

BAB II TINJAUAN PUSTAKA

A. Penelitian Terkait

Peneliti mengacu pada beberapa penelitian sebelumnya yang relevan dengan topik menggunakan metode algoritma apriori. Penelitian – penelitian ini di gunakan untuk mendalami bagaimana metode tersebut dapat membantu dalam pengambilan keputusan yang lebih baik, sebagai referensi dalam penyusunan penelitian ini, berikut beberapa penelitian terdahulu yang relevan dalam kurun waktu lima tahun terakhir

Tabel 1. Penelitian Terkait

No.	Peneliti	Keterangan Penelitian
(1)	(2)	(3)
1	Natasya Adi Pradipta, Rr. Dewi Handayani Untari N (2024) [8] [SINTA 4]	Judul : Implementasi Algoritma Apriori Untuk Analisis Pola Pembelian Produk Donat Bolong Metode : Algoritma Apriori Hasil Penelitian : Dari hasil pengujian dengan program python yang telah dilakukan dengan menggunakan Algoritma Apriori, maka dapat menarik beberapa kesimpulan pola pembelian konsumen pada hasil penelitian yaitu menentukan berdasarkan variasi rasa, ditemukan bahwa variasi rasa yang paling diminati adalah Oreo dan Taro.
2	Sonia Marselina, Jajam Haerul Jaman, Dwi Ely Kurniawan (2023) [2] [SINTA 3]	Berdasarkan hasil dan pembahasan yang dilakukan maka diperoleh kesimpulan bahwa produk-produk yang sering dibeli dari hasil pengujian association rule yang diperoleh menghasilkan kombinasi produk dengan nilai confidence 83% dan nilai support 14% yaitu, jika membeli Green tea maka akan membeli Kentang Goreng. Artinya seorang konsumen yang membeli green tea, maka 100% akan membeli kentang goreng. Aturan ini cukup

No.	Peneliti	Keterangan Penelitian
(1)	(2)	(3)
		signifikan karena mewakili 14% dari catatan transaksi yang digunakan.
3	Yumaris Alfi Alhillah, Wowon Priatna, Aida Fitriyani (2023) [9] [SINTA 3]	Judul: Implementation of Apriori Algorithm for Determining Spare Parts Product Recommendation Packages Metode: Algoritma Apriori Hasil Penelitian: Hasil data dari penelitian tersebut didapatkan dari jumlah nilai minimum support dan minimum confidence produk suku cadang, yang dihasilkan pada jenis produk suku cadang kendaraan bermotor. Karena produk yang sering dibutuhkan oleh pelanggan pada kendaraan bermotor memiliki variasi yang sangat beragam, jadi diperlukan sebuah strategi guna mengetahui pola kombinasi produk yang diminati pelanggan. Hasil yang diperoleh menunjukkan bahwa dua set item yang sering muncul adalah filter oli dengan nilai kepercayaan sebesar 68% dan filter udara dengan nilai kepercayaan sebesar 63%. Dikarenakan setiap pelanggan memiliki perbedaan pemakaian kendaraan bermotor maka 2 produk tersebut yang paling banyak diminati pelanggan.
4	Reza Aditya Permana, Primandani Arsi2, Pungkas Subarkah (2024) [10] [SINTA 3]	Judul: Penerapan Algoritma Apriori untuk Rekomendasi Produk Spare Part Mobil Metode: Algoritma Apriori Hasil Penelitian : Algoritma Apriori dapat diterapkan untuk mengidentifikasi suku cadang yang sering dibeli oleh konsumen dengan memperhatikan kebiasaan

No.	Peneliti	Keterangan Penelitian
(1)	(2)	(3)
		transaksi konsumen. Hasil analisis dengan menetapkan batas minimum support 8% dan minimum confidence 60% menghasilkan dua puluh delapan aturan asosiasi. Contohnya, jika seorang konsumen melakukan pembelian Clearance Sensor Set Q HV & V HV (Platinum White Pearl MC) maka 100% akan membeli Surround Monitor. Dari hasil aturan yang didapatkan dari data, dapat disimpulkan suku cadang apa yang sering dibeli bersamaan oleh setiap konsumen.
5	I Dewa Ayu Indah Saraswati, I Made Agus Oka Gunawan, I Made Agus Widiana Putra (2024) [11] [SINTA 4]	Judul: Analisis Keranjang Belanja pada Data Ritel Non-Toko menggunakan Algoritma FP-Growth Metode: Algoritma FP-Growth Hasil Penelitian: Analisis keranjang belanja pada dataset transaksi penjualan ritel non-toko dilakukan melalui tahapan CRISP-DM dan diproses secara menyeluruh dalam tools google colab. Dengan membandingkan berbagai nilai minimum support, ditemukan bahwa model dengan minimum support 1% (0,01) memberikan keseimbangan terbaik. Model ini membentuk 980 jumlah frequent itemsets dengan panjang kombinasi 1-itemset sebanyak 621, 2-itemset sebanyak 304, 3-itemset sebanyak 53, dan 4-itemset sebanyak 2, yang menghasilkan 954 aturan asosiasi. Evaluasi aturan asosiasi berdasarkan parameter nilai confidence menunjukkan nilai maksimum sebesar 93,91% dan minimum sebesar 9,39%. Lalu berdasarkan evaluasi parameter nilai

No.	Peneliti	Keterangan Penelitian
(1)	(2)	(3)
		lift, semua aturan asosiasi memiliki nilai > 1 , dengan maksimum sebesar 66,38 dan minimum sebesar 1,29. Hal ini menunjukkan bahwa semua aturan asosiasi kuat dan sering terjadi, di mana item B (consequent) sering dibeli bersamaan dengan item A (antecedent) dalam satu keranjang belanja oleh pelanggan.
6	Agustinus Ardiyanto Nenat, Yoseph Pius Kurniawan Kelen, Leonard Peter Gelu Budiman Baso (2025) [12] [SINTA 4]	Judul: IMPLEMENTASI ALGORITMA FP-GROWTH UNTUK MENGETAHUI POLA PENEMPATAN BUKU DI PERPUSTAKAAN DAERAH TTU Metode: ALGORITMA FP-GROWTH Hasil Penelitian: Hasil visualisasi menunjukkan bahwa buku-buku berhasil dikelompokkan ke dalam enam rak berdasarkan kesamaan karakteristik tertentu. Setiap rak memuat kumpulan buku yang homogen, menunjukkan efektivitas sistem dalam mengorganisir koleksi secara sistematis dan memudahkan proses pencarian. Selain itu, algoritma FP-Growth terbukti membantu dalam menyusun tata letak buku yang sering dipinjam sehingga proses peminjaman menjadi lebih nyaman, cepat, dan efisien. Dengan menempatkan buku-buku yang memiliki hubungan erat pada rak yang berdekatan, pengunjung dapat lebih mudah menemukan dan meminjam buku yang dibutuhkan.
7	Sherina Aulia Miranda, Fahrullah,	Judul: Implementasi Association Rule Dalam Menganalisis Data Penjualan Sheshop dengan Menggunakan Algoritma Apriori

No.	Peneliti	Keterangan Penelitian
(1)	(2)	(3)
	Deddy Kurniawan (2022) [13] [SINTA 3]	Metode: Algoritma Apriori Hasil Penelitian: produk-produk yang sering dibeli bersamaan dapat dilihat dari 6 aturan terbaik dengan nilai confidence 58% sampai dengan 75%. Berdasarkan 6 association rule yang diperoleh menghasilkan 2 produk yang sering dibeli dalam waktu bersamaan yaitu Alquran dan tasbih dengan nilai confidence 75%. Aturan kombinasi produk telah ditemukan menggunakan association rule dan telah diuji menggunakan perangkat lunak RapidMiner. Association rule dapat diterapkan untuk pertimbangan pengadaan produk oleh Sheshop.

Gap Penelitian

Meskipun algoritma apriori telah banyak digunakan dalam analisis market basket untuk mengidentifikasi pola pembelian konsumen di berbagai sektor seperti ritel modern, supermarket, dan e-commerce, penerapannya dalam konteks lingkungan Pendidikan berbasis pondok pesantren masih sangat terbatas.

Sebagian Besar penelitaian sebelumnya lebih menekankan pada konsumen umum di lingkungan perkotaan atau pasar komersial, tanpa memperhatikan karakteristik unik dari perilaku belanja santri uang di pengaruhi oleh budaya, kebijakan pondok, serta keterbatasan akses dan pilihan produk.

Selain itu, belum banyak studi yang mengkaji bagaimana pola pembelian santri dapat di manfaatkan untuk optimalisasi pengelolaan kantin pondok pesantren, baik dari segi stok barang, penyesuaian produk yang di tawarkan, maupun strategi penjualan berbasis data. Dengan demikian, penelitian ini mencoba mengisi celah tersebut dengan menerapkan algoritma apriori untuk menganalisis data transaksi di kantin pondok pesantren , sehingga dapat di temukan pola – pola belanja sepesifik santri yang berguna untuk pengambilan keputusan yang lebih efektif oleh pengelola kantin.

B. Landasan Teori

Manusia memperoleh pengetahuan dan keterampilan melalui pengalaman yang didapat sepanjang hidup, baik dari dalam dirinya sendiri maupun dari lingkungan sekitarnya. Proses pembelajaran ini juga membentuk perilaku konsumen. Ketika seseorang membeli barang, keputusan tersebut didasarkan pada kebutuhan dan keinginan yang telah dipelajari. Informasi yang diterima tentang suatu produk dapat memengaruhi tindakan dan pilihan konsumen.

Perilaku konsumen adalah ilmu yang membahas bagaimana konsumen memahami bahasa marketing, mempertimbangkan produk, melakukan keputusan beli, dan mengulang. Studi Perilaku Konsumen merupakan suatu proses ketika individu atau kelompok melakukan seleksi, membeli atau menggunakan dan membuang produk, pelayanan, ide dan pengalaman untuk memuaskan kebutuhannya, juga mendalami dampak proses tersebut pada konsumen dan masyarakat.[14]

Dalam konteks kantin pondok pesantren, perilaku santri dalam membeli produk sangat di pengaruhi oleh bebrapa faktor, seperti kebutuhan sehari-hari, harga produk, ketersediaan barang, serta kebijakan pondok yang mengatur pola konsumsi santri.

1. Pola Pembelian Konsumen

Menurut Kamus Besar Bahasa Indonesia (KBBI), pola adalah sebuah sistem atau cara kerja.[15] pola membantu kita memahami dan memprediksi fenomena yang terjadi, serta dapat di gunakan untuk membuat keputusan yang lebih baik. Dalam konteks yang lebih luas, pola dapat membantu kita mengidentifikasi hubungan antara elemen-elemen yang berbeda dan memahami bagaimana mereka berinteraksi. Dengan demikian pola dapat menjadi alat yang sangat berguna dalam berbagai aspek kehidupan.

Pembelian merupakan sinonim dari pengadaan, yang diartikan pengadaan adalah proses bisnis memilih sumber, pemesanan dan memperoleh barang dan jasa.[16] pembelian dan pengadaan sering kali di gunakan sebagai sinonim, keduanya merujuk pada proses memperoleh barang atau jasa yang di butuhkan oleh suatu organisasi atau individu. Dalam konteks bisnis, pengadaan sangat penting karena dapat mempengaruhi efisiensi dan efektivitas operasional perusahaan. Dengan melakukan pengadaan yang efektif, perusahaan dapat memperoleh barang dan jasa yang di butuhkan dnegan harga yang kompetitif dan kualitas yang baik. Meskipun keduanya memiliki makna yang mirip, ada sedikit perbedaan antara pembelian dan pengadaan. Pembelian lebih fokus pada proses transaksi, sedangkan pengadaan lebih luas dan mencakup seluruh proses bisnis, mulai dari identifikasi kebutuhan hingga pembayaran. Dalam prakteknya, kedua istilah ini sering kali di gunakan secara bergantian, tetapi pengadaan lebih umum di gunakan dalam konteks bisnis dan organisasi, sedangkan pembelian lebih umum di gunakan dalam konteks individu.

Pola pembelian adalah kebiasaan atau kecenderungan konsumen dalam memilih dan membeli produk berdasarkan faktor-faktor tertentu, seperti kebutuhan, preferensi, harga, dan promosi.[14]

Penelitian menunjukkan bahwa lingkungan Pendidikan, seperti sekolah maupun pesantren, cenderung membuat santri memilih produk yang bersifat praktis, mengenyangkan, dan memiliki harga terjangkau.

2. Data Dan Informasi

Dalam melakukan data mining untuk menemukan pola dalam dataset, penggunaan data sangatlah penting sebagai bahan rujukan bagi algoritma itu sendiri. Menurut Lubis[17] data adalah fakta-fakta yang menggambarkan suatu kejadian yang sebenarnya pada waktu tertentu. Ketika berurusan dengan analisis asosiasi dalam data mining, data merupakan Kumpulan fakta-fakta yang mencerminkan berbagai kejadian atau transaksi yang terjadi. Fakta-fakta ini kemudian diorganisir dan dikelompokkan ke dalam sebuah dataset, yang kemudian akan menjadi landasan bagi algoritma untuk menemukan pola asosiasi di antara berbagai item dalam data tersebut. Hasil akhir dari pengolahan tersebut menjadi informasi baru bagi orang yang membacanya.

Informasi merupakan hal yang sangat mendasar yang diperlukan dalam pengambilan keputusan untuk mencegah bias dari pengambil keputusan dan mengurangi potensi kesalahan. Informasi juga dapat diartikan sebagai data yang telah diolah menjadi bentuk yang lebih bermakna dan berguna bagi penerima informasi tersebut. Sebagaimana yang diungkapkan oleh Anggraeni [18], informasi didefinisikan sebagai sekumpulan data atau fakta yang diorganisasi atau diolah dengan cara tertentu sehingga memiliki arti bagi penerima. Dengan pengolahan informasi, data menjadi bahan mentah yang kemudian diolah menjadi informasi yang lebih berguna sebagai landasan untuk mengambil keputusan yang lebih baik.

3. Konsep Data Mining

Data mining merupakan proses iterative dan interaktif untuk menemukan pola atau model baru yang sempurna, bermanfaat dan dapat dimengerti dalam suatu database yang sangat besar (massive database). Data mining berisi pencarian trend atau pola yang diinginkan dalam database besar untuk membantu pengambil keputusan diwaktu yang akan datang, pola-pola ini dikenali perangkat tertentu yang dapat memberikan suatu analisa data yang berguna dan berwawasan yang kemudian dapat dipelajari dengan lebih teliti, yang mungkin saja menggunakan perangkat pendukung keputusan yang lain.[19]

Data mining sering juga disebut sebagai KDD, karena prosesnya meliputi pengumpulan, pengelompokan, pemakaian data untuk menemukan suatu keteraturan, pola atau hubungan dalam dataset yang sangat besar . Menurut Budi Santosa pada bukunya yang berjudul "Data Mining : Teknik Pemanfaatan

Data Untuk Keperluan Bisnis” secara garis besar, tahapan – tahapan dalam data mining dapat diuraikan sebagai berikut [20]:

a. *Data Cleaning*

Data Cleaning adalah Proses membersihkan data dari data noise dan tidak konsisten [21]. Data cleaning bertujuan untuk membuang data-data yang tidak cocok terhadap teknik/metode data mining yang akan digunakan juga melengkapi data yang hilang dari dataset/database tersebut supaya performa data mining dapat meningkat.

b. *Data Integration*

Data integration adalah proses untuk menggabungkan data dari berbagai sumber yang berbeda [21]. Data yang akan diproses dalam data mining berasal dari banyak database. Integrasi data diperlukan untuk menyatukan database menjadi satu kesatuan database baru. Integrasi data yang hasilnya diteliti dapat meningkatkan akurasi dan kecepatan dari proses data mining.

c. *Data Transformation*

Data transformation adalah proses merubah bentuk data menjadi data yang sesuai untuk proses mining [21]. Pada tahap transformasi data, data yang digunakan akan diubah menjadi format yang sesuai dengan metode yang digunakan pada proses data mining.

d. *Data Mining*

Data mining adalah proses penting yang menggunakan sebuah metode tertentu untuk memperoleh sebuah pola dari data [21]. Tahap ini merupakan tahapan dimana kita bisa mulai melakukan penerapan metode data mining terhadap data yang telah diproses dan disortir sebelumnya untuk mencari pola dan informasi yang tersembunyi di dalam data tersebut.

e. *Pattern Evaluation*

Pattern Evaluation adalah proses mengidentifikasi pola [21]. Pola yang dihasilkan dari proses data mining bisa jadi berbeda dari hipotesis yang telah dibuat. Jika hal ini terjadi, maka hal yang harus kita lakukan adalah melakukan evaluasi pola terhadap hasil yang ada dan memperbaikinya. Contoh solusinya bisa dengan mengubah metode yang digunakan ataupun menerima hasil yang diberikan sistem sebagai pengetahuan baru.

f. *Knowledge Presentation*

Knowledge Presentation adalah yang dapat merepresentasikan informasi yang di butuhkan, proses di mana informasi yang telah di dapatkan kemudian digunakan oleh pemilik data [21]. Pada tahap akhir proses data mining ini, informasi atau pengetahuan baru yang telah dihasilkan akan disajikan terhadap stakeholder yang bersangkutan. Hal ini dikarenakan tidak setiap stakeholder memahami hasil yang ditemukan, maka yang perlu kita lakukan adalah menyajikannya dengan baik supaya dapat dipahami dengan baik juga oleh

stakeholder. Salah satunya adalah memvisualisasikan data pada saat presentasi hasil proses data mining itu sendiri.

Dalam konteks kantin pondok pesantren, data mining dapat di gunakan untuk mengidentifikasi kebiasaan pembelian santri sehingga pengelola kantin dapat mengoptimalkan persediaan barang dan strategi pemasaran.

4. Analisis Asosiasi

Analisis asosiasi adalah salah satu teknik dalam data mining yang menjadi dasar dari salah satu teknik data mining lainnya untuk menemukan aturan asosiasi antar suatu kombinasi item [22]. Analisis asosiasi juga didefinisikan sebagai suatu proses untuk menemukan pola yang sering muncul, hubungan antara suatu kombinasi item dan saling membuat korelasi antara item yang telah dikelompokkan (itemset) atau “muncul” secara bersamaan kemudian mengambil kesimpulan berdasarkan hubungan yang terbentuk dari beberapa item juga itemset tadi. Dibatasi dengan aturan “If antecedent, then consequent” bersama dengan ukuran support dan confidence yang menjadi parameter untuk mendapatkan aturan yang optimal [23].

Terdapat dua metode utama dalam merepresentasikan tipe data dalam MBA, yaitu format data transaksi dan format data tabular. Format data transaksi membutuhkan dua field, berupa field ID dan field item, dimana tiap record merepresentasikan hanya satu item. Berikut contoh format data transaksi sebagaimana ditampilkan pada Tabel 2.1 (Yusuf et al., 2006).

Tabel 2 Format Item transaksi

Field ID	Field Item
1	Item 1
1	Item 2
2	Item n
....	
..	..

Pada format data tabular, setiap catatan mewakili transaksi secara biner dengan mengandung nilai 0 atau 1 sesuai dengan keberadaan atau ketiadaan item tersebut. Format data tabular dapat diilustrasikan melalui Tabel di bawah.

Tabel 3 contohh format tabular

ID_Transaksi	Item 1	Item 2		Item n
1	1	1	0	0
2	0	0	0	1
....	...	...	...	...

n	...	...	...	...
---	-----	-----	-----	-----

pada record terdapat ID_Transaksi 1 yang memilih item 1 dan 2, diindikasikan dengan nilai 1 pada keduanya. ID_Transaksi 3 yang tidak memilih item 1 ditunjukkan dengan nilai 0 pada kolom tersebut, tetapi pada kolom item 2 untuk ID_Transaksi 3, terdapat nilai 1 yang menunjukkan bahwa item tersebut diambil. Dengan demikian, nilai 0 atau 1 di sini mencerminkan apakah seorang pelanggan (ID_Transaksi) mengambil item tersebut atau tidak dalam transaksi belanja mereka[23].

5. Aturan Asosiasi

Aturan Asosiasi atau disebut dengan association rules merupakan salah satu teknik data mining yang berguna dalam mencari aturan asosiatif dari suatu barang. Aturan Asosiasi dibentuk dengan cara menganalisis pola data yang sering muncul (frequent pattern) dan dengan menggunakan parameter support dan confidence untuk mengidentifikasi hubungan yang paling penting dan sesuai.

Support adalah indikasi yang menunjukkan seberapa sering suatu item muncul di dalam suatu database. Sedangkan Confidence menunjukkan berapa kali pernyataan tersebut benar. [5]

Berikut merupakan penjelasan mengenai support dan confidence pada aturan asosiasi :

a. Support

Support adalah suatu ukuran yang menunjukkan seberapa besar tingkat dominasi suatu item/item set dari keseluruhan transaksi. Ukuran ini akan menentukan apakah suatu item/item set layak untuk dicari confidence selanjutnya atau tidak.

$$support(A) = \frac{\text{Jumlah Transaksi mengandung A}}{\text{Total transaksi}}$$

Sedangkan nilai support dari 2 item di peroleh dari rumus berikut

$$support(A \cap B) = \frac{\text{Jumlah transaksi mengandung A dan B}}{\text{Total Transaksi}}$$

Pada tahapan analisis asosiasi, terdapat istilah minimum support. Minimum support didefinisikan sebagai nilai persentase yang digunakan untuk menyaring aturan-aturan yang memiliki nilai support di bawah ambang batas tersebut. Penentuan nilai minimum support perlu disesuaikan dengan tujuan analisis dan karakteristik dataset. Jika nilai minimum support terlalu tinggi, analisis dapat kehilangan aturan-aturan yang mungkin memiliki nilai

informatif. Sebaliknya, jika nilai minimum support terlalu rendah, dapat menyebabkan banyak aturan yang tidak relevan atau terlalu umum.

Oleh karena itu, pemilihan nilai minimum support memerlukan kebijaksanaan dan pemahaman yang baik terhadap konteks analisis dan karakteristik data yang sedang diteliti. Batasan yang tepat untuk minimum support memungkinkan analisis asosiasi memberikan aturan-aturan yang bermakna dan sesuai dengan kebutuhan penulis.

b. Confidence

Confidence adalah suatu ukuran yang menunjukkan hubungan antar 2 (dua) item secara conditional seperti: seberapa sering item A dibeli jika orang membeli item B

$$\text{confidence } p(B/A) = \frac{\text{jumlah transaksi mengandung A dan B}}{\text{Jumlah Transaksi mengandung A}}$$

Dalam gambaran umum, Association Rule Mining memiliki dua tahap proses yaitu:

- 1) Temukan semua frequent item set : anggota frequent item set setidaknya memiliki jumlah dalam database sama dengan minimum support yang telah ditentukan di awal.
- 2) Temukan semua aturan asosiasi yang kuat dari frequent item set aturan asosiasi yang dihasilkan harus memenuhi minimum support dan minimum confidence.

Dengan menggunakan persamaan tersebut, kita dapat mengevaluasi seberapa kuat asosiasi antara item A dan B berdasarkan frekuensi kemunculan keduanya dalam transaksi. Semakin tinggi nilai confidence, semakin besar kemungkinan item B akan dibeli bersamaan dengan A dalam suatu transaksi.

Pada tahapan analisis asosiasi, penetapan nilai minimum confidence sangat penting karena membantu penulis untuk mengontrol kualitas dari aturan asosiasi yang dihasilkan. Confidence mengukur seberapa kuat hubungan antara antecedent dan consequent pada suatu aturan. Nilai confidence yang tinggi menunjukkan bahwa aturan tersebut cenderung lebih kuat dan dapat diandalkan.

Penetapan nilai minimum confidence dilakukan oleh penulis berdasarkan batasan atau standar tertentu. Nilai ini dapat bervariasi tergantung pada tujuan penelitian, domain data, atau preferensi peneliti. Sebagai contoh, jika penelitian memiliki tujuan untuk mendapatkan aturan yang sangat kuat, nilai minimum confidence dapat ditetapkan lebih tinggi. Sebaliknya, jika penelitian ingin mempertahankan sejumlah besar aturan termasuk yang mungkin kurang kuat, nilai minimum confidence dapat ditetapkan lebih rendah.

c. *Lift*

Nilai Lift merupakan parameter yang digunakan untuk menilai kekuatan dari aturan asosiasi yang terbentuk berdasarkan nilai support dan confidence. Selain itu, nilai lift juga berperan sebagai indikator validitas suatu aturan asosiasi. Sebagai contoh, Lift (A, B) menunjukkan kekuatan hubungan antara item A dan B. Nilai lift juga menjadi suatu ukuran untuk mengetahui kekuatan aturan asosiasi yang telah terbentuk [24]. Rentang nilai lift mulai dari 0 sampai ∞ (tak hingga). Dengan menggunakan rumus sebagai berikut.

$$Lift(A, B) = \frac{Confidence(A, B)}{Benchmark\ Confidence(A, B)} \cdot 100\%$$

Nilai Lift (A, B) didapatkan dengan pembagian nilai Confidence (A, B) dan Benchmark Confidence (A, B). Selanjutnya, nilai benchmark confidence dapat dihitung dengan rumus sebagai berikut.

$$\begin{aligned} & \text{Benchmark Confidence (A, B)} \\ &= \frac{\text{Jumlah transaksi mengandung B}}{\text{Total Transaksi}} \cdot 100 \end{aligned}$$

Benchmark confidence adalah perbandingan dari jumlah item consequent dengan jumlah transaksi dalam dataset. Dengan menggunakan persamaan tersebut, kita dapat mengevaluasi seberapa besar peningkatan kemungkinan kemunculan item B ketika item A sudah ada dalam transaksi, dan nilai benchmark confidence memberikan gambaran referensi terkait kemungkinan pembelian item B secara umum.

6. Algoritma Apriori

Algoritma apriori merupakan suatu algoritma dasar yang diusulkan oleh Agrawal dan Srikant pada tahun 1994 dalam penentuan frequent itemsets untuk membuat aturan asosiasi [25]. Apriori adalah algoritma yang sering digunakan untuk menambang itemset menggunakan aturan asosiasi yang bersifat boolean [26]. Output dari algoritma apriori adalah aturan asosiasi yang menggambarkan hubungan atau pola yang ditemukan dalam dataset. Aturan asosiasi ini biasanya terdiri dari dua bagian utama, yaitu antecedent (jika) dan consequent (maka). Aturan ini menyatakan bahwa jika suatu kondisi tertentu terpenuhi (antecedent), maka ada kecenderungan atau asosiasi untuk kondisi lain terpenuhi juga (consequent).

Aturan asosiasi adalah pola-pola asosiatif yang memenuhi syarat nilai minimal dari dua parameter yaitu support dan confidence [27]. Aturan asosiasi akan menyatakan asosiasi dari item yang ada dalam aturan tersebut. Aturan asosiasi berbentuk “If antecedent, then consequent”, dilengkapi dengan tingkat

support dan confidence aturan tersebut [23]. Algoritma apriori adalah sebuah algoritma yang digunakan untuk menemukan item dan itemset yang relevan dengan tujuan untuk menentukan aturan asosiasi. Dalam data mining, algoritma apriori termasuk dalam kategori teknik analisis asosiasi yang digunakan untuk menemukan sebuah kombinasi tiap item. Dengan teknik ini, penulis dapat menghitung seberapa sering kombinasi item muncul dan kemudian menggunakan informasi ini untuk tujuan prediksi dari pola yang muncul [28].

Output lainnya dari algoritma apriori bisa mencakup informasi mengenai nilai support dan confidence dari setiap aturan asosiasi. Nilai support mengukur seberapa sering aturan tersebut ditemukan dalam dataset, sementara confidence mengukur seberapa kuat keterkaitan antara antecedent dan consequent pada aturan tersebut.

7. Cara kerja Algoritma Apriori

Menurut Han [29] dalam penelitiannya, algoritma aprori memiliki dua proses utama yang harus dilakukan, yaitu:

a. Join (pengabungan)

Pada proses ini setiap item dikombinasikan dengan item yang lainnya secara terus menerus hingga tidak dapat terbentuk kombinasi baru kembali.

b. Prune (Pemangkasan)

Pada proses ini hasil dari item yang telah dikombinasikan sebelumnya, akan dipangkas dengan menggunakan parameter minimum support yang telah ditentukan oleh penulis. Minimum support adalah batas ambang tertentu nilai support yang ditentukan oleh penulis untuk mem-filter apakah suatu itemset termasuk dari frequent itemset dengan besaran nilai support tertentu.

Langkah – langkah algoritma Apriori

Langkah pertama algoritma apriori adalah support dari setiap item dihitung dengan cara melakukan scanning pada database/dataset. Setelah support dipilih sebagai frequent itemset dengan panjang 1 atau bisa disingkat sebagai 1-itemset. Kita melakukan iterasi yang menghasilkan k-itemset yang berarti satu set yang terdiri dari k item. (k) disini mewakili dari jumlah item dalam satu set.

Pada iterasi kedua menghasilkan 2-itemset yang setiap set-nya terdiri dari dua item. Prosesnya yaitu pertama dibuatkan kandidat 2-item dari kombinasi semua 1-itemset. Lalu, untuk setiap kandidat 2 itemset ini dihitung nilai support-nya dengan melakukan scanning pada database/dataset. Nilai support ini adalah jumlah transaksi dalam database/dataset yang mengandung kedua item dalam kandidat 2-itemset. Setelah nilai support dari setiap kandidat 2-itemset didapatkan, kandidat 2-itemset yang memenuhi syarat minimum support dapat ditetapkan sebagai 2-itemset yang juga merupakan frequent

itemset dengan panjang 2. Hal ini akan terus berulang sampai iterasi ke-k yang dapat dibagi menjadi beberapa bagian sebagai berikut:

a. Pembentukan Kandidat Item set

Kandidat k-itemset dihasilkan dengan menggabungkan (k-1) itemset yang diperoleh dari iterasi sebelumnya. Salah satu karakteristik algoritma apriori adalah terdapat pemangkasan kandidat k-itemset yang memiliki subset (k-1) item yang tidak termasuk dalam pola frequent itemset dengan panjang (k-1).

b. Perhitungan nilai support dari tiap kandidat k-itemset

Nilai support untuk setiap kandidat k-itemset dihitung dengan memeriksa seluruh database/dataset untuk menghitung berapa banyak transaksi yang memuat semua item yang ada di dalam kandidat k-itemset tersebut. Algoritma apriori dicirikan dengan kebutuhan untuk melakukan perhitungan dengan melakukan scanning pada seluruh database/dataset sebanyak jumlah item yang ada pada kandidat k-itemset terpanjang.

c. Penetapan frequent itemsets

Frequent itemset yang berisi k-item atau k-itemset dipilih dari kandidat k-itemset yang memiliki nilai support lebih tinggi dari batas minimum support yang telah ditetapkan sebelumnya. Kemudian nilai confidence dari setiap kombinasi item dihitung.

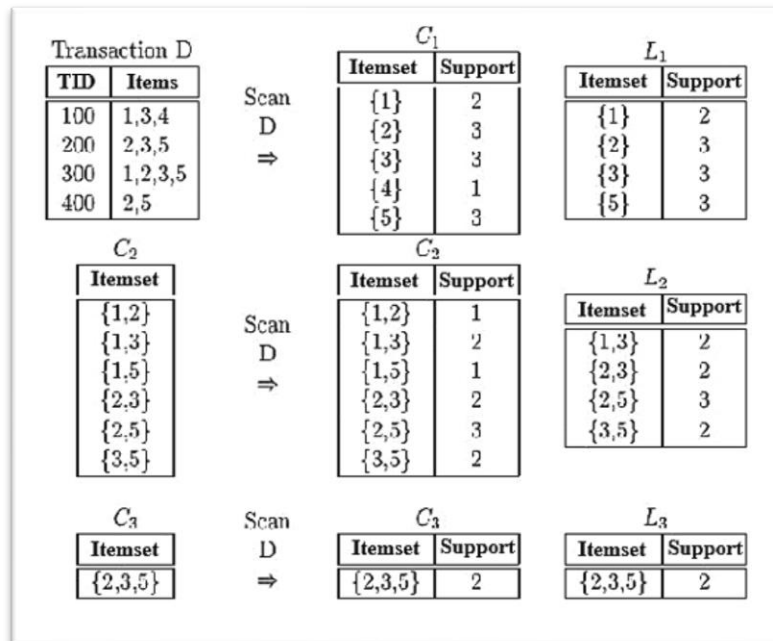
d. Akhir Iterasi

Jika tidak ditemukan frequent itemset baru maka seluruh proses dihentikan. Sebaliknya, jika masih ada, maka k ditambah satu dan kembali ke bagian 1.

Secara ringkasnya langkah langkah algoritma apriori adalah sebagai berikut :

1. Scanning dataset / Dataset
2. menghitung nilai support dari masing-masing item
3. Pruning, yaitu membuang itemset yang tidak memenuhi minimum support, dan melanjutkan kandidat itemset yang memenuhi batas minimum support pada scan berikutnya.
4. Kombinasi semua Kandidat
5. Mengulangi langkah 1 sampai 4 hingga tidak ditemukannya lagi frequent itemset.

Ilustrasi penerapan algoritma apriori



Gambar 1 ilustrasi penerapan algoritma apriori

Gambar 2 ilustrasi penerapan algoritma apriori

Berdasarkan ilustrasi yang buat oleh Agrawal dan Srikant , batas minimum support yang digunakan dalam contoh ini adalah sebesar 50% atau setara dengan nilai 2. Pada tahap iterasi pertama, item yang memiliki support kurang dari nilai 2 dihapuskan dari himpunan 1-itemset L_1 . Setelah itu, kandidat 2-itemset C_2 pada iterasi kedua terbentuk melalui permutasi dari item-item yang tersisa di L_1 . Setelah menghitung support untuk kandidat 2-itemset dari database, terbentuklah 2-itemset L_2 . Proses serupa diulangi pada iterasi selanjutnya, dan langkah-langkah tersebut akan berlanjut hingga tidak ada kandidat baru.

Berdasarkan ilustrasi yang buat oleh Agrawal dan Srikant , batas minimum support yang digunakan dalam contoh ini adalah sebesar 50% atau setara dengan nilai 2. Pada tahap iterasi pertama, item yang memiliki support kurang dari nilai 2 dihapuskan dari himpunan 1-itemset L_1 . Setelah itu, kandidat 2-itemset C_2 pada iterasi kedua terbentuk melalui permutasi dari item-item yang tersisa di L_1 . Setelah menghitung support untuk kandidat 2-itemset dari database, terbentuklah 2-itemset L_2 . Proses serupa diulangi pada iterasi selanjutnya, dan langkah-langkah tersebut akan berlanjut hingga tidak ada kandidat baru.