

Komparasi Algoritma *Equivalence Class Transformation (ECLAT)* dan *FP-Growth* pada Analisis Data Penjualan

Fahrul Ardian Nugroho^{*1}, Bambang Pramono², Isnawaty³, Statiswaty⁴

^{1,2,3,4}Jurusan Teknik Informatika, Fakultas Teknik, Universitas Halu Oleo

e-mail: ^{*1}fahrul.sultra@gmail.com, ²bambang.pramono@uho.ac.id, ³isnawaty@uho.ac.id,

⁴statiswaty@uho.ac.id

Abstrak

Minimarket Sinar Mart Kendari menghadapi persaingan ketat dalam sektor ritel. Masalah pemahaman perilaku konsumen menjadi hal yang harus menjadi perhatian. Saat ini, data penjualan yang berlimpah merupakan sumber yang belum dimanfaatkan secara maksimal. Pemanfaatan data dapat dilakukan dengan teknik *data mining*, khususnya analisis keranjang belanja. Terdapat beragam algoritma *data mining*, namun fokus penelitian ini adalah perbandingan algoritma *ECLAT* dan *FP-Growth*. Algoritma ini dipilih karena sesuai karakteristik data dan tujuan penelitian. Hasil penelitian menunjukkan bahwa sistem analisis menggunakan algoritma *ECLAT* dan *FP-Growth* berhasil dibuat. Hal ini terlihat dari suksesnya pengujian *black box*, manual, dan validitas. Dalam eksperimen menggunakan data 76.168 transaksi, dengan nilai *minimum support* 0.005%, 0.010%, 0.015%, 0.020%, 0.025% dan 0.030%, serta *minimum confidence* sebesar 60%, didapati jumlah aturan yang dihasilkan sama. Selain itu pada eksperimen tersebut dapat disimpulkan bahwa *ECLAT* lebih efisien dalam hal konsumsi memori, sementara *FP-Growth* lebih cepat dalam menganalisis pola asosiasi baik ketika data yang diproses kecil maupun besar.

Kata kunci—Penambangan Data, Aturan Asosiasi, Analisis Belanja, *ECLAT*, *FP-Growth*

Abstract

Sinar Mart Kendari Minimarket faces tough competition in the retail sector. The problem of understanding consumer behavior is something that must be paid attention to. Currently, the abundance of sales data is a resource that has not been utilized optimally. Utilization of data can be done using data mining techniques, especially market basket analysis. There are various data mining algorithms, but the focus of this research is a comparison of the ECLAT and FP-Growth algorithms. This algorithm was chosen because it suits the characteristics of the data and research objectives. The research results show that the analysis system using the ECLAT and FP-Growth algorithms was successfully created. This can be seen from the success of black box, manual, and validity testing. In experiments using data from 76,168 transactions, with minimum support values of 0.005%, 0.010%, 0.015%, 0.020%, 0.025%, and 0.030%, and a minimum confidence of 60%, it was found that the number of rules produced was the same. Apart from that, from this experiment, it can be concluded that ECLAT is more efficient in terms of memory consumption, while FP-Growth is faster in analyzing association patterns both when the data being processed is small and large.

Keywords—Data Mining, Association Rule, Market Basket Analysis, *ECLAT*, *FP-Growth*

1. PENDAHULUAN

Minimarket Sinar Mart telah mengadopsi teknologi komputer untuk merekam transaksi harian pelanggan, dengan

data akumulatif disimpan dalam *database* kasir. Meskipun, pemanfaatan data transaksi terbatas pada laporan mingguan atau bulanan untuk menghitung keuntungan. Menurut manajer, rata-rata terdapat sekitar 400

transaksi harian atau 12.000 transaksi per bulan. Sayangnya, potensi pemanfaatan data ini belum maksimal, seperti mengidentifikasi pola pembelian konsumen.

Cara untuk mendapatkan pola pembelian dari *database* dapat menggunakan suatu yang disebut *data mining*. *Data mining* digunakan untuk mendapatkan informasi terkini yang berharga dan bermanfaat dalam pengumpulan data yang melibatkan antara manusia dan perangkat komputer, baik melalui cara otomatis ataupun manual [1].

Sebagaimana tujuan dari sistem analisis yang akan dibangun adalah menemukan asosiasi antar barang pada toko ritel, terdapat banyak kategori algoritma *data mining* seperti *classification*, *regression*, *time series*, *clustering*, *summarization*, *association rule*, dan *sequence discovery* [2]. Algoritma *association rule mining* sangat sesuai untuk tujuan ini karena mampu mengidentifikasi hubungan atau pola menarik antar variabel dalam *dataset* besar.

Association rule mining memiliki beberapa pendekatan, antara lain *frequent pattern*, *sequential pattern*, dan *structured pattern* [2]. Setelah dilakukan *review* dengan kegunaan setiap pendekatan, karena hasil akhir yang ingin didapatkan adalah pola asosiasi antar barang, pendekatan yang paling cocok di antara ketiganya adalah *frequent itemset mining*. Hal ini karena metode tersebut dapat menemukan pola pembelian dari pelanggan.

ECLAT dan *FP-Growth* adalah algoritma *frequent itemset mining* populer yang lebih baik daripada Apriori dalam hal waktu eksekusi dan penggunaan memori [2]. Kedua algoritma ini mengadopsi pendekatan yang berbeda, *ECLAT* menggunakan pendekatan format vertikal sementara *FP-Growth* menggunakan pendekatan *pattern growth*. *FP-Growth* adalah salah satu algoritma alternatif untuk menemukan himpunan item yang sering muncul dari kumpulan data [3]. Karakteristik dari algoritma ini adalah struktur data yang digunakan dalam pohon yaitu *FP-Tree*. Melalui *FP-Tree*, algoritma *FP-Growth* dapat secara langsung mengekstrak himpunan item yang sering muncul dari *FP-Tree*. Dalam penelitian ini juga menggunakan algoritma lainnya yaitu *ECLAT* yang merupakan

pencarian data menggunakan algoritma *depth-first* untuk menentukan *intersection* yang spesifik. *ECLAT* menggunakan *database* dengan tata letak vertikal. Setiap item disimpan bersama dengan penutupnya (*tidlist*) dan menggunakan pendekatan berbasis *intersection* untuk menghitung dukungan dari suatu *itemset* [4].

Tantangan dalam menganalisis data transaksi penjualan di minimarket Sinar Mart adalah volume data yang besar. *ECLAT* dan *FP-Growth*, meskipun berbeda pendekatan, perbedaan dalam berbagai aspek seperti teknik, strategi, pemindaian data, metodologi, dan struktur penyimpanan. Namun, keduanya memenuhi syarat dukungan kapasitas data dan sesuai dengan tujuan analisis.

Terdapat beberapa penelitian terdahulu yang berkaitan dengan perbandingan algoritma *frequent-itemset mining*. Penelitian yang dilakukan oleh Krisna et al. [5] mengkaji analisis pola data transaksi penjualan dengan menerapkan algoritma *FP-Growth* dan *ECLAT*. Fokus utama perbandingan tersebut adalah pada metrik waktu proses dan jumlah aturan yang dihasilkan. Berbeda dengan penelitian sebelumnya, penelitian yang penulis lakukan menambahkan metrik baru dalam perbandingan, yaitu analisis penggunaan ruang memori. Hal ini adalah aspek penting yang perlu dianalisis untuk mendapatkan pemahaman yang lebih komprehensif terkait performa algoritma. Penelitian selanjutnya yaitu oleh Risna et al. [6], penelitian tersebut menganalisis jumlah aturan yang dihasilkan dari data transaksi dengan menerapkan algoritma *FP-Growth* dan *ECLAT*. Namun, perbandingan hanya berdasarkan jumlah aturan, sedangkan penulis memasukkan metrik tambahan seperti kecepatan eksekusi proses dan konsumsi ruang memori. Selain itu, penelitian sebelumnya hanya menggunakan 6.038 data transaksi, dimana kurang representatif untuk memahami kinerja algoritma dalam berbagai situasi. Sedangkan dalam penelitian ini, penulis menggunakan data lebih besar, yaitu 76.168 transaksi. Penelitian lainnya yaitu oleh Sudarsono et al. [7] melakukan perbandingan dari algoritma *ECLAT* dan *FP-Growth*. Peneliti menetapkan nilai *minimum support* 0.001 dan untuk pengujian pertama dengan nilai *minimum*

confidence 0.8 dan pengujian kedua dengan minimum confidence 1.0. Hal yang membedakan dengan yang penulis lakukan adalah pada penelitian tersebut uji coba dilakukan dengan menetapkan parameter confidence yang berbeda, sedangkan yang penulis lakukan dalam penelitian ini adalah menggunakan parameter minimum support yang berbeda serta lebih banyak sehingga memungkinkan untuk merepresentasikan pada berbagai situasi.

2. METODE PENELITIAN

2.1. Tahap penelitian

Tahap penelitian yang diterapkan dalam penelitian ini terdiri dari studi pustaka, pengumpulan data, analisis data, dan penarikan kesimpulan.

a. Studi Pustaka

Pengumpulan data, informasi tertulis atau data sekunder yang bersumber melalui referensi buku maupun internet, untuk mendukung penelitian semakin relevan.

b. Pengumpulan Data

Peneliti mengambil data yang berasal dari minimarket Sinar Mart untuk kemudian diolah dan implementasikan.

c. Analisis Data

Metode yang digunakan untuk menganalisis data menggunakan Knowledge Discovery in Database (KDD). Proses ini dimulai dari data selection sampai evaluasi.

d. Penarikan Kesimpulan

Tahap akhir pada penelitian dari algoritma ECLAT dan FP-Growth kemudian dibandingkan sebagai tujuan untuk menentukan algoritma yang paling baik saat di implementasikan pada sistem.

2.2. Pengumpulan Data

Dataset yang digunakan dalam penelitian ini untuk analisis pola asosiasi adalah data transaksi dari minimarket Sinar Mart. Jumlah total transaksi yang dianalisis mencakup penjualan selama periode 6 bulan, yaitu dari bulan Mei hingga Oktober 2023, dengan total mencapai 76.168 transaksi. Informasi struktur data terkait catatan setiap transaksi penjualan dapat dilihat pada Gambar 1.

Sinar Mart
 Jl. D I Panjaitan No 88 A, Lepo - Lepo
 PERIODE : 01/05/2023 S/D 30/05/2023
 TANGGAL CETAK 04/11/2023

NO.	TANGGAL NO. FAKTUR	PELANGGAN	OPERATOR	PEMBAYARAN	SUBTOTAL	CARD	DISKON	PPN	JUMLAH
***	KODE BARANG	NAMA BARANG		SATUAN	QTY	HARGA			SUBTOTAL
1	01/05/2023-010523001		DELFI	LIANG PAS	6.000	0			6.000
	8886008101091	AQUA 1500 ML		BTL	1	6.000			6.000---
2	01/05/2023-010523002		DELFI	LIANG PAS	5.000	0			5.000
	8996001600269	LE MINERALE 600 ML		PCS	2	2.500			5.000---
3	01/05/2023-010523003		DELFI	LIANG PAS	47.500	0			47.500
	8999999535667	PEPSODENT ACTION CENGEKHE 160 G		PCS	1	20.500			20.500---
	899999574963	LIFEBUOY BW SHISO & MINERAL 450ML		PCS	1	27.000			27.000---
4	01/05/2023-010523004		DELFI	LIANG PAS	40.500	0			40.500
	899999948167	SUNSLK SHP BLK SHINE 10ML		PCS	1	10.500			10.500---
	899866602549	BAR SOAP NUVO FAMILY BIRU 76GR		PCS	2	3.000			6.000---
	8998009010620	ULTRA MILK COKLAT 1000 ML		PCS	1	18.500			18.500---
	8885013131727	CHOCOLATE MAX CONE		PCS	1	5.500			5.500---
5	01/05/2023-010523005		DELFI	LIANG PAS	196.500	0			196.500
	AMF	AIR MINUM FLOW GELAS 240ML isi 40ML		DOS	1	26.500			26.500---
	899866601429	POWDET DAIA BUNGA BAG 290GR		PCS	5	5.000			25.000---
	8992628022156	BIMOLI REFILL 1LTR		PCS	2	23.000			46.000---
	8992745140962	HIT AER LILY BLOSSOM 600 ML + 20 PE		PCS	1	42.000			42.000---
	899886618243	BODYWASH GIV MULBERRY 825 ML		PCS	1	36.000			36.000---
	8998866100700	LIQUISHW MAMALEMON POUCH 680ML		PCS	2	10.500			21.000---
6	01/05/2023-010523006		DELFI	LIANG PAS	44.500	0			44.500
	8992959508541	SWEETY BABY WIPES 80+4S PERFUMED		PCS	1	18.500			18.500---
	8993365132535	MADU TJ 250 GR MURNI		PCS	1	26.000			26.000---

Gambar 1 Dataset Transaksi Penjualan

2.3. Analisis Data

Metode yang digunakan sebagai langkah untuk menganalisis data menggunakan metode Knowledge Discovery in Database atau KDD, dimana dalam metode ini terdiri dari lima langkah [8], adapun alur prosesnya seperti pada Gambar 2.



Gambar 2 Alur Proses Knowledge Discovery in Database

a. Data Selection

Langkah ini memiliki peran dalam memilih data yang relevan dan sesuai dengan tujuan proyek. Tahap ini melibatkan proses pemilihan atribut yang diperlukan untuk analisis, dan menghilangkan yang kurang atau bahkan tidak berpengaruh terhadap hasil analisis. Dari beberapa atribut dataset, atribut yang akhirnya digunakan dalam proses selanjutnya yakni no. faktur, tanggal transaksi, kode barang, dan nama barang. Lebih detail dapat dilihat pada Gambar 3.

	A	B	C	D
1	No. Faktur	Tanggal Transaksi	Kode Barang	Nama Barang
2	010523001	01/05/2023	8886008101091	AQUA 1500 ML
3	010523002	01/05/2023	8996001600269	LE MINERALE 600 ML
4	010523003	01/05/2023	8999999535667	PEPSODENT ACTION CENGEKHE 160 G
5	010523003	01/05/2023	8999999574963	LIFEBUOY BW SHISO & MINERAL 450ML
6	010523004	01/05/2023	8999999048167	SUNSLK SHP BLK SHINE 10ML
7	010523004	01/05/2023	899866602549	BAR SOAP NUVO FAMILY BIRU 76GR
8	010523004	01/05/2023	8998009010620	ULTRA MILK COKLAT 1000 ML
9	010523004	01/05/2023	8885013131727	CHOCOLATE MAX CONE
10	010523005	01/05/2023	AMF	AIR MINUM FLOW GELAS 240ML isi 40ML
11	010523005	01/05/2023	899866601429	POWDET DAIA BUNGA BAG 290GR
12	010523005	01/05/2023	8992628022156	BIMOLI REFILL 1LTR
13	010523005	01/05/2023	8992745140962	HIT AER LILY BLOSSOM 600 ML + 20 PE
14	010523005	01/05/2023	899886618243	BODYWASH GIV MULBERRY 825 ML
15	010523005	01/05/2023	8998866100700	LIQUISHW MAMALEMON POUCH 680ML
16	010523006	01/05/2023	8992959508541	SWEETY BABY WIPES 80+4S PERFUMED
17	010523006	01/05/2023	8993365132535	MADU TJ 250 GR MURNI
18	010523007	01/05/2023	8886008101053	AQUA 600 ML

Gambar 3 Hasil Seleksi Atribut Data

b. *Data Pre-Processing*

Secara garis besar tahap ini berguna untuk meningkatkan kualitas *dataset*. Ini dicapai dengan mengikuti langkah-langkah seperti melengkapi nilai yang hilang, menghapus data duplikat, dan menyelesaikan inkonsistensi data. Namun, berdasarkan data yang telah dianalisis peneliti, *dataset* yang tersedia sudah baik dan lengkap. Sehingga langkah-langkah tersebut tidak diperlukan.

c. *Data Transformation*

Tahap selanjutnya adalah proses mengubah data ke dalam bentuk yang mampu digunakan untuk perhitungan pada sebuah algoritma [9]. Pada penelitian ini peneliti menggunakan platform sistem berbasis *website* untuk implementasi algoritma sehingga data *spreadsheet* perlu ditransformasi ke bentuk *relational database*. Dari data berbentuk *row spreadsheet* kemudian dinormalisasi menjadi ke bentuk 3 tabel database yakni *transaction*, *product*, dan *transactionProducts*. Termasuk melakukan juga penyesuaian tipe data pada atribut.

2.4. Association Rule

Dalam penambangan aturan asosiasi, terdapat dua pengukuran dasar yang penting, yaitu *support* dan *confidence*. Karena basis data berskala besar dan pengguna hanya tertarik pada item yang sering dibeli, biasanya pengguna menentukan ambang batas untuk *support* dan *confidence* untuk menyingkirkan aturan yang tidak begitu menarik atau berguna. Kedua ambang tersebut disebut *minimum support* dan *minimum confidence*, secara berurutan. Ini merupakan dua parameter dasar dalam *Association Rule Mining (ARM)*.

a. *Support*

Support dari suatu aturan asosiasi didefinisikan sebagai persentase atau fraksi dari catatan yang mengandung himpunan $X \cup Y$ terhadap total jumlah *record* dalam basis data. Jumlah untuk setiap item ditingkatkan satu setiap kali item tersebut ditemukan dalam transaksi yang berbeda di dalam basis data selama proses pemindaian. Ini berarti *support* tidak memperhitungkan kuantitas item. Dari definisi tersebut, dapat dilihat bahwa dukungan suatu item adalah signifikansi statistik dari aturan asosiasi. Misalkan dukungan suatu item adalah 0,1%, ini berarti hanya 0,1 persen dari transaksi yang

mengandung pembelian item tersebut. Rumus menghitung nilai *support* dari aturan asosiasi diperoleh menggunakan Rumus 1 [10].

$$\text{Support}(X, Y) = \frac{\sum \text{Transaksi Mengandung } XY}{\sum \text{Transaksi}} \quad (1)$$

b. *Confidence*

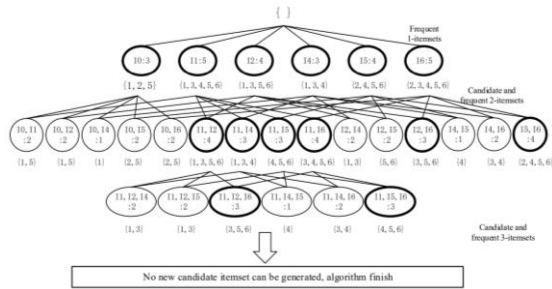
Confidence atau kepercayaan dari sebuah aturan asosiasi didefinisikan sebagai persentase atau fraksi dari jumlah transaksi yang mengandung $X \cup Y$ terhadap total jumlah catatan yang mengandung X . Dengan kata lain, kepercayaan mengukur sejauh mana aturan asosiasi $X \Rightarrow Y$ dianggap kuat atau relevan. Jika jumlah transaksi yang mengandung $X \cup Y$ relatif tinggi dibandingkan dengan jumlah transaksi yang hanya mengandung X , dan persentasenya melebihi ambang batas kepercayaan yang ditentukan, maka aturan asosiasi tersebut dianggap menarik. *Confidence* dihitung dengan menggunakan Rumus 2 [10].

$$\text{Confidence}(X|Y) = \frac{\text{Support } XY}{\text{Support } X} \quad (2)$$

2.5. Algoritma ECLAT

Algoritma *ECLAT* diperkenalkan oleh Zaki yaitu algoritma yang berbasis kedalaman (*depth-first*) [11] yang menggunakan format data vertikal; dengan demikian, dapat memperoleh dukungan dari suatu *itemset* melalui interseksi daripada mengakses *dataset* secara berulang-ulang. Berdasarkan teori *lattice* [12], algoritma *ECLAT* menggunakan strategi *divide and conquer* untuk mendekomposisi ruang pencarian asli menjadi subspace yang lebih kecil, yang telah terbukti efektif [11].

ECLAT membagi ruang pencarian menjadi beberapa subspace yang tidak tumpang tindih. *Itemset* dengan awalan yang sama diklasifikasikan ke dalam kelas yang sama, dan pembentukan kandidat *itemset* selalu dilakukan dalam setiap kelas secara terpisah. Untuk memudahkan pemahaman terkait algoritma ini, berikut pada Gambar 4 [13] adalah contoh visualisasi proses *mining* dari algoritma *ECLAT* jika digambarkan dalam *search space*.



Gambar 4 Contoh Search Space ECLAT

Pada Gambar 4, dengan menetapkan *minimum support* = 3, disitu menggambarkan proses penambangan *frequent itemset* menggunakan ECLAT. Pada Gambar 4, data dalam elips adalah *itemset* dan dukungannya, dan data di bawah elips adalah *TIDset* dari *itemset* tersebut. *Itemset* dalam elips yang ditebalkan adalah *frequent itemset*, dan *itemset* dalam elips yang tidak ditebalkan adalah kandidat *itemset* yang tidak *frequent*.

Semua *itemset* yang sering muncul dapat dihasilkan oleh *1-itemset* yang sering {I0, I1, I2, I4, I5, I6}; artinya, setiap *itemset* yang sering harus memiliki awalan panjang 1 yang sama dengan salah satu *1-itemset* yang sering. Selanjutnya, kita dapat membagi seluruh ruang pencarian menjadi 6 *subspace* yang tidak tumpang tindih. Kami melakukan operasi yang sama untuk *k-itemset* yang sering dan membagi ruang pencarian secara rekursif. Kelas ekuivalensi secara jelas dapat meningkatkan efisiensi pembentukan kandidat *itemset* dan mengurangi konsumsi memori.

2.6. Algoritma FP-Growth

Salah satu algoritma yang saat ini yang cepat dan populer untuk penambangan *frequent itemset* adalah algoritma *FP-growth* yang diperkenalkan oleh Han, Pei, dan Yin untuk sepenuhnya menghindari pembentukan kandidat [14]. Hal ini dilakukan dengan menggunakan *tree* untuk menyimpan keranjang belanja, sehingga dapat menghemat jumlah memori yang signifikan, daripada menyimpan kandidat seperti yang dilakukan Apriori dan ECLAT. Apriori adalah algoritma yang menggunakan tata letak horizontal dan pencarian *breadth-first*. Begitu pula dengan ECLAT, tetapi pada algoritma ini menggunakan tata letak vertikal dan pencarian *depth-first* [15].

Struktur *tree* dalam *FP-Growth* memberikan pandangan vertikal terhadap data.

Namun, *FP-Growth* juga menambahkan *header table* untuk setiap item individu yang memiliki dukungan di atas dukungan ambang batas. *Header table* ini berisi *linked-list* melalui *tree* untuk menghubungkan setiap *node* dari jenis yang sama. *Header table* memberikan *FP-Growth* pandangan horizontal terhadap data, selain pandangan vertikal yang diberikan oleh *tree* [15].

a. Konstruksi FP-Tree

Frequent-pattern tree (atau *FP-tree*) adalah struktur pohon yang didefinisikan sebagai berikut [16].

- Terdiri dari satu *root* yang diberi label "null", *item-prefix subtrees* sebagai anak dari *root*, dan *frequent-item-header table*.
- Setiap *node* dalam *item-prefix subtree* terdiri dari tiga bidang: nama-item, *support count*, dan *node-link*, dimana nama-item mendaftarkan item yang diwakili oleh *node* ini, *support count* adalah jumlah transaksi yang diwakili oleh bagian dari jalur yang mencapai *node* ini, dan *node-link* menghubungkan item yang sama ke *node* berikutnya dalam *FP-tree*, atau *null* jika tidak ada.
- Setiap entri yang terdapat dalam *header table frequent-item* terdiri dari dua bagian utama, yakni nama item dan kepala *node-link*. Komponen ini menggambarkan informasi kunci terkait dengan item tertentu dan menyediakan penunjuk ke *node* pertama dalam *FP-tree* yang membawa item tersebut. Dengan struktur ini, *header table* menjadi instrumen penting dalam mencatat dan mengorganisir pola-pola yang sering muncul dalam *dataset* transaksi.

b. Penambangan FP-Tree

Proses ini adalah tahapan menemukan *frequent itemset* berdasarkan *FP-Tree* yang dibuat sebelumnya. Dalam konteks analisis *frequent item i* pada algoritma *FP-Growth*, seluruh pola yang mengandung *i* dapat diidentifikasi dengan mengikuti *node-link* yang terkait dengan *i* di *header FP-Tree*. Konsep ini memiliki signifikansi penting karena memungkinkan untuk mengekstrak dengan efisien semua pola yang berkaitan dengan *i*. Contoh, jika ingin mengidentifikasi semua pola yang mengandung *i*, dapat dimulai dengan *node-link* yang terkait dengan *i* di *header FP-Tree*. Dengan melakukan itu, semua pola yang mencakup *i* dapat dijelajahi dan dikumpulkan. Ini merupakan fitur kunci

dalam algoritma *FP-Growth* yang memfasilitasi identifikasi dan pengumpulan semua pola yang terkait dengan item *i* dalam *FP-Tree*. Ini menjadi salah satu aspek penting dalam proses penambangan *frequent itemset* menggunakan algoritma *FP-Growth*.

Berikut adalah langkah pembentukan *frequent itemset* algoritma *FP-Growth* [17].

- Pembangkitan *conditional pattern base*, dilakukan dengan melihat *FP-Tree* yang berisi akhiran *i*. Setiap lintasan yang tidak mengandung *i* dibuang. Jadi *support* untuk setiap titik adalah nilai *support* untuk titik tersebut yang muncul bersama *i*. Dalam menyebutkan *conditional pattern base*, hanya *prefix* nya saja yang disebutkan.
- Pembangkitan *conditional tree*, *support count* dari setiap item pada setiap *conditional pattern base* dijumlahkan, lalu item yang memiliki jumlah *support count* lebih besar atau sama dengan *minimum support count* yang telah ditentukan akan dibangkitkan dengan *conditional tree*.
- Pencarian *frequent itemset*, jika *conditional tree* merupakan *single path*, maka untuk mendapatkan *frequent itemset* dilakukan kombinasi untuk setiap item yang berada pada *conditional tree*. Jika *conditional tree* bukan merupakan *single path* maka *frequent itemset* dibangkitkan secara rekursif.

2.7. Lift

Meskipun *minimum support* dan *minimum confidence* membantu menyaring atau mengesampingkan eksplorasi sejumlah besar *rules* yang tidak menarik, banyak *rules* yang dihasilkan masih tidak menarik bagi pengguna. Sayangnya, hal ini terutama terjadi saat penambangan dengan *minimum support* yang rendah atau penambangan untuk pola yang panjang. Ini menjadi hambatan dari penambangan aturan asosiasi.

Suatu aturan menarik atau tidak bersifat subjektif, berbeda antara satu pengguna dengan pengguna lainnya. Namun, metode menarik objektif, yang didasarkan pada statistik "di balik" data, dapat digunakan sebagai langkah menuju tujuan menyaring aturan yang tidak menarik yang mungkin akan ditampilkan kepada pengguna. Salah satu caranya adalah dengan menggunakan nilai *lift*.

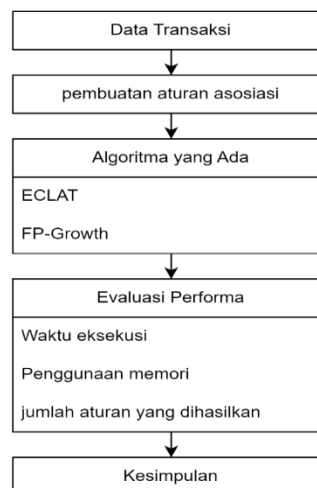
Lift adalah ukuran korelasi sederhana yang dinyatakan bahwa kejadian himpunan

item A bersifat independen terhadap kejadian himpunan item B jika $P(X \cup Y) = P(X)P(Y)$; sebaliknya, himpunan item X dan Y saling tergantung dan berkorelasi sebagai suatu peristiwa. Definisi ini dapat dengan mudah diperluas untuk lebih dari dua himpunan item. *Lift* antara kejadian X dan Y dapat diukur dengan menggunakan Rumus 3 [18].

$$Lift(X, Y) = \frac{P(X \cup Y)}{P(X)P(Y)} \quad (3)$$

2.8. Metode Komparasi Algoritma

Berikut pada Gambar 5 adalah diagram yang menunjukkan alur metode dalam menganalisis atau mengkomparasi algoritma *ECLAT* dan *FP-Growth*.



Gambar 5 Metode Komparasi Algoritma

Dataset merupakan data penjualan 6 bulan dimana berisi 76.168 transaksi dan 5.871 item. Metrik kinerja seperti total waktu eksekusi dihitung dalam detik, jumlah aturan asosiasi, dan penggunaan memori dihitung dalam *megabyte*.

Pada evaluasi kinerja algoritma dilakukan eksperimen dengan cara memberi parameter nilai *minimum support* yang berbeda yaitu 0.005%, 0.010%, 0.015%, 0.020%, 0.025%, dan 0.030% agar dapat dilihat bagaimana performa setiap algoritma ketika ambang batas atau banyaknya data yang analisis berbeda, sedangkan *minimum confidence* yaitu 60%. Nilai-nilai tersebut didapat dengan mencari nilai *support* tertinggi tetapi tetap dapat menghasilkan aturan

asosiasi, dan nilai *support* terendah dimana *support count* / frekuensi-nya lebih dari 1. Setelah dilakukan evaluasi kinerja, akhirnya dari menghasilkan suatu perbandingan antara algoritma *ECLAT* dan *FP-Growth*.

3. HASIL DAN PEMBAHASAN

3.1. Hasil Evaluasi Algoritma *ECLAT*

Evaluasi dilakukan dengan cara melakukan perbandingan antara hasil dari perhitungan manual dan hasil sistem. Hal ini dilakukan untuk membuktikan apakah jalannya algoritma pada sistem sudah sesuai dengan hasil perhitungan manual sebagaimana mestinya. Pada pengujian ini peneliti menggunakan contoh data transaksi penjualan dengan total sebanyak 20 transaksi. Dapat dilihat secara detail pada Tabel 1.

Tabel 1 Data Pengujian Algoritma

Kode Transaksi	Daftar Barang Dalam Transaksi
T1	Susu, Roti, Biskuit
T2	Roti, Susu, Biskuit, Sereal Jagung
T3	Roti, Teh, Minuman Cokelat
T4	Selai, Mie Instan, Roti, Susu
T5	Mie Instan, Teh, Biskuit
T6	Roti, Teh, Minuman Cokelat
T7	Mie Instan, Teh, Sereal Jagung
T8	Mie Instan, Roti, Teh, Biskuit
T9	Selai, Mie Instan, Roti, Teh
T10	Roti, Susu
T11	Kopi, Saus, Biskuit, Sereal Jagung
T12	Kopi, Saus, Biskuit, Sereal Jagung
T13	Kopi, Gula, Minuman Cokelat
T14	Roti, Kopi, Saus
T15	Roti, Gula, Biskuit
T16	Kopi, Gula, Sereal Jagung
T17	Roti, Gula, Minuman Cokelat
T18	Roti, Kopi, Gula
T19	Roti, Kopi, Gula
T20	Teh, Susu, Kopi, Sereal Jagung

Tabel 2 Daftar Barang pada Data Pengujian

Kode Barang	Nama Barang
I1	Susu
I2	Roti
I3	Biskuit
I4	Sereal Jagung
I5	Teh

I6	Minuman Cokelat
I7	Selai
I8	Mie Instan
I9	Kopi
I10	Saus
I11	Gula

Perhitungan manual pada algoritma *ECLAT* menggunakan data uji menghasilkan 7 asosiasi kuat. Hasil asosiasi ini tersaji secara detail pada Tabel 3.

Tabel 3 Hasil Perhitungan Manual Algoritma *ECLAT*

<i>Antec.</i>	<i>Conseq.</i>	<i>Sup. AB</i>	<i>Conf.</i>	<i>Lift</i>
I10	I9	0,15	1	2,5
I6	I2	0,15	0,75	1,15
I1	I2	0,20	0,8	1,23
I8	I5	0,20	0,8	2,29
I4	I9	0,20	0,67	1,67
I11	I9	0,20	0,67	1,67
I11	I2	0,20	0,67	1,03

Hasil analisis pada sistem menggunakan algoritma *ECLAT* juga menghasilkan 7 asosiasi kuat dan menarik, serta semua asosiasi tersebut menunjukkan hasil yang sama persis. Fakta ini menunjukkan bahwa implementasi algoritma pada sistem yang dibangun telah berjalan dengan baik dan sesuai dengan apa yang diharapkan. Hasil asosiasi tersaji secara detail pada Gambar 6.

No.	Aturan Asosiasi	Nilai Support (%)	Nilai Confidence (%)	Lift Ratio
1	Jika membeli I1 - Susu Maka berpotensi membeli I2 - Roti	15,00	100,00	2,50
2	Jika membeli I4 - Sereal Jagung Maka berpotensi membeli I2 - Roti	20,00	80,00	1,23
3	Jika membeli I8 - Mie Instan Maka berpotensi membeli I5 - Teh	20,00	80,00	2,29
4	Jika membeli I6 - Minuman Cokelat Maka berpotensi membeli I2 - Roti	15,00	75,00	1,15
5	Jika membeli I4 - Sereal Jagung Maka berpotensi membeli I9 - Kopi	20,00	66,67	1,67
6	Jika membeli I11 - Gula Maka berpotensi membeli I9 - Kopi	20,00	66,67	1,67
7	Jika membeli I11 - Gula Maka berpotensi membeli I2 - Roti	20,00	66,67	1,03
8	Jika membeli I10 - Saus Maka berpotensi membeli I2 - Roti	15,00	60,00	0,92

Gambar 6 Hasil Perhitungan Sistem Algoritma *ECLAT*

3.2. Hasil Evaluasi Algoritma *FP-Growth*

Sama seperti algoritma *ECLAT*, evaluasi dilakukan dengan cara melakukan perbandingan antara hasil perhitungan manual dan hasil pada sistem menggunakan data transaksi penjualan pada Tabel 1. Perhitungan manual pada algoritma *FP-Growth* menggunakan data uji menghasilkan 7 asosiasi kuat, tersaji secara detail pada Tabel 4.

Tabel 4 Hasil Perhitungan Manual Algoritma *FP-Growth*

<i>Antec.</i>	<i>Conseq.</i>	<i>Sup. AB</i>	<i>Conf.</i>	<i>Lift</i>
I10	I9	0,15	1	2,5
I6	I2	0,15	0,75	1,15
I8	I5	0,20	0,8	2,29
I1	I2	0,20	0,8	1,23
I4	I9	0,20	0,67	1,67
I11	I9	0,20	0,67	1,67
I11	I2	0,20	0,67	1,03

Hasil analisis pada sistem menggunakan algoritma *FP-Growth* juga menghasilkan 7 asosiasi kuat dan semuanya menunjukkan hasil yang sama. Fakta ini menunjukkan bahwa implementasi algoritma pada sistem sudah sesuai, tersaji secara detail pada Gambar 7.

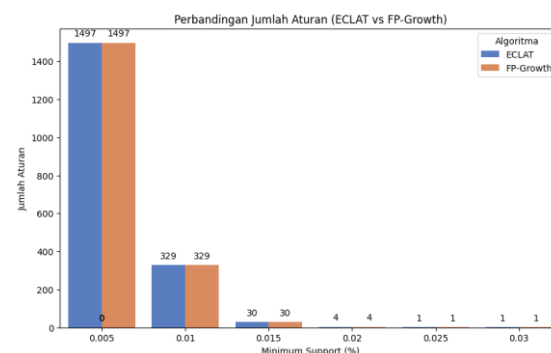
No.	Aturan-Asosiasi	Nilai Support (%)	Nilai Confidence (%)	Lift Ratio
1	Jika membeli I-1-Bata Maka berpotensi membeli I-1-Kopi	15.00	100.00	2.50
2	Jika membeli H-1-Me Instan Maka berpotensi membeli E-1-Teh	20.00	80.00	2.39
3	Jika membeli A-1-Susu Maka berpotensi membeli B-1-Ragi	20.00	80.00	1.23
4	Jika membeli F-1-Misuman Cokelat Maka berpotensi membeli B-1-Ragi	15.00	75.00	1.15
5	Jika membeli D-1-Sereal Jagung Maka berpotensi membeli I-1-Kopi	20.00	80.00	1.67
6	Jika membeli K-1-Gula Maka berpotensi membeli B-1-Ragi	20.00	66.67	1.03
7	Jika membeli K-1-Gula Maka berpotensi membeli I-1-Kopi	20.00	66.67	1.67
8	Jika membeli H-1-Me Instan Maka berpotensi membeli B-1-Ragi	15.00	80.00	0.92

Gambar 7 Hasil Perhitungan Sistem Algoritma *FP-Growth*

3.3. Analisis Perbandingan Hasil Eksperimentasi Algoritma *ECLAT* dan Algoritma *FP-Growth*

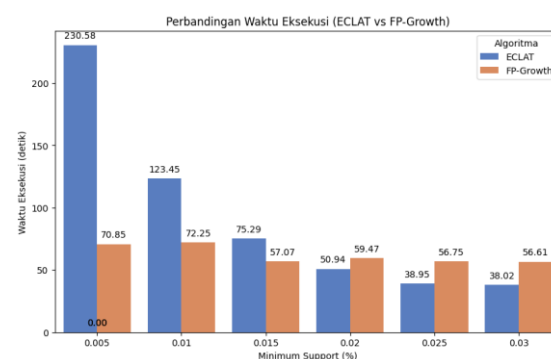
Pada penelitian ini dilakukan eksperimen untuk menguji bagaimana pola asosiasi ditemukan pada kedua algoritma yaitu *ECLAT* dan *FP-Growth*, ketika nilai *minimum support*-nya berbeda. Nantinya, *output* dari hasil dari skenario digunakan untuk analisis kinerja berdasarkan 3 parameter yaitu jumlah asosiasi yang terbentuk, waktu eksekusi, dan konsumsi memori.

Grafik pada Gambar 8 menunjukkan perbandingan aturan asosiasi yang terbentuk dari kedua algoritma. Keduanya menghasilkan jumlah asosiasi yang sama. Aturan terbanyak ketika *minimum support* 0.005% yaitu 1.497 aturan. Semakin tinggi *minimum support*, semakin sedikit asosiasi yang akan terbentuk.



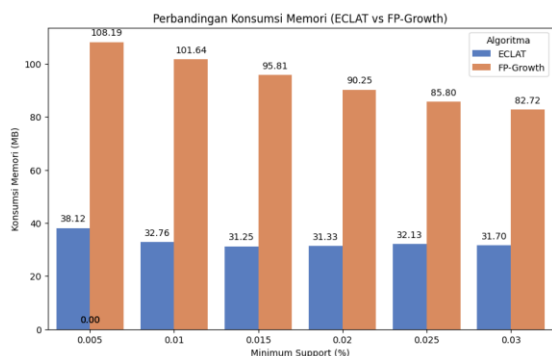
Gambar 8 Grafik Perbandingan Jumlah Aturan Asosiasi

Grafik pada Gambar 9 menunjukkan perbandingan waktu eksekusi pada proses analisis. Setelah melakukan eksperimen, proses analisis *FP-Growth* lebih cepat. Perbedaan signifikan ketika *minimum support* 0.005%, *FP-Growth* hanya membutuhkan waktu 70.85 Detik sedangkan *ECLAT* membutuhkan 230.58 Detik.



Gambar 9 Grafik Perbandingan Waktu Eksekusi

Grafik pada Gambar 10 menunjukkan perbandingan konsumsi memori yang dibutuhkan pada analisis masing-masing algoritma. Hasilnya *ECLAT* membutuhkan lebih sedikit memori. Perbedaan besar ketika menetapkan *minimum support* 0.005%, algoritma *ECLAT* membutuhkan sekitar 38,12 MB sedangkan *FP-Growth* membutuhkan memori sekitar 108,19 MB.



Gambar 10 Grafik Perbandingan Konsumsi Memori

Berdasarkan hasil dari eksperimen yang diperoleh dari penerapan algoritma *FP-Growth* dan *ECLAT* menggunakan ambang batas atau nilai *minimum support* yang berbeda-beda yaitu 0.005%, 0.010%, 0.015%, 0.020%, 0.025%, dan 0.030%, serta batasan asosiasi kuat dengan *minimum confidence* yaitu 60% mengungkapkan sejumlah perbedaan mendasar, yakni jumlah aturan asosiasi yang dihasilkan, tingkat kecepatan eksekusi, dan konsumsi memori. Informasi lengkap mengenai perbandingan kinerja masing-masing algoritma ini dapat dilihat lengkap pada Tabel 5.

Tabel 5 Perbandingan Hasil Eksperimen

Support (%)	ECLAT			FP-Growth		
	Jumlah Aturan	Waktu (s)	Memori (MB)	Jumlah Aturan	Waktu (s)	Memori (MB)
0.005	1.497	230,58	38,12	1.497	70,85	108,19
0.010	329	123,45	32,76	329	72,25	101,64
0.015	30	75,29	31,25	30	57,07	95,81
0.020	4	50,94	31,33	4	59,47	90,25
0.025	1	38,95	32,13	1	56,75	85,80
0.030	1	38,02	31,70	1	56,61	82,72

4. KESIMPULAN

Berdasarkan hasil dari beberapa eksperimen dengan menggunakan data yang sama dan ambang batas yang berbeda menunjukkan bahwa algoritma *ECLAT* dan *FP-Growth* menghasilkan jumlah aturan asosiasi yang sama. Namun, dari aspek lain, *ECLAT* lebih unggul dari sisi konsumsi memori, sedangkan *FP-Growth* lebih unggul dari sisi kecepatan proses analisis, *FP-Growth* bekerja lebih baik dari *ECLAT* baik itu ketika memproses dengan ambang batas besar, artinya data yang dianalisis pada setiap proses oleh algoritma sedikit, maupun ketika menganalisis data yang besar.

5. SARAN

Mencoba melakukan analisis ketika mengimplementasi fitur adaptif berdasarkan kriteria khusus. Fitur ini memanfaatkan kriteria khusus pada *dataset* yang dapat ditentukan pengguna, seperti utilitas item, sehingga memberikan personalisasi yang lebih besar pada proses pembuatan aturan asosiasi sesuai dengan kebutuhan.

DAFTAR PUSTAKA

- [1] P. M. S. Tarigan, J. T. Hardinata, H. Qurniawan, M. Safii, and R. Winanjaya, "Implementasi Data Mining Menggunakan Algoritma

- Apriori Dalam Menentukan Persediaan Barang (Studi Kasus: Toko Sinar Harahap),” *Just IT*, vol. 12, no. 2, pp. 51–61, 2022.
- [2] A. Sharma and A. Ganpati, “Association Rule Mining Algorithms: a Comparative Review,” *International Research Journal of Engineering and Technology*, vol. 8, no. 11, pp. 848–853, 2021, [Online]. Available: www.irjet.net
- [3] Fadlina, “Data Mining Untuk Analisa Tingkat Kejahatan Jalanan Dengan Algoritma Association Rule Metode Apriori,” *Informasi dan Teknologi Ilmiah (INTI)*, vol. 3, no. 1, pp. 144–154, 2014, [Online]. Available: <http://upi-yptk.ac.id>
- [4] J. E. Samodra, B. Susanto, and W. S. Raharjo, “Implementasi Algoritma ECLAT Untuk Frequent Pattern Mining Pada Penjualan Barang,” *Jurnal Teknologi Media Teknika*, vol. 10, no. 2, pp. 1–10, 2015.
- [5] K. N. Wijaya, R. Firsandaya Malik, S. Nurmaini, M. I. Komputer, F. Ilmu, and K. Unsri, “Analisa Pola Frekuensi Keranjang Belanja Dengan Perbandingan Algoritma Fp-Growth (Frequent Pattern Growth) Dan Eclat Pada Minimarket,” vol. 7, no. 2, pp. 2407–4322, 2020, [Online]. Available: <http://jurnal.mdp.ac.id>
- [6] R. L. Najmi, M. Irsyad, F. Insani, A. Nazir, and . P., “Analisis Pola Asosiasi Data Transaksi Penjualan Minuman Menggunakan Algoritma FP-Growth dan Eclat,” *Building of Informatics, Technology and Science (BITS)*, vol. 5, no. 1, Jun. 2023, doi: 10.47065/bits.v5i1.3592.
- [7] A. Wijaya, “Perbandingan Algoritma ECLAT dan FP-Growth pada Penjualan Barang (Studi Kasus: Minimarket 212 Mart Veteran Utama).”
- [8] H. M. Safhi, B. Frikh, and B. Ouhbi, “Assessing reliability of Big Data Knowledge Discovery process,” in *Procedia Computer Science*, Elsevier B.V., 2019, pp. 30–36. doi: 10.1016/j.procs.2019.01.005.
- [9] I. W. Saputro and B. W. Sari, “Uji Performa Algoritma Naïve Bayes untuk Prediksi Masa Studi Mahasiswa,” *Citec Journal*, vol. 6, no. 1, pp. 1–11, 2019.
- [10] Q. Zhao and S. S. Bhowmick, “Association Rule Mining: A Survey,” 2003.
- [11] M. J. Zaki, “Scalable Algorithms for Association Mining,” *IEEE Transaction on Knowledge and Data Engineering*, vol. 12, no. 3, pp. 372–390, 2000.
- [12] L. Xiao *et al.*, “Strongly correlated quantum walks in optical lattices,” *Science (1979)*, vol. 347, no. 6227, pp. 1229–1233, Mar. 2015, doi: 10.1126/science.1259866.
- [13] C. Zhang, P. Tian, X. Zhang, Z. L. Jiang, L. Yao, and X. Wang, “Fast Eclat algorithms based on minwise hashing for large scale transactions,” *IEEE Internet Things J*, vol. 6, no. 2, pp. 3948–3961, Apr. 2019, doi: 10.1109/JIOT.2018.2885851.
- [14] J. Han, J. Pei, and Y. Yin, “Mining Frequent P attens without Candidate Generation,” *ACM*, vol. 29, no. 2, pp. 1–12, 2000.
- [15] J. Heaton, “Comparing Dataset Characteristics that Favor the Apriori, Eclat or FP-Growth Frequent Itemset Mining Algorithms,” *Institute of Electrical and Electronics Engineers*, pp. 1–7, 2016.
- [16] J. Han, J. Pei, Y. Yin, and R. Mao, “Mining Frequent Patterns without Candidate Generation: A Frequent-Pattern Tree Approach,” in *Data Mining and Knowledge Discovery*, 2004, pp. 53–87.
- [17] D. Samuel, “Penerapan Struktur FP-Tree dan Algoritma FP-Growth dalam Optimasi Penentuan Frequent Itemset,” Institut Teknologi Bandung, 2008.
- [18] J. Han, M. Kamber, and J. Pei, *Data Mining Concepts and Techniques*, 3rd ed. Waltham: Morgan Kaufmann Publisher, 2011.