

Penerapan Metode *Data Mining* terhadap Data Transaksi Penjualan Menggunakan Algoritma Apriori pada Toko Metro Akustik

Lia Aprita ^{1*)}, Andreas Perdana ²⁾

¹⁾²⁾ Teknik Informatika, STMIK Dharma Wacana

Correspondence author: liaaprita5@gmail.com, Kota Metro, Indonesia

DOI: <https://doi.org/10.37012/jtik.v9i1.1381>

Abstrak

Toko Metro Akustik adalah Toko yang menjual berbagai jenis alat musik akustik. Selama ini pemanfaatan data transaksi penjualan pada Toko Metro Akustik hanya disimpan sebagai arsip. Padahal, data tersebut dapat dimanfaatkan dan diolah menjadi informasi yang berguna untuk peningkatan penjualan produk maupun inovasi produk. Dalam hal ini, perlu dilakukan analisis data transaksi untuk mendapatkan pola penjualan dengan menerapkan Algoritma Apriori. Dengan adanya informasi mengenai pola penjualan, Toko Metro Akustik dapat mengetahui apa yang paling sering dibeli oleh konsumen. Hasil dari penelitian ini adalah Algoritma Apriori mampu digunakan untuk menentukan produk yang paling sering dibeli konsumen dengan melihat kecenderungan konsumen dalam melakukan transaksi. Hasil analisis yang didapatkan setelah menggunakan perbandingan minimum *support* dan minimum *confidence* yang berbeda-beda berdasarkan data transaksi yang ada adalah dengan menggunakan minimum *support* 4% dan minimum *confidence* 60% menghasilkan dua aturan asosiasi. Informasi ini dapat memudahkan dalam menyiapkan produk yang paling sering dibeli konsumen serta menjadikan paket promo dalam menarik minat konsumen dan meningkatkan pemasaran.

Kata Kunci: *Data Mining, Support, Confidence, Algoritma Apriori, Transaksi Penjualan*

Abstract

Metro Akustik Store is a store that sells various types of acoustic musical instruments. So far, the use of sales transaction data at Metro Akustik Stores has only been stored as an archive. In fact, this data can be utilized and processed into useful information for increasing product sales and product innovation. In this case, it is necessary to analyze transaction data to obtain sales patterns by applying the Apriori Algoritma. With information about sales patterns, Toko Metro Akustik can find out what consumers buy the most. The results of this study are that the Apriori Algorithm can be used to determine the products that consumers buy the most by looking at consumer tendencies in making transactions. The results of the analysis obtained after using different minimum support and minimum confidence comparisons based on existing transaction data are using a minimum support of 4% and a minimum confidence of 60% to produce two association rules. This information can make it easier to prepare products that consumers buy most often and make promo packages to attract consumer interest and improve marketing.

Keywords: *Data Mining, Support, Confidence, Apriori Algorithm, Sales Transactions*

PENDAHULUAN

Penjualan merupakan proses jual beli baik produk maupun jasa yang kerap dilakukan pada kehidupan kita sehari-hari. Saat ini banyak sekali jenis barang yang sering diperjual belikan, contohnya pada Toko Metro Akustik yaitu Toko yang menjual berbagai jenis alat musik akustik seperti gitar, ukulele, karon dan aksesoris pendukung bermusik lainnya.

Selama ini pemanfaatan data transaksi penjualan pada Toko Metro Akustik hanya disimpan sebagai arsip. Padahal, data tersebut dapat dimanfaatkan dan diolah menjadi

informasi yang berguna untuk peningkatan penjualan produk maupun inovasi produk. Dalam hal ini, perlu dilakukan analisis data transaksi untuk mendapatkan pola penjualan. Dengan adanya informasi mengenai pola penjualan, Toko Metro Akustik dapat mengetahui apa yang paling sering dibeli oleh konsumen. Sehingga dari pola pembelian konsumen juga, pengambilan keputusan bisa dilakukan oleh pihak Toko yang berhubungan dengan produk yang akan dijual.

Kegiatan penjualan pada Toko Metro Akustik terus berjalan dan semakin banyak data yang dihasilkan. Data penjualan semakin lama akan semakin banyak. Jika dibiarkan tidak akan bermanfaat. Pengolahan data dengan cara sederhana tidak mendapatkan hasil yang efektif karena besarnya volume data yang diolah dan kesulitan untuk melihat asosiasi antara penjualan barang yang satu dengan yang lain. Pemanfaatan informasi dan pengetahuan yang terkandung dalam banyaknya data tersebut, sering disebut dengan *data mining*.

Data Mining adalah proses pengumpulan sekaligus menyaring data dengan memanfaatkan kumpulan data yang sangat besar tentu saja dengan serangkaian proses guna mendapatkan informasi yang penting dari kumpulan data tersebut. Maka dari itu agar data penjualan berguna akan dilakukan kegiatan pengumpulan data untuk menemukan jenis produk apa yang sering dibeli bersamaan dalam suatu waktu oleh konsumen dengan menggunakan Algoritma Apriori.

Algoritma Apriori pertama kali diperkenalkan oleh Agrawal dan Shrikant (1994) yang berguna untuk menentukan *frequent itemset* pada sekumpulan data. Algoritma Apriori adalah algoritma yang paling terkenal untuk menentukan pola frekuensi tinggi. Pola frekuensi tinggi merupakan pola item-item didalam suatu database yang memiliki frekuensi atau support diatas ambang batas tertentu yang disebut istilah minimum support. Pola frekuensi ini digunakan untuk menyusun aturan asosiatif dan juga beberapa teknik *data mining* lainnya. Algoritma Apriori ini sangat membantu dalam pembentukan suatu kombinasi item yang dapat dikelompokkan berdasarkan parameter yang nantinya akan menghasilkan nilai untuk membantu menentukan dalam strategi penjualan produk di toko tersebut.

Penerapan Algoritma Apriori ini membantu dalam menentukan suatu kombinasi item yang sering dibeli secara bersamaan. Kombinasi item ini dapat digunakan sebagai acuan proses penjualan produk di toko. Tujuan menggunakan Algoritma Apriori ini agar bisa membantu dalam pengambilan keputusan dalam menyusun tata letak produk, supaya produk

yang banyak dibeli diletakan ditempat yang mudah dicari dan begitu pula produk yang sering dibeli secara bersamaan. Penerapan Algoritma Apriori pada teknik data mining ini sangat mudah dipahami dalam proses pembentukan kombinasi itemset pada hasil produk-produk barang di Toko Metro Akustik.

METODE

Metode penelitian adalah seluruh metode yang dipakai selama proses penelitian. Penjelasan berikutnya tentang metode penelitian akan dijelaskan dalam bagian selanjutnya. Metode pengumpulan data yang dilakukan adalah:

1. Observasi

Melakukan observasi langsung dan mencatat secara cermat sistem kerja untuk mengumpulkan informasi dan mendapatkan informasi yang dibutuhkan, berlokasi di Toko Metro Akustik, Jl. Ki Hajar Dewantara No.14, RT.22/RW.10, Yosorejo, Kota Metro, Lampung 34124.

2. Wawancara

Teknik pengumpulan data dalam interaksi atau komunikasi langsung dengan mengajukan pertanyaan sesuai topik yang dipilih, dan data yang digunakan dalam penelitian ini adalah data sekunder. Data sekunder yang diperoleh peneliti berasal dari transaksi penjualan Toko Metro Akustik. Penggunaan data pada penelitian ini berjumlah 458 record transaksi. Atribut yang digunakan berupa item penjualan yakni Ukulele, Lignum-CE17 Original, Lignum-AE Elektrk, Lignum- B Original, Lignum-C17 Original, Lignum- D Original, Lignum- D17 Original, Lignum- E17 Original, Lignum- E17 Ukulele Original, Peking Kayu, Senar Gitar, Stand Gitar, Tali Strap Gitar, Tas Gitar, Tuner, Pick Gitar, Capo Gitar, Drayer Gitar, Equalizer, Kajian, Ketipung, Kunci L Gitar dengan format dataset .xls.

3. Studi Pustaka

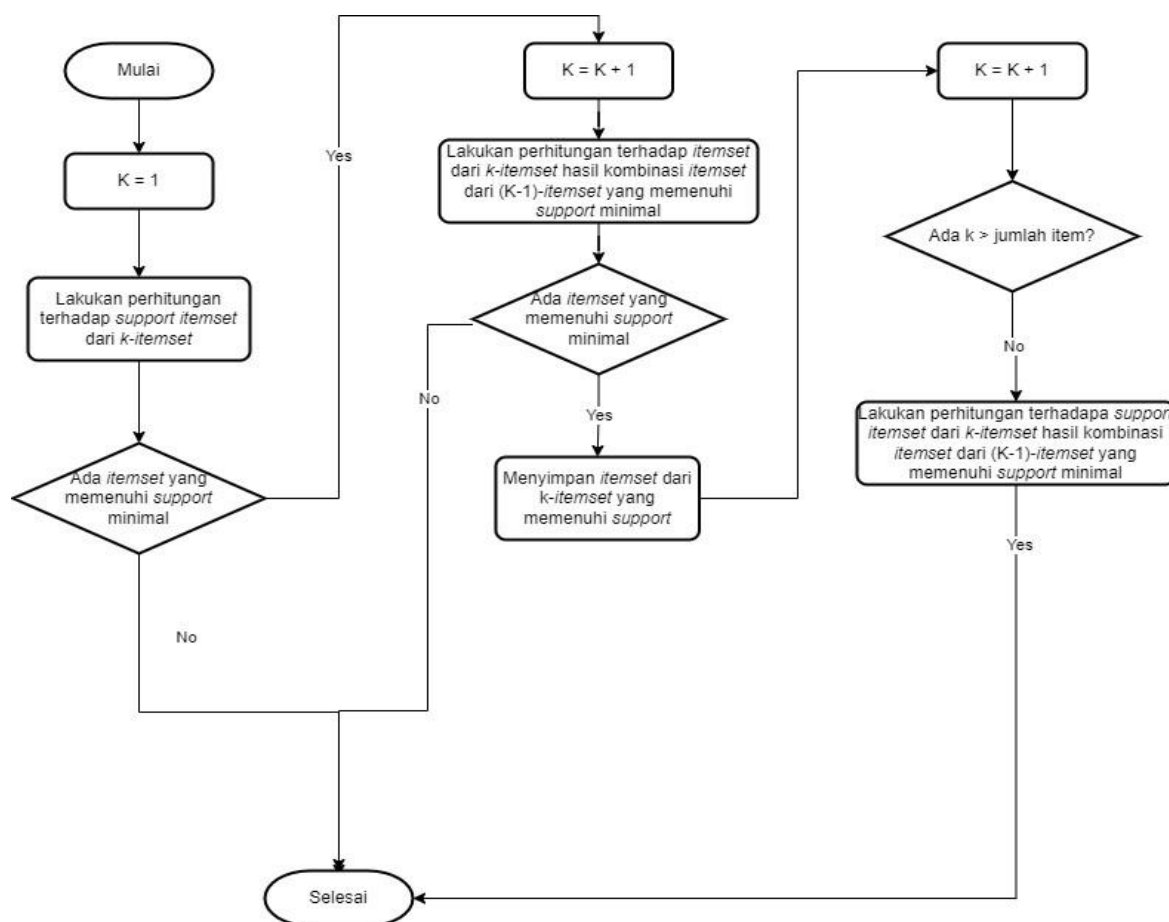
Metode ini diimplementasikan dengan mempelajari buku, jurnal akademik, dan sumber dari berbagai website terkait pemodelan data.

Metode pengolahan data adalah Algoritma Apriori. Prinsip yang digunakan dalam penelitian ini adalah prinsip Apriori Cut, dimana barang yang jarang muncul tidak diuji.

Prosedur berlanjut sesuai dengan langkah-langkah berikut:

1. Mengolah data excel dengan total 458 record kedalam bentuk tabel biner guna

- menemukan kandidat 1-itemset (C1) dan juga menghitung nilai *support*-nya. Setelah itu bandingkan antara nilai *support* dengan *minimum support* yang sebelumnya telah ditentukan, apabila nilai *support* lebih besar atau nilainya sama dengan *minimum support*, itemset terhitung dalam *large-itemset set 1* (L1).
2. *Itemset* yang tidak terhitung dalam *large-itemset* tidak dipakai untuk melakukan iterasi berikutnya. (Proses *pruning*).
 3. *Large-itemset set 1* (L1) digunakan untuk proses iterasi yang berikutnya. Pada *large-itemset set 1* (L1) dilakukan proses join pada dirinya sendiri untuk menghasilkan kandidat 2-itemset (C2). Setelah itu bandingkan nilai *support* dari semua item yang ada pada C2 dengan *minimum support*, jika nilainya lebih atau sama dengan *minimum support* maka akan masuk kedalam *large-itemset L2*. Ulangi langkah yang sama seperti mencari *large-itemset* yang sebelumnya.
 4. Analisa Algoritma Apriori menghasilkan persentase kombinasi item tersebut dengan nilai kepastian nya adalah kuatnya hubungan item kedalam aturan asosiasi produk yang akan menjadi target penjualan.



Gambar 1. Flowchart Metode Apriori

HASIL DAN PEMBAHASAN

Pada proses pemahaman data, dilakukan pengumpulan data awal. Adapun data yang digunakan untuk diproses melalui *Data Mining* adalah transaksi data penjualan konsumen pada bulan Oktober tahun 2022.

Ada 3 atribut pada *dataset* transaksi, yaitu:

1. Nomor Transaksi, atribut nomor transaksi ialah atribut yang menunjukkan nomor nomor transaksi penjualan yang terjadi.
2. Tanggal, atribut tanggal ialah atribut yang menunjukkan waktu terjadinya transaksi penjualan.
3. Nama Produk, atribut nama produk ialah atribut yang menunjukkan seluruh transaksi produk yang terjadi.

Tahap pengolahan data merupakan kegiatan untuk menyusun *dataset* akhir yaitu dengan cara memilih data, membersihkan data, menentukan atribut data yang diperlukan dan yang terakhir ialah melakukan transformasi terhadap data.

Tabel 1. Sampel data transaksi yang sudah ditransformasi

No	Produk
1	Lignum-E17 Original, Senar Gitar, Pick Gitar
2	Lignum-E17 Original, Senar Gitar, Capo Gitar
3	Lignum-D Original
4	Capo Gitar, Lignum-D Original
5	Lignum-D Original, Capo Gitar
6	Lignum-D Original
7	Lignum-C17 Original, Tali Strap Gitar, Pick Gitar, Senar Gitar
8	Lignum-C17 Original
...
...
458	Tas Gitar, Lignum-E17 Original

Data yang didapat dari Toko Metro Akustik sebelumnya berbentuk file excel (.xlsx) dan selanjutnya data dibersihkan dengan mengeliminasi atribut yang tidak digunakan untuk penelitian. Setelah data telah dibersihkan, tersisa satu atribut yang akan digunakan dalam penelitian ini yaitu Produk.

Proses pemodelan ialah memilih teknik *Data Mining* yang akan digunakan. Pada penelitian ini teknik *Data Mining* yang digunakan adalah teknik asosiasi. Proses pemodelan memiliki tujuan untuk mencari aturan asosiasi, yang mana aturan asosiasi nantinya dijadikan tolok ukur untuk melihat beberapa kombinasi item yang paling sering dibeli oleh konsumen. Hal tersebut dapat mempermudah pihak toko untuk membuat keputusan bisnis.

Langkah-langkah untuk membuat model *Data Mining* dengan menggunakan Algoritma Apriori adalah sebagai berikut:

1. Pertama, tentukan data apa saja yang ingin diproses.
2. Kedua, tentukan nilai *minimum support* dan *minimum confidence*-nya.
3. Ketiga, menyusun aturan asosiasi yang terbentuk.

Sebagai ilustrasi, dilakukan proses pencarian aturan asosiasi dengan ketentuan nilai *minimum support* adalah 8% dan *minimum confidence* 60%. Iterasi-1 dilakukan untuk membentuk kandidat 1-itemset (C1) dari data transaksi tersebut dan cari jumlah *support*-nya. Caranya ialah dengan membagi antara jumlah kemunculan item dengan jumlah semua transaksi sebanyak 458 transaksi.

$$Support (Item A) = \frac{Jumlah\ Transaksi\ Mengandung\ Item\ A}{Total\ Transaksi\ (458)}$$

Tabel 2. Kandidat 1-itemset (C1)

Nama Itemset	Support %	Banyak Transaksi
Ukulele	0,22	1
Lignum-CE17 Original	15,50	71
Lignum-AE Elektrk	0,87	4
Lignum- C17 Original	13,10	60
Lignum- D Original	2,40	11
Lignum- D17 Original	13,76	63
Lignum- E17 Original	48,25	221
Lignum- E17 Ukulele Original	2,40	11
Peking Kayu	1,09	5
Senar Gitar	15,07	69
Stand Gitar	4,80	22
Tali Strap Gitar	0,66	3
Tas Gitar	12,01	55
Tuner	0,66	3
Pick Gitar	4,37	20
Capo Gitar	8,73	40
Drayer Gitar	0,66	3
Equalizer	0,22	1
Kajon	0,87	4
Ketipung	0,44	2
Kunci L Gitar	1,75	8

Nilai *minimum support* yang telah ditentukan sebelumnya adalah 4%, jadi item yang nilai *support*-nya dibawah 4% akan dieliminasi. Large-itemset 1 yang terbentuk adalah:

Tabel 3. Large-itemset 1 (L1)

Nama Itemset	Support %
Lignum-CE17 Original	15,50
Lignum- C17 Original	13,10
Lignum- D17 Original	13,76
Lignum- E17 Original	48,25
Senar Gitar	15,07
Stand Gitar	4,80
Tas Gitar	12,01
Pick Gitar	4,37
Capo Gitar	8,73

Pada Iterasi-2 dilakukan proses cross item L1 untuk membentuk kandidat C2 (memiliki dua itemset) setelah itu cari support-nya. Pada kandidat yang item-nya sama maka dihitung satu, contoh saat itemset {Lignum-CE17 Original} digabung dengan {Lignum-CE17 Original}, hasilnya {Lignum-CE17 Original} saja bukan {Lignum-CE17 Original, Lignum-CE17 Original}. Jadi kombinasi itemset yang memiliki elemen yang sama hanya dihitung satu kali. Lakukan Iterasi berikutnya dengan cara yang sama dengan iterasi-1, dan yang didapat adalah sebagai berikut:

$$\text{Support (Item A, B)} = \frac{\text{Jumlah Transaksi Mengandung Item A dan B}}{\text{Total Transaksi (458)}}$$

Tabel 4. Kandidat 2-itemset (C2)

Nama Itemset	Support %	Banyak Transaksi
Lignum-CE17 Original, Lignum-C17 Original	0,00	0
Lignum-CE17 Original, Lignum-D17 Original	0,00	0
Lignum-CE17 Original, Lignum-E17 Original	0,00	0
Lignum-C17 Original, Lignum-D17 Original	0,44	2
Lignum-C17 Original, Lignum-E17 Original	0,44	2
Lignum-D17 Original, Lignum-E17 Original	0,22	1
Lignum-CE17 Original, Senar Gitar	0,00	0
Lignum-CE17 Original, Tas Gitar	0,00	0
Lignum-C17 Original, Senar Gitar	5,24	24
Lignum-C17 Original, Tas Gitar	3,06	14
Lignum-D17 Original, Senar Gitar	1,09	5
Lignum-D17 Original, Tas Gitar	0,00	0
Lignum-E17 Original, Senar Gitar	0,22	1
Lignum-E17 Original, Tas Gitar	0,00	0
Tas Gitar, Senar Gitar	5,68	26
Tas Gitar, Pick Gitar	0,22	1
Tas Gitar, Capo Gitar	3,06	14
Lignum-CE17 Original, Capo Gitar	0,00	0
Lignum-C17 Original, Capo Gitar	2,18	10
Lignum-D17 Original, Capo Gitar	1,09	5
Lignum-E17 Original, Capo Gitar	3,06	14
Lignum-CE17 Original, Stand Gitar	0,00	0
Lignum-C17 Original, Stand Gitar	0,66	3
Lignum-D17 Original, Stand Gitar	1,09	5
Lignum-E17 Original, Stand Gitar	1,09	5
Stand Gitar, Senar Gitar	0,44	2
Stand Gitar, Capo Gitar	1,31	6
Stand Gitar, Tas Gitar	0,87	4
Lignum-CE17 Original, Pick Gitar	0,00	0
Lignum-C17 Original, Pick Gitar	0,87	4
Lignum-D17 Original, Pick Gitar	0,00	0
Lignum-E17 Original, Pick Gitar	2,18	10
Senar Gitar, Capo Gitar	4,15	19
Senar Gitar, Pick Gitar	4,15	19
Pick Gitar, Capo Gitar	0,87	4
Pick Gitar, Stand Gitar	0,44	2

Tabel 5. *Large-itemset 2 (L2)*

Nama Itemset	Support %
Lignum-C17 Original, Senar Gitar	5,24
Tas Gitar, Senar Gitar	5,68
Senar Gitar, Capo Gitar	4,15
Senar Gitar, Pick Gitar	4,15

Pada iterasi-3 dilakukan proses cross item L2 untuk membentuk kandidat C3 (memiliki 3 *itemset*).

$$\text{Support (Item A, B, C)} = \frac{\text{Jumlah Transaksi Mengandung Item A, B, C}}{\text{Total Transaksi (458)}}$$

Tabel 6. Kandidat 3-*itemset* (C3)

Nama Itemset	Support %	Banyak Transaksi
Lignum-C17 Original, Senar Gitar, Tas Gitar	0,00	0
Lignum-C17 Original, Senar Gitar, Pick Gitar	1,53	7
Lignum-C17 Original, Tas Gitar, Pick Gitar	0,00	0
Lignum-C17 Original, Senar Gitar, Capo Gitar	0,44	2
Lignum-C17 Original, Tas Gitar, Capo Gitar	0,22	1
Lignum-C17 Original, Pick Gitar, Capo Gitar	0,00	0
Senar Gitar, Capo Gitar, Pick Gitar	0,44	2
Senar Gitar, Tas Gitar, Pick Gitar	0,00	0
Tas Gitar, Capo Gitar, Pick Gitar	0,00	0
Tas Gitar, Senar Gitar, Pick Gitar	0,00	0
Tas Gitar, Senar Gitar, Capo Gitar	0,44	2

Pada iterasi-4 tidak dilakukan proses cross item pada L3 karena tidak ada satu kandidat yang terbentuk. Jadi proses iterasi dihentikan dan tidak ada himpunan L3 yang terbentuk. Pada tabel dibawah adalah semua *large-itemset* hasil dari proses iterasi yang memenuhi nilai *minimum support* yang telah ditentukan.

Tabel 7. Tabel semua *large-itemset* hasil iterasi

Nama Itemset	Support %
Lignum-C17 Original, Senar Gitar	5,24
Tas Gitar, Senar Gitar	5,68
Senar Gitar, Capo Gitar	4,15
Senar Gitar, Pick Gitar	4,15

Setelah semua *itemsets* telah terbentuk, *itemsets* tersebut dipisah menjadi dua posisi yaitu *antecedent* yaitu sebab dari terjadinya sesuatu dan *consequent* yaitu akibat yang terjadi dari peristiwa *antecedent*, agar bisa menentukan semua kemungkinan asosiasi yang akan terbentuk. Sebagai contoh, diketahui dari salah satu contoh yang diambil, konsumen yang membeli Lignum-C17 Original memiliki kemungkinan untuk membeli Senar Gitar, akan tetapi tidak berarti konsumen yang membeli Senar Gitar juga memiliki kemungkinan untuk

membeli Lignum-C17 Original. Lignum-C17 Original disini berposisi sebagai *antecedent* dan Senar Gitar berposisi sebagai *consequent*. Berikut merupakan cara untuk menghitung *confidence*:

$$Confidence (Item A \Rightarrow B) = \frac{Support\ yang\ Mengandung\ (Item\ A,\ B)}{Support\ (Item\ A)}$$

Tabel 8. Hasil Perhitungan *Confidence*

Nama Itemset	Support Antecedent %	Support Item %	Confidence %
Lignum-C17 Original => Senar Gitar	14,8	15,5	95,48
Tas Gitar => Senar Gitar	5,68	12,01	47,29
Senar Gitar => Capo Gitar	7,86	15,07	52,16
Senar Gitar => Pick Gitar	12,88	15,07	85,47

Nilai *Minimum confidence* yang telah ditentukan sebelumnya adalah 60%, jadi item yang nilai *confidence*-nya dibawah 60% akan dieliminasi.

Tabel 9. Final Perhitungan *Confidence*

Nama Itemset	Confidence %
Lignum-C17 Original => Senar Gitar	95,48
Senar Gitar => Pick Gitar	85,47

Hasil akhir prosesnya ialah ada 2 aturan asosiasi berdasarkan parameter yang sebelumnya telah ditentukan yaitu nilai *minimum support*-nya 4% dan *minimum confidence*-nya 60%. Asosiasi aturan yang terbentuk dapat dilihat pada tabel dibawah ini.

Tabel 10. Asosiasi aturan terbentuk

Asosiasi Aturan	Confidence %
Jika konsumen membeli “Lignum-C17 Original” maka 95,48% konsumen juga membeli “Senar Gitar”	95,48
Jika konsumen membeli “Senar Gitar” maka 85,47% konsumen juga membeli “Pick Gitar”	85,47

KESIMPULAN DAN REKOMENDASI

Berdasarkan penerapan Metode Apriori pada data transaksi penjualan Toko Metro Akustik maka diperoleh kesimpulan :

1. Analisis data transaksi yang dilakukan pada Toko Metro Akustik dari rentang waktu bulan Oktober 2022 dengan jumlah transaksi sebanyak 458 data transaksi penjualan berhasil dianalisis menggunakan Algoritma Apriori untuk menentukan produk yang paling sering dibeli konsumen dengan melihat kecenderungan konsumen dalam

melakukan transaksi yaitu kombinasi (Lignum-C17 Original, Senar Gitar) dan (Senar Gitar, Pick Gitar) .

2. Hasil analisis yang didapatkan setelah menggunakan perbandingan *minimum support* dan *minimum confidence* yang berbeda-beda berdasarkan data transaksi yang ada adalah dengan menggunakan *minimum support* 4% (kuatnya kombinasi item tersebut dalam database) dan *minimum confidence* 60% (kuatnya hubungan antar item dalam aturan asosiasi) menghasilkan dua aturan asosiasi. Informasi ini dapat berguna untuk menaikkan penjualan yaitu dengan mengetahui produk apa saja yang sering dibeli oleh konsumen, sehingga dengan hal tersebut pihak Toko dapat membuat keputusan bisnis dengan membuat rekomendasi paket penjualan produk.

REFERENSI

- Badrul, M. (2016). Algoritma Asosiasi Dengan Algoritma Apriori Untuk Analisa Data Penjualan. Bodon, F. (2003). A fast APRIORI implementation. FIMI, 3, 63.
- Chang, R., & Liu, Z. (2011). An improved apriori algorithm. Proceedings of 2011 International Conference on Electronics and Optoelectronics, 1, V1-476.
- Chen, M.-S., Han, J., & Yu, P. S. (1996). Data mining: an overview from a database perspective. IEEE Transactions on Knowledge and Data Engineering, 8(6), 866–883.
- Hegland, M. (2007). The apriori algorithm—a tutorial. Mathematics and Computation in Imaging Science and Information Processing, 209–262.
- Iswandi, P., Permana, I., Salisah, F. N., Studi, P., Informasi, S., Pendahuluan, A., & Apriori, B. A. (2020). Penerapan Algoritma Apriori Pada Data Transaksi Tata Letak Barang.
- Jha, J., & Ragha, L. (2013). Educational data mining using improved apriori algorithm. International Journal of Information and Computation Technology, 3(5), 411–418.
- S. J. Tamba and E. Bu’ulolo, (2019). “Implementasi Algoritma Apriori Pada Sistem Persediaan Buah-buahan (Studi Kasus : Lotte Mart Wholesale Medan),” Oct. 2019.
- Suntoro, J. (2019). Data Mining : Algoritma dan Implementasi dengan pemrograman. PHP. Jakarta: PT Elex Media Komputindo.
- Witten Ian H., Eibe Frank, and Hall Mark A., (2011). Data Mining : Practical Machine Learning Tools and Techniques 3rd Edition. Elsevier.