

PENGUNAAN METODE K-MEANS CLUSTERING UNTUK MENENTUKAN PEMINATAN PRODUK BERDASARKAN STOK BARANG PADA TOKO AYYSEE COLLECTION MENGUNAKAN RAPIDMINER

Ach. Tri Hamsyah ^{*1)}, Rokhmatul Insani ²⁾, dan Anisa Dzulkarnain ³⁾

¹⁾ Sistem Informasi, Fakultas Teknologi Informasi dan Bisnis, Institut Teknologi Telkom Surabaya, Jl. Ketintang
No. 156, Surabaya, 60231, Indonesia
hamsyah@student.ittelkom-sby.ac.id

Abstrak

Dalam menentukan stok barang yang sesuai, toko Ayysee Collection memerlukan metode efektif dalam peminatan produk dan pola penjualan. Data yang diperoleh diolah dengan *Cross-Industry Standart Process for Data Mining* (CRISP-DM). Penelitian ini menggunakan RFM sebagai inputan untuk proses K-Means clustering menggunakan Rapidminer. Hasil K-Means clustering mendapatkan tiga kategori produk, yakni kategori produk yang paling diminati 4 produk, cukup diminati 12 produk, dan kurang diminati 8 produk. Penentuan pola penjualan berdasarkan produk yang terjual menggunakan metode Apriori. Dengan syarat minimum *support* 10% dan *confidence* 45% didapatkan 3 kombinasi. Kombinasi pertama dengan *confidence* 63%. Kombinasi kedua dengan *confidence* 55%. Sedangkan Kombinasi ketiga dengan *confidence* 51%. Hasil dari penelitian ini bisa dimanfaatkan agar owner Ayysee Collection bisa mengetahui produk mana yang harus di prioritaskan serta produk mana yang perlu diperbaiki atau dihentikan produksinya. Serta hasil dari pola penjualan bisa bermanfaat untuk owner yang nantinya akan dibuatkan *promo bundling* pada produk yang dijual.

Kata kunci: Apriori, Clustering, CRISP-DM, K-means, Pola penjualan, RFM, Rapidminer.

1. Pendahuluan (Introduction)

Penumpukan barang yang terjadi bisa menimbulkan beberapa kerugian yang dialami oleh pelaku usaha. Salah satu dari kerugian yang dapat dialami yakni terjadi ketidakseimbangan antara jumlah barang yang paling banyak peminatnya konsumen dan jumlah jenis produk yang kurang peminatnya. Contoh permasalahan yang dapat terjadi yaitu barang yang paling banyak peminatnya memiliki stok yang sedikit, tetapi jumlah stok produk yang kurang diminati oleh konsumen tertumpuk banyak sehingga mengakibatkan kerugian pada toko tersebut. Permasalahan seperti ini pernah terjadi pada *online shop* yang mana di toko tersebut melakukan pemesanan yang berlebihan dalam jumlah banyak agar nantinya tidak perlu melakukan pemesanan lagi karena stok sudah ada. Namun hal ini menyebabkan beberapa stok barang banyak yang menumpuk dan tidak habis terjual dikarenakan mereka tidak mempertimbangkan dahulu produk mana yang seharusnya diperbanyak jumlah stoknya dan produk mana yang hanya perlu stok secukupnya (Lutfi and Nilogiri, 2018).

Permasalahan serupa terjadi juga pada toko Ayysee Collection. Hal ini dikarenakan stok barang merupakan sesuatu yang perlu diperhatikan dengan benar apabila memiliki sebuah usaha yang bergerak dibidang penjualan suatu barang. Toko Ayysee Collection adalah toko yang menjual berbagai macam hijab dan beberapa jenis pakaian muslimah. Ketersediaan barang pada toko Ayysee Collection ini merupakan suatu elemen yang sangat penting. Perlu adanya upaya dalam mengelompokkan barang yang paling diminati, cukup diminati, dan kurang diminati. Selain itu, mengetahui pola penjualan suatu produk diperlukan karena nantinya akan menjadi patokan untuk *promo bundling* pada toko Ayysee Collection. Upaya-upaya tersebut dilakukan untuk menghindari penumpukan barang yang kurang diminati oleh pembeli.

Pada penelitian lain terdapat juga yang mengalami permasalahan terkait barang-barang pakaian yang sering diminati pembeli seringkali habis terjual dan kekurangan stok. Maka dari itu, untuk memudahkan dalam mengkategorikan produk pakaian dapat dilakukan dengan mencari berdasarkan

nilai yang didapat dari RFM (*Recency, Frequency, dan Monetary*) untuk digunakan sebagai data set. Lalu dilanjutkan dengan menggunakan metode *K-Means Clustering* sebagai cara untuk mengelompokkan data yang terbagi menjadi 3 kelompok, yakni tertinggi, sedang dan rendah. Hasil yang di dapat dari penelitian ini, penjualan tertinggi ada 2 produk, penjualan sedang ada 5 produk, sedangkan penjualan rendah mendapatkan 15 produk (Firmansah, Dedy Irawan and Vendyansyah, 2021).

Pola penjualan merupakan hal yang penting karena untuk kemajuan bisnis pada suatu usaha. Oleh karena itu, memerlukan proses pengolahan data yang bisa berguna untuk memberi informasi terkait produk mana yang mempunyai potensi di beli secara bersamaan. Untuk mengetahui pola penjualan produk merupakan hal yang cukup penting dikarenakan apabila mengetahui hal tersebut, kita bisa menerapkan strategi untuk menjual produk kedepannya. Penelitian ini menggunakan Algoritma *Apriori* karena metode tersebut cocok digunakan untuk kumpulan item dalam jumlah besar dan sederhana untuk di implementasikan (Syahril, Erwansyah and Yetri, 2020).

Dari penelitian-penelitian sebelumnya, dapat disimpulkan *K-Means* bisa diartikan sebagai metode pengelompokan (*Clustering*) yang mana dalam prosesnya terdapat pembagian data ke dalam beberapa kelompok. Algoritma ini hanya dapat berfungsi untuk atribut *numeric*. Algoritma *K-Means* merupakan algoritma yang paling sering digunakan karena keefektifannya dalam mengelompokkan berbagai macam data, termasuk data yang besar (Aulia, 2021). Setelah dilakukan proses *K-Means* akan dilakukan proses menentukan pola penjualan dengan menggunakan algoritma *Apriori* yang nantinya bisa mengetahui terkait hasil dari pola penjualan produk berdasarkan riwayat stok barang penjualan produk yang ada pada toko *Ayysee Collection*.

Tujuan dari penelitian ini yaitu untuk melakukan pengelompokkan produk berdasarkan 3 kategori, yakni produk yang paling diminati, cukup diminati, dan kurang diminati. Selain itu, penelitian ini juga bertujuan untuk mengetahui pola penjualan terkait produk yang dibeli oleh konsumen. Dalam hal ini, peneliti menggunakan CRISP-DM sebagai proses mengolah data-data tersebut. Upaya ini dilakukan untuk membantu pemilik usaha Toko *Ayysee Collection* dalam mengetahui produk mana yang harus diprioritaskan dari hasil *K-Means Clustering*. Tidak hanya itu, hasil dari pola penjualan yang didapat dengan metode *Apriori* bisa membantu pemilik usaha untuk *promo bundling*.

2. Metode Penelitian (Methods)

2.1 Business Understanding

Business Understanding yakni suatu tahapan yang mana permasalahan terkait bisnis akan dimaknai serta tepat sarasannya. Tahapan Business Understanding di penelitian ini adalah tahapan dimana peneliti berusaha memahami dan belajar terkait masalah-masalah yang ditemukan, lalu mencari solusi untuk melakukan penyelesaian masalah tersebut.

2.2 Data Understanding

Data Understanding memiliki makna yaitu suatu tahapan yang dimana terjadi proses mengumpulkan data. Setelah terjadi proses pengumpulan data, peneliti juga melakukan pemahaman terkait data-data yang sudah dikumpulkan terkait kualitas data tersebut.

2.3 Data Preparation

Data Preparation merupakan suatu tahapan yang mana terjadi proses mengolah data yang dilakukan sebelum diolah di *Data Mining*. Dalam tahapan ini, dilakukannya penyeleksian data yang akan dipakai, pembersihan data agar tidak terdapat adanya duplikasi pada data, serta transformasi data supaya data-data tersebut sudah siap diproses dan dilakukan pengolahan ke tahap selanjutnya.

2.4 Modeling

Modeling merupakan suatu tahapan yang mana peneliti harus menentukan untuk menggunakan algoritma *Data Mining* yang sesuai dengan permasalahan yang ada untuk mengoptimalkan hasil yang diperoleh.

2.5 Evaluation

Evaluation merupakan suatu tahapan yang mana dilaksanakannya suatu evaluasi pada hasil dari data yang diperoleh, dan menentukan apabila model telah meraih tujuan yang sudah ditentukan.

3. Hasil dan Pembahasan (Result And Discussion)

3.1 Business Understanding

tahapan ini, peneliti mengidentifikasi masalah yang didapat dari permasalahan yang terjadi pada toko Ayysee Collection. Tujuan dilakukannya hal ini adalah agar peneliti bisa memperoleh informasi dan pengetahuan terkait permasalahan apa yang dialami oleh Toko Ayysee Collection. Permasalahan yang dialami yaitu Toko Ayysee Collection mengalami kesulitan dalam mengelompokkan produk mana yang masuk kategori paling diminati, cukup diminati, dan kurang diminati. Proses selanjutnya dari tahapan tersebut yakni melakukan pencarian pola penjualan agar dapat difungsikan sebagai sarana mengetahui produk mana yang bisa dijadikan untuk *promo bundling*.

3.2 Data Understanding

tahapan ini peneliti melakukan pengambilan data di toko Ayysee Collection. Data yang digunakan merupakan beberapa kumpulan data, yang nantinya akan menjadi patokan dalam mencari informasi dari data yang ada dan data tersebut akan dipilih untuk dilanjutkan ke tahap selanjutnya. Data – data tersebut bisa dilihat pada Tabel 1.

Tabel 1. *Data Understanding*

No	Atribut	Deskripsi
1	No. Pesanan	Digunakan Untuk mengetahui nomor produk yang di pesan
2	Status Pesanan	Digunakan Untuk mengetahui status pada produk yang di pesan
3	Status Pembatalan/ Pengembalian	Digunakan untuk mengetahui apakah status pembatalan diterima atau tidak
4	No. Resi	Digunakan untuk mengetahui nomor yang diberikan oleh ekspedisi
5	Opsi Pengiriman	Digunakan untuk mengetahui opsi apa yang dipilih dalam pengiriman
6	Antar ke counter/ pick-up	Digunakan untuk mengetahui apakah produk yang di pesan sudah di pickup

No	Atribut	Deskripsi
7	Pesanan Harus Dikirimkan Sebelum (Menghindari keterlambatan)	Digunakan untuk mengetahui apakah produk yang di pesan sudah di kirim
8	Waktu Pengiriman Diatur	Digunakan untuk mengetahui waktu pengiriman yang sudah ditentukan
9	Waktu Pesanan Dibuat	Digunakan untuk mengetahui waktu pesanan yang sudah kita buat
10	Waktu Pembayaran Dilakukan	Digunakan untuk mengetahui waktu pembayaran yang sudah kita lakukan
11	SKU Induk	Digunakan untuk mengidentifikasi suatu produk
12	Nama Produk	Digunakan untuk mengetahui nama produk
13	Nomor Referensi SKU	Digunakan untuk kode unik pada suatu produk
14	Nama Variasi	Digunakan untuk mengetahui variasi pada produk
15	Harga Awal	Digunakan untuk mengetahui harga asli dari produk
16	Harga Setelah Diskon	Digunakan untuk mengetahui harga setelah dapat potongan harga
17	Jumlah	Digunakan untuk mengetahui jumlah produk
18	Total Harga Produk	Digunakan untuk mengetahui total dari produk
19	Total Diskon	Digunakan untuk mengetahui diskon yang didapat
20	Diskon Dari Penjual	Digunakan untuk mengetahui diskon yang didapat dari penjual
21	Diskon Dari Shopee	Digunakan untuk mengetahui diskon yang didapat dari shopee
22	Berat Produk	Digunakan untuk mengetahui berat produk
23	Jumlah Produk di Pesanan	Digunakan untuk mengetahui jumlah yang didapat di pesan
24	Total Berat	Digunakan untuk mengetahui total berat

No	Atribut	Deskripsi
25	Voucher Ditanggung Penjual	Digunakan untuk mengetahui voucher yang didapat ditanggung penjual
26	Cashback Koin	Digunakan untuk mengetahui cashback koin
27	Voucher Ditanggung Shopee	Digunakan untuk mengetahui voucher yang didapat ditanggung shopee
28	Paket Diskon	Digunakan untuk mengetahui paket diskon
29	Paket Diskon (Diskon dari Shopee)	Digunakan untuk mengetahui paket yang didapat dari shopee)
30	Paket Diskon (Diskon dari Penjual)	Digunakan untuk mengetahui paket yang didapat dari penjual)
31	Potongan Koin Shopee	Digunakan untuk mengetahui potongan yang didapat dari koin shopee
32	Diskon Kartu Kredit	Digunakan untuk mengetahui diskon yang didapat jika memakai kartu kredit
33	Ongkos Kirim Dibayar oleh Pembeli	Digunakan untuk mengetahui ongkos yang didapatkan pembeli
34	Estimasi Potongan Biaya Pengiriman	Digunakan untuk mengetahui estimasi yang didapat dari biaya pengiriman
35	Ongkos Kirim Pengembalian Barang	Digunakan untuk mengetahui ongkos yang didapat dari pengembalian barang
36	Total Pembayaran	Digunakan untuk mengetahui jumlah total pembayaran
37	Perkiraan Ongkos Kirim	Digunakan untuk mengetahui perkiraan yang didapat ongkos kirim
38	Catatan dari Pembeli	Digunakan untuk mengetahui catatan yang didapat dari pembeli
39	Catatan	Digunakan untuk mengetahui catatan yang didapat dari pembeli
40	Username (Pembeli)	Digunakan untuk mengetahui username pembeli

No	Atribut	Deskripsi
41	Nama Penerima	Digunakan untuk mengetahui nama penerima
42	No. Telepon	Digunakan untuk mengetahui no telepon pembeli
43	Alamat Pengiriman	Digunakan untuk mengetahui alamat pengiriman
44	Kota/Kabupaten	Digunakan untuk mengetahui tempat tinggal kota pembeli
45	Provinsi	Digunakan untuk mengetahui tempat tinggal provinsi pembeli
46	Waktu Pesanan Selesai	Digunakan untuk mengetahui pesanan selesai di proses

3.3 Data Preparation

a. Seleksi data

tahap ini data yang sudah didapatkan harus disaring dahulu dan diambil atribut yang akan dipakai untuk di proses. Terdapat 4 atribut yang digunakan untuk proses pengelompokan produk menggunakan K-Means *Clustering*, yaitu waktu pesanan dibuat, nama produk, jumlah produk dipesan dan total pembayaran yang nantinya atribut tersebut akan dicari terkait nilai RFM nya. Data diambil berdasarkan atribut yang digunakan seperti Tabel 2.

Tabel 2. Seleksi data

No	Waktu Pesanan Dibuat	Nama Produk	Jumlah Produk di Pesan	Total Pembayaran
1	2022-06-01 01:42	INNER BASIC By Ayysee	4	98.000
2	2022-06-01 01:42	INNER BASIC By Ayysee	4	98.000
3	2022-06-01 01:42	INNER BASIC By Ayysee	4	98.000
4	2022-06-01 04:00	INNER NINJA HATORY By Ayysee	1	28.750
5	2022-06-01 08:21	INNER NINJA HATORY By Ayysee	3	86.400
6	2022-06-01 08:21	INNER NINJA HATORY By Ayysee	3	86.400
7	2022-06-01 08:21	INNER NINJA HATORY By Ayysee	3	86.400

No	Waktu Pesanan Dibuat	Nama Produk	Jumlah Produk di Pesan	Total Pembayaran
8	2022-06-01 09:04	Manset Gamis Lengan Panjang / Long sleeve dress/ Custom Size/ By Ayysee	1	122.313
9	2022-06-01 09:14	INNER NINJA HATORY By Ayysee	1	43.936
10	2022-06-01 10:09	INNER NINJA HATORY By Ayysee	1	37.904
11	2022-06-01 11:39	Manset Baju Lengan Panjang (Long sleeve)/ Custom Size By Ayysee	9	353.496
12	2022-06-01 11:39	Inner Leher By Ayysee	9	353.496
13	2022-06-01 11:39	Inner Leher By Ayysee	9	353.496
14	2022-06-01 11:39	Inner Leher By Ayysee	9	353.496
15	2022-06-01 11:39	Inner Leher By Ayysee	9	353.496
....				
6037	2022-11-30 23:28	Compact Hijab kit / Set Jarum Hijab By Ayysee	4	127.836

b. Pembersihan data

Setelah data diseleksi sesuai dengan atribut yang nantinya akan dilakukan pre-processing data. Pembersihan data dilakukan agar tidak terjadi duplikasi data. Untuk mengetahui apabila ada data yang duplikat menggunakan fitur yang ada pada excel yaitu menggunakan *duplicate value*, selanjutnya untuk mengetahui data tidak terjadi *missing value* menggunakan rumus pada excel countblank. Dari total 6037 data tidak ditemukan terjadinya *duplicate value* dan *missing value* setelah dilakukan proses pembersihan data.

c. Transformasi data

Transformasi data bertujuan agar data yang sudah di proses melalui tahap sebelumnya bisa diolah dengan menggunakan algoritma K-means clustering. Untuk proses ini menggunakan metode RFM (*Recency, Frequency, Monetary*) Nilai RFM didapatkan melalui *Recency*, yang berarti waktu sebuah produk yang terjual dikurangi dengan tanggal periode. *Frequency* didapatkan dari total berapa kali produk itu dibeli. *Monetary* didapatkan dari total harga terhadap sebuah produk yang dibeli. Berikut pada Tabel 3 merupakan pengkodean produk dan hasil *transformasi data* bisa dilihat pada Tabel 4.

Tabel 3 Pengkodean produk

No produk	Nama Produk
1	INNER BASIC By Ayysee
2	INNER NINJA HATORY By Ayysee
3	Manset Gamis Lengan Panjang / Long sleeve dress/ Custom Size/ By Ayysee
4	Manset Baju Lengan Panjang (Long sleeve)/ Custom Size By Ayysee
5	Inner Leher By Ayysee
6	Manset Gamis Tanpa Lengan (Inner Dress Sleeveless)/ Custom Size By Ayysee
7	Compact Hijab kit / Set Jarum Hijab By Ayysee
8	Inner Skirt (Custom Size) By Ayysee
9	Manset Baju Knit Inner Knit Manset Rajut Turtle Neck Knit / Baju Atasan Rajut
10	HIJAB INSTAN WIYATI By Ayysee
11	Voal Meera Voal Lasercut Square Voal Basic By Ayysee
12	Scrunchie POP By Ayysee
13	Inner Silang By Ayysee
14	Instant Fleta Hijab Instan Lengan By Ayysee
15	Crochet Vest Knit Vest Korean Crochet Vest
16	ROK PLISKET PREMIUM / ROK MAYUNG PLISKET / GRADE A / Pleated Skirt By Ayysee
17	Manset Baju Tanpa Lengan (Sleeveless)/ Custom Size By Ayysee
18	Penelope Skirt By Ayysee
19	KULOT PLISKET PREMIUM Pleated Kulot By Ayysee
20	Lily Vest (Regular Size) By Ayysee
21	Premium Textured Silk Scarf By Ayysee

No produk	Nama Produk
22	Pashmina Sophia (Pashmina Ceruty Premium) 180x75Cm
23	Lily Vest (Big Size) By Ayysee
24	Ella Shirt (Off White & Bishop Sleeve) By Ayysee

Tabel 4 *Transformasi data*

No produk	Recency	Frequency	Monetary
1	0	4144	117065879
2	0	8358	255767989
3	1	669	74473190
4	0	272	15633959
5	0	2262	59196517
6	4	133	14748330
7	0	498	19321710
8	5	99	6705921
9	1	62	4167990
10	67	40	1833515
11	3	50	2339752
12	5	201	5792698
13	4	510	14696619
14	50	14	833650
15	114	4	137700
16	175	8	503000
17	11	60	2606088
18	107	20	1328803
19	48	7	437501
20	43	6	367356
21	8	169	8881483
22	11	29	1333439
23	47	1	70720
24	5	6	288271

Setelah mendapatkan nilai RFM selanjutnya dilakukan *Z-score normalization* dengan cara menghitung nilai rata-rata dan deviasi standar. Sehingga pada saat proses menggunakan K-Means Clustering bisa lebih optimal dan tidak ada kendala dikarenakan nilai data tidak berjauhan. Setelah dilakukan normalisasi, data tersebut akan digunakan sebagai data set untuk proses modeling menggunakan K-Means Clustering. Hasil tersebut bisa dilihat pada Tabel 5.

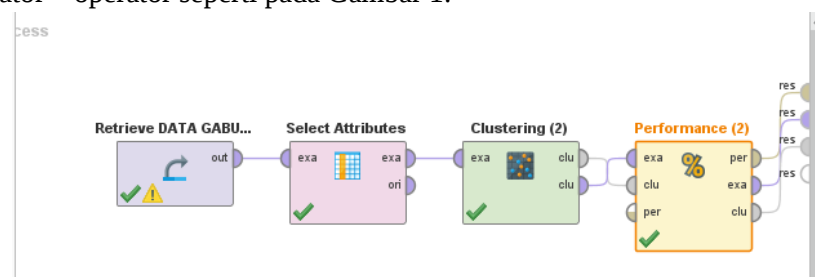
Tabel 5. Normalisasi RFM

No produk	normalisasi R	normalisasi F	normalisasi M
1	0.25870649	0.965949303	0.94705622
2	0.25870649	0.999977382	0.999975718
3	0.26584587	0.486075343	0.806754928
4	0.25870649	0.402330333	0.431953937
5	0.25870649	0.793149947	0.724631167
6	0.287846316	0.373843874	0.425823891
7	0.25870649	0.44970686	0.457639415
8	0.295366285	0.366973098	0.37114848
9	0.26584587	0.359544263	0.354365969
10	0.794123913	0.355151943	0.339173776
11	0.280418007	0.357146108	0.342447142
12	0.295366285	0.387704333	0.365079419
13	0.287846316	0.45224887	0.425466502

No produk	normalisasi R	normalisasi F	normalisasi M
14	0.673030853	0.349985764	0.332744311
15	0.967894507	0.348006081	0.328297839
16	0.99928218	0.348797461	0.330628783
17	0.342261678	0.359144183	0.344174073
18	0.955182195	0.351175539	0.335922309
19	0.657067296	0.348599554	0.330210353
20	0.615968199	0.348401688	0.329762478
21	0.318450341	0.38116272	0.385733896
22	0.342261678	0.352962936	0.33595212
23	0.648976844	0.347412979	0.327871166
24	0.295366285	0.348401688	0.329257815

3.4 Tahapan *Modeling K-Means* menggunakan Rapidminer

Hasil akhir pada pengolahan data menggunakan rapidminer didalam canvas yang sudah ditambahkan operator – operator seperti pada Gambar 1.



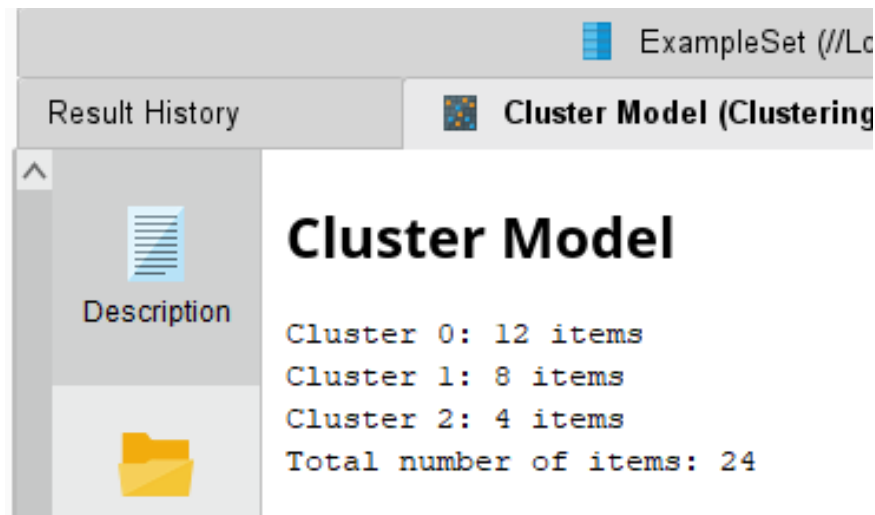
Gambar 1. *Modeling K-Means*

Penjelasan mengenai proses pada gambar diatas adalah awalnya data akan di-import selanjutnya menggunakan select attrbutes yang berfungsi untuk pemilihan atribut yang akan digunakan untuk proses k means secara langsung. Pada operator clustering merupakan proses data diolah dengan menggunakan metode K-Means clustering, setelah itu operator performance digunakan untuk

mengetahui performa dan hasil perhitungan cluster yang dihasilkan berdasarkan data riwayat stok barang penjualan.

3.5 Evaluasi

Berikut Hasil cluster model setelah melakukan proses k-means clustering dengan rapidminer. Didapatlan cluster 0 = 12 produk, cluster 1 = 8 produk dan cluster 2 = 4 produk. Hasil tersebut bisa dilihat pada Gambar 2.



Gambar 2. Hasil Cluster

3.6 Data preparation apriori dengan rapidminer

a. Seleksi data

Pada tahap ini data yang sudah didapatkan harus disaring dahulu dan diambil atribut yang akan dipakai untuk di proses. Terdapat 3 atribut yang digunakan untuk proses pola penjualan menggunakan *Apriori*, yaitu *Username*, nama produk dan jumlah produk dipesan. Data diambil berdasarkan atribut yang digunakan contoh seperti Tabel 6.

Tabel 6. Seleksi data

no	Jumlah Produk di Pesan	Nama Produk	Username (Pembeli)
1	4	INNER BASIC By Ayysee	brillyananda
2	4	INNER BASIC By Ayysee	brillyananda
3	4	INNER BASIC By Ayysee	brillyananda
4	1	INNER NINJA HATORY By Ayysee	tasya64
5	3	INNER NINJA HATORY By Ayysee	nanda_hafizah
6	3	INNER NINJA HATORY By Ayysee	nanda_hafizah

no	Jumlah Produk di Pesan	Nama Produk	Username (Pembeli)
7	3	INNER NINJA HATORY By Ayysee	nanda_hafizah
8	1	Manset Gamis Lengan Panjang / Long sleeve dress/ Custom Size/ By Ayysee	sandrawidiaa
9	1	INNER NINJA HATORY By Ayysee	autumnmadness
...			
6037	4	Compact Hijab kit / Set Jarum Hijab By Ayysee	kadidiri

b. Pembersihan data

Sesudah data yang dibutuhkan dipilih dan diseleksi sesuai dengan atribut yang nantinya akan dilakukan pre-processing data, tahap ini dilakukan karena nantinya tidak terjadi duplikasi data dan missing value pada data yang nantinya akan diproses untuk tahap selanjutnya. Tujuan dari tahap pembersihan data adalah agar data tersebut bisa diolah dan dilakukan proses *data mining*. Untuk mengetahui apabila ada data yang duplikat menggunakan fitur yang ada pada excel yaitu menggunakan *duplicate value*. Selanjutnya untuk mengetahui data tidak terjadi missing value menggunakan rumus pada excel countblank.

c. Transformasi data

tahap ini dilakukan transformasi data agar bisa dioperasikan ke dalam rapidminer menggunakan algoritma apriori. Transformasi yang dilakukan adalah melakukan pivot table yang ada di Excel untuk mengetahui setiap username belanja apa saja seperti pada Tabel 7.

Tabel 7. Transformasi data

No	Username	Compact Hijab kit / Set Jarum Hijab By Ayysee	Crochet Vest Knit Vest Korean Crochet Vest	...	Lily Vest (Big Size) By Ayysee
1	12ryanti	3	0		0
2	23.05.98	0	0		0
3	8brempw8zz	0	0		0

No	Username	Compact Hijab kit / Set Jarum Hijab By Ayysee	Crochet Vest Knit Vest Korean Crochet Vest	...	Lily Vest (Big Size) By Ayysee
4	aan_thomas	0	0		0
5	abaabi968	0	0		0
6	adenf12	0	0		0
7	adindaalifiaputri	0	0		0
8	adindankhz	0	0		0
9	affiance	0	0		0
10	afifahfebriana	0	0		0
...					
432	dwintasusanti	2	0		0

modeling apriori membutuhkan jumlah produk dipesan menjadi bilangan biner 0 sampai 1. Angka 1 menandakan produk dibeli sedangkan 0 produk tidak dibeli seperti pada Tabel 8.

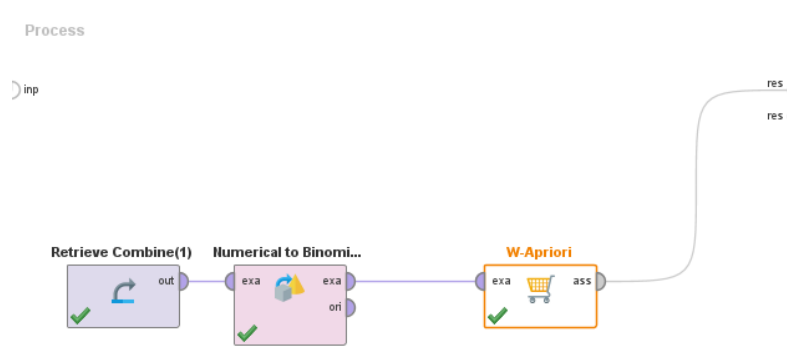
Tabel 8. Bilangan biner

No	Username	Compact Hijab kit / Set Jarum Hijab By Ayysee	Crochet Vest Knit Vest Korean Crochet Vest	...	Lily Vest (Big Size) By Ayysee
1	12ryanti	1	0		0
2	23.05.98	0	0		0
3	8bremppw8zz	0	0		0
4	aan_thomas	0	0		0
5	abaabi968	0	0		0
6	adenf12	0	0		0
7	adindaalifiaputri	0	0		0
8	adindankhz	0	0		0

No	Username	Compact Hijab kit / Set Jarum Hijab By Ayysee	Crochet Vest Knit Vest Korean Crochet Vest	...	Lily Vest (Big Size) By Ayysee
9	affiance	0	0		0
10	afifahfebriana	0	0		0
...					
432	dwintasusanti	1	0		0

3.7 Modeling apriori

Hasil akhir pada pengolahan apriori menggunakan rapidminer didalam *canvas* yang telah ditambahkan operator – operator yang digunakan seperti pada Gambar 3.



Gambar 3. Modeling apriori

3.8 Evaluasi

Hasil yang didapatkan untuk produk dengan ketentuan minimum *support* 10%, dan minimum *confidence* 45%. Dikarenakan jika menggunakan diatas presentase tersebut, hasil untuk kombinasi produk tidak muncul. Dikarenakan kondisi data yang digunakan maksimal untuk presentase support 10% dan confidence 45%. Berikut hasil yang telah memenuhi syarat *association rules* yang sudah ditentukan minimum support dan confidence oleh peneliti, seperti Gambar 4.

W-Apriori

```

Apriori
=====

Minimum support: 0.1 (43 instances)
Minimum metric <confidence>: 0.45
Number of cycles performed: 18

Generated sets of large itemsets:

Size of set of large itemsets L(1): 7

Size of set of large itemsets L(2): 2

Best rules found:

1. Inner Leher | By Ayysee=1 120 ==> INNER NINJA HATORY | By Ayysee=0 76   conf: (0.63)
2. INNER NINJA HATORY | By Ayysee=0 177 ==> INNER BASIC | By Ayysee=1 97   conf: (0.55)
3. INNER BASIC | By Ayysee=1 192 ==> INNER NINJA HATORY | By Ayysee=0 97   conf: (0.51)
  
```

Gambar 4. Hasil Evaluasi

3.9 Kesimpulan

K-means clustering digunakan untuk menentukan peminatan produk pada toko Ayysee Collection dengan menggunakan Rapidminer. Hasil yang didapatkan terpadat 3 kategori produk, cluster 0 (cukup diminati) = 12 produk, cluster 1 (kurang diminati) = 8 produk dan cluster 2 (paling diminati) = 4 produk. Nilai davies bouldin indeks yang diperoleh sebesar 0,417. Hal ini menunjukkan pengelompokan produk menggunakan K-Means menghasilkan hasil cluster yang baik karena mendekati angka 0.

Penentuan pola penjualan berdasarkan produk yang terjual menggunakan metode Apriori. Dengan syarat minimum support 10% dan confidence 45% didapatkan 3 kombinasi. Kombinasi pertama dengan confidence 63% pada produk Inner Leher | By Ayysee dan INNER NINJA HATORY | By Ayysee. Kombinasi kedua dengan confidence 55%, pada produk INNER NINJA HATORY | By Ayysee dan INNER BASIC | By Ayysee. Kombinasi ketiga dengan confidence 51% INNER BASIC | By Ayysee dan INNER NINJA HATORY | By Ayysee. Sehingga produk yang bisa dijadikan untuk *promo bundling* adalah produk Inner Leher | By Ayysee, INNER NINJA HATORY | By Ayysee dan INNER BASIC | By Ayysee.

Daftar Pustaka

- Aulia, S. (2021) 'Klasterisasi Pola Penjualan Pestisida Menggunakan Metode K-Means Clustering (Studi Kasus Di Toko Juanda Tani Kecamatan Hutabayu Raja)', *Djtechno: Jurnal Teknologi Informasi*, 1(1), pp. 1–5. Available at: <https://doi.org/10.46576/djtechno.v1i1.964>.
- Firmansah, R.Y., Dedy Irawan, J. and Vendyansyah, N. (2021) 'Analisis Rfm (Recency, Frequency and Monetary) Produk Menggunakan Metode K-Means', *JATI (Jurnal Mahasiswa Teknik Informatika)*, 5(1), pp. 334–341. Available at: <https://doi.org/10.36040/jati.v5i1.3282>.
- Lutfi, M.A. and Nilogiri, A. (2018) 'Implementasi Algoritma K-Means Clustering Untuk Pengelompokan Minat Konsumen Pada Produk Online Shop'.
- Syahril, M., Erwansyah, K. and Yetri, M. (2020) 'Penerapan Data Mining Untuk Menentukan Pola Penjualan Peralatan Sekolah Pada Brand Wigglo Dengan Menggunakan Algoritma Apriori', *J-SISKO TECH (Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD)*, 3(1), p. 118. Available at: <https://doi.org/10.53513/jsk.v3i1.202>.