan pada data dalam mendapatkan ikatan yang terbukti serta merumuskan yang sebelumnya tidak diketahui dengan cara terbaru disimpulkan dan bermanfaat untuk pemilik data tersebut [7], [8].

Secara garis besar, data mining dapat dibagi menjadi 2 kelompok utama, yaitu: [1]

- 1) *Descriptive mining*, yaitu: metode yang digunakan dalam menemukan ciri-ciri penting dari data dalam satu basis data. Ada beberapa Teknik dalam data mining termasuk *descriptive mining* adalah *clustering*, *association*, dan *sequential mining*.
- 2) *Predictive mining*, yaitu: metode untuk mendapatkan model dari data dengan beberapa variabel lain di masa yang akan datang. Salah satu proses yang ada dalam *predictive mining* adalah klasifikasi.

### B. Clustering

*Clustering* merupakan suatu proses pengelompokan *record*, observasi, atau mengelompokkan kelas yang memiliki kesamaan objek. *Clustering* sering digunakan sebagai tahap pertama dalam metode data mining. Ada banyak algoritma *clustering* yang telah dimanfaatkan oleh peneliti sebelumnya seperti *K-Means*, *Improved K-Means*, *Fuzzy C-Means*, *DBSCAN*, *K-Medoids (PAM)*, *CLARANS* dan *Fuzzy Subtractive* [9].

Tujuan dari adanya data *clustering* ini adalah meminimumkan *objektif function* yang di-*set* ke dalam metode *clustering*, yang biasanya berupaya meminimumkan keragaman antar *cluster* [1]. Analisis pengelompokan atau biasa disebut *clustering* adalah cara memisahkan data dalam suatu gabungan ke dalam beberapa grup yang datanya sama

dalam suatu kelompok lebih besar daripada kesamaan data tersebut dengan data dalam kelompok lain. Potensi *clustering* yaitu bisa dimanfaatkan dalam mengetahui bentuk dalam data yang bisa digunakan lebih lanjut dalam berbagai aplikasi secara umum seperti klasifikasi, pengerjaan gambar, dan pemahaman pola [10]. Penerapan metode *clustering* mendapatkan pemahaman berupa penetapan beberapa *cluster* catatan data yang mempunyai kemiripan atribut [11].

### C. Metode K-Means

Algoritma *K-means Clustering* merupakan suatu proses pengkajian data atau metode *data mining* yang menggunakan model tanpa supervisi (*unsupervised*) dan merupakan satu dari banyak model yang melakukan penggabungan data dengan sistem partisi [2]. Metode *K-Means* berupaya menggabungkan data yang ada kedalam beberapa kategori, dimana data dalam satu kategori memiliki ciri yang sama satu dengan yang lainnya dan memiliki ciri yang berbeda dengan data yang ada di dalam kategori yang lain [4].

Dalam metode *k-means*, kita perlu menentukan jumlah *cluster k*. Pada titik itu, fokus pengelompokan *k* dipilih secara acak [12].

Dasar algoritma *K-means* adalah [4]:

- 1) Tentukan nilai *k* menjadi jumlah *cluster* yang diperlukan
- 2) Inisialisasi *k* menjadi *centroid* yang dapat dibangkitkan secara random.
- 3) Hitung jarak data ke masing-masing *centroid* menggunakan persamaan *Euclidean Distance* sebagai berikut:
 
$$D(i, j) = \sqrt{(x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2} \quad (1)$$
- 4) Gabungkan setiap data berdasarkan jarak terdekat antara data dengan *centroid*-nya.
- 5) Tentukan posisi *centroid* baru (*k*)
- 6) Kembali ke langkah 3 jika posisi *centroid* baru dengan *centroid* lama tidak sama.

Pengelompokan *k-means* digunakan sebagai metode pengelompokan dengan alasan bahwa algoritma *K-means* dapat menangani sejumlah data set besar secara efektif. Algoritma *k-means* adalah sejenis metode pengelompokan berdasarkan partisi, yang biasanya menilai kesamaan dengan menghitung jarak. Inti dari algoritma *K-means* adalah memilih titik pusat *k* secara acak dan mempartisi data sesuai dengan jarak antara data dan titik tengah *k* [13]-[15].

## III. METODE PENELITIAN

Metode penelitian menjelaskan langkah yang digunakan dalam melakukan implementasi *clustering k-means* untuk mengelompokkan data penjualan barang menjadi tiga *cluster*. Tahapan metodologi penelitian ditunjukkan pada Gambar 2.

Penelitian ini menggunakan jenis data kuantitatif, yaitu: berupa angka atau nominal data yang dapat dihitung. Dalam penelitian ini menggunakan *data set* historis penjualan dan merupakan jenis data kuantitatif karena berupa angka atau nominal yang dapat dihitung. Sumber data yang digunakan merupakan data sekunder karena data yang didapat tidak secara langsung dan melalui media perantara yang telah dicatat dan diperoleh oleh pihak lain. Data yang dibutuhkan dalam penelitian ini adalah data penjualan sebelumnya, pengumpulan

data dalam penelitian ini berupa *data set* yang didapat dari *website kaggle.com* yang kemudian dianalisis dan diterapkan untuk mendapatkan hasil kesimpulan penjualan pada *souq.com*.



Gambar 2. Tahapan Metode Penelitian

#### IV. HASIL

Data penelitian ini didapat dari situs *website Kaggle.com*, dimana data yang digunakan sebanyak 50 Data. dimana terdapat 10 produk dari setiap kategori yang ada, *output* yang diinginkan adalah mendapatkan hasil 3 *cluster* yaitu data penjualan dari yang paling laris ( $C_1$ ), data penjualan yang laris ( $C_2$ ) dan data penjualan kurang laris ( $C_3$ ). Variabel atau atribut yang dipakai untuk pengelompokan penjualan ini terdiri atas harga produk, dan jumlah produk yang terjual, seperti pada Tabel 1.

Tabel 1. Sampel Data

No	Kategori Produk	Nama Produk	Harga produk (SAR)	Jumlah Barang Terjual
1.	Casual & Dress Shoes	Lacoste Carnaby Lace Up Shoes For Men Size 42 EU, White	295,98	66
2.	Casual & Dress Shoes	Geox U KAVEN Fashion Sneakers For Men , size 43 EU , Black	215,48	67
3.	Casual & Dress Shoes	Hugo Boss Casual Shoe for Men Size 42 EU - Dark Navy	545,99	71
4.	Casual & Dress Shoes	Polo Ralph Lauren Fashion Sneakers For Men Size 44 EU, Black	222,48	80
5.	Casual & Dress Shoes	Lacoste Casual Shoe For Women Size 39 EU - White	225,98	71

6.	Casual & Dress Shoes	Lacoste Casual Shoe for Men Size 44 EU - White	254,49	83
7.	Casual & Dress Shoes	Lacoste Casual Shoe for Men Size 42 EU - Black	309,98	87
8.	Casual & Dress Shoes	Puma Fashion Sneakers Casual Shoe For Men, Navy, 43 EU	187,49	87
9.	Casual & Dress Shoes	Geox U FEDERICO Fashion Sneakers For Men , size 44 EU , Black	250,98	88
10.	Casual & Dress Shoes	Jack and Jones Shoes for Men - Navy - 44 EU	90,49	94
11.	Athletic Shoes	adidas ASWEERUN Sneaker for Mens, Grey - Size 43.3 EU	194,99	94
12.	Athletic Shoes	Nike SB Check Solar Cnvs Skateboarding Shoes for Men	211,48	94
13.	Athletic Shoes	Nike Tanjun Running Shoe For Men (Black - 42 EU)	259,98	92
14.	Athletic Shoes	Nike Court Royale Sport Sneakers for Women - White/Metallic Silver Size - 38 EU	234,99	92
15.	Athletic Shoes	Nike Sb Check Solar Cnvs Sneaker For Men (Black - 43 EU)	211,48	91
16.	Athletic Shoes	Adidas Asweerun Running Shoes For Men - Core Black	194,99	91
17.	Athletic Shoes	Ellesse Classic Shoe For Women (Black - 39 EU)	144,98	90
18.	Athletic Shoes	Adidas STRUTTER Sneaker for Men , Size 42.7 EU, White	214,99	90
19.	Athletic Shoes	Nike Fitness Shoes For Men , 42.5 EU , Black	211,48	87
20.	Athletic Shoes	Adidas VS Pace Nubuck Contrast Side Stripe Sport Sneakers for Men - White, 43 1/3	186,98	86
21.	Eyewear	Polaroid Sunglasses For Unisex, Black- Pld 2054/S-003 M9, Square Frame	168,99	96
22.	Eyewear	Polaroid Sunglasses for Men, Black- PLD 7005/S 223783YYV63RC	134,99	94
23.	Eyewear	Ray-Ban Aviator Unisex Sunglasses - Rb3025-W0879-58-14-135, Green Lens	302,98	96
24.	Eyewear	Ray-Ban Sunglasses For Men - Grey, 3553I, 59, 6, 8G	282,99	98
25.	Eyewear	Ray-Ban Aviator Men's Sunglasses - Rb 3549-002/T3 - 61-16-145 mm, Grey Lens	453,99	90
26.	Eyewear	Ray-Ban Aviator Sunglasses For Men - Rb3449-003/8G 59, Grey Lens	464,49	99

27.	<i>Eyewear</i>	Ray Ban Sunglasses For Unisex, Rb3025 001/51 58-14, Brown Lens, Aviator Frame	397,97	89	48.	<i>Sportswear</i>	Adidas Design 2 Move Solid Sport T-Shirt For Women - Black, Ds8722, M	77,98	76
28.	<i>Eyewear</i>	Ray-Ban Rectangular Unisex Sunglasses - Rb4165 601/71 54-16, Green Lens, Rectangle Fram	372,49	94	49.	<i>Sportswear</i>	Adidas FL_SPR Z FT 3ST BLACK for MALE	69,98	83
29.	<i>Eyewear</i>	Esprit Women's Sunglasses - Brown Frame, Brown Gradient Lens- ET19461/535, size 56-16-142	143,99	92	50.	<i>Sportswear</i>	Adidas Must Haves 3-Stripes Sport T-Shirt for Men - Black & White	69,98	88
30.	<i>Eyewear</i>	Carrera Aviator Sunglasses for Unisex , 8010/S-R80-59-12-140-WJ	228,98	89					
31.	<i>Perfumes &amp; Fragrances</i>	Versace Eros Pour Femme for Women - Eau de Parfum, 100ml	216,48	92					
32.	<i>Perfumes &amp; Fragrances</i>	Jovan Musk for Women - Eau de Cologne, 59ml	21,40	93					
33.	<i>Perfumes &amp; Fragrances</i>	Davidoff Cool Water For Men - Eau de Toilette, 125ml	69,90	95					
34.	<i>Perfumes &amp; Fragrances</i>	Davidoff The Game for Men - Eau de Toilette, 100ml	81,62	95					
35.	<i>Perfumes &amp; Fragrances</i>	Nikos Sculpture Homme for Men - Eau de Toilette, 100ml	49,95	95					
36.	<i>Perfumes &amp; Fragrances</i>	Oud Sharqia for Unisex - Eau de Parfum, 80ml	22,00	96					
37.	<i>Perfumes &amp; Fragrances</i>	Calvin Klein Beauty for Women - Eau de Parfum, 100ml	84,98	97					
38.	<i>Perfumes &amp; Fragrances</i>	Eternity by Calvin Klein for Men - Eau de Toilette, 100ml	83,90	97					
39.	<i>Perfumes &amp; Fragrances</i>	Burberry London for Women - Eau de Parfum, 100ml	127,97	98					
40.	<i>Perfumes &amp; Fragrances</i>	Mancera Roses Vanille for Unisex - Eau de Parfum, 120ml	197,98	99					
41.	<i>Sportswear</i>	Under Armour Rival Fleece Training Sports Pants for Men Black, L	176,48	66					
42.	<i>Sportswear</i>	Adidas Essentials Linear Sport Tank Top For Women - Black White, M	51,99	66					
43.	<i>Sportswear</i>	adidas Essentials Plain Sport T-shirt for Men - White	51,99	69					
44.	<i>Sportswear</i>	Adidas D2M HR 34 T Sport Tight for Women - Black	90,49	75					
45.	<i>Sportswear</i>	Bodytalk Training T-shirt for Men, Size M, Cotton - White	38,48	75					
46.	<i>Sportswear</i>	Nike 23 Alpha Dry Graphic Basketball Shorts For Men - Multi Color, XL	122,99	75					
47.	<i>Sportswear</i>	Under Armour Baseline Fleece Jogger7 Pant For Men, Size M, Brown	172,48	75					

A. Penerapan K-means Clustering

1) Langkah awal dalam memutuskan total cluster yang ingin dibuat. Dalam penelitian ini cluster akan disusun sebanyak 3 cluster. Kemudian pilih pusat pertama dari semua cluster, dimana pusat pertama yang dipilih secara random ditunjukkan pada Tabel 2.

Tabel 2.  
Titik Pusat Awal Cluster

Cluster	Harga produk	Jumlah barang terjual
$C_1$	211,48	94
$C_2$	453,99	90
$C_3$	84,98	97

2) Posisikan semua data/objek ke cluster yang paling dekat dan hitung jarak dari data ke-1 kepada fokus cluster

$$Cluster 1 = \sqrt{(295,98 - 211,48)^2 + (66 - 94)^2} = 89,02$$

$$Cluster 2 = \sqrt{(295,98 - 453,99)^2 + (66 - 90)^2} = 159,82$$

$$Cluster 3 = \sqrt{(295,98 - 84,98)^2 + (66 - 97)^2} = 213,27$$

Jarak dari data ke-2 terhadap pusat cluster

$$Cluster 1 = \sqrt{(215,48 - 211,48)^2 + (67 - 94)^2} = 27,29$$

$$Cluster 2 = \sqrt{(215,48 - 453,99)^2 + (67 - 90)^2} = 239,62$$

$$Cluster 3 = \sqrt{(215,48 - 84,98)^2 + (67 - 97)^2} = 133,90$$

Berdasarkan hasil rekapitulasi jarak dari data pertama kepada inti cluster di atas dapat diambil kesimpulan bahwa jarak data awal paling dekat dengan cluster 1. Hasil perhitungan sepenuhnya untuk 50 data produk dapat dilihat pada Tabel 3.

Tabel 3.  
Hasil Perhitungan Setiap Data pada Iterasi 1

Data Ke-i	Jarak Ke			Cluster
	$C_1$	$C_2$	$C_3$	
1	89,02	159,82	213,27	1
2	27,29	239,62	133,90	1
3	335,30	93,94	461,74	2
4	17,80	231,73	138,55	1
5	27,19	228,80	143,38	1
6	44,39	199,62	170,09	1
7	98,75	144,04	225,22	1
8	24,99	266,52	103	1

9	39,95	203,02	166,24	1
10	120,99	363,52	6,27	3
11	16,49	259,03	110,05	1
12	0	242,54	126,54	1
13	48,54	194,02	175,07	1
14	23,59	219,01	150,09	1
15	3	242,51	126,64	1
16	16,76	259	110,17	1
17	66,62	309,01	60,41	3
18	5,32	239	130,20	1
19	7	242,53	126,89	1
20	24,58	267,04	102,59	1
21	42,54	285,06	84,02	1
22	76,49	319,03	50,10	3
23	91,52	151,13	218	1
24	71,62	171,19	198,01	1
25	242,54	0	369,08	2
26	253,06	13,83	379,52	2
27	186,56	56,03	313,09	2
28	161,01	81,60	287,53	2
29	67,52	310,01	59,22	3
30	18,20	225,01	144,22	1
31	5,39	237,52	131,60	1
32	190,08	432,60	63,71	3
33	141,58	384,12	15,21	3
34	129,86	372,40	3,91	3
35	161,53	404,07	35,09	3
36	22,09	432,03	62,99	1
37	126,54	369,08	0	3
38	127,62	370,16	01,08	3
39	83,61	326,12	43	3
40	14,40	256,17	113,02	1
41	44,82	278,55	96,61	1
42	161,93	402,72	45,27	3
43	161,44	402,55	43,27	3
44	122,47	363,81	22,68	3
45	174,04	415,78	51,44	3
46	90,51	331,34	43,92	3
47	43,38	281,91	90,22	1
48	134,71	376,27	22,14	3
49	141,93	384,07	20,52	3
50	141,63	384,02	17,49	3

3) Setelah semua data ditempatkan ke dalam cluster yang terdekat, kemudian hitung kembali pusat cluster yang baru

berdasarkan rata-rata anggota yang ada pada cluster tersebut. Untuk mencari centroid baru ( $C_k$ ) dihitung dengan persamaan:

$$C_k = \left(\frac{1}{n_k}\right) \sum d_1 \tag{2}$$

Berdasarkan Tabel 3, titik pusat cluster baru untuk parameter stok awal memiliki 26 anggota:

$$C_1 = (295,98 + 215,48 + 222,48 + 225,98 + 254,49 + 309,98 + 187,49 + 250,98 + 194,99 + 211,48 + 259,98 + 234,99 + 211,48 + 194,99 + 214,99 + 211,48 + 186,98 + 168,99 + 302,98 + 282,99 + 228,98 + 216,48 + 22 + 197,98 + 176,48 + 172,48) / 26 = 22$$

$$C_1 = (66 + 67 + 80 + 71 + 83 + 87 + 87 + 88 + 94 + 92 + 92 + 91 + 91 + 90 + 87 + 86 + 96 + 96 + 98 + 89 + 92 + 96 + 99 + 66 + 75) / 26 = 86,65$$

Cluster kedua memiliki 5 anggota:

$$C_2 = (545,99 + 453,99 + 464,49 + 397,97 + 372,49) / 5 = 437,99$$

$$C_2 = (71 + 90 + 99 + 89 + 94) / 5 = 88,6$$

Selanjutnya untuk cluster ketiga memiliki 18 anggota:

$$C_3 = (90,49 + 144,98 + 134,99 + 143,99 + 21,4 + 69,9 + 81,62 + 49,95 + 84,98 + 83,9 + 127,97 + 51,99 + 90,49 + 38,48 + 122,99 + 77,98 + 69,98 + 69,98) / 18 = 82,54$$

$$C_3 = (94 + 90 + 94 + 92 + 93 + 95 + 95 + 95 + 97 + 97 + 98 + 66 + 75 + 75 + 75 + 76 + 83 + 88) / 18 = 87,67$$

Ulangi perhitungan cluster baru untuk parameter terjual, sehingga hasil dari rekapitulasi didapatkan titik pusat baru dari setiap cluster seperti Tabel 4.

Tabel 4.  
Titik Pusat Baru Cluster

Cluster	Harga Produk	Jumlah Barang Terjual
$C_1$	22	86.65
$C_2$	437.99	88.6
$C_3$	82.54	87.67

4) Setelah ditemukan titik pusat yang baru tiap cluster, ulang kembali dari tahap sebelumnya sampai titik pusat cluster tidak berganti lagi dan datanya tidak lagi berpindah dari satu cluster ke cluster lain.

Berdasarkan hasil penggabungan dari keseluruhan data memanfaatkan proses *k-means clustering*, hasilnya adalah bahwa akhir pengelompokan berada di iterasi ke-6, dimana pusat cluster tidak berubah lagi dan tidak adanya data yang berpindah dari cluster satu ke cluster lainnya. Hasil cluster yang didapatkan dari tiap data tersebut terdapat pada Tabel 5.

Tabel 5.  
Hasil Akhir Pengelompokan

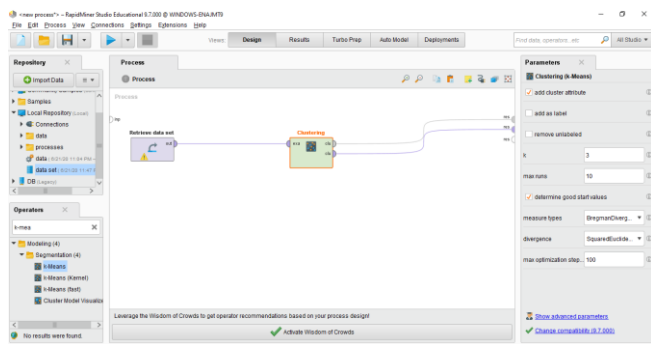
Nama Cluster	Data Barang
Cluster 1	10, 22, 32, 33, 34, 35, 36, 37, 38, 39, 42, 43, 44, 45, 46, 48, 49, 50
Cluster 2	3, 7, 25, 26, 27, 28
Cluster 3	1, 2, 4, 5, 6, 8, 9, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 23, 24, 29, 30, 31, 40, 41, 47

Berdasarkan temuan *cluster* yang terbentuk pada tabel 4 dapat dilihat bahwa hasil penggabungan dari *cluster* dapat dilihat sebagai berikut.

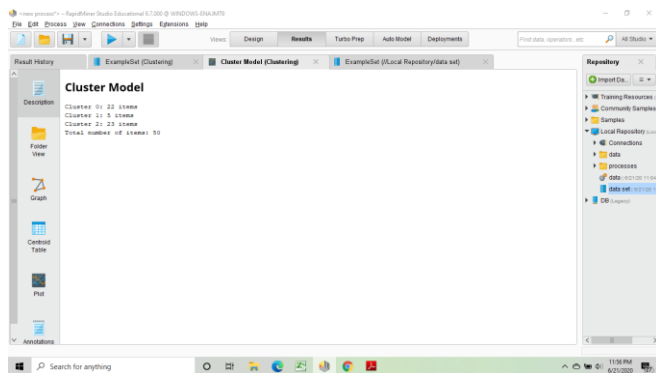
Barang paling laris ( $C_1$ ) = 18 Barang  
 Barang yang laris ( $C_2$ ) = 6 barang  
 Barang kurang laris ( $C_3$ ) = 26 barang

**B. Implementasi Tool Rapid Miner**

Pada tahap pengetesan algoritma ini untuk menunjukkan bukti pada tahap analisis sebelumnya dan pengetesan secara manual, maka dari itu perlu dilakukan pengetesan lagi untuk mengkategorikan data untuk menaikkan jumlah penjualan menggunakan algoritma *k-means*.



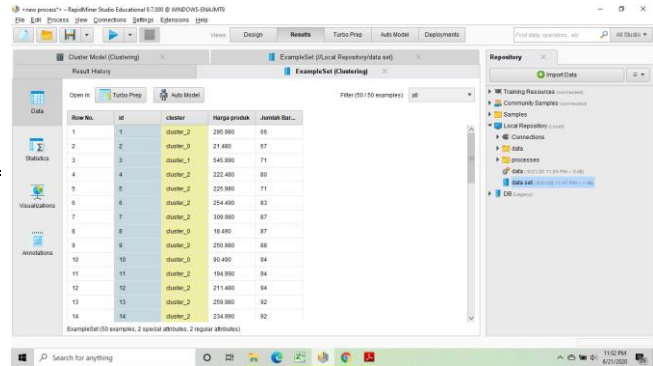
Gambar 3. Penggunaan Algoritma K-Means dengan nilai K=3



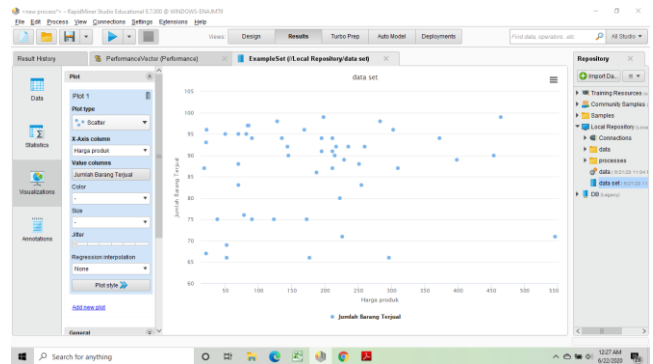
Gambar 4. Hasil Pengelompokan

Hasil *cluster* pada Rapidminer diawali dengan *cluster* 0 (nol), tetapi untuk teori *clustering* diawali menggunakan *cluster* 1. Jadi dalam *cluster* 0 pada Rapidminer sama dengan *cluster* 1 ( $C_1$ ) pada perhitungan skema. Hasil yang diperoleh

dari perhitungan menggunakan *Rapidminer* seperti dalam Tabel 6.



Gambar 5. Data dalam setiap Cluster

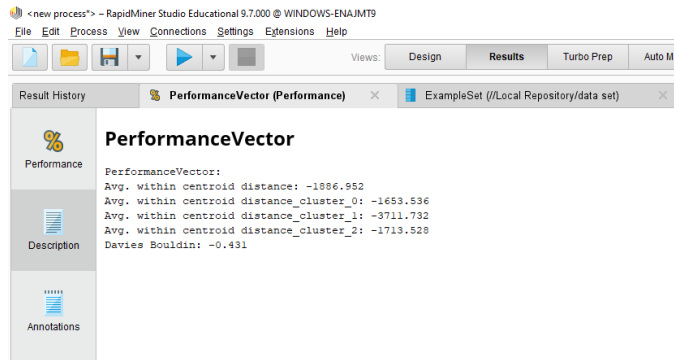


Gambar 6. Hasil Grafik Rapidminer

Tabel 6.  
Hasil Pengelompokan dengan RapidMiner

Nama Cluster	Data Barang	Jumlah Data Barang
Cluster 1	2, 8, 10, 17, 22, 29, 32, 33, 34, 35, 36, 37, 38, 39, 42, 43, 44, 45, 46, 48, 49, 50	22 Barang
Cluster 2	3, 25, 26, 27, 28	5 Barang
Cluster 3	1, 4, 5, 6, 7, 9, 11, 12, 13, 14, 15, 16, 18, 19, 20, 21, 23, 24, 30, 31, 40, 41, 47	23 Barang

Untuk mengetahui performa dari metode *K-means* dalam mengelompokan data penjualan produk-produk Souq.com dilakukan pengetesan *performance* pada tools *Rapidminer*, dengan memanfaatkan *Cluster Distance Performance*, sehingga hasil yang didapat sebagai berikut:



Gambar 1. Hasil Evaluasi Cluster

Avg. within centroid distance adalah rata-rata dalam jarak cluster yang dihitung dengan rata-rata jarak antara centroid dan semua contoh cluster. Avg. within centroid distance yang diperoleh adalah -1886.952. Sedangkan untuk evaluasi cluster digunakan metode Davies-Bouldin Index, dimana semakin kecil nilai Davies Bouldin Index (DBI) yang diperoleh (non negatif >=0), maka semakin baik pula nilai cluster yang didapatkan dari pengelompokan dengan algoritma clustering. Nilai Davies Bouldin Index yang diperoleh adalah 0.431.

### C. Evaluasi Confusion Matrix

Confusion matrix merupakan evaluasi performa dari algoritma yang dihasilkan.

Tabel 7.  
Confusion matrix

Nilai prediksi	Nilai Aktual	
	TP	FN
	FP	TN

#### 1) Accuracy

Digunakan untuk mengetahui akurasi dari hasil perhitungan algoritma. Berikut persamaan yang digunakan:

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \times 100\% \quad (3)$$

Berikut akurasi dari hasil perhitungan algoritma yang menggunakan Ms. Excel (a) dan Rapidminer (b).

$$a = \frac{16 + 15}{50} \times 100\% = 62\%$$

$$b = \frac{22 + 20}{50} \times 100\% = 64\%$$

#### 2) Recall

Recall digunakan dalam mengetahui tingkat kesesuaian prediksi dan data sebenarnya.

$$Recall = \frac{TP}{TP + FN} \times 100\% \quad (4)$$

Berikut hasil perhitungan recall.

$$a = \frac{16}{24} \times 100\% = 67\%$$

$$b = \frac{22}{27} \times 100\% = 81\%$$

#### 3) Precision

Precision digunakan untuk mengetahui kesesuaian rasio prediksi dan keseluruhan prediksi positif.

$$Precision = \frac{TP}{TP + FP} 100\% \quad (5)$$

Berikut hasil perhitungan Precision.

$$a = \frac{16}{27} \times 100\% = 59\%$$

$$b = \frac{22}{25} \times 100\% = 88\%$$

## V. KESIMPULAN

Dengan adanya pengelompokan data ini, bagian organisasi dapat memahami produk paling laris, laris dan tidak laris. Sehingga tidak akan ada lagi penumpukan barang di gudang. Output yang dihasilkan dari penelitian ini yaitu, barang yang paling laris terdapat 18 produk, barang yang laris terdapat 6 produk dan barang kurang laris terdapat 26 produk. Hasil implementasi data menggunakan tools Rapidminer didapat bahwa barang paling laris terdapat 22 produk, barang yang laris terdapat 5 produk, dan barang yang kurang laris terdapat 23 produk. Dalam evaluasi cluster yang dilakukan didapatkan nilai Davies Bouldin Index (DBI) yaitu 0,431, sehingga bisa dikategorikan cluster yang didapatkan sudah cukup baik karena kemiripan antar anggota cluster berada pada >=0. Peneliti juga menggunakan evaluasi confusion matrix dalam mengukur accuracy, recall dan precision. Untuk perhitungan menggunakan Ms.Excel didapatkan hasil akurasi sebanyak 62%, recall 67% dan precision 59%. Sedangkan untuk perhitungan menggunakan Rapidminer hasil akurasi yang didapatkan sebanyak 64%, recall 81% dan precision 88%. Dari hasil yang didapatkan maka dapat ditarik kesimpulan bahwa perhitungan menggunakan Rapidminer memiliki nilai akurasi, recall dan precision yang lebih baik dibandingkan dengan perhitungan menggunakan Ms.Excel sehingga peneliti merekomendasikan untuk menggunakan tools Rapidminer dalam menghindari kesalahan perhitungan.

Penelitian ini memiliki beberapa batasan diantaranya data yang didapat adalah data sekunder yang sudah dipublikasikan di internet, sampel data yang dianalisis pada penelitian ini sebanyak 50 produk dan cluster yang digunakan hanya 3. Untuk penelitian selanjutnya dapat menambahkan parameter atau variabel lainnya dalam metode clustering, sampel yang digunakan bisa lebih banyak agar hasilnya lebih representative. Pada penelitian selanjutnya dapat dilanjutkan dengan membandingkan dengan metode K-medoids dan Fuzzy C-means untuk pengelompokan atau klasifikasi data.

## REFERENSI

- [1] Y. Darmi and A. Setiawan, "Penerapan Metode Clustering K-Means Dalam Pengelompokan Penjualan Produk," *J. Media Infotama*, vol. 12, no. 2, pp. 148–157, 2016.
- [2] S. P. Tamba, F. T. Kesuma, And Feryanto, "Penerapan Data Mining Untuk Menentukan Penjualan Sparepart Toyota Dengan Metode K-Means Clustering," *J. Sist. Inf. Ilmu Komput. Prima (Jusikom Prima)*, vol. 2, no. 2, pp. 67–72, 2019.
- [3] M. H. Siregar, "Data Mining Klasterisasi Penjualan Alat-Alat Bangunan Menggunakan Metode K-Means (Studi Kasus Di Toko Adi Bangunan)," *J. Teknol. Dan Open Source*, vol. 1, no. 2, pp. 83–91, 2018, doi: 10.36378/Jtos.V1i2.24.
- [4] F. Nasari And S. Darma, "Penerapan K-Means Clustering Pada Data Penerimaan Mahasiswa Baru (Studi Kasus : Universitas Potensi Utama)," *Semin. Nas. Teknol. Inf. Dan Multimed.*, vol. 2, no. 1, pp. 73–78, 2015.
- [5] F. Indriyani And E. Irfiani, "Clustering Data Penjualan Pada Toko Perlengkapan Outdoor Menggunakan Metode K-Means," *Juita J. Inform.*, vol. 7, no. 2, pp. 109–114, 2019, doi: 10.30595/Juita.V7i2.5529.
- [6] N. A. Hasibuan *Et Al.*, "Implementasi Data Mining Untuk Pengaturan Layout Minimarket Dengan Menerapkan Association Rule," *J. Ris. Komput.*, Vol. 4, No. 4, Pp. 6–11, 2017, doi: 10.30865/Jurikom.V4i4.686.
- [7] M. Anjelita, A. P. Windarto, And A. Wanto, "Analisis Metode K-Means Pada Kasus Ekspor Barang Perhiasan Dan Barang Berharga Berdasarkan Negara Tujuan," *Semin. Nas. Sains Teknol. Inf.*, vol. 2, no. 8, pp. 476–482, 2019.
- [8] C. Wirawan, "Teknik Data Mining Menggunakan Algoritma Decision Tree C4.5 untuk Memprediksi Tingkat Kelulusan Tepat Waktu," *Applied Information System and Management (AISM)*, vol. 3, no. 1, pp. 43-52, May 2020. doi: 10.15408/aism.v3i1.13033
- [9] D. F. Pramesti, Lahan, M. Tanzil Furqon, And C. Dewi, "Implementasi Metode K-Medoids Clustering Untuk Pengelompokan Data Potensi Kebakaran Hutan/Lahan Berdasarkan Persebaran Titik Panas (Hotspot)," *J. Pengemb. Teknol. Inf. Dan Ilmu Komput.*, vol. 1, no. 9, pp. 723–732, 2017, doi: 10.1109/Eumc.2008.4751704.
- [10] R. W. Sari, A. Wanto, And A. P. Windarto, "Implementasi Rapidminer Dengan Metode K-Means (Study Kasus: Imunisasi Campak Pada Balita Berdasarkan Provinsi)," *Komik (Konferensi Nas. Teknol. Inf. Dan Komputer)*, vol. 2, no. 1, pp. 224–230, 2018, Doi: 10.30865/Komik.V2i1.930
- [11] Q. Aini and E. Khudzaeva, "Potential Halal Tourism Destinations with Applying K-Means Clustering," *I.J. Intelligent Systems and Applications*, vol. 9, pp. 9-17, 2019.
- [12] N. Arunkumar *Et Al.*, "K-Means Clustering And Neural Network For Object Detecting And Identifying Abnormality Of Brain Tumor," *Soft Comput.*, vol. 23, no. 19, pp. 9083–9096, 2019, doi: 10.1007/S00500-018-3618-7.
- [13] W. Wu And M. Peng, "A Data Mining Approach Combining K-Means Clustering With Bagging Neural Network For Short-Term Wind Power Forecasting," *J. Internet Things*, vol. 4662, Pp. 1–8, 2017, doi: 10.1109/Jiot.2017.2677578.
- [14] A. N. Yusril, I. Larasati, and Q. Aini, "Implementasi Text Mining untuk Advertising dengan Menggunakan Metode K-Means Clustering pada Data Tweets Gojek Indonesia," *SISTEMASI: Jurnal Sistem Informasi*, vol. 9, no. 3, pp. 586-596, Sep 2020.
- [15] W. Afifi, D. R. Nastiti, and Q. Aini, "Clustering K-Means pada Data Ekspor (Studi Kasus: PT. Gaikindo)," *Jurnal SIMETRIS*, vol. 11, no. 1, Apr 2020.