



ANALISIS ALGORITMA FP-GROWTH DAN APRIORI UNTUK MENEMUKAN MODEL ASOSIASI TERBAIK PADA *DATASET* ONLINE RETAIL

Meirynda Lastika Rahimsyah, Yudi Ramdhani
Universitas Adhirajasa Reswara Sanjaya, SATU University
meirynda31@gmail.com, yudi@ars.ac.id

Abstract— In the digital era, the online retail industry is growing rapidly and is becoming an important sector. However, challenges arise in the analysis of sales transaction data on the Online Retail dataset. This study aims to overcome problems in the analysis of sales transaction data in the Online Retail dataset. The main focus includes selecting the optimal association algorithm between FP-Growth and Apriori, identifying relevant association models on complex datasets, and the efficiency and performance of algorithms in processing sales transaction data. The method used is association data processing using the FP-Growth and Apriori algorithms. Implementation of the association rule involves adding a lift metric as a measure of association strength. Measurement of processing time is also carried out to determine the efficiency of implementation. The results showed that FP-Growth and Apriori could produce an association model with the same frequent itemset and value matrix, namely a support value of 0.12 and a confidence value of 0.96, but there were differences in the resulting model order. The Apriori algorithm produces a model with the highest support value at index 18, while FP-Growth at index 10. In addition, the FP-Growth algorithm shows an advantage in faster processing time (0.004 seconds) compared to Apriori (0.007 seconds). This research provides a better understanding of the use of association algorithms in the context of the online retail industry.

Keywords— *Apriori Algorithm, Association, Association rule, FP-Growth Algorithm, Retail online.*

I. INTRODUCTION

Dalam sebuah perusahaan *retail*, terjadi banyak sekali transaksi setiap harinya. Sejumlah besar data yang dihasilkan selama seminggu atau bahkan satu tahun menyimpan jutaan catatan dalam *database* perusahaan [1], [2]. Untuk menentukan strategi bisnis yang tepat, faktor penting yang harus diperhatikan adalah kecepatan dan ketepatan pengambilan keputusan. Akibatnya, eksekutif tidak bisa lagi hanya mengandalkan intuisi dan harus mempertimbangkan data yang ada. Setiap kegiatan operasional perusahaan selalu dilakukan pencatatan dan didokumentasikan, sehingga membentuk data yang tersimpan dalam *database*. Data tersebut akan terus bertambah seiring waktu. Dalam konteks ini, penting bagi perusahaan untuk melaksanakan praktik pengelolaan data yang baik guna memastikan akurasi dan kepercayaan data yang disimpan. Hal ini bertujuan agar data dapat diandalkan



sebagai dasar yang *valid* dalam pengambilan keputusan yang tepat. Dengan data yang akurat dan dapat dipercaya, perusahaan dapat mengurangi risiko kesalahan dalam pengambilan keputusan, meningkatkan efisiensi operasional, dan memaksimalkan hasil yang diinginkan [2].

Salah satu aspek penting dalam informasi transaksional adalah pemahaman terhadap pola belanja pelanggan saat melakukan pembelian barang. Dengan pemahaman ini, perusahaan dapat mengidentifikasi produk-produk yang dapat ditawarkan kepada pelanggan secara bersamaan. Hal ini memungkinkan perusahaan untuk meningkatkan strategi penjualan, mengoptimalkan penawaran produk, dan memaksimalkan kesempatan penjualan kepada pelanggan [3].

Salah satu cara untuk mengumpulkan informasi tentang kebiasaan belanja pelanggan untuk produk ritel adalah dengan mengamati pola pembelian kolektif mereka. Dengan metode ini, dimungkinkan untuk memahami kecenderungan pembelian setiap pelanggan pada waktu tertentu. Pengetahuan ini dapat membantu perusahaan mengatasi masalah yang sering muncul ketika barang yang biasanya dibeli bersamaan kehabisan *stock*, yang berujung pada keluhan pelanggan [3]. Namun, identifikasi pola dari data penjualan *retail online* merupakan tugas yang kompleks karena melibatkan banyak informasi yang terdapat dalam data penjualan. Proses ini memerlukan teknik yang tepat agar hasil yang diperoleh konsisten dan relevan bagi pelaku bisnis, tidak hanya penjual *retail* saja. Oleh karena itu, diperlukan pendekatan atau metode yang efektif untuk mengelola dan menganalisis kumpulan data penjualan tersebut [4], [5].

Analisis data penjualan oleh pelaku bisnis bertujuan untuk mendapatkan pola asosiasi antara *item* yang sering secara bersamaan dibeli dalam transaksi penjualan, proses ini umumnya dikenal sebagai *data mining*. Tergantung pada tujuan yang ingin dicapai, *data mining* dapat diklasifikasikan ke dalam beberapa kategori, seperti eksplorasi deskriptif, estimasi, prediksi, klasifikasi, pengelompokan, dan asosiasi [6]. *Data mining* merupakan proses analisis data yang menerapkan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* ke berbagai sumber data besar untuk mengekstraksi dan mengidentifikasi informasi berharga dan pengetahuan terkait [7], [8]. Proses ini memungkinkan untuk mengungkap pola, hubungan, dan asosiasi laten yang dapat memberikan wawasan penting untuk pengambilan keputusan, perencanaan strategis, dan pemahaman yang lebih dalam tentang data yang dimiliki. Dengan bantuan *data mining*, dapat mengoptimalkan kinerja bisnis, meningkatkan efisiensi operasional, dan mengidentifikasi peluang baru di berbagai sektor, termasuk bisnis, kesehatan, pendidikan, dan lainnya, dengan bantuan *data mining* [8]. *Data mining* merupakan proses yang serbaguna dengan kemampuan untuk diarahkan ke berbagai tujuan. Bergantung pada sasaran yang ingin dicapai, *data mining* dapat diklasifikasikan menjadi beberapa kategori utama, antara lain eksplorasi deskripsi, estimasi, prediksi, klasifikasi, pengelompokan, dan asosiasi [6].



MBA (*Market Basket Analysis*) adalah metode *data mining* yang sering dipakai untuk menemukan koneksi di antara *set data* yang besar. Tujuan utama dari metode ini adalah mengidentifikasi kombinasi produk yang dibeli oleh pelanggan secara bersamaan [9], [10]. Dalam *data mining*, teknik asosiasi digunakan dengan tujuan untuk mengidentifikasi aturan asosiasi yang melampaui minimum *support* dan *confidence* [3], [11]. Teknik asosiasi berperan dalam meningkatkan efektivitas penjualan dengan menghubungkan data transaksi pembeli dan memungkinkan pendeteksian tren pembelian konsumen. Informasi ini memungkinkan pelaku bisnis untuk melakukan langkah bisnis yang tepat [12]. Dalam teknik asosiasi, algoritma *frequent itemset* digunakan untuk mencari hubungan barang sebelum membentuk aturan asosiasi. Algoritma ini mengidentifikasi *item* atau barang yang sering muncul bersama dalam *dataset*, membantu mengungkap pola-pola asosiasi yang relevan. Penggunaan algoritma *frequent itemset* penting dalam memahami hubungan antarbarang yang kuat dan mendukung pengambilan keputusan yang cerdas serta pengembangan strategi bisnis yang efektif [3]. Pada penelitian ini, algoritma *FP-Growth* dan algoritma *Apriori* digunakan sebagai metode analisis data. Dengan penerapan kedua algoritma tersebut, dilakukan identifikasi *itemset* yang muncul bersamaan dalam himpunan data, yang selanjutnya digunakan sebagai dasar untuk menentukan *rule association* dalam analisis *market basket* [13].

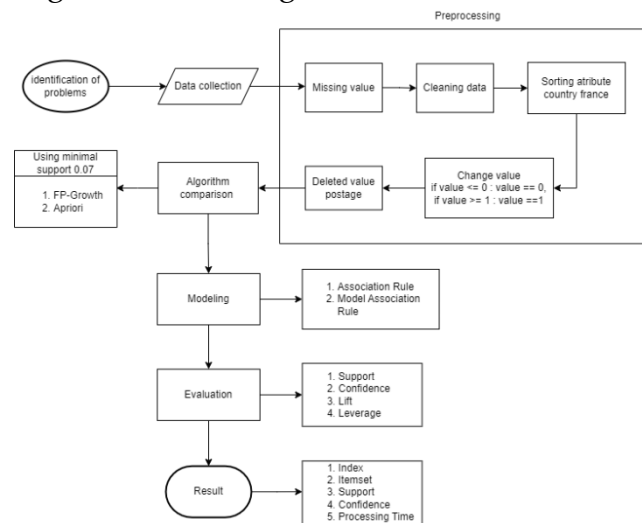
Metode *Apriori* merupakan sebuah algoritma yang digunakan untuk menemukan pola asosiasi antara *item* pada sebuah himpunan data. Algoritma *Apriori* sebagian besar digunakan dalam pemrosesan data dalam bisnis pasar, umumnya diterapkan untuk menemukan *frequent itemset* dengan pola tertinggi dalam kumpulan data. *Frequent itemset* merupakan pola *item* yang dapat ditemukan dalam kumpulan data dengan melampaui nilai *support* dan *confidence* yang telah ditetapkan sebagai *threshold* atau batasan minimal, *Itemset* ini juga sering digunakan dalam pembentukan aturan asosiasi [4], [14]. Dengan memanfaatkan nilai minimum *support* yang sebelumnya sudah ditentukan, algoritma *Apriori* mampu mengidentifikasi pola frekuensi tinggi yang dapat memberikan analisis data yang mendalam [15].

Metode *FP-Growth* adalah perkembangan dari algoritma *Apriori*, pada algoritma ini tidak menghasilkan kandidat karena menggunakan pendekatan pembangunan pohon (*tree*) untuk mencari *frequent itemset*. Dengan mengeliminasi langkah *generate candidate*, algoritma *FP-Growth* memberikan efisiensi yang lebih baik dalam *data mining* asosiasi, terutama pada *dataset* dengan jumlah *item* yang besar [5]. Faktor ini menghasilkan algoritma *FP-Growth* memiliki kecepatan yang lebih tinggi dibandingkan dengan algoritma *Apriori*. Keuntungan *FP-Growth* terletak pada jumlah pemindaian *database* yang hanya dilakukan satu atau dua kali, sedangkan *Apriori* memerlukan pemindaian basis data yang berulang. Hal ini mengurangi waktu eksekusi dan membuat algoritma *FP-Growth* lebih cepat dan efisien [16]. *FP-Growth* juga memanfaatkan kompresi data dalam *FP-Tree*, meningkatkan kecepatan pemrosesan dan efektivitas dalam penelitian dan aplikasi praktis, terutama dalam konteks *dataset retail online*. Dalam Algoritma

Apriori, akurasi aturan yang dihasilkan lebih tinggi dibandingkan *FP-Growth*. Akan tetapi, karena *Apriori* melakukan pemindaian berulang, performanya melambat. Tetapi, tujuan dari algoritma tersebut ini tetap sama, yakni untuk mengidentifikasi *frequent itemset* [17]

II. METHODOLOGY

Metode penelitian ini terdiri dari serangkaian tahapan yang diuraikan dalam sebuah bagan atau *flowchart*. Tahapan-tahapan penelitian yang digunakan dalam penelitian ini dapat digambarkan sebagai berikut:



Gambar 1. Design Penelitian

A. Identification of problem

Penelitian ini bertujuan untuk menangani beberapa masalah yang muncul dalam analisis data transaksi penjualan pada *dataset Online Retail*. Masalah-masalah yang diidentifikasi meliputi pemilihan algoritma asosiasi yang optimal antara *FP-Growth* dan *Apriori*, identifikasi model asosiasi yang relevan pada *dataset Online Retail* yang kompleks, dan efisiensi serta kinerja algoritma dalam mengolah data transaksi penjualan dengan cepat dan efektif. Dengan mengatasi masalah-masalah ini, penelitian ini diharapkan dapat memberikan wawasan dan pemahaman yang lebih baik tentang penggunaan algoritma asosiasi untuk menemukan model asosiasi terbaik pada industri ritel *online*.

Hal paling penting untuk penelitian adalah kumpulan data untuk bahan pemrosesan data yang menggunakan algoritme tertentu. Data *Ritel Online* yang diambil dari situs web UCI antara 12-01-2010 dan 12-09-2011 digunakan untuk menyusun *dataset* untuk penelitian ini. Kumpulan data ini terdiri dari 541.910 elemen dengan atribut berikut: *no voice*, *stock code*, *description*, *quantity*, *invoice date*, *unit price*, dan *country*. Namun, untuk memfokuskan penelitian pada objek yang lebih spesifik, penelitian ini hanya menggunakan data transaksi yang memiliki nilai atribut *country*



"France". Dengan demikian, *dataset* yang digunakan dalam penelitian ini terdiri dari 392 elemen dengan hanya menggunakan 3 atribut, yaitu *invoice no*, *description*, dan *quantity*. Sebelum proses pengolahan data dilakukan, atribut-atribut tersebut akan disortir pada tahap *preprocessing* untuk mempermudah analisis yang akan dilakukan.

B. Data Collection

Penelitian ini berfokus pada penggunaan kumpulan data Ritel Online yang diambil dari situs web UCI dalam periode tertentu. Dataset ini awalnya terdiri dari 541.910 elemen dengan beragam atribut. Namun, penelitian ini mempersempit fokusnya dengan hanya mempertimbangkan data transaksi yang memiliki atribut "country" dengan nilai "France", menghasilkan dataset yang terdiri dari 392 elemen dengan 3 atribut kunci, yaitu *invoice number*, *description*, dan *quantity*. Sebelum melakukan analisis, atribut-atribut tersebut akan disortir dalam tahap *preprocessing*. Pengolahan data ini dilakukan menggunakan platform Google Colab dan bahasa pemrograman Python, dengan dukungan berbagai library seperti NumPy, Pandas, Matplotlib, dan Time. Penelitian ini mengimplementasikan algoritma FP-Growth, Apriori, dan *association_rules* untuk mengidentifikasi pola frekuensi tinggi dalam dataset dan membangun aturan asosiasi yang relevan.

C. Preprocessing Data

Dalam tahap *preprocessing data*, dilakukan langkah-langkah untuk mengeliminasi outlier, memperbaiki data yang bermasalah, menangani data yang kosong, serta memastikan kelengkapan dan konsistensi atribut dalam dataset. *Preprocessing data* sangat krusial karena berkontribusi secara signifikan terhadap kualitas mining yang baik. Dengan melakukan *preprocessing* yang tepat, dapat memastikan bahwa data yang digunakan dalam analisis memiliki integritas yang tinggi, akurat, dan representatif. Hal ini memberikan dasar yang kuat untuk menghasilkan hasil mining yang relevan, valid, dan dapat diandalkan [18]. Terdapat beberapa metode *preprocessing data* yang diterapkan pada penelitian ini guna menghasilkan data yang telah diproses, diantaranya:

1. Missing value

Proses pengecekan missing value merupakan tahap yang penting dalam analisis data. Pada tahap ini, dilakukan pengecekan terhadap atribut-atribut yang terdapat dalam dataset untuk mengidentifikasi atribut yang tidak memiliki nilai atau informasi yang tersedia. Missing value dapat mempengaruhi hasil analisis data atau pemodelan, sehingga perlu dilakukan strategi penanganan seperti penghapusan baris atau kolom yang mengandung missing value. Langkah ini memastikan integritas dan kualitas data yang digunakan dalam analisis, sehingga hasil analisis yang diperoleh lebih tepat dan reliabel.

2. Cleaning data



Pada tahap cleaning data merupakan langkah penting dalam rangka meningkatkan kualitas data dengan mengidentifikasi dan mengoreksi kesalahan serta data yang tidak selaras [18] Tujuan dari cleaning data adalah untuk memastikan bahwa data yang digunakan dalam pengolahan memiliki integritas yang tinggi dan sesuai dengan kebutuhan analisis. Proses ini mencoba untuk menghilangkan tipe data yang berlebihan dan tidak sesuai. Dengan melakukan cleaning data yang cermat, dapat meminimalkan dampak dari data yang tidak valid atau tidak sesuai, sehingga memungkinkan pengolahan data yang lebih akurat dan menampilkan hasil yang optimal. Pada penelitian ini dilakukan cleaning data dari beberapa atribut, diantaranya :

- a) Pada atribut description dilakukan cleaning data untuk memastikan bahwa setiap nilai dalam atribut tersebut tidak memiliki spasi diawal maupun di akhirnya. Kegunaannya untuk membersihkan data dan memastikan konsistensi dalam pemrosesan selanjutnya.
- b) Pada atribut invoice no juga dilakukan perubahan pada type data yang asalnya object menjadi string. Kegunaannya agar sewaktu dieksekusi, atribut tersebut akan berisi nilai-nilai yang direpresentasikan sebagai string.
- c) Pada atribut invoice no juga membuat subset baru dari dataset yang telah di input yang berisi baris-baris di mana atribut invoice no tidak mengandung karakter "C".

3. Sorting attribute

Dalam tahap sorting atribut, peneliti melakukan seleksi data dengan mempertimbangkan atribut country. Hanya data dengan nilai atribut country yang sama dengan "France" yang dipilih untuk penelitian ini, sehingga penelitian dapat difokuskan pada satu kelompok data tertentu. Selain itu, dilakukan juga sorting atribut untuk menentukan atribut mana yang akan digunakan dalam proses pengolahan data. Pada dataset yang digunakan dalam penelitian ini, dilakukan proses sorting terhadap 8 atribut yang tersedia menjadi 3 atribut yang dipilih, yaitu invoice no, description, dan quantity. Dengan melakukan sorting ini, peneliti dapat menyederhanakan dan memfokuskan data yang akan digunakan sehingga memudahkan proses pengolahan data selanjutnya.

4. Change value

Dalam tahap change value, dilakukan proses transformasi pada atribut data quantity. Nilai-nilai data quantity yang kurang dari 1 diubah menjadi nilai 0, sedangkan nilai-nilai yang lebih besar atau setidaknya sama dengan 1 akan diubah menjadi nilai 1. Hal ini dilakukan untuk memberikan konsistensi dan mempermudah interpretasi data quantity.

5. Deleted Postage Column

Dalam proses penelitian ini, dilakukan penghapusan kolom data yang tidak relevan atau tidak akan digunakan dalam analisis. Hal ini dilakukan untuk menjaga kelancaran proses perhitungan dan mencegah gangguan yang disebabkan oleh data yang tidak terpakai. Dengan menghilangkan kolom-kolom data yang tidak relevan, fokus analisis dapat diberikan



pada atribut-atribut yang lebih penting dan berkaitan langsung dengan tujuan penelitian. Tindakan ini juga membantu mengoptimalkan penggunaan sumber daya komputasi dan mempercepat proses pengolahan data secara keseluruhan.

D. Algorithm Comparison

Dalam proses penelitian ini, dilakukan perbandingan perbandingan antara algoritma FP-Growth dan Apriori dalam penelitian ini adalah untuk mengidentifikasi kelebihan dan kelemahan masing-masing algoritma dalam konteks pengolahan dataset retail online. Dengan membandingkan kinerja keduanya, penelitian ini bertujuan untuk memberikan pemahaman yang lebih mendalam tentang bagaimana algoritma-algoritma ini dapat memengaruhi proses analisis data dan hasil yang dihasilkannya

E. Modeling

Pada tahap pemodelan dalam metode penelitian ini, fokus diberikan pada pengembangan *association rule* dan model yang dihasilkannya. Dengan menggunakan algoritma asosiasi, seperti *Apriori* dan *FP-Growth*, peneliti akan mengeksplorasi pola asosiasi yang signifikan antara *item-item* dalam *dataset*. Model asosiasi yang dihasilkan akan memberikan pemahaman yang lebih baik tentang hubungan dan keterkaitan antara *item-item* tersebut dalam konteks penjualan dan transaksi data.

1. Association rule

Metode *data mining* yang bertujuan untuk mengidentifikasi hubungan atau *rule* antara kombinasi *item* merupakan *Association rule*. Evaluasi *association rule* dilakukan berdasarkan nilai *confidence* dan *support*. Nilai *support* mengindikasikan seberapa sering kombinasi *item* muncul pada *dataset*, sementara nilai *confidence* mengukur seberapa sering kombinasi *item* terjadi bersamaan dengan *item* lain. Pemahaman yang baik terhadap nilai *support* dan *confidence* sangat penting dalam mengidentifikasi aturan asosiasi yang relevan [19]. Aturan asosiasi bukanlah *dataset* tunggal atau aturan klasifikasi, melainkan beragam aturan yang ditentukan untuk pengaturan yang berbeda dalam kumpulan data, dan biasanya menghasilkan prediksi yang berbeda. Oleh karena itu, *dataset* kecil pun dapat menghasilkan banyak aturan asosiasi yang berbeda. Untuk membatasi jumlah aturan dan meningkatkan akurasi, digunakan ukuran *support* dan *confidence* [20].

Untuk memangkas atau mengatur batasan dalam menentukan hasil *itemset* dibutuhkan nilai *support* dan *confidence* yang minimal agar peneliti dapat menentukan jumlah kemunculan *itemset* dalam data. Pada minimum *support* jika nilai minimum *support* yang sudah ditetapkan yaitu 0.09, sehingga *itemset* yang muncul dengan frekuensi di atas atau sama dengan 0.09 akan dianggap sering dan dianggap sebagai



kandidat dalam pembentukan aturan asosiasi. Sama hal nya dengan nilai minimum *confidence* digunakan untuk memfilter aturan asosiasi yang memiliki tingkat kepercayaan dibawah ambang batas yang ditentukan. Nilai minimum *confidence* yang lebih besar akan manampilkan aturan yang lebih kuat dan relevan, tetapi jumlah aturan yang ditemukan bisa menjadi lebih sedikit. Sebaliknya, nilai minimum *confidence* yang lebih rendah akan menghasilkan lebih banyak aturan, tetapi dapat menghasilkan aturan yang lemah atau *noise*.

F. Evaluasi

Dalam tahap evaluasi penelitian ini, digunakan beberapa metrik evaluasi, yaitu support, confidence, lift, dan leverage, bersama dengan waktu proses. Support mengukur sejauh mana itemset mendominasi dalam kumpulan data dan digunakan secara bersamaan, digunakan dengan nilai minimal 0.07. Confidence mengukur kepastian asosiasi antara dua item dalam itemset. Lift mengukur signifikansi dan relevansi asosiasi item dalam itemset, dengan nilai di atas 1 menunjukkan hubungan yang signifikan. Leverage mengukur tingkat korelasi antara item dalam analisis asosiasi, dengan nilai mendekati 1 menunjukkan korelasi yang kuat. Evaluasi ini bertujuan untuk menghasilkan model yang efektif dalam menganalisis data dan mengidentifikasi aturan asosiasi yang relevan dan signifikan.

G. Result

Hasil penelitian ini mencakup komponen penting seperti indeks, itemset, nilai support, nilai confidence, dan waktu pemrosesan. Indeks digunakan untuk identifikasi posisi dan urutan itemset dalam dataset, mencerminkan kombinasi item terkait dengan frekuensi tertentu. Support mengukur kemunculan itemset dalam dataset, sedangkan confidence mengindikasikan tingkat kepercayaan dalam hubungan antara itemset. Waktu pemrosesan mencatat durasi yang dibutuhkan algoritma dalam mengolah data. Kesemua komponen ini memberikan wawasan penting pada hasil analisis asosiasi dalam penelitian ini..

III. RESULT AND DISCUSSION

Hasil penelitian ini meliputi *Preprocessing Data*, *Implementation of Apriori and FP-Growth Algorithms* dan *Association Rule Implementation*.

A. Preprocessing Data

Setelah pengumpulan data dan melakukan *preprocessing data*, *dataset* telah disaring sehingga hanya atribut dari *invoice no*, *description*, dan *quantity* yang tersisa untuk dilakukan pengolahan data dengan 392 elemen yang terdokumentasi dalam Tabel 1. Tindakan ini dilakukan untuk memfokuskan analisis pada informasi yang relevan dan penting untuk tujuan penelitian. Dengan mempertahankan atribut tersebut, penelitian



ini dapat memulai langkah-langkah berikutnya dalam analisis data dengan lebih terarah dan efisien.

Tabel 1. Dataset retail online hasil preprocessing

Invoice No	Description	Quantity
581587	PACK OF 20 SPACEBOY NAPKINS	12
581587	CHILDREN'S APRON DOLLY GIRL	6
581587	CHILDRENS CUTLERY DOLLY GIRL	4
581587	CHILDRENS CUTLERY CIRCUS PARADE	4
581587	BAKING SET 9 PIECE RETROSPOT	3

B. Implementation of Apriori and FP-Growth Algorithms

Langkah berikutnya yaitu mengimplementasikan Apriori terhadap pengolahan data dengan min *support* 0.07. Hasil pemodelan menggunakan algoritma Apriori menghasilkan 51 model yang terdokumentasikan dalam Tabel 2.

Tabel 2. Model frequent itemset algoritma Apriori

index	support	itemsets
0	0.07	4 TRADITIONAL SPINNING TOPS
1	0.1	ALARM CLOCK BAKELIKE GREEN
...	...	...
49	0.12	SET/6 RED SPOTTY PAPER PLATES', 'SET/6 RED SPOTTY PAPER CUPS
50	0.1	SET/20 RED RETROSPOT PAPER NAPKINS', 'SET/6 RED SPOTTY PAPER CUPS

Selanjutnya Implementasi Algoritma FP-Growth yang juga berhasil menghasilkan sebanyak 51 model *frequent itemset* yang telah terdokumentasikan secara rinci dalam Tabel 3. Hal ini menunjukkan bahwa algoritma FP-Growth mampu mengidentifikasi pola-pola yang sering muncul dalam dataset yang digunakan.



Tabel 3. Model frequent itemset algoritma FP-Growth

index	support	itemsets
0	0.18	RED TOADSTOOL LED NIGHT LIGHT
1	0.16	ROUND SNACK BOXES SET OF4 WOODLAND
...	...	...
49	0.1	PLASTERS IN TIN WOODLAND ANIMALS, PLASTERS IN TIN CIRCUS PARADE
50	0.1	SET/20 RED RETROSPOT PAPER NAPKINS', 'SET/6 RED SPOTTY PAPER CUPS

Berdasarkan hasil pemodelan menggunakan algoritma FP-Growth dan Apriori dengan minimal *support* 0.07, ditemukan bahwa kedua algoritma menghasilkan 51 model *frequent itemset*. Namun, perlu diperhatikan bahwa terdapat perbedaan dalam indeks dan *itemset* yang dihasilkan oleh kedua algoritma, meskipun jumlah *itemset* atau indeksnya sama. Hal ini menunjukkan bahwa implementasi kedua algoritma dapat menghasilkan model yang unik dalam analisis asosiasi.

C. Association Rule Implementation

Proses selanjutnya dilakukan implementasi *association rule* dengan menggunakan model yang dihasilkan oleh masing-masing algoritma. Pada langkah ini, dilakukan penambahan metrik *lift* sebagai ukuran kekuatan asosiasi antara dua *item* atau kelompok *item* dalam *association rule*. Dalam pembentukan model ini, hanya nilai *lift* yang melebihi 1 yang dianggap. Selain itu, proses ini juga mencakup pengukuran waktu perhitungan pada setiap algoritma untuk menentukan algoritma mana yang lebih cepat dalam proses perhitungan. Dengan demikian, penelitian ini tidak hanya memberikan hasil asosiasi yang relevan, tetapi juga memberikan informasi tentang efisiensi waktu dalam penerapan algoritma.

Implementasi *association rule* penting dalam analisis data dan pengambilan keputusan. Hal ini membantu mengungkap hubungan antara item dalam *dataset*, memungkinkan rekomendasi yang personal, meningkatkan pengalaman pelanggan, dan mendukung strategi pemasaran yang efektif. Selain itu, *association rule* juga membantu dalam pengambilan keputusan yang informasional dan strategis, dengan memahami pola yang ada dan merancang tindakan yang tepat. Implementasi *association rule* juga dapat meningkatkan kualitas data dan mengoptimalkan proses analisis.

Dalam Tabel 4, terdapat hasil pemodelan menggunakan algoritma Apriori yang menghasilkan 26 model dengan minimum *support* 0.07. Dalam model tersebut, ditemukan nilai *support* tertinggi yaitu 0.12 dan nilai *confidence*-nya 0.96.

Tabel 4. Hasil model algoritma Apriori

index	antec	Conse	as	cs	s	c	l	le	co
-------	-------	-------	----	----	---	---	---	----	----



0	ALARM CLOCK BAKELIKE GREEN	ALARM CLOCK BAKELIKE PINK	0.10		0.10	0.07	0.76	7.48	0.06	3.79
1	ALARM CLOCK BAKELIKE PINK	ALARM CLOCK BAKELIKE GREEN	0.10		0.10	0.07	0.73	7.48	0.06	3.28
2	ALARM CLOCK BAKELIKE GREEN	ALARM CLOCK BAKELIKE RED	0.10		0.09	0.08	0.82	8.64	0.07	4.92
...	...	...	...		...	...	...	...	...	...
23	SET/6 RED SPOTTY PAPER PLATES	SET/6 RED SPOTTY PAPER CUPS', 'SET/20 RED RETROSPOT PAPER NAPKINS	0.13		0.10	0.10	0.78	7.64	0.09	4.08
24	SET/6 RED SPOTTY PAPER CUPS	SET/6 RED SPOTTY PAPER PLATES', 'SET/20 RED RETROSPOT PAPER NAPKINS	0.14		0.10	0.10	0.72	7.08	0.09	3.23
25	SET/20 RED RETROSPOT PAPER NAPKINS	SET/6 RED SPOTTY PAPER PLATES', 'SET/6 RED SPOTTY PAPER CUPS	0.13		0.12	0.10	0.75	6.13	0.08	3.51

Penulis memutuskan untuk mempersingkat nama atribut dalam penelitian ini agar lebih ringkas dan mudah dipahami. Berikut adalah arti dari setiap atribut yang telah dipersingkat:



Tabel 5. Dataset atribut

No	Attributes	Information
1	Index	Index
2	antenc	Antecedents
3	conse	Consequents
4	as	Antecedents Consequents
5	cs	Consequents Antecedents
6	s	Support
7	c	Confidence
8	l	Lift
9	le	leverage
10	con	Conviction

Selanjutnya, pada Tabel 6 terdapat hasil pemodelan menggunakan algoritma *FP-Growth* yang juga menghasilkan 26 model dengan minimum *support* 0.07. Dalam tabel tersebut, terlihat bahwa terdapat nilai *support* tertinggi sebesar 0.12 dan nilai *confidence* sebesar 0.96.

Tabel 6. Hasil model algoritma *FP-Growth*

index	antec	conse	as	cs	s	c	l	le	co
0	ALARM CLOCK BAKELIKE GREEN	ALARM CLOCK BAKELIKE PINK	0.10	0.10	0.07	0.76	7.48	0.06	3.79
1	ALARM CLOCK BAKELIKE PINK	ALARM CLOCK BAKELIKE GREEN	0.10	0.10	0.07	0.73	7.48	0.06	3.28
2	ALARM CLOCK BAKELIKE GREEN	ALARM CLOCK BAKELIKE RED	0.10	0.09	0.08	0.82	8.64	0.07	4.92
...	...	...	...	...	...	...	...	...	...
23	PLASTERS IN TIN CIRCUS PARADE	PLASTERS IN TIN WOODLAND ANIMALS	0.17	0.17	0.10	0.61	3.55	0.07	2.10
24	SET/20 RED RETROSPOT	SET/6 RED SPOTTY PAPER CUPS	0.13	0.14	0.10	0.77	5.58	0.08	3.74



	PAPER NAPKINS								
25	SET/6 RED SPOTTY PAPER CUPS	SET/20 RED RETROSPOT PAPER NAPKINS	0.14	0.13	0.10	0.74	5.58	0.08	3.35

Langkah selanjutnya adalah melakukan filter untuk memilih model yang memiliki nilai *confidence* minimal 0.9 pada kedua model dari masing-masing algoritma. Hasil pemodelan algoritma Apriori pada tabel 7 menunjukkan adanya 3 model asosiasi yang memenuhi kriteria tersebut.

Tabel 7. Hasil filter model min support 0.1 pada algoritma Apriori

index	antec	conse	as	cs	s	c	l	le	co
18	SET/6 RED SPOTTY PAPER PLATES	SET/6 RED SPOTTY PAPER CUPS	0.13	0.14	0.12	0.96	6.97	0.10	21.56
21	SET/6 RED SPOTTY PAPER PLATES, SET/20 RED RETROSPOT PAPER NAPKINS	SET/6 RED SPOTTY PAPER CUPS	0.10	0.14	0.10	0.98	7.08	0.09	34.49
22	SET/6 RED SPOTTY PAPER CUPS, SET/20 RED RETROSPOT PAPER NAPKINS	SET/6 RED SPOTTY PAPER PLATES	0.10	0.13	0.10	0.98	7.64	0.09	34.90

Hasil filter pemodelan algoritma *FP-Growth* pada Tabel IV.12 menunjukkan adanya 3 model asosiasi yang memenuhi kriteria . Dengan menggunakan nilai *confidence* yang tinggi, kita dapat memastikan bahwa aturan asosiasi yang terbentuk memiliki tingkat kepercayaan yang kuat, sehingga dapat memberikan panduan yang lebih handal dalam pengambilan keputusan dan strategi bisnis.

Tabel 8. Hasil filter model min support 0.1 pada algoritma FP-Growth

index	antec	Conse	as	cs	s	c	l	le	co
10	SET/6 RED SPOTTY PAPER PLATES	SET/6 RED SPOTTY	0.13	0.14	0.12	0.96	6.97	0.10	21.56



		PAPER CUPS							
14	SET/20 RED RETROSPOT PAPER NAPKINS, SET/6 RED SPOTTY PAPER PLATES	SET/6 RED SPOTTY PAPER CUPS	0.10	0.14	0.10	0.98	7.08	0.09	34.49
15	SET/20 RED RETROSPOT PAPER NAPKINS, SET/6 RED SPOTTY PAPER CUPS	SET/6 RED SPOTTY PAPER PLATES	0.10	0.13	0.10	0.98	7.64	0.09	34.90

Dapat disimpulkan bahwa langkah selanjutnya setelah pemodelan menggunakan algoritma *Apriori* dan *FP-Growth* adalah melakukan *filter* untuk memilih model yang memiliki nilai *confidence* minimal 0.9. Hasil pemodelan kedua algoritma menunjukkan adanya 3 model asosiasi yang memenuhi kriteria tersebut dengan *itemset* dan *index* yang berbeda pada tiap model yang dihasilkan, sebagaimana terdokumentasikan dalam Tabel 7 dan Tabel 8.

Pada penelitian ini dipilih satu model dengan nilai *support* tertinggi dari algoritma *FP-Growth* dan *Apriori* yang terdokumentasi dalam Tabel 9.

Tabel 9. Hasil model dengan nilai *support* tertinggi pada tiap algoritma

	antecedents & consequents	
	SET/6 RED SPOTTY PAPER PLATES & SET/6 RED SPOTTY PAPER CUPS	
Algoritma	FP-Growth	Apriori
index	10	18
support	0.122449	0.122449
confidence	0.96	0.96
waktu proses	0.004317522048950195 detik	0.007213592529296875 detik

Hasil dari penelitian ini menghasilkan pengetahuan yang Meskipun kedua algoritma menghasilkan model asosiasi dengan *frequent itemset* dan matriks nilai yang sama yaitu nilai *support* 0.12 dan nilai *confidence* 0.96, dan juga terdapat perbedaan dalam nilai indeks yang dihasilkan oleh masing-masing algoritma. Algoritma *Apriori* menghasilkan model dengan nilai *support* tertinggi pada *index* ke-18, sementara *FP-Growth* menghasilkan model dengan nilai tertinggi pada indeks ke-10. Dapat disimpulkan bahwa urutan model yang dihasilkan oleh kedua algoritma berbeda,



meskipun *frequent itemset* yang dihasilkan tetap sama. Dalam penelitian ini, *FP-Growth* menghasilkan model dengan nilai indeks yang lebih rendah, menunjukkan keunggulan algoritma tersebut. Untuk waktu proses pada setiap algoritma pun berbeda. Pada penelitian ini waktu proses algoritma *FP-Growth* juga unggul dengan waktu 0.004 detik, sebaliknya untuk waktu proses algoritma *Apriori* lebih lambat yaitu 0.007 detik.

IV. CONCLUSIONS

Berdasarkan hasil penelitian ini, dapat disimpulkan bahwa terdapat perbedaan yang signifikan antara model asosiasi yang dihasilkan oleh algoritma *Apriori* dan *FP-Growth* pada dataset retail online. Meskipun terdapat satu model frequent itemset yang sama dengan nilai support 0.12 dan confidence 0.96, urutan model yang dihasilkan oleh kedua algoritma berbeda. Algoritma *Apriori* menghasilkan model dengan nilai support tertinggi pada indeks ke-18, sementara algoritma *FP-Growth* menghasilkan model dengan nilai support tertinggi pada indeks ke-10. Hasil ini menunjukkan bahwa kedua algoritma memiliki kualitas yang baik dalam merepresentasikan hubungan antara item-item dalam dataset. Selain itu, algoritma *FP-Growth* menawarkan keunggulan dalam efisiensi dan kecepatan dalam mengolah data transaksi pada dataset retail online. Pengukuran waktu proses menunjukkan bahwa algoritma *FP-Growth* menyelesaikan proses analisis data dengan waktu eksekusi sebesar 0.004 detik, sedangkan algoritma *Apriori* memerlukan waktu eksekusi yang lebih lama, yaitu 0.007 detik. Meskipun algoritma *Apriori* memiliki keakuratan yang lebih tinggi dalam menghasilkan aturan asosiasi, perbedaan waktu eksekusi tersebut menunjukkan bahwa algoritma *FP-Growth* lebih cepat dalam pemrosesan data transaksi pada dataset retail online. Oleh karena itu, terutama dalam menghadapi data dengan volume yang besar dan memerlukan kecepatan dalam pengambilan keputusan, algoritma *FP-Growth* merupakan pilihan yang lebih optimal untuk analisis data dalam lingkungan retail online.

REFERENCES

- [1] M. S. Sandy, H. Setiawan, U. Indahyanti, F. Sains, and D. Teknologi, "Analisis Data Mining Produk Retail Menggunakan Metode Asosiasi Dengan Menerapkan Algoritma *Apriori*," 2023.
- [2] A. R. Wibowo and A. Jananto, "Implementasi Data Mining Metode Asosiasi Algoritma *FP-Growth* Pada Perusahaan Ritel," *Inspiration: Jurnal Teknologi Informasi dan Komunikasi*, vol. 10, no. 2, pp. 200–212, 2020.
- [3] I. Made *et al.*, "Aplikasi Data Mining Asosiasi Barang Menggunakan Algoritma *Apriori-TID*," 2022.



- [4] M. Badaruddin and R. Rayendra, "Penerapan Algoritma Apriori Pada Analisa Data Penjualan Ecommerce," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 6, no. 2, p. 1032, Apr. 2022, doi: 10.30865/mib.v6i2.3976.
- [5] C. Na. Dewi, F. Putrawansyah, and D. Puspita, "Implementasi Algoritma FP-Growth Pada E-Commerce Kopi Pagar Alam Menggunakan Framework Codeigniter," *Indonesian Journal of Computer Science*, vol. 10, no. 2, pp. 447–458, 2021.
- [6] E. D. Sikumbang, "Penerapan Data Mining Penjualan Sepatu Menggunakan Metode Algoritma Apriori," *Jurnal Teknik Komputer*, vol. 4, no. 1, 2018.
- [7] N. Syahputri, "Penerapan Data Mining Asosiasi pada Pola Transaksi dengan Metode Apriori," 2020.
- [8] K. F. Irnanda and A. P. Windarto, *Seminar Nasional Teknologi Komputer & Sains (SAINTEKS) Penerapan Klasifikasi C4.5 Dalam Meningkatkan Kecakapan Berbahasa Inggris dalam Masyarakat*. 2020.
- [9] A. Silvanie, "PENCARIAN FREQUENT ITEMSET DENGAN ALGORITMA APRIORI DAN PYTHON.," *Jurnal Nasional Informatika*, vol. 1, No. 2, pp. 103–113, 2020.
- [10] Y. Kurnia, Y. Isharianto, Y. C. Giap, A. Hermawan, and Riki, "Study of application of data mining market basket analysis for knowing sales pattern (association of items) at the O! Fish restaurant using apriori algorithm," in *Journal of Physics: Conference Series*, Institute of Physics Publishing, Jun. 2019. doi: 10.1088/1742-6596/1175/1/012047.
- [11] M. H. Prayitno and Rasim, "Analisis Penjualan Produk Retail Dengan Metode Data Mining Asosiasi," *Jurnal Kajian Ilmiah Universitas Bhayangkara Jakarta Raya*, vol. 18, no. 3, pp. 231–237, 2018.
- [12] M. Badrul, "ALGORITMA ASOSIASI DENGAN ALGORITMA APRIORI UNTUK ANALISA DATA PENJUALAN," *Jurnal Pilar Nusa Mandiri*, vol. 12, no. 2, pp. 121–129, 2016.
- [13] A. Junaidi, "Implementasi Algoritma Apriori dan FP-Growth Untuk Menentukan Persediaan Barang," *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, vol. 8, no. 1, pp. 61–67, 2019.
- [14] A. H. Maulana, C. Adinda Hartawan, F. Della Irawan, and P. Seta Ananta, "Pemanfaatan Big Data dalam Bisnis E-commerce OLX," *JURNAL INTECH*, vol. 2, no. 2, p. 1, 2021.
- [15] F. A. Sianturi, "PENERAPAN ALGORITMA APRIORI UNTUK PENENTUAN TINGKAT PESANAN," *Jurnal Mantik Penusa*, vol. 2, no. 1, pp. 50–57, 2018, [Online]. Available: <http://bowmasbow.blogspot.com/20>
- [16] S. Sumirat and Y. Ramdhani, "Implementasi Algoritma Fp-Growth Untuk Penentuan Paket Hemat Produk Skincare," vol. 2, no. 2, 2021.



- [17] R. Fitria, W. Nengsih, and D. H. Qudsi, "Implementasi Algoritma FP-Growth Dalam Penentuan Pola Hubungan Kecelakaan Lalu Lintas," *Jurnal Sistem Informasi*, vol. 13, no. 2, p. 118, Oct. 2017, doi: 10.21609/jsi.v13i2.551.
- [18] F. Winda Sari and N. Jannah, "Analisis Profil Mahasiswa Politeknik Negeri Batam dengan Teknik Data Mining Asosiasi dan Clustering," 2016.
- [19] B. Septia Pranata and D. Putro Utomo, "Bulletin of Information Technology (BIT) Penerapan Data Mining Algoritma FP-Growth Untuk Persediaan Sparepart Pada Bengkel Motor (Study Kasus Bengkel Sinar Service)," *Bulletin of Information Technology (BIT)*, vol. 1, no. 2, pp. 83–91, 2020.
- [20] R. Rachman, "Penentuan Pola Penjualan Media Edukasi dengan Menggunakan Metode Algoritme Apriori dan FP-Growth," *Paradigma*, vol. 23, no. 1, 2021.