

Prediksi Pembelian Barang Pada Distributor Lampu Menggunakan Metode Apriori pada PT. XYZ

Rudolfo Rizki Damanik¹, Moyo Hady Poernomo²

^{1,2}Fakultas Sains dan Teknologi, Universitas Pignatelli Triputra, Indonesia

Email: ¹echos.damanik@gmail.com, ²moyohadi@gmail.com

ABSTRAK

Apriori merupakan algoritma yang banyak digunakan untuk menentukan pola hubungan antar produk yang sering dibeli bersamaan dalam suatu toko. Algoritma Apriori ini akan sesuai diterapkan dalam bidang penentuan strategi ataupun promosi. PT. XYZ adalah salah satu Perusahaan Distributor Lampu. Salah satu masalah dari perusahaan adalah ketidakseimbangan stok di gudang. Sehingga tujuan dari penelitian ini adalah untuk mendapatkan prediksi pembelian produk dari supplier untuk menjaga keseimbangan stok dengan produk yang terjual ke konsumen. Objek dalam penelitian ini yaitu penerapan algoritma apriori. Data penelitian ini berupa data transaksi penjualan di PT. XYZ. Analisis data menggunakan Software RapidMiner dengan menentukan nilai support dan confidence. Hasil analisis menunjukkan terdapat 22 aturan asosiasi dengan minimum support 20% dan minimum confidence 50% dari informasi ini diharapkan dapat membantu pihak perusahaan dalam menyusun strategi pembelian produk dari supplier sesuai dengan itemset yang terbentuk.

Kata Kunci: Data Mining, Association Rules, Algoritma Apriori

ABSTRACT

Apriori is an algorithm that is widely used to determine the pattern of relationships between products that are often bought together in a store. This Apriori algorithm will be suitable to be applied in the field of determining strategy or promotion. PT. XYZ is one of the Lighting Distributor Companies. One of the problems of the company is the imbalance of stock in the warehouse. So, the purpose of this research is to obtain product purchase predictions from suppliers to maintain stock balance with product solds to customers. The object of this research is the application of apriori algorithm. This research data is in the form of sales transaction data at PT. XYZ. Data analysis using RapidMiner Software by determining the value of support and confidence. The analysis shows that there are 22 Association Rules with a minimum support 20% and a minimum confidence 50%. This information is expected to help the company in developing a product purchasing strategy from the supplier.

Keywords: Data Mining, Association Rules, Apriori Algorithm

Penulis Korespondensi:

Rudolfo Rizki Damanik

Email: echos.damanik@gmail.com

Article Info

Diterima: 14 Februari 2023

Direvisi: 18 Februari 2023

Disetujui: 22 Februari 2023

This is an open access article under the [CC BY](https://creativecommons.org/licenses/by/4.0/) license.



1. PENDAHULUAN

Pada masa sekarang teknologi informasi telah berkembang sangat pesat dan memberikan dampak yang besar bagi kehidupan masyarakat. Dampak tersebut berupa beberapa aspek yaitu aspek perekonomian, bisnis, pendidikan, dan lain-lain. Dalam aspek bisnis, seorang pebisnis khususnya di bidang penjualan akan mengumpulkan berbagai data atau informasi untuk mendapatkan keuntungan yang maksimal dan mencegah kerugian. Perkembangan pasar modern yang semakin maju dapat kita lihat dari banyaknya pusat perbelanjaan, distributor, maupun grosir yang dibangun untuk melayani kebutuhan konsumen.

Salah satu permasalahan yang sering dihadapi oleh distributor adalah stok barang. Jika jumlah stok barang terlalu sedikit dan permintaan pasar tidak dapat dipenuhi, maka akan mengakibatkan kekecewaan pada konsumen dan dapat menyebabkan kerugian yang besar pada distributor itu sendiri. Begitu pula dengan stok barang yang terlalu besar mengakibatkan kerugian karena terjadinya penyusutan nilai guna barang dan distributor harus mengeluarkan biaya tambahan seperti biaya inventaris dan pemeliharaan.

PT. XYZ adalah suatu perusahaan distributor yang bergerak di bidang penjualan lampu. Lampu-lampu yang dijual berupa lampu sorot, lampu jalan, dan lain-lain. PT. XYZ memiliki beberapa masalah salah satunya adalah stok barang dimana perusahaan tidak dapat memperhitungkan stok barang yang akan dibeli atau dipesan dari *supplier* untuk memenuhi permintaan pasar sehingga beresiko mengganggu keseimbangan stok barang. Oleh karena itu pihak perusahaan harus bisa segera mengetahui bagaimana cara menentukan pesanan barang pada *supplier* agar barang yang masuk dan keluar dapat seimbang. Oleh sebab itu, pihak distributor PT. XYZ memerlukan suatu sistem yang dapat menghasilkan informasi prediksi tentang keinginan dan kebiasaan pembelian barang oleh konsumen pada umumnya, sehingga perusahaan dapat memastikan pembelian atau pemesanan barang dari *supplier*.

Saat ini sistem yang masih digunakan oleh PT. XYZ hanya mencatat data-data transaksi penjualan setiap harinya dan tidak dapat memprediksi pembelian barang kepada *supplier*. Sistem untuk memprediksi pembelian barang seperti ini dapat dibuat dengan mencatat data penjualan dan diproses dengan menggunakan metode pada *data mining*. Salah satu metode Data Mining yang dapat digunakan untuk memprediksi pembelian barang adalah metode Apriori.

2. METODE PENELITIAN

Data Mining

Data mining adalah salah satu kedisiplinan cabang ilmu komputer sebagai proses komputasi penejelajahan pola dalam kumpulan data dengan menyertakan metode kecerdasan buatan, *machine learning*, statistik dan sistem *database*, dengan tujuan keseluruhan prosesnya adalah untuk mengesktrak informasi dari kumpulan data dan mengubahnya menjadi sekumpulan struktur data yang dapat dimengerti dan digunakan [1].

Definisi secara umum dari *Data Mining* yaitu ekstraksi informasi atau pola yang menarik atau penting dari data yang ada di *database*. Dalam jurnal ilmiah, *Data Mining* juga disebut sebagai *Knowledge Discovery in Database* (KDD). Karakteristik *Data Mining* sebagai berikut:

- Data Mining* berhubungan dengan penemuan sesuatu yang tersembunyi dan pola data tertentu yang tidak diketahui sebelumnya.
- Data Mining* biasa menggunakan data yang sangat besar. Biasanya data yang besar digunakan untuk membuat hasil lebih dapat dipercaya.
- Data Mining* berguna untuk membuat keputusan kritis.

Teknik *Data Mining*

Teknik *Data Mining* digunakan untuk memeriksa basis data berukuran besar sebagai cara untuk menemukan pola yang baru dan berguna. Ada beberapa jenis teknik analisa yang dapat dikategorikan dalam *Data Mining*. Ini adalah tiga teknik *Data Mining* yang sering digunakan yaitu:

1. *Association Rule Mining*

Association rule mining adalah teknik *Data Mining* untuk menemukan aturan asosiasi antara suatu kombinasi item. *Association rule* memiliki 2 bagian yaitu *if* dan *then*. *If* adalah variabel yang ditemukan pada data, lalu *then* ditemukan dalam kombinasi dengan *if*. Beberapa algoritma yang populer adalah Apriori, Partition, FP-Growth, dan ECLAT. Contoh aturan asosiasi dari analisa pembelian di suatu pasar swalayan adalah dapat diketahuinya berapa besar kemungkinan seorang pelanggan membeli roti bersamaan dengan susu [2].

2. *Clustering*

Clustering juga disebut sebagai *segmentation*. Metode ini digunakan untuk mengidentifikasi kelompok alami dari sebuah kasus yang didasarkan pada sebuah kelompok atribut, mengelompokkan data yang memiliki kemiripan atribut. *Clustering* adalah metode *Data Mining* yang *unsupervised*, karena tidak ada satu atributpun yang digunakan untuk memandu proses

pembelajaran, jadi seluruh atribut input diperlakukan sama. Kebanyakan algoritma *clustering* membangun sebuah model melalui serangkaian pengulangan dan berhenti ketika model tersebut telah memusat atau berkumpul [2]. Metode *clustering* secara umum dapat dibagi menjadi dua yaitu *hierarchical clustering* dan *partitional clustering*. Contoh dari metode *hierarchical clustering* yaitu: *Single Linkage*, *Complete Linkage*, *Average Linkage*, *Average Group Linkage*. Sedangkan contoh dari metode *partitional clustering* yaitu: *K-Means*, *Fuzzy K-Means*, dan *Mixture Modelling*.

3. Klasifikasi

Klasifikasi adalah suatu pengelompokan data dimana data yang digunakan tersebut mempunyai kelas label atau target. Sehingga algoritma-algoritma untuk menyelesaikan masalah klasifikasi dikategorisasikan ke dalam *supervised learning* atau pembelajaran yang diawasi. Maksud dari pembelajaran yang diawasi adalah data label atau target ikut berperan sebagai 'supervisor' atau 'guru' yang mengawasi proses pembelajaran dalam mencapai tingkat akurasi atau presisi tertentu. Beberapa metode umum yang dapat digunakan dalam klasifikasi yaitu: *Backpropagation* yang dapat digunakan untuk memprediksi harga saham [3], *Support Vector Classification (SVC)*, *Extreme Learning Machine (ELM)* yang dapat digunakan untuk mengklasifikasi penyakit jantung aritmia [4], *K-NN*, *Naive Bayes* yang dapat digunakan untuk memprediksi *video performance* [5], dan lain-lain.

4. Algoritma Apriori

Algoritma Apriori adalah salah satu algoritma yang melakukan pencarian *frequent itemset* dengan memakai aturan asosiasi. Algoritma Apriori adalah algoritma yang digunakan untuk menghasilkan *association rules* dengan pola *if-then*, dimana *k-itemset* digunakan untuk mengeksplorasi *itemset* dengan melakukan pendekatan *iterative* yang dikenal pencarian *level-wise*. *Association rules* sering juga disebut dengan *Market Basket Analysis* [6] karena kegunaannya dapat membantu distributor untuk memesan barang dari *supplier* dan dapat menganalisis perilaku konsumen. Algoritma Apriori menggunakan pengetahuan frekuensi atribut yang telah diketahui sebelumnya untuk memproses informasi selanjutnya. Pada Algoritma Apriori menentukan kandidat yang mungkin muncul dengan cara memerhatikan minimum *support* & *minimum confidence*. *Support* adalah nilai pengujung atau persentase kombinasi sebuah item dalam *database*.

Rumus support adalah sebagai berikut:

$$Support(A,B) = \frac{(\text{Jumlah transaksi A,B})}{\text{Total transaksi}} \times 100\% \quad (1)$$

Sedangkan *confidence* adalah nilai kepastian yaitu kuatnya hubungan antar item dalam sebuah Apriori. *Confidence* dapat dicari setelah pola frekuensi munculnya sebuah item ditemukan.

Rumus untuk menghitung *confidence* adalah sebagai berikut

Contoh misalnya ditemukan aturan $A \rightarrow B$ maka:

$$Confidence P(B|A) = \frac{\text{Total transaksi mengandung A dan B}}{\text{Transaksi mengandung A}} \times 100\% \quad (2)$$

Prinsip dari Algoritma Apriori antara lain :

1. Mengumpulkan item yang tunggal kemudian mencari item yang tersebar.
2. Dapatkan *candidate pairs* kemudian hitung *large pairs* dari masing – masing item.
3. Temukan *candidate triplets* dari setiap item dan seterusnya.
4. Setiap subset dari sebuah *frequent itemset* harus menjadi *frequent*

Proses dalam Algoritma Apriori. Proses utama yang dilakukan dalam Algoritma Apriori untuk mendapat *frequent itemset* yaitu :

a. *Join* (Penggabungan)

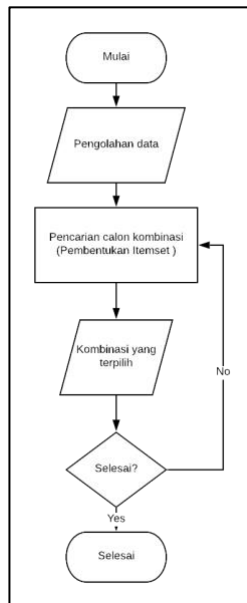
Proses ini dilakukan dengan cara pengkombinasian item dengan item yang lainnya hingga tidak dapat terbentuk kombinasi lagi.

b. *Prune* (Pemangkasan)

Proses pemangkasan yaitu hasil dari item yang telah dikombinasikan kemudian dipangkas dengan memakai minimum *support* yang telah ditentukan oleh user.

Algoritma Apriori dibagi menjadi beberapa tahap yang disebut iterasi atau *pass*. Tiap iterasi menghasilkan pola frekuensi tinggi dengan panjang yang sama dimulai dari *pass* pertama yang menghasilkan pola frekuensi tinggi dengan panjang satu. Di iterasi pertama ini, *support* dari setiap item dihitung dengan memindai *database*. Setelah *support* dari setiap item didapat, item yang memiliki *support* di atas minimum *support* dipilih sebagai pola frekuensi tinggi dengan panjang 1 atau sering disingkat *1-itemset*. Singkatan *k-itemset* berarti satu set yang terdiri dari k item [7].

Untuk melakukan proses pengolahan data transaksi pembelian perlu digambarkan tahapan kerja yang dapat dilakukan pada algoritma Apriori seperti gambar 2.1 berikut:



Gambar 1. Ilustrasi dari penerapan Algoritma Apriori

3. HASIL DAN PEMBAHASAN

Analisa Data

Pada tahap analisa data, penulis melakukan langkah-langkah sebagai berikut:

a. Tahap preprocessing

Tahap ini melakukan seleksi data, yaitu pemilihan data dari sekumpulan data sehingga data yang digunakan tidak memiliki duplikasi data dan transformasi data yaitu proses transformasi pada data yang telah dipilih yaitu transaksi penjualan lampu, sehingga data tersebut sesuai untuk proses *data mining*.

Berikut Tabel 1 ini adalah contoh data transaksi penjualan lampu yang masih belum dilakukan penyeleksian data.

No faktur	Kode barang	Nama barang	Jumlah item
1	X.008.02	XGP500 100-240V K2 OBSTIVISION	1
2	D.010.5.41	DN027B LED20/WW D200 RD	1
3	H.010.1.20	HALOGEN ESS 50W/12V/36D CLOSED	4
	L.010.5.13B	LED BULB 3.5-25W E27 3000K 230V P45 [APR]	
	P.010.1.06	PLC 13W/865/2P	
	T.010.1.20	TLD 36W/54	
4	L.010.5.20	LEDTUBE 1200MM 16W 765 T8 AP C G ECOVIT	1
5	T.010.1.20	TLD 36W/54	3
	P.010.1.18	PL-S 7W/865/2P	
	T.010.3.27	TCW097 2X36WI HFS	
6	T.010.1.20	TLD 36W/54	3
	L.010.5.23A	LED BULB 8-70W E27 3000K 230V A60 INDO	
	T.010.3.27	TCW097 2X36WI HFS	
7	T.010.1.20	TLD 36W/54	3
	M.010.1.10	ML 500W E40	
	T.010.3.27	TCW097 2X36WI HFS	

8	T.010.1.20	TLD 36W/54	1
9	L.010.5.12	LED DRIVER 220-240V 60VA	1
10	T.010.1.20	TLD 36W/54	2
	T.010.3.27	TCW097 2X36WI HFS	
11	T.010.1.20	TLD 36W/54	3
	M.010.5.54	MAS LEDTUBE STD 1200MM 10.5/865 T8 I	
	T.010.3.27	TCW097 2X36WI HFS	
12	L.010.5.11A	LED BULB 14-100W E27 3000K 230V A67	4
	L.010.5.13B	LED BULB 3.5-25W E27 3000K 230V P45 [APR]	
	L.010.5.23A	LED BULB 8-70W E27 3000K 230V A60 INDO	
	T.010.1.09	TLD 15W/54	
13	L.010.5.10B	LED BULB 4-40W E27 6500K 230V P45 (APR)	1
14	S.010.2.05	SN 56	1
15	E.010.1.09	ESS 23W/82 WW E27	4
	M.010.1.10	ML 500W E40	
	P.010.1.18	PL-S 7W/865/2P	
	T.010.1.20	TLD 36W/54	
16	M.010.5.54	MAS LEDTUBE STD 1200MM 10.5/865 T8 I	1
17	H.010.1.02	HPI-T PLUS 400W E40	1
18	M.010.5.54	MAS LEDTUBE STD 1200MM 10.5/865 T8 I	2
	T.010.3.21A	TCW097 2X36WE [KOSONGAN]	
	P.010.1.18	PL-S 7W/865/2P	3
19	T.010.1.20	TLD 36W/54	
	T.010.3.27	TCW097 2X36WI HFS	

Dari data tabel 1 diatas, Dilakukan penyeleksian data dimana setiap transaksi yang jumlah item dibawah 2 item maka akan dihapus data transaksi tersebut seperti tabel 2 berikut.

Tabel 2. Data Transaksi yang terseleksi

No faktur	Kode barang	Nama barang	Jumlah item
3	H.010.1.20	HALOGEN ESS 50W/12V/36D CLOSED	4
	L.010.5.13B	LED BULB 3.5-25W E27 3000K 230V P45 [APR]	
	P.010.1.06	PLC 13W/865/2P	
	T.010.1.20	TLD 36W/54	
5	T.010.1.20	TLD 36W/54	3
	P.010.1.18	PL-S 7W/865/2P	
	T.010.3.27	TCW097 2X36WI HFS	
6	T.010.1.20	TLD 36W/54	3
	L.010.5.23A	LED BULB 8-70W E27 3000K 230V A60 INDO	
	T.010.3.27	TCW097 2X36WI HFS	
7	T.010.1.20	TLD 36W/54	3
	M.010.1.10	ML 500W E40	
	T.010.3.27	TCW097 2X36WI HFS	
10	T.010.1.20	TLD 36W/54	2
	T.010.3.27	TCW097 2X36WI HFS	
11	T.010.1.20	TLD 36W/54	3
	M.010.5.54	MAS LEDTUBE STD 1200MM 10.5/865 T8 I	
	T.010.3.27	TCW097 2X36WI HFS	
12	L.010.5.11A	LED BULB 14-100W E27 3000K 230V A67	4

	L.010.5.13B	LED BULB 3.5-25W E27 3000K 230V P45 [APR]	
	L.010.5.23A	LED BULB 8-70W E27 3000K 230V A60 INDO	
	T.010.1.09	TLD 15W/54	
15	E.010.1.09	ESS 23W/82 WW E27	4
	M.010.1.10	ML 500W E40	
	P.010.1.18	PL-S 7W/865/2P	
	T.010.1.20	TLD 36W/54	
18	M.010.5.54	MAS LEDTUBE STD 1200MM 10.5/865 T8 I	2
	T.010.3.21A	TCW097 2X36WE [KOSONGAN]	
	P.010.1.18	PL-S 7W/865/2P	3
19	T.010.1.20	TLD 36W/54	
	T.010.3.27	TCW097 2X36WI HFS	

b. Analisa Pola Frekuensi tinggi

Tahap ini mencari kombinasi item yang memenuhi syarat minimum dari nilai *support* dalam *database*. Nilai *support* sebuah item diperoleh dengan rumus berikut

$$Support(A) = \frac{\text{Jumlah transaksi A}}{\text{Total transaksi}} \tag{3}$$

Sedangkan nilai *support* dari 2 item didapat dengan menggunakan rumus:

$$Support(A,B) = \frac{\text{Jumlah transaksi A,B}}{\text{Total transaksi}} \tag{4}$$

Mengambil contoh data dari tabel 1 kemudian dilakukan pembentukan item set. Sebuah item set adalah himpunan item-item yang ada dalam I, dan k-item set adalah item set yang k item. Misalkan {Halogen, Led Bulb} adalah sebuah 2-item set dan {Halogen, Led Bulb, TCW} merupakan 3-item set. Untuk menghitung persentase *support* dari masing-masing item digunakan rumus sebagai berikut beserta contoh perhitungannya:

$$Support(A) = \frac{\text{Jumlah transaksi A}}{\text{Total transaksi}} * 100\% \tag{5}$$

Total transaksi = 10

Banyak transaksi untuk setiap item:

Tabel 3. Tabel Data Tabular

Transaksi	A	B	C	D	E	F	G	H	I	J	K	L	M
1	1	1	1	0	0	0	1	0	0	0	0	0	0
2	0	0	1	1	0	0	0	0	0	0	0	1	0
3	0	0	1	1	0	0	0	0	1	0	0	0	0
4	0	0	1	1	1	0	0	0	0	0	0	0	0
5	0	0	1	1	0	0	0	0	0	0	0	0	0
6	0	0	1	1	0	1	0	0	0	0	0	0	0
7	0	0	0	0	0	0	1	1	1	1	0	0	0
8	0	0	1	0	1	0	0	0	0	0	1	1	0
9	0	0	0	0	0	1	0	0	0	0	0	0	1
10	0	0	1	1	0	0	0	0	0	0	0	1	0
Total	1	1	8	6	2	2	2	1	2	1	1	3	1

Support(A) = ((1))/10*100%=10%

Support(B) = ((1))/10*100%=10%

Support(C) = ((8))/10*100%=80%

Support(D) = ((6))/10*100%=60%

Support(E) = ((2))/10*100%=20%

Support(F) = ((2))/10*100%=20%

$$\begin{aligned} \text{Support(G)} &= ((2))/10 * 100\% = 20\% & \text{Support(H)} &= ((1))/10 * 100\% = 10\% \\ \text{Support(I)} &= ((2))/10 * 100\% = 20\% & \text{Support(J)} &= ((1))/10 * 100\% = 10\% \\ \text{Support(K)} &= ((1))/10 * 100\% = 10\% & \text{Support(L)} &= ((3))/10 * 100\% = 30\% \\ \text{Support(M)} &= ((1))/10 * 100\% = 10\% \end{aligned}$$

Jumlah transaksi dan *support %* untuk tiap *item* ditunjukkan pada tabel 4 berikut.

Tabel 4. Tabel Data Tabular

Kode	Nama item	Jumlah transaksi	Support %
A	HALOGEN ESS 50W/12V/36D CLOSED	1	10%
B	PLC 13W/865/2P	1	10%
C	TLD 36W/54	8	80%
D	TCW097 2X36WI HFS	6	60%
E	ML 500W E40	2	20%
F	MAS LEDTUBE STD 1200MM 10.5/865 T8 I	2	20%
G	LED BULB 3.5-25W E27 3000K 230V P45 [APR]	2	20%
H	LED BULB 14-100W E27 3000K 230V A67	1	10%
I	LED BULB 8-70W E27 3000K 230V A60 INDO	2	20%
J	TLD 15W/54	1	10%
K	ESS 23W/82 WW E27	1	10%
L	PL-S 7W/865/2P	3	30%
M	TCW097 2X36WE [KOSONGAN]	1	10%

Pembentukan itemset dimulai dengan menentukan $\Phi = 2$, maka kita dapat menemukan jumlah dari *itemset* yang lebih besar dari Φ terdapat 7 *item* $F1 = \{C, D, E, F, G, I, L\}$. Maka akan terbentuk kombinasi 2 *item set* yaitu = $\{C,D, C,E, C,F, C,G, C,I, C,L, D,E, D,F, D,G, D,I, D,L, E,F, E,G, E,I, E,L, F,G, F,I, F,L, G,I, G,L, I,L\}$.

Tabel 5. Support Kombinasi 2-itemset

Kode	Kode	Jumlah	Support%
C	D	6	60%
C	E	2	20%
C	F	1	10%
C	G	1	10%
C	I	1	10%
C	L	3	30%
D	E	1	10%
D	F	1	10%
D	G	0	0
D	I	1	10%
D	L	2	20%
E	F	0	0
E	G	0	0
E	I	0	0
E	L	1	10%
F	G	0	0
F	I	0	0
F	L	0	0
G	I	1	10%
G	L	0	0
I	L	0	0

Dari kombinasi 2-itemset pada tabel 5 maka terlihat bahwa ada 4 kombinasi 2-itemset yang memenuhi Φ yang ditentukan yaitu $F2 = \{C,D, C,E, C,L, D,L\}$.

Tabel 6. Support Kombinasi 3-itemset

Kode	Kode	Kode	Jumlah	Support%
------	------	------	--------	----------

C	D	E	1	10%
C	E	L	1	10%
C	D	L	2	20%

Dari kombinasi 3-itemset pada tabel 6 maka terlihat bahwa hanya ada 1 itemset yang memenuhi syarat Φ yang ditentukan Pembentukan aturan asosiasi

Setelah semua pola frekuensi tinggi ditemukan, barulah dicari aturan asosiasi yang memenuhi syarat minimum untuk confidence dengan menghitung confidence aturan asosiasi A-> B. Nilai confidence dari aturan A-> B diperoleh dengan rumus berikut:

$$\text{Confidence } P(B | A) = (\text{Total transaksi mengandung A dan B}) / (\text{Transaksi mengandung A})$$

Dari proses pembentukan aturan asosiasi maka akan diperoleh nilai confidence dari setiap itemset, dan kemudian tentukan minimum confidence untuk mendapatkan rule association.

Setelah perhitungan confidence pada kombinasi 2-itemset maka hasil dapat dilihat pada tabel 7.

$$\text{Confidence } P(B | A)\% = (\text{Total transaksi mengandung A dan B}) / (\text{Transaksi mengandung A}) * 100\% \quad (6)$$

$$\text{Confidence } P(D | C) = 6/8 * 100\% = 75\%$$

$$\text{Confidence } P(E | C) = 2/8 * 100\% = 25\%$$

$$\text{Confidence } P(F | C) = 1/8 * 100\% = 12,5\%$$

$$\text{Confidence } P(G | C) = 1/8 * 100\% = 12,5\%$$

$$\text{Confidence } P(I | C) = 1/8 * 100\% = 12,5\%$$

$$\text{Confidence } P(L | C) = 3/8 * 100\% = 37,5\%$$

$$\text{Confidence } P(E | D) = 1/6 * 100\% = 16,6\%$$

$$\text{Confidence } P(F | D) = 0/6 * 100\% = 0$$

$$\text{Confidence } P(G | D) = 0/6 * 100\% = 0$$

$$\text{Confidence } P(I | D) = 1/6 * 100\% = 16,6\%$$

$$\text{Confidence } P(L | D) = 2/6 * 100\% = 33,3\%$$

$$\text{Confidence } P(F | E) = 0/2 * 100\% = 0$$

$$\text{Confidence } P(G | E) = 0/2 * 100\% = 0$$

$$\text{Confidence } P(I | E) = 0/2 * 100\% = 0$$

$$\text{Confidence } P(L | E) = 1/2 * 100\% = 50\%$$

$$\text{Confidence } P(G | F) = 0/2 * 100\% = 0$$

$$\text{Confidence } P(I | F) = 0/2 * 100\% = 0$$

$$\text{Confidence } P(L | F) = 0/2 * 100\% = 0$$

$$\text{Confidence } P(I | G) = 1/2 * 100\% = 50\%$$

$$\text{Confidence } P(L | G) = 0/2 * 100\% = 0$$

$$\text{Confidence } P(L | I) = 0/2 * 100\% = 0$$

Tabel 7. Hasil perhitungan Confidence

Kode	Kode	Support%	Confidence%
C	D	60%	75%
C	E	20%	25%
C	F	10%	12,50%
C	G	10%	12,50%
C	I	10%	12,50%
C	L	30%	37,50%
D	E	10%	16,60%
D	F	0	0%
D	G	0	0%
D	I	10%	16,60%
D	L	20%	33,30%
E	F	0	0,00%
E	G	0	0,00%
E	I	0	0,00%
E	L	10%	50,00%
F	G	0	0,00%
F	I	0	0,00%
F	L	0	0,00%
G	I	10%	50%
G	L	0	0,00%
I	L	0	0,00%

Dari hasil perhitungan confidence di tabel 7 diatas maka ditentukan minimum confidence sebesar 50 % maka akan menghasilkan aturan asosiasi seperti pada tabel 8.

Tabel 8. Hasil Aturan Asosiasi

Aturan Asosiasi	Confidence %
Jika terjual lampu "TLD 36W/54" maka akan terjual lampu "TCW097 2X36WI HFS"	75%
Jika terjual lampu "ML 500W E40" maka akan terjual lampu "PL-S 7W/865/2P"	50%

Jika terjual lampu "LED BULB 14-100W E27 3000K 230V A67" maka akan terjual lampu "LED BULB 8-70W E27 3000K 230V A60 INDO 2X36WI HFS"

50%

Implementasi RapidMiner

Pada bagian ini akan dibahas proses *RapidMiner* dalam menemukan pola pembelian produk oleh *customer* untuk dapatkan prediksi pembelian produk dari aturan asosiasi yang dihasilkan. Data yang digunakan adalah data penjualan dalam bentuk excel [8]. Hasil dari data mining ini adalah informasi yang berupa pola/aturan asosiasi penjualan produk. Di bawah ini merupakan langkah-langkah kerja pengimplementasian data mining menggunakan *RapidMiner*, yaitu sebagai berikut:

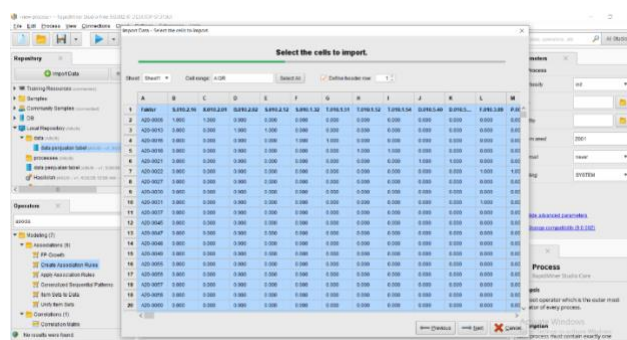
1. Pembuatan Format Tabular

Pada tahap ini yang dilakukan adalah mempersiapkan data yang akan diolah yaitu data transaksi penjualan produk pada tahun 2019. Data transaksi penjualan memiliki 595 data. Untuk pembuatan data format tabular menggunakan *Microsoft Excel*, dimana data yang digunakan dapat dilihat pada lampiran. Kemudian dilakukan *Importing* tabel *Microsoft Excel* yang telah dibuat kedalam *Local Repository*. Pertama di klik tombol *import data* setelah itu akan muncul form *import data step 1* untuk memilih lokasi file yang digunakan untuk proses penelitian seperti gambar 2 [9].



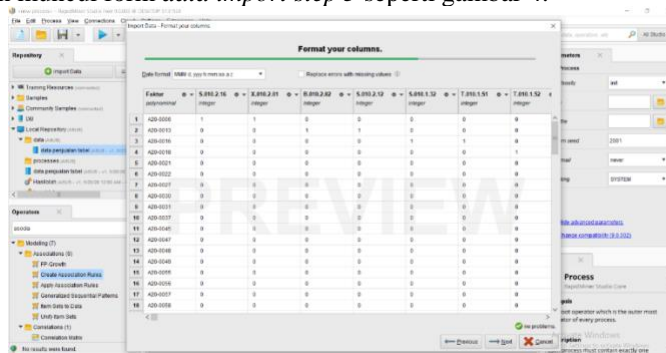
Gambar 2. Alur proses *import data*

Setelah data telah dipilih maka selanjutnya klik *Next* dan akan muncul form *data import step 2* seperti gambar 3.



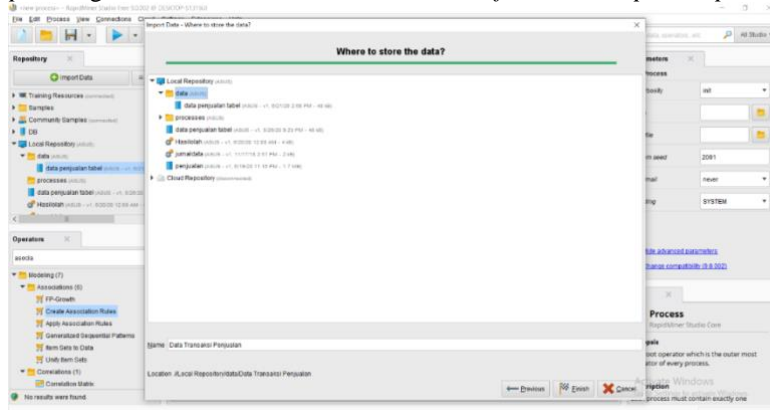
Gambar 3. Alur proses *import data*

Setelah muncul *form step* ke 2 seperti gambar diatas lalu dipilih tombol *select all*. Untuk melanjutkan ke step selanjutnya klik *Next* maka akan muncul form *data import step 3* seperti gambar 4.



Gambar 4. Alur proses *import data*

Pada *step* ke 3 ini tidak ada dilakukan apapun hanya dapat melihat *preview* dari data yang diimport maka dari itu langsung ke *data import step 4* dengan cara klik *Next* maka akan muncul *form data import step 4* seperti gambar 5.

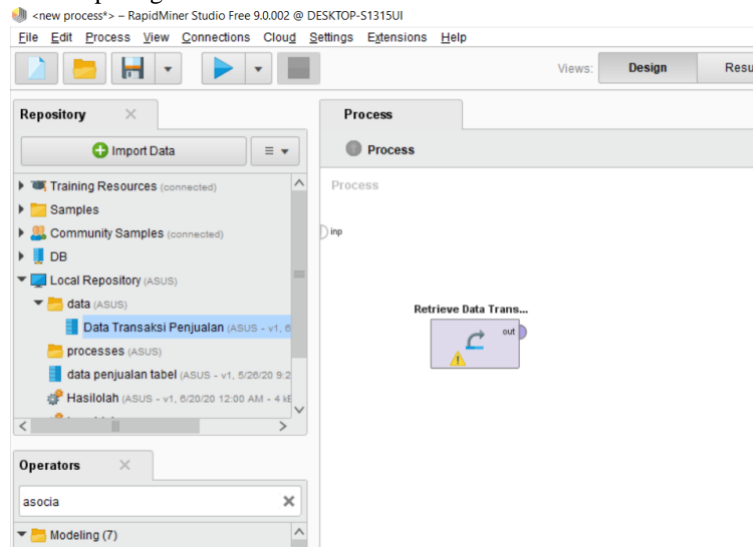


Gambar 5. Alur proses *import data*

Setelah muncul *form data import step 4* seperti gambar diatas lalu pilih lokasi penyimpanan dan isi nama file yang ingin disimpan. Setelah itu klik *Finish*.

2. Susunan Operator Algoritma Apriori

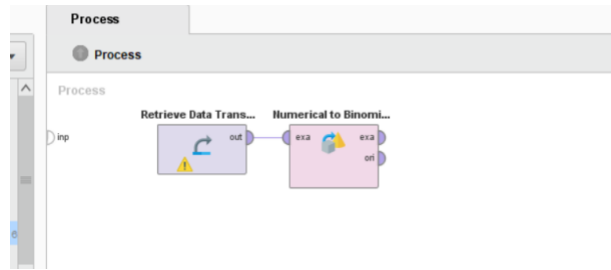
Pada tahap ini yang dilakukan adalah *drag and drop* tabel data penjualan ke dalam *process*. Sehingga data penjualan tabel muncul dalam *main process* seperti gambar 6.



Gambar 6. *Drag and drop* data transaksi penjualan

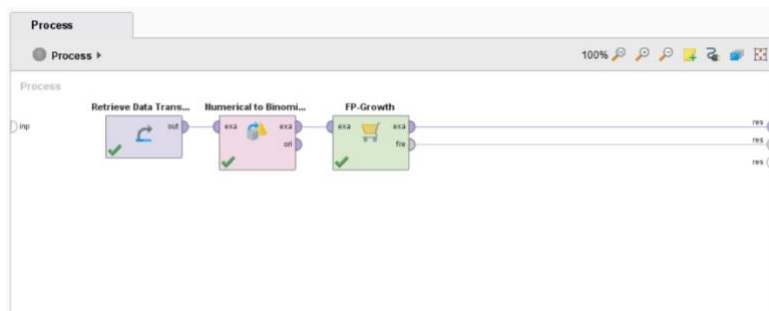
Dalam penelitian ini membutuhkan setidaknya tiga buah operator, yaitu *Association Rule*, *FP-Growth*, dan *Numerical to Binomial*. *Association rules* dilakukan dengan menganalisis data pada *frequent if/then patterns* menggunakan kriteria *support* dan *confidence* untuk mengidentifikasi suatu relasi antar item. *Frequent if/then pattern* didapatkan menggunakan operator *FP-Growth*. *Frequent itemsets* merupakan kelompok item yang sering muncul bersama-sama dalam data. Operator *FP-Growth* mengkalkulasikan semua *frequent itemset* dari input yang diberikan menggunakan struktur data *FP-tree* dan wajib bahwa semua atribut dari masukan merupakan bilangan binominal. Operator *Numerical to Binomial* untuk mengubah nilai atribut yang berbeda pada tabel transaksi penjualan menjadi bentuk binominal [10].

Gunakan pencarian filter untuk memudahkan dalam menemukan operator yang dibutuhkan. Setelah itu cari Operator *Numerical to Binomial* di pencarian. Operator ini terdapat pada hirarki: *Data Transformation – Type Conversion*. Kemudian hubungkan tabel data penjualan dengan operator *Numerical to Binomial*. Proses ini akan membuat nilai dari tabel transaksi penjualan menjadi *Binominal Attributes* dapat dilihat di Gambar 7.



Gambar 7. Proses menghubungkan operator

Selanjutnya hubungkan operator *Numerical to Binominal* dengan operator *FP-Growth* pada *example output* seperti Gambar 8.

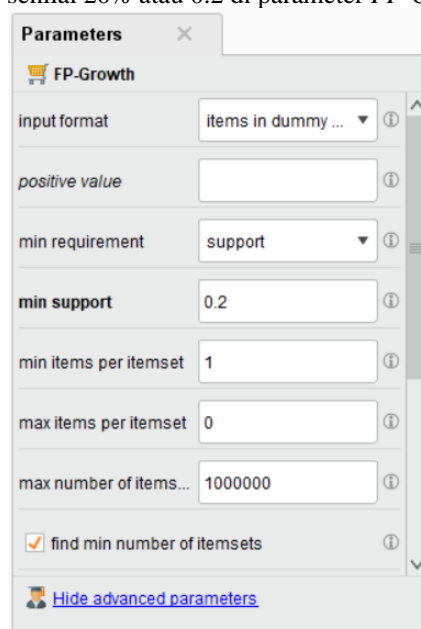


Gambar 8. Menghubungkan Operator *Numerical to Binominal* dengan Operator *FP-Growth*

Output dari *Numerical to Binominal* terdapat 2 buah, yaitu *example* dan *original*.

- *Example*, *numeric attributes* dikonversikan menjadi *binominal attributes* melalui output ini.
- *Original*, *numeric attributes* dilewatkan tanpa konversi. Biasanya digunakan untuk proses tertentu saat dibutuhkan. Untuk proses ini dilewatkan output pada *example*.

Setelah itu diisi minimum *support* senilai 20% atau 0.2 di parameter *FP-Growth* seperti gambar 9.

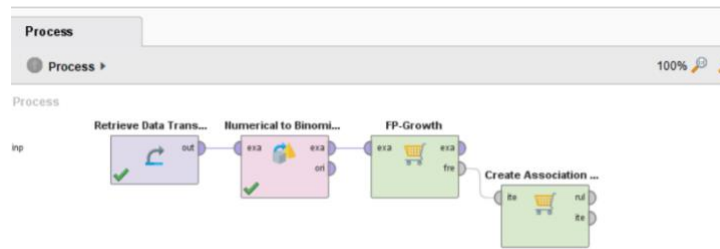


Gambar 9. Parameter *FP-Growth*

Kemudian menghubungkan operator *FP-Growth* dengan operator *Association Rules*. Pada operator *FP-Growth* terdapat 2 buah output, yaitu *example* dan *frequent*.

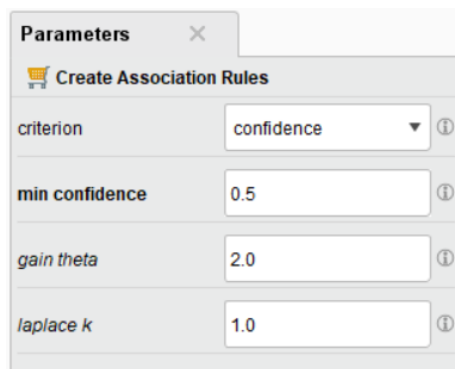
- *Example*, input yang diberikan dilewatkan tanpa adanya perubahan. Biasanya digunakan untuk proses tertentu saat dibutuhkan.

- *Frequent, frequent itemset* dikirimkan melalui *output* ini. Untuk proses ini dilewatkan *output* pada *frequent* seperti gambar 10.



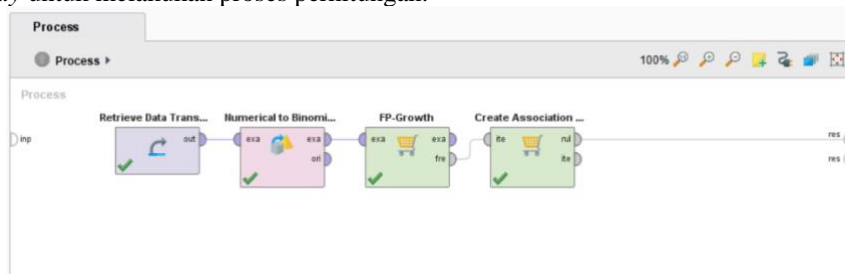
Gambar 10. Menghubungkan *Operator FP-Growth* dengan *Operator Create Association Rules*.

Setelah itu diisi minimum *confidence* senilai 50% atau 0,5 di Parameter *Association Rules* seperti Gambar 11 dibawah.



Gambar 11. *Parameter Create Association Rules*

Tahap berikutnya hubungkan *Association Rules* pada *result*. Sehingga seluruhnya membentuk seperti gambar 12 Kemudian klik ikon *Play* untuk melakukan proses perhitungan.

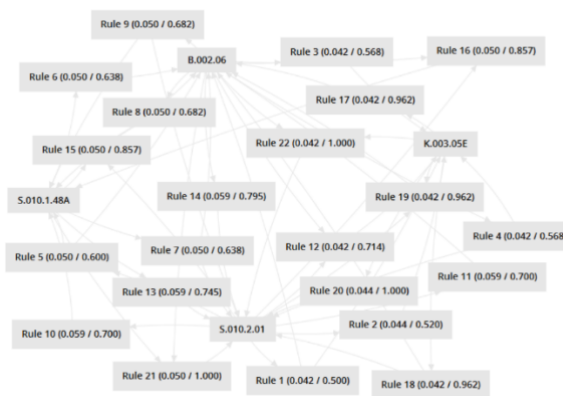


Gambar 12. Susunan *Operator Association Rules*.

Setelah itu akan muncul sebuah tab *Association Rules* yang baru, yang isinya berisi seluruh *itemset* yang memenuhi parameter *FP-Growth* dan *Association Rules*. Totalnya terdapat 22 aturan yang ditemukan.

No.	Premises	Conclusion	Support	Confidence
1	S.010.2.01	B.002.06, K.003.05E	0.042	0.500
2	S.010.2.01	K.003.05E	0.044	0.520
3	B.002.06	K.003.05E	0.042	0.568
4	B.002.06	S.010.2.01, K.003.05E	0.042	0.568
5	S.010.2.01	S.010.1.48A, B.002.06	0.050	0.600
6	S.010.1.48A	B.002.06	0.050	0.638
7	S.010.1.48A	S.010.2.01, B.002.06	0.050	0.638
8	B.002.06	S.010.1.48A	0.050	0.682
9	B.002.06	S.010.2.01, S.010.1.48A	0.050	0.682
10	S.010.2.01	S.010.1.48A	0.059	0.700
11	S.010.2.01	B.002.06	0.059	0.700
12	S.010.2.01, B.002.06	K.003.05E	0.042	0.714
13	S.010.1.48A	S.010.2.01	0.059	0.745
14	B.002.06	S.010.2.01	0.059	0.795
15	S.010.2.01, S.010.1.48A	B.002.06	0.050	0.857
16	S.010.2.01, B.002.06	S.010.1.48A	0.050	0.857
17	K.003.05E	B.002.06	0.042	0.962
18	K.003.05E	S.010.2.01, B.002.06	0.042	0.962
19	S.010.2.01, K.003.05E	B.002.06	0.042	0.962

Gambar 13. Hasil Association Rules



Gambar 14. Hasil dalam bentuk Grafik

AssociationRules

```

Association Rules
[S.010.2.01] --> [B.002.06, K.003.05E] (confidence: 0.500)
[S.010.2.01] --> [K.003.05E] (confidence: 0.520)
[B.002.06] --> [K.003.05E] (confidence: 0.568)
[B.002.06] --> [S.010.2.01, K.003.05E] (confidence: 0.568)
[S.010.2.01] --> [S.010.1.48A, B.002.06] (confidence: 0.600)
[S.010.1.48A] --> [B.002.06] (confidence: 0.638)
[S.010.1.48A] --> [S.010.2.01, B.002.06] (confidence: 0.638)
[B.002.06] --> [S.010.1.48A] (confidence: 0.682)
[B.002.06] --> [S.010.2.01, S.010.1.48A] (confidence: 0.682)
[S.010.2.01] --> [S.010.1.48A] (confidence: 0.700)
[S.010.2.01] --> [B.002.06] (confidence: 0.700)
[S.010.2.01, B.002.06] --> [K.003.05E] (confidence: 0.714)
[S.010.1.48A] --> [S.010.2.01] (confidence: 0.745)
[B.002.06] --> [S.010.2.01] (confidence: 0.795)
[S.010.2.01, S.010.1.48A] --> [B.002.06] (confidence: 0.857)
[S.010.2.01, B.002.06] --> [S.010.1.48A] (confidence: 0.857)
[K.003.05E] --> [B.002.06] (confidence: 0.962)
[K.003.05E] --> [S.010.2.01, B.002.06] (confidence: 0.962)
[S.010.2.01, K.003.05E] --> [B.002.06] (confidence: 0.962)
[K.003.05E] --> [S.010.2.01] (confidence: 1.000)
[S.010.1.48A, B.002.06] --> [S.010.2.01] (confidence: 1.000)
[B.002.06, K.003.05E] --> [S.010.2.01] (confidence: 1.000)
    
```

Gambar 15. Hasil dalam bentuk Description

Aturan asosiasi yang didapatkan dari hasil data mining di *Software Rapid Miner* adalah :

1. Jika membeli SI 51 maka akan membeli BHL 250L200 dan KAP HPIT 250W/ 400WE T.DLM ZX (TBL), dengan nilai persentase *confidence* adalah sebesar 50%
2. Jika membeli SI 51 maka akan membeli KAP HPIT 250W/ 400WE T.DLM ZX (TBL), dengan nilai persentase *confidence* adalah sebesar 52%

3. Jika membeli BHL 250L200 maka akan membeli KAP HPIT 250W/ 400WE T.DLM ZX (TBL), dengan nilai persentase *confidence* adalah sebesar 56,8%
4. Jika membeli BHL 250L200 maka akan membeli SI 51 dan KAP HPIT 250W/ 400WE T.DLM ZX (TBL), dengan nilai persentase *confidence* adalah sebesar 56,8%
5. Jika membeli SI 51 maka akan membeli HPI-T ESSENTIAL 250W E E40 1SL/12 dan BHL 250L200 dengan nilai persentase *confidence* adalah sebesar 60%
6. Jika membeli HPI-T ESSENTIAL 250W E E40 1SL/12 maka akan membeli BHL 250L200 dengan nilai persentase *confidence* adalah sebesar 63,8%
7. Jika membeli HPI-T ESSENTIAL 250W E E40 1SL/12 maka akan membeli SI 51 dan BHL 250L200 dengan nilai persentase *confidence* adalah sebesar 63,8%
8. Jika membeli BHL 250L200 maka akan membeli HPI-T ESSENTIAL 250W E E40 1SL/12 dengan nilai persentase *confidence* adalah sebesar 68,2%
9. Jika membeli BHL 250L200 maka akan membeli SI 51 dan HPI-T ESSENTIAL 250W E E40 1SL/12 dengan nilai persentase *confidence* adalah sebesar 68,2%
10. Jika membeli SI 51 maka akan membeli HPI-T ESSENTIAL 250W E E40 1SL/12 dengan nilai persentase *confidence* adalah sebesar 70%
11. Jika membeli SI 51 maka akan membeli BHL 250L200 dengan nilai persentase *confidence* adalah sebesar 70%
12. Jika membeli SI 51 dan BHL 250L200 maka akan membeli KAP HPIT 250W/ 400WE T.DLM ZX (TBL), dengan nilai persentase *confidence* adalah sebesar 71,4%
13. Jika membeli HPI-T ESSENTIAL 250W E E40 1SL/12 maka akan membeli SI 51 dengan nilai persentase *confidence* adalah sebesar 74,5%
14. Jika membeli BHL 250L200 maka akan membeli SI 51 dengan nilai persentase *confidence* adalah sebesar 79,5%
15. Jika membeli SI 51 dan HPI-T ESSENTIAL 250W E E40 1SL/12 maka akan membeli BHL 250L200, dengan nilai persentase *confidence* adalah sebesar 85,7%
16. Jika membeli SI 51 dan BHL 250L200 maka akan membeli HPI-T ESSENTIAL 250W E E40 1SL/12, dengan nilai persentase *confidence* adalah sebesar 85,7%
17. Jika membeli KAP HPIT 250W/ 400WE T.DLM ZX (TBL) maka akan membeli BHL 250L200, dengan nilai persentase *confidence* adalah sebesar 96,2%
18. Jika membeli KAP HPIT 250W/ 400WE T.DLM ZX (TBL) maka akan membeli SI 51 dan BHL 250L200, dengan nilai persentase *confidence* adalah sebesar 96,2%
19. Jika membeli SI 51 dan KAP HPIT 250W/ 400WE T.DLM ZX (TBL) maka akan membeli BHL 250L200, dengan nilai persentase *confidence* adalah sebesar 96,2%
20. Jika membeli KAP HPIT 250W/ 400WE T.DLM ZX (TBL) maka akan membeli SI 51, dengan nilai persentase *confidence* adalah sebesar 100%
21. Jika membeli HPI-T ESSENTIAL 250W E E40 1SL/12 dan BHL 250L200 maka akan membeli SI 51 dengan nilai persentase *confidence* adalah sebesar 100%
22. Jika membeli BHL 250L200 dan KAP HPIT 250W/ 400WE T.DLM ZX (TBL) maka akan membeli SI 51 dengan nilai persentase *confidence* adalah sebesar 100%

Semakin tinggi nilai *support* dan *confidence* maka semakin dapat dipercaya *rules* yang dihasilkan. Tetapi harus diperhitungkan juga hasilnya. Terkadang jika nilai minimum *support* ataupun *confidence* terlalu tinggi, maka akan muncul kemungkinan tidak ditemukannya *rules* yang memenuhi. Terdapat 4 produk dengan kombinasi item set yang paling sering muncul adalah BHL 250L200, KAP HPIT 250W/ 400WE T.DLM ZX (TBL), SI 51, dan HPI-T ESSENTIAL 250W E E40 1SL/12. Dari hasil aturan asosiasi tersebut dapat diambil untuk dijadikan dalam strategi pembelian produk dari *supplier* agar tetap menjaga keseimbangan stok di *Warehouse*.

4. KESIMPULAN

Data mining Apriori dapat memproses dan diimplementasikan dengan menggunakan *database* penjualan produk, karena dapat menemukan pola kombinasi antar itemset sehingga dapat dijadikan informasi yang sangat penting dalam pengambilan keputusan yang berguna untuk mempersiapkan jumlah stok produk apa yang diperlukan kedepannya.

Dari hasil perhitungan data mining menggunakan algoritma *apriori* dengan *Software* RapidMiner, data transaksi penjualan produk dengan batasan *minimum support* 20% dan *minimum confidence* 50%, membentuk 22 *rules*. Salah satu *rules* yang terbentuk adalah jika membeli BHL 250L200 dan KAP HPIT 250W/ 400WE T.DLM ZX (TBL) maka akan membeli SI 51 dengan nilai *confidence* 100%. Terdapat 4 produk dengan kombinasi *itemset* yang paling sering muncul adalah BHL 250L200, KAP HPIT 250W/ 400WE T.DLM ZX (TBL), SI 51, dan HPI-T ESSENTIAL 250W E E40 1SL/12 dari informasi ini dapat menjadi rekomendasi kepada pihak perusahaan untuk mengatur dalam pembelian stok produk dari *itemset* yang telah didapatkan.

REFERENSI

- [1] M. J. Zaki and W. Meira Jr., *Data Mining and Analysis: Fundamental Concepts and Algorithms*, United State of America: Cambridge University Press, 2014.
- [2] R. T. Vulandari, *Data Mining Teori dan Aplikasi Rapidminer*, Yogyakarta: Penerbit Gava Media, 2017.
- [3] J. J. Pangaribuan and M. Lestari, "Perbandingan Metode Moving Average (MA) Dan Neural Network Yang Berbasis Algoritma Backpropagation Dalam Prediksi Harga Saham," *Journal Information System Development (ISD)*, vol. 5, no. 1, pp. 26-34, 2020.
- [4] O. P. Barus and N. Surantha, "The Classification of Arrhythmia Using the Method of Extreme Learning Machine," *ICIC Express Letters*, vol. 14, no. 12, pp. 1147-1154, 2020.
- [5] V. Leonie and J. J. Pangaribuan, "Prediksi Video Performance Akun Youtube Buzzfeed Menggunakan Metode Naïve Bayes," *Riset dan E-Jurnal Manajemen Informatika Komputer*, vol. 5, no. 1, pp. 184-191, 2020.
- [6] G. Gunadi and D. I. Sensuse, "Penerapan Metode Data Mining Market Basket Analysis Terhadap Data Penjualan Produk Buku Dengan Menggunakan Algoritma Apriori dan Frequent Pattern Growth (FP-Growth): Studi Kasus Percetakan PT. Gramedia," *Jurnal Telematika MKOM M*, vol. 4, no. 1, pp. 118-132, 2012.
- [7] P. Bhatia, *Data Mining and Data Warehousing: Principles and Practical Techniques*, United State of America: Cambridge University Press, 2019.
- [8] G. Shmueli, P. C. Bruce, I. Yahav, N. R. Patel and K. C. Lichtendahl Jr., *Data Mining for Business Analytics: Concepts, Techniques, and Applications in R*, United State of America: Wiley, 2017.
- [9] E. Pimpler, *Data Visualization and Exploration with R: A practical guide to using R, RStudio, and Tidyverse for data visualization, exploration, and data science applications*, United State of America: CreateSpace Independent Publishing Platform, 2018.
- [10] D. O. Embarak, *Data Analysis and Visualization Using Python*, United State of America: Apress Berkeley, CA, 2018.