

ANALISIS PERBANDINGAN ALGORITMA FP-GROWTH DAN TPQ-APRIORI DALAM MENENTUKAN RULE BASED TERBAIK UNTUK SISTEM REKOMENDASI PRODUK

Lalu Zazuli Azhar Mardedi¹, Kartarina², Moch. Syahrir^{3*}

Universitas Bumigora Mataram^{1,2,3}

muhammadsyahriralfath@gmail.com¹, pujutsega@gmail.com², kartarina@universitasbumigora.ac.id³

Abstrak – Algoritma association rule yang populer adalah apriori dan fp-growth, kedua algoritma tersebut tersebut sudah sangat familiar di kalangan peneliti data mining, akan tetapi ada beberapa kelemahan yang ditemukan dari algoritma association rule tersebut, diantaranya scan dataset yang lama dalam proses pencarian frekuensi itemset, penggunaan memori yang besar serta base rule yang dihasilkan terkadang kurang optimal. Dalam penelitian ini penulis melakukan komparasi algoritma fp-growth dan TPQ-apriori untuk menganalisa hasil base rule dari kedua algoritma tersebut. TPQ-Apriori adalah algoritma hasil pengembangan dari algoritma apriori, dimana kinerja algoritma TPQ-Apriori lebih baik dari algoritma apriori tradisional dalam hal proses scan dataset dalam pencarian frekuensi itemset. Untuk eksperimen algoritma fp-growth menggunakan tools rapidminer sementara algoritma TPQ-apriori menggunakan program aplikasi yang dibangun sendiri. Sementara dataset yang digunakan adalah data penjualan pada CV. Charandita Kusuma NTB yang telah di upload pada situs kaggle. Adapun hasil pengujian base rules dari keseluruhan hasil pengujian rule dengan dataset penjualan CV. Charandita Kusuma NTB tersebut bisa ditarik suatu kesimpulan bahwa semakin besar dataset yang akan diolah maka hasilnya akan semakin optimal apabila menggunakan algoritma fp-growth rapidminer akan tetapi tidak optimal bila dataset yang akan diolah adalah dataset kecil. Beberapa rule tidak tampil pada algoritma fp-growth dengan tools rapidminer. Sementara algoritma TPQ-Apriori yang telah dikembangkan mampu menghasilkan rule yang optimal baik untuk dataset yang besar maupun dataset yang kecil.

Kata kunci: *Association Rule, Fp-Growth, TPQ-Apriori, Data Mining, Dataset;*

Abstract - The popular association rule algorithms are a priori and fp-growth, these two algorithms are very familiar among data mining researchers, however there are several weaknesses found in the association rule algorithm, including scanning the dataset for a long time in the process of searching for itemset frequencies, the use of large memory and the resulting base rules are sometimes less than optimal. In this research, the author compared the fp-growth and TPQ-apriori algorithms to analyze the base rule results of the two algorithms. TPQ-Apriori is an algorithm resulting from the development of the apriori algorithm, where the performance of the TPQ-Apriori algorithm is better than the traditional apriori algorithm in terms of the dataset scanning process in searching for itemset frequencies. For experiments, the fp-growth algorithm used the rapidminer tool while the TPQ-apriori algorithm used an application program that was built by ourselves. Meanwhile, the dataset used is sales data on CV. Charandita Kusuma NTB which has been uploaded to the Kaggle site. The base rules testing results are from the overall rule testing results with the CV sales dataset. Charandita Kusuma NTB can draw a conclusion that the larger the dataset to be processed, the more optimal the results will be if using the fp-growth rapidminer algorithm, but it is not optimal if the dataset to be processed is a small dataset. Some rules do not appear in the fp-growth algorithm with the rapidminer tool. Meanwhile, the TPQ-Apriori algorithm that has been developed is able to produce optimal rules for both large datasets and small datasets.

Keyword : *Association Rule, Fp-Growth, TPQ-Apriori, Data Mining, Dataset;*

1. Pendahuluan

Data mining adalah disiplin ilmu yang bertujuan untuk mengekstrak pengetahuan serta menemukan pola dari suatu data yang besar dengan cara mempelajari dan mengembangkan algoritma[1][2]. Dalam perannya data mining terdiri dari estimasi, forecasting, klasifikasi, klasterisasi, dan asosiasi[3], dalam menggunakan lima peranan

data mining untuk mengekstrak pengetahuan dan menemukan pola harus disesuaikan dengan karakteristik data, sebab semakin baik data yang akan diolah oleh algoritma data mining maka hasil yang didapat akan semakin optimal. Peran asosiasi banyak diimplementasikan dalam bidang usaha berupa e-commers, retail dan restoran namun bisa juga diterapkan pada bidang lain seperti

analisa bug software dan analisa data biologi serta data medis[4] sehingga kita memastikan bahwa manfaat cukup banyak digunakan dalam berbagai bidang. Asosiasi atau yang umum disebut dalam istilah data mining adalah association rule merupakan salah satu teknik data mining untuk mencari dan menemukan sekumpulan aturan asosiasi antara suatu kombinasi item[5]. Atau dengan pengertian lain adalah suatu aturan asosiatif dari implikasi kombinasi hubungan antara suatu item[4][6]. Algoritma-algoritma association rule yang umum digunakan dan dikembangkan adalah Apriori, Fp-Growth, Eclat, dan Hash Based[5][7]. Dalam algoritma association rule yang dihasilkan adalah berupa rule atau aturan yang bisa diukur dengan menggunakan support, confidence, liftrasio, leverage, conviction dan certainty factor. Support adalah persentase kombinasi item tersebut dalam database, confidence adalah kuatnya hubungan antara item dalam aturan asosiasi, liftrasio adalah untuk menguji nilai kevalidan hubungan antara item begitupun dengan leverage dan conviction untuk menguji seberapa pengaruh dan keyakinan antara antecedent dan consequent.

Cikal bakal algoritma-algoritma asosiasi adalah dari algoritma Apriori yang pertama kali diseminarkan[8] lalu diperbaiki setahun setelahnya[9] dalam implementasinya Apriori bisa menghasilkan rule yang optimal akan tetapi waktu yang digunakan untuk scan dataset sangat lama sebab pendekatan yang digunakan oleh Apriori dalam mencari frekuensi menggunakan generasi kandidat dimana semua item harus ditelusuri untuk menentukan kandidat k-itemset, k-itemset maksudnya adalah berapa iterasi yang akan terbentuk dari seluruh penelusuran dan mengakibatkan banyak menggunakan memori[5][7]. Perbaikan yang dilakukan oleh[10] dengan pendekatan fp-tree yang sekarang kita kenal dengan nama algoritma fp-growth dari sisi waktu scan dataset sudah sangat baik karena hanya melakukan dua kali scan dataset, akan tetapi rule yang dihasilkan oleh algoritma Fp-Growth tidak seoptimal Apriori dan juga penggunaan memori masih cukup besar[11], yang sekarang kita kenal dengan nama Frequent Pattern Growth(Fp-Growth). Kemudian yang terakhir yang menjadi rujukan penulis adalah algoritma TPQ-apriori yang dikembangkan oleh[12], artikel rujukan ini telah tembus pada salah satu jurnal Q2 di jepang, teknik yang dilakukan dalam penelitian ini adalah mentransformasi format horizontal yang selama ini di pakai oleh algoritma apriori ke bentuk format vertical,

dimana teknik ini bisa mengurangi tuple sehingga dimensi dataset berkurang, serta mempartisi dataset yang akan diolah ke dalam beberapa partisi yang mengakibatkan scan dataset lebih cepat. Proses kerja metode yang diusulkan oleh peniliti[12], pada tahap awal dilakukan preprosesing dimana proses ini bertujuan supaya dataset yang diolah sesuai dengan format yang dibutuhkan sebelum dilakukan proses untuk pada tahap selanjutnya. Secara teknis dilakukan perubahan struktur tabel yang berkaitan dengan TID(Transaction Identifier) dan item, dengan tujuan untuk mempermudah proses partisi dataset dan pencarian frekuensi itemset. Tahap selanjutnya adalah dataset yang sebelumnya telah dipreprosesing akan dilakukan partisi berdasarkan record sesuai kebutuhan dengan tujuan untuk mempercepat pencarian frekuensi itemset dalam dataset dengan volume yang besar, dan tiap-tiap partisi tersebut diterapkan teknik TID-List Vertical dimana teknik ini bisa mengurangi record dataset sehingga mempercepat pencarian frekuensi itemset

2. Kajian Pustaka

Untuk menentukan state of the art dari penelitian ini penulis melakukan review terhadap beberapa paper yang relevan dengan topik yang di bahas. Ada beberapa artikel yang membahas tentang penerapan, hibrid dan komparasi algoritma association rules di antaranya:

a. Asosiasi Data Mining Menggunakan Algoritma Apriori Dan Fpgrowth Sebagai Dasar Pertimbangan Penentuan Paket Sembako.

Pada tahun 2020 penelitian ini[13] penulis menggunakan 1805 data transaksi, namun tidak di jelaskan berapa banyak item yang yang saling berelasi, dalam penelitian ini penulis mengklaim apriori lebih cepat 0.03s dari pada fp-growth 0.09s akan tetapi apriori mendapatkan 8 rule hanya pada tahap 2-itemset dan fp-growth mendapatkan 14 rule sampai pada tahap 3-itemset. Meskipun pada dasarnya model pencarian frekuensi itemset apriori dan fp-growth berbeda seharusnya itemset yang didapat sama-sama bisa sampai 3-itemset yang berbeda harusnya rulenya saja, tapi kenyataan berbeda, di samping itu data yang digunakan kurang baik sebab minimum support yang menjadi acuan sangat rendah.

b. Penerapan Data Mining dalam Analisis Pola Pembelian Minuman dan Makanan

di Kantin SMAN 4 Langsa Menggunakan Algoritma FP-Growth.

Pada tahun 2020 penelitian ini[14] meskipun data yang digunakan dalam rentang waktu 15 hari akan tetapi jelas item barangnya sangat sedikit. Penulis mengklaim fp-growth dalam penelitian ini mampu menampilkan hasil yang baik dari apriori padahal tidak ada komparasi dalam penelitian ini sementara klaimnya tidak bisa di buktikan.

c. Development of an inventory management system using association rule.

Pada tahun 2021 dalam penelitian[15] hanya berupa pengembangan sistem yang mana di dalamnya diterapkan algoritma association rule, tidak ada standart pengujian sistem yang di lakukan berupa apakah layak atau belum layak, apakah benar memberikan solusi dan lain sebainya.

d. Perbandingan Algoritma Apriori Dan Algoritma Fp-Growth Untuk Rekomendasi Item Paket Pada Konten Promosi.

Pada tahun 2021 penelitian ini[16] penulis menggunakan 33 paket penjualan, dengan jumlah total item 17 menggunakan 15% minimum support dan 20% minumum confidence, sementara rule yang terbentuk hanya 2-itemset, kelemahan dalam penelitian hanya sebatas 2-itemset saja, pada paket rekomendasi bisa 3-itemset atau bahka 4-itemset, data yang digunakan masih sedikit, minimum support dan confidence juga masih sedikit, sehingga untuk mengukurnya kurang optimal, dan yang terakhir tidak menghitung lifrasio.

e. Knowledge discovery in manufacturing datasets using data mining techniques to improve business performance.

Pada tahun 2022 dalam penelitian[17] bertujuan untuk mendeteksi korelasi antara kerusakan dengan penjualan, serta mampu memprediksi sehingga bisa melakukan atisipasi, namun tidak dijelaskan secara rinci jenis datanya dan lain-lain.

f. Data mining technique for grouping products using clustering based on association.

Pada tahun 2023 dalam penelitian ini[18] penulis melakukan hibrid algoritma yakni fp-growth dengan k-menas cluster, setelah di dapat rule base, barang-barang tersebut akan di kelompokan lagi bedasarkan cluster yang terbentuk. Dalam penelitian ini terdapat 29 produk dalam 12 transaksi dan separohnya

berda di atas minumum support dan minimum confidence.

3. Perancangan Sistem / Metode Penelitian

Bagian ini berisi mengenai tahapan-tahapan metode penelitian. Tahapan ini juga digunakan untuk menjelaskan tentang solusi yang diutarakan pada permasalahan penelitian dan untuk mencapai tujuan dari penelitian.

a. Desain Penelitian

Metode penelitian yang dilakukan adalah metode penelitian eksperimen, dengan tahapan penelitian sebagai berikut:

- Koleksi Dataset

Dataset yang digunakan bisa didownload adalah dataset yang diperoleh dari tosbera data transaksi penjualan CV. Charandita Kusuma NTB, dataset tersebut telah diupload pada situs www.kaggle.com/datasets/syahrir12345678/data-transaksi-cv-charandita-kusuma-ntb

- Pengolahan Awal Data(Pre-prosesing)

Pengolahan data awal meliputi, pembersihan data serta penyesuain dengan format dan ekstensi yang dibutuhkan.

- Metode Yang Digunakan Untuk Komparasi

Pada tahapan ini data akan dianalisa, setelah data dianalisa lalu diterapkan dengan metode-metode yang sesuai dengan jenis data tersebut.

- Eksperimen Dan Pengujian Metode

Pada tahap ini dijelaskan langkah-langkah eksperimen komprasi untuk menganalisa rule dari metode mana yang paling baik dan optimal.

- Evaluasi

Pada tahapan ini, dilakukan evaluasi hasil penerapan metode yang diusulkan dalam mengetahui pembentukan rule yang optimal berdasarkan hasil support, confidence, dan lifrasio.

b. Koleksi Data

Pengumpulan atau koleksi data merupakan tahap awal yang dilakukan, dimana dari masalah yang ada, akan dikumpulkan data terkait berupa dataset yang diperoleh dari data minimarket koperasi pegawai telkom kota Mataram NTB Indonesia.

c. Pengolahan Awal Data(Pre-prosesing)

Sebelum masuk pada model algoritma data harus dipreprosesing terlebih dahulu untuk memastikan format yang akan masuk pada model sesuai dengan yang diharapkan. Semakin baik data yang akan diolah maka akan semakin optimal hasil dari proses algoritma tersebut, dalam penelitian ini proses pre-prosessing mengganti formatnya ke bentuk yang dibutuhkan.

d. Komparasi Metode

Untuk komparasi kita menggunakan tiga metode yakni algoritma apriori, algoritma fp-growth dan algoritma TPQ-apriori, sementara tools yang digunakan adalah weka, rapidminer dan program aplikasi yang rancangan sendiri. Ada tiga algoritma association rule yang akan digunakan dalam penelitian ini untuk menganalisa rule yang terbentuk hasil dari mengkomparasi ke tiga algoritma tersebut, sementara data yang akan diolah untuk pengujian adalah data transaksi penjualan CV. Charandita Kusuma NTB.

- Algoritma Fp-Growth

Algoritma fp-growth di kembangkan dari algoritma apriori, tentu keduanya saling melengkapi, fp-growth dalam proses pencarian frekuensi itemset sudah sangat baik dari algoritma apriori namun rule yang dihasilkan tidak sebaik algoritma apriori, juga penggunaan memori masih cukup besar. Algoritma fp-growth adalah salah satu algoritma dari teknik association rules yang dapat digunakan untuk menentukan himpunan data yang paling sering muncul (frequent itemset) dalam sebuah kumpulan data dengan pendekatan pada konsep fp-tree[10]. Penting tidaknya suatu aturan asosiatif dapat diketahui dengan dua parameter yaitu support dan confidence[4],[10]. Support(nilai penunjang) adalah persentase kombinasi item tersebut dalam database. Confidence(nilai kepastian) adalah kuatnya hubungan antar item dalam aturan asosiasi. Sebuah aturan asosiasi dikatakan interesting jika nilai support adalah lebih besar dari minimum support dan juga nilai confidence adalah lebih besar dari minimum confidence. Sementara untuk menguji nilai kevalidan hubungan antar item menggunakan lifrasio.

Formula Support:

(1)

$$Support_{A,B} = \frac{Jml\ Transaksi\ A\ dan\ B}{Total\ Transaksi}$$

Formula Confidence:

(2)

$$Confidence_{A,B} = \frac{Jml\ Transaksi\ A\ dan\ B}{Jml\ Transaksi\ A}$$

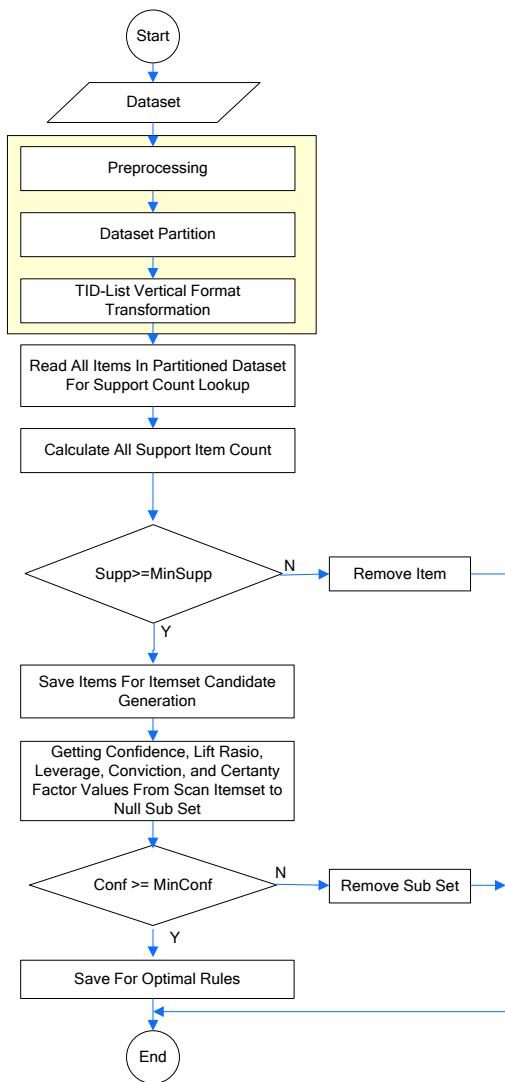
Formula Lifrasio:

(3)

$$Lift\ Rasio = \frac{Support_{A\ dan\ B}}{Support\ A \times Support\ B}$$

- Algoritma TPQ-Apriori

TPQ adalah singkatan dari Tid-list vertical Partitioning Query, menggunakan teknik format tid-list vertical dengan sistem partisi berbasis query, algoritma di klaim mampu lebih cepat untuk proses scan dataset dalam pencarian frekuensi itemset serta mampu menghasilkan rule yang optimal[12]. Adapun tahap awal dilakukan yakni pre-prosesing untuk melakukan penyesuaian format sesuai yang dibutuhkan dengan ekstensi dot sql(.sql). Tahap selanjutnya dataset dipartisi, setelah proses partisi dataset selesai, tiap partisi akan di terapkan pendekatan Tid-list vertical yang tujuannya dengan pendekatan ini record-recordnya bisa terkurangi dan proses pencarian frekuensi itemset bisa lebih cepat. Alur lengkap dari algoritma TPQ-Apriori kita lihat pada Gambar 1



Gambar 1. Alur Algoritma Tpq-Apriori

e. Eksperimen Dan Pengujian Metode

Analisa perbandingan rule hasil dari TPQ-Apriori dengan Fp-Growth dalam tools rapidminner menggunakan dataset data transaksi penjualan CV. Charandita Kusuma NTB, untuk datasetnya bisa di download <https://www.kaggle.com/datasets/syahrir12345678/data-transaksi-cv-charandita-kusuma-ntb>.

Untuk pengujian rule hasil dari TPQ-Apriori dilakukan menggunakan tools yang dikembangkan sendiri dalam penelitian ini menggunakan tools Microsoft Visual Studio 2013 dan Mysql server 5.5 yang mendukung paralelisasi.

f. Evaluasi

Proses evaluasi akan memberikan hasil tentang nilai support, confidence, liftrasio, conviction, leverage dan certainty factor.

Analisis Perbandingan Algoritma Fp-Growth dan Tpq-Apriori Dalam Menentukan Rule Based Terbaik Untuk Sistem Rekomendasi Produk

Support menentukan seberapa sering aturan tersebut diterapkan dalam dataset[1][2][3][4]. Support atau dukungan adalah indikasi seberapa sering itemset muncul dalam dataset. Support dapat diformulasikan seperti Rumus 4.

(4)

$$\text{Support } A, B = \frac{\text{Jml Transaksi } A \text{ dan } B}{\text{Total Transaksi}}$$

Dimana :

N=Total Transaksi

X=Antecedent

Y=Consequent

Confidence menentukan frekuensi item dalam Y muncul dalam transaksi yang mengandung X. Confidence atau kepercayaan seberapa sering aturan atau rule itu tersebut terbukti benar. Confidence dapat diformulasikan seperti pada Rumus 5.

(5)

$$\text{Confidence } A, B = \frac{\text{Jml Transaksi } A \text{ dan } B}{\text{Jml Transaksi } A}$$

Untuk menguji valid dan tidaknya suatu aturan atau hubungan antara item menggunakan liftrasio. Liftrasio adalah nilai yang mengukur besarnya hubungan antara antecedent dan consequent yang tidak saling bergantung (independent) [1][2][3][4]. Liftrasio memiliki range mulai dari 0 sampai dengan ∞ . Nilai yang mendekati 1 mengindikasikan bahwa antecedent dan consequent tidak memiliki ketergantungan. Nilai yang jauh dari 1 mengindikasikan bahwa antecedent menyediakan informasi tentang consequent. Atau dengan pemahaman lain jika lift adalah > 1 itu memungkinkan kita mengetahui sejauh mana dua kejadian itu bergantung satu sama lain, dan membuat aturan itu berpotensi berguna untuk memprediksi consequent dalam dataset. Dan jika lift < 1 itu membuat kami tahu bahwa item-item itu saling menggantikan. Ini berarti bahwa kehadiran satu item memiliki efek negatif pada keberadaan item lain dan sebaliknya.

(6)

$$\text{Lift Rasio} = \frac{\text{Support}(A \text{ dan } B)}{\text{Support } A \times \text{Support } B}$$

Dimana :

∞ =Infinity Atau Null

σ = Support Count

\cup =Berhubungan

Selain liftrasio kita bisa menguji dengan formula lain yakni conviction atau nilai keyakinan[1][2][3][4]. Conviction adalah nilai yang mengukur tingkat implikasi dari suatu aturan. Conviction sangat memperhatikan arah dari suatu association rule. Conviction mengindikasikan bahwa $\text{conviction}(X \rightarrow Y) \neq \text{conviction}(Y \rightarrow X)$.

$$\text{Conv}(X - Y) = (1 - \sigma(Y)) / (1 - \text{conf}(X - Y)) \quad (7)$$

Untuk banyaknya item antecedent dan consequence yang dijual secara bersamaan dalam suatu dataset yang lebih dari yang diharapkan kita bisa menggunakan Leverage. Nilai 0 menunjukkan antecedent dan consequent independent. Leverage memiliki jangkauan nilai dari -0,25 sampai dengan 0,25. Dapat diformulasi seperti pada rumus 8.

(8)

$$\text{Leve}(X - Y) = \sigma(X \cup Y) - (\sigma(X) * \sigma(Y))$$

Dalam penelitian ini untuk tambahan evaluasi untuk mengambil suatu keputusan dalam menilai hubungan atau korelasi rule yang terbentuk menggunakan metode certainty factor. Metode certainty factor menurut David McAllister “ adalah suatu metode untuk membuktikan apakah suatu fakta itu pasti atau tidak pasti berbentuk, metrik yang biasanya digunakan dalam sistem pakar”.

(9)

$$\text{CF}(X - Y) = (\text{conf}(X - Y) - \sigma(Y)) / (1 - \sigma(Y))$$

4. Implementasi Sistem dan Hasil

Eksperimen untuk melihat hasil pembentukan rule optimal atau tidak. Dasar yang digunakan dalam mencari frekuensi itemset adalah algoritma apriori yang berbasis query, untuk mempercepat proses tersebut TPQ-Apriori melakukan split atau partisi dataset. Masing-masing dataset yang telah dipartisi, hasil partisi format tabel awalnya berbentuk horizontal akan ditransformasikan ke bentuk vertical dan dengan dua pendekatan ini, proses pencarian frekuensi pada algoritma Apriori yang

dikembangkan menjadi lebih cepat. Untuk proses pengujian runtime dilakukan berulang-ulang sebanyak 10 kali untuk memastikan rule yang dihasilkan dari pengujian tersebut mampu menghasilkan rule yang konsisten.

a. Dataset

Sumber data yang di peroleh adalah data public yang di download pada link (<https://www.kaggle.com/datasets/syahrir1234/5678/data-transaksi-cv-charandita-kusuma-ntb>). Total record data lebih kurang 796 record, sementara total item data terdiri dari 27 item, dan data ini terdiri dari 110 transaksi. Data yang di ambil adalah data proses transaksi belanja pada supermarket. Sebelum data digunakan data dilakukan preprosesing, model preprosesing dilakukan dengan membentuk tabel binari binominal serta partisi dataset.

Pada Tabel 1 terdapat 110 transaksi dengan 27 item, item dalam algoritma apriori akan menjadi atribut yang akan di lakukan pencarian frekuensinya. Data ini merupakan data transaksi penjualan.

b. Pengujian Pembentukan Rule

Pada tahapan ini yakni tahapan ini di lakukan secara bersamaan dan rule yang dihasilkan oleh algoritma fp-growth dengan tpq-apriori, sehingga tidak dibuatkan tabel terpisah. Untuk ambang batas tresholdpun di buat sama. Dengan menggunakan dataset yang sama, atribut yang sama, serta ambang treshold yang sama. Pengujian program di lakukan dengan acuan perbandingan rule yang terbentuk, dengan nilai support, confidence, nilai liftrasio dan conviction. Dengan ukuran data yang sama untuk record maupun fitur. Sebagai bahan pengujian menggunakan data public yang telah di uraikan pada pembahasan sebelumnya. Pada Tabel 2 kita bisa melihat data yang digunakan terdiri dari 27 atribut. Selanjutnya data ini ada di olah sehingga bisa mendapatkan frekuensi dan menghitung support untuk kandidat 1-Itemset. Adapun persamaan untuk menghitung support :
Support (A) = (jumlah transaksi mengandung A) / (total transaksi) x 100.

Tabel 1. Tabel Tabular Dataset

TRANS	OB AJA 10 W-T-BU LB CLA SSI C	KAC AN G ASI N	KAC AN G KO RO	EM PIN G PED AS	RENG GINA NG PUTIH	KAC AN G TOB A BAL AD O	KUE BA WA NG	KAC AN G POL ON G TEP UN G	OB AJA 30 W-T-BU LB CLA SSI C	AMB YAR 6W PRE MIU M GEL AS	PA NG - PA NG PE DA S	KAC AN G TEL UR	KR UP UK EM PIN G BEL INJ O	MAC ARO NI	KRI PIK PIS AN G	KAC AN G KUL IT OV EN	KR UP UK PA ND A	AMB YAR 4,5 W PRE MIU M GEL AS	KRU PUK KEN TAN G PUTI H	KRI PIK SU KU N	AMB YAR 12W PRE MIU M GEL AS	DO DOL NA NG KA BES AR	KRU PUK TEN GGI RI	KAC AN G AT OM	FITI NG GAN TUN G HITA M
ABDI JAYA BAN GUN AN	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ABIL A	0	1	1	1	0	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
AINI	0	0	0	1	0	1	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0
AL-RIZKI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
AL-USRO H	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
...
...
...
VEDI NA	0	0	1	1	0	0	0	0	1	0	0	1	0	1	1	1	1	0	0	0	1	0	0	0	0
W. FAMI LY	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
WAHI DAH	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
WILL Y	1	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	1	0	0	1	0	0	0	0
WR INSA N	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Analisis Perbandingan Algoritma Fp-Growth dan Tpq-Apriori Dalam Menentukan Rule Based Terbaik
Untuk Sistem Rekomendasi Produk

Tabel 1. Rule 2-Itemset.

No	Rule		FP-Growth				TPQ Apriori				
	Premises	Conclusion	Support (0,5)	Confidence (0,8)	Lift Ratio	Conviction	Support (0,5)	Confidence (0,8)	Lift Ratio	Conviction	Leverage
1	Kacang Koro	Kacang Asin	0.18	0.87	3.67	5.85	0.18	0.83	3.39	4.44	0.13
2	Emping Pedas	Kacang Asin	0	0	0	0	0.17	0.83	3.36	4.44	0.12
3	Kue Bawang	Kacang Asin	0.15	0.94	3.99	13.74	0.16	1	4.07	∞	0.12
4	Rengginang Putih	Kacang Asin	0.16	0.85	3.62	5.34	0.16	0.86	3.49	5.39	0.12
5	Kacang Polong Tepung	Kacang Asin	0.14	0.9	3.98	12.98	0.15	1	4.07	∞	0.12
6	Kacang Toba Balado	Kacang Asin	0.14	0.8	3.38	3.81	0.14	0.8	3.26	3.77	0.1
7	Stik Keju	Kacang Asin	0.12	0.8	3.48	4.32	0.13	0.88	3.59	6.29	0.1
8	Kacang Telur	Kacang Asin	0.13	0.9	3.96	12.21	0.13	0.94	3.82	12.58	0.1
9	Krupuk Panda	Kacang Asin	0.11	0.9	3.92	10.69	0.12	1	4.07	∞	0.1
10	Pang-Pang Pedas	Kacang Asin	0.10	0.8	3.38	3.81	0.12	0.88	3.56	6.29	0.09
----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
165	Macaroni	Krupuk Emping Belinjo	0	0	0	0	0.9	0.67	4.888	2.62	0.07
166	Kripik Pisang	Krupuk Emping Belinjo	0	0	0	0	0.9	0.71	5.238	2.98	0.07
167	Kacang Atom	Krupuk Emping Belinjo	0	0	0	0	0.8	0.75	5.499	3.45	0.07
168	Kripik Sukun	Krupuk Emping Belinjo	0	0	0	0	0.8	0.69	5.076	2.79	0.07
169	Sumpia	Krupuk Emping Belinjo	0	0	0	0	0.7	0.62	4.511	2.27	0.06
170	Krupuk Tenggiri	Krupuk Emping Belinjo	0	0	0	0	0.6	0.58	4.275	2.06	0.05
171	Dodol Nangka Besar	Krupuk Emping Belinjo	0	0	0	0	0.6	0.54	3.946	1.88	0.05
172	AMBYAR 4,5W PREMIUM GELAS	Krupuk Emping Belinjo	0	0	0	0	0.09	0.08	0.565	0.94	-0.01
173	AMBYAR 12W PREMIUM GELAS	Krupuk Emping Belinjo	0	0	0	0	0.09	0.08	0.565	0.94	-0.01

Analisis Perbandingan Algoritma Fp-Growth dan Tpq-Apriori Dalam Menentukan Rule Based Terbaik Untuk Sistem Rekomendasi Produk

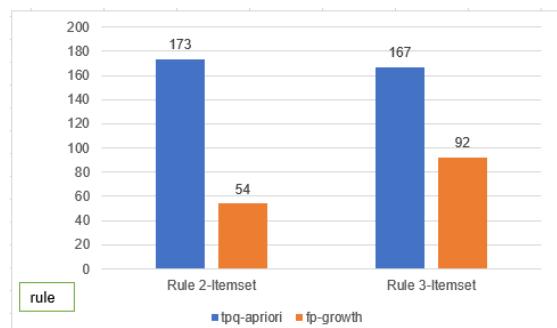
Tabel 2. Rule 3-Itemset.

No	Rule		FP-Growth				TPQ Apriori				
	Premises	Conclusion	Support (0,5)	Confidence (0,8)	Lift Ratio	Conviction	Support (0,5)	Confidence (0,8)	Lift Ratio	Conviction	Leverage
1	Rengginang Putih, Kacang Toba Balado	Kacang Asin	0.10	0.8	∞	5	0.10	0.8	3.26	3.77	0.08
2	Emping Pedas, Kacang Toba Balado	Kacang Asin	0.12	0.8	∞	5.66	0.13	0.88	3.596	6.29	0.1
3	Emping Pedas, Rengginang Putih	Kacang Asin	0.10	0.8	∞	7	0.11	0.93	3.784	10.78	0.09
4	Emping Pedas, Stik Keju	Kacang Asin	0.10	0.8	∞	7	0.10	0.86	3.492	5.39	0.08
5	Kacang Koro, Stik Keju	Kacang Asin	0.10	0.8	∞	7	0.11	0.87	3.531	5.8	0.08
6	Emping Pedas, Kacang Koro	Kacang Asin	0.13	0.8	∞	8.5	0	0	0	0	0
7	Kacang Koro, Kacang Kulit Oven	Kacang Asin	0.10	0.9	∞	13	0.11	0.93	3.784	10.78	0.09
8	Emping Pedas, Kacang Telur	Kacang Asin	0.11	0.9	∞	14	0.12	1	4.075	∞	0.1
9	Kacang Koro, Rengginang Putih	Kacang Asin	0.15	0.9	∞	18	0.15	0.94	3.847	12.58	0.11
10	Kacang Koro, Emping Pedas	Kacang Asin	0	0	0	0	0.14	0.89	3.623	6.86	0.11
----	----	----	----	----	----	----	----	----	----	----	----
----	----	----	----	----	----	----	----	----	----	----	----
----	----	----	----	----	----	----	----	----	----	----	----
160	Kacang Koro, Krupuk Panda	Kacang Polong Tepung	0	0	0	0	0.10	0.73	4.745	3.13	0.08
161	Kacang Koro, Kacang Atom	Kacang Polong Tepung	0	0	0	0	0.10	0.85	5.475	5.64	0.08
162	Emping Pedas, Kacang Toba Balado	Kacang Polong Tepung	0.13	0.8	∞	8.5	0	0	0	0	0
163	Kacang Asin, Kacang Toba Balado	Kacang Polong Tepung	0.12	0.8	∞	8	0	0	0	0	0
164	Kacang Asin, Stik Keju	Kacang Polong Tepung	0.10	0.8	∞	7	0	0	0	0	0
165	Kacang Asin, Emping Pedas	Kacang Polong Tepung	0.13	0.8	∞	6	0	0	0	0	0
166	Emping Pedas, Kacang Koro	Kacang Polong Tepung	0.12	0.8	∞	5.66	0	0	0	0	0
167	Kacang Koro, Kacang Toba Balado	Kacang Polong Tepung	0.11	0.8	∞	5.33	0	0	0	0	0

Analisis Perbandingan Algoritma Fp-Growth dan Tpq-Apriori Dalam Menentukan Rule Based Terbaik
 Untuk Sistem Rekomendasi Produk

Pada Tabel 2 kita bisa melihat warna hijau adalah rule base dari algoritma fp-growth dengan tools rapidminer sementara warna kuning adalah rule base dari algoritma tpq-apriori, untuk rule yang di beri warna merah muda adalah rule yang seharusnya ada tapi tidak di temukan pada algoritma fp-growth menggunakan tools rapidminer sementara pada tpq-apriori ditemukan, total rule 2-itemset yang terbentuk adalah 173 rule menggunakan algoritma tpq-apriori, sementara 54 rule menggunakan algoritma fp-growth. Pada Tabel 3 kita menggunakan pemberian warna yang sama dengan Tabel 2. Total rule 3-itemset yang terbentuk adalah 167 rule dengan menggunakan algoritma tpq-apriori sementara 92 rule dari algoritma fp-growth, untuk data lengkap hasil pengujian rule di dilihat di <https://www.kaggle.com/datasets/syahrir12345678/hasilpenquiantpq-apriorivsfp-growth>.

Untuk mempermudah kita bisa lihat pada Gambar 2 grafik perbandingan rule



Gambar 2. Grafik Perbandingan Rule 2-Itemset dan 3-Itemset Dengan Algoritma Tpq-Apriori dan Fp-Growth.

Pada Gambar 2 kita bisa melihat rule yang dihasilkan oleh algoritma tpq-apriori jauh lebih banyak dari algoritma fp-growth, baik untuk 2-itemset maupun 3-itemset, ini mengindikasikan bahwa algoritma tpq-apriori mampu menghasilkan rule yang optimal apabila dibandingkan dengan fp-growth, sebab pendekatan kedua algoritma ini dalam mencari frekuensi itemset berbeda, tpq apriori mencari berdasarkan kumpulan item yang di dapat dalam sebuah dataset, sementara fp-growth menggunakan pendekatan fp-tree dalam bagan pohon, sehingga node utama adalah item yang memiliki frekuensi paling tinggi dan node utama dalam fp tree tidak dimasukan dalam lintasan kombinasi, sehingga kombinasi rule untuk item tersebut tidak terbentuk.

5. Kesimpulan

Algoritma tpq-apriori adalah salah satu bagian dari algoritma association rule dengan pendekatan asosiatif hasil pengembangan dari algoritma apriori tradisional, algoritma apriori mudah diimplementasikan dengan kelebihan mampu menghasilkan kombinasi rule yang optimal. Akan tetapi kelemahahnya adalah waktu scan dataset yang sangat lama. Scan dataset yang lama disebabkan oleh proses pencarian frekuensi item atau support count dalam kumpulan dataset. Item dalam algoritma association rule merupakan atribut atau fitur atau dimensi yang akan dilakukan pencarian frekuensi atau support countnya. Sehingga semakin besar volume dan semakin banyak item-itemnya maka waktu yang dibutuhkan akan semakin lama. Lamanya waktu scan dataset akan berimbas pada besarnya penggunaan memori dan prosesor itulah yang menjadi dasar algoritma tpq-apriori.

Masalah pada penelitian ini adalah algoritma association rule khususnya fp-growth terkadang tidak optimal dalam menghasilkan rule base sementara tpq-apriori mampu optimal. Prioritas dalam penelitian ini adalah melakukan komparasi dua algoritma yakni fp-growth, dan tpq-apriori dengan dataset yang sama untuk melihat dan memastikan terbentuknya rule yang optimal dan konsisten. Metode yang digunakan untuk bahan komparasi dan perbandingan adalah algoritma fp-growth serta tpq-apriori. Algoritma fp-growth adalah metode yang dikembangkan oleh jia wei han tahun 2000, serta tpq-apriori yang dikembangkan oleh Moch. Syahir tahun 2023. Untuk pengujian rule, hasil rule tpq-apriori dibandingkan dengan algoritma fp-growth dalam tools rapidminer. Dari hasil pengujian tpq-apriori mampu lebih optimal dan konsisten dari fp-growth dalam tools rapidminer. Untuk mempermudah kita bisa lihat seperti pada grafik di bawah ini. Kita bisa melihat data yang ditujukan oleh grafik pada Gambar 1, data yang digambarkan dengan warna biru adalah rules hasil dari tpq-apriori sementara yang warna orange adalah rules hasil dari fp-growth.

6. Pustaka

- [1] C. Aggarwal, S. Mita, Y. Yamazoe, T. Kamataki, and R. Kato, *Data Mining The Textbook*, vol. 14, no. 3. 2015. doi: 10.1016/0304-3835(81)90152-X.
- [2] H. Witten, Frank, *Data Mining Practical Machine Learning*. Burlington USA: Morgan Koufmann, 2011.
- [3] D. T. Larose, *Data Mining Methods and*

- [4] Models, vol. 28. A JOHN WILEY & SONS, INC PUBLICATION, 2007. doi: 10.1017/S0068245400011199.
- [5] J. H. Caru C Aggarwal, *Frequent Pattn Mining*, vol. 9783319078. Springer Cham Heidelberg NewYork Dordrecht London, 2014. doi: 10.1007/978-3-319-07821-2_7.
- [6] Fitria, Nengsih, and Qudsi, "Implementasi Algoritma FP-Growth Dalam Penentuan Pola Hubungan Kecelakaan Lalu Lintas," *J. Sist. Inf.*, vol. 13, no. 2, p. 118, 2017, doi: 10.21609/jsi.v13i2.551.
- [7] C. Kumar, "A Survey on Association Rule Mining using Apriori Algorithm," *Int. Conf. Adv. Comput. Commun. Technol. ACCT*, vol. 2015-April, no. 5, pp. 212–216, 2015, doi: 10.1109/ACCT.2015.69.
- [8] H. Xue, Q. Yang, and S. Chen, *The Top Ten Algorithms in Data Mining*, vol. 6, no. SVM. New York: CRC PRESS TAYLOR & FRANCIS GROUP, 2009. doi: 10.1007/s10115-007-0114-2.
- [9] R. Agrawal, T. Imieliński, and A. Swami, "Mining Association Rules Between Sets of Items in Large Databases," *ACM SIGMOD Rec.*, vol. 22, no. 2, pp. 207–216, 1993, doi: 10.1145/170036.170072.
- [10] S. Agrawal, "Fast Algorithms for Mining Association Rules," *Jisuanji Gongcheng/Computer Eng.*, vol. 30, no. 15, p. 34, 2004.
- [11] Y. Y. Jiawei Han, Jian Pei, "Mining Frequent Patterns without Candidate Generation," *Sch. Comput. Sci.*, vol. 22, no. 3, pp. 1–12, 2000.
- [12] E. Darrab, "Vertical Pattern Mining Algorithm for Multiple Support Thresholds," *Procedia Comput. Sci.*, vol. 112, pp. 417–426, 2017, doi: 10.1016/j.procs.2017.08.051.
- [13] M. Syahrir and A. Anggrawan, "Improvement of Apriori Algorithm Performance Using the TID-List Vertical Approach and Data Partitioning," *Int. J. Intell. Eng. Syst.*, vol. 16, no. 2, pp. 176–191, 2023, doi: 10.22266/ijies2023.0430.15.
- [14] H. Maulidiya and A. Jananto, "Asosiasi Data Mining Menggunakan Algoritma Apriori dan FP-Growth sebagai Dasar Pertimbangan Penentuan Paket Sembako," *Proceeding SENDIU 2020*, vol. 6, pp. 36–42, 2020.
- [15] R. F. Haya and D. Ramadani, "Penerapan Data Mining dalam Analisis Pola Pembelian Minuman dan Makanan di Kantin SMAN 4 Langsa Menggunakan Algoritma FP-Growth," *InfoTekjar J. Nas. Inform. dan Teknol. Jar.*, vol. 4, no. 2, pp. 331–334, 2020, [Online]. Available: <https://doi.org/10.30743/infotekjar.v4i2.1826>
- [16] T. O. Oladele, R. O. Ogundokun, A. A. Adegun, E. A. Adeniyi, and A. T. Ajanaku, "Development of an inventory management system using association rule," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 21, no. 3, pp. 1868–1876, 2021, doi: 10.11591/ijeeecs.v21.i3.pp1868-1876.
- [17] M. Mariko, "Perbandingan Algoritma Apriori Dan Algoritma Fp-Growth Untuk Rekomendasi Item Paket Pada Konten Promosi," *Explore*, vol. 11, no. 2, p. 24, 2021, doi: 10.35200/explore.v11i2.438.
- [18] A. G. Shaaban, M. H. Khafagy, M. A. Elmasry, H. El-Beih, and M. H. Ibrahim, "Knowledge discovery in manufacturing datasets using data mining techniques to improve business performance," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 26, no. 3, pp. 1736–1746, 2022, doi: 10.11591/ijeeecs.v26.i3.pp1736-1746.
- [19] E. P. W. Mandala and D. E. Putri, "Data mining technique for grouping products using clustering based on association," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 31, no. 2, pp. 835–844, 2023, doi: 10.11591/ijeeecs.v31.i2.pp835-844.