

Penerapan Metode *Association Rule* Menggunakan Algoritma *Fp-Growth* dan *Eclat* pada Pola Konsumsi Masyarakat Perkotaan di Indonesia

Aldiansyah¹, Eri Setiawan², Agus Sutrisno³

^{1,2,3} Jurusan Matematika, Universitas Lampung, Bandar Lampung, Indonesia
Jl. Prof. Dr. Ir. Sumantri Brojonegoro No.1, Kota Bandar Lampung, Lampung 35141
Email: aldy80267@gmail.com¹, aldiansyah.104419@students.unila.ac.id²

*Korespondensi penulis : aldy80267@gmail.com

Abstrak

Pola konsumsi masyarakat di Indonesia merupakan salah satu tolak ukur dalam menentukan indeks perekonomian masyarakat Indonesia. Daya konsumsi masyarakat Indonesia memiliki kontribusi yang sangat penting dalam kenaikan indeks perekonomian di Indonesia. Dengan melihat faktor yang dapat menentukan pola konsumsi masyarakat Indonesia, pola konsumsi mampu mempertahankan daya konsumsi masyarakat. Dalam penelitian ini akan dilakukan pencarian pola konsumsi masyarakat perkotaan di Indonesia menggunakan dua algoritma untuk mendapatkan algoritma terbaik yang memiliki nilai *support*, *confidence* dan *rule* terbesar, yaitu *FP-Growth* dan *Eclat*. Hasil yang diperoleh digunakan untuk melihat pola konsumsi masyarakat perkotaan di Indonesia. Nilai *support* dan *confidence* terbesar dihasilkan oleh algoritma *Eclat* sebesar 10% dan 100% dengan *rule* sebanyak 15 dan pola terbaik yang dihasilkan adalah masyarakat yang mengkonsumsi makanan berdasarkan faktor agama adalah masyarakat yang berusia diantara 30 dan 50 tahun.

Kata Kunci: *Association Rule*, *FP-Growth*, *Eclat*, *Support*, dan *Confidence*.

Abstract

The consumption pattern of the people in Indonesia is one of the benchmarks in determining the economic index of the Indonesian people. The consumption power of the Indonesian people has a very important contribution in increasing the economic index in Indonesia. By looking at the factors that can determine the consumption pattern of the Indonesian people, it is able to maintain the consumption power of the people. In this study, a search for consumption patterns of urban communities in Indonesia will be carried out using two algorithms, to get the best algorithm that has the largest support, confidence and rule values, namely FP-Growth and Eclat. The results obtained are used to see the consumption patterns of urban communities in Indonesia. The largest support and confidence values are generated by the Eclat algorithm by 10% and 100% with 15 rules and the best pattern produced is that people who consume food based on religious factors are people between the ages of 30 and 50 years.

Keywords: *Association Rule*, *FP-Growth*, *Eclat*, *Support*, and *Confidence*.

1. Pendahuluan

Data mining adalah analisis sekumpulan data dalam pengamatan untuk menemukan hubungan atau pola yang tidak terduga dan untuk meringkas data dengan cara baru yang dapat dipahami dan bermanfaat bagi pemilik data [1]. Salah satu pola data yang bisa dilihat adalah pola konsumsi masyarakat yang mana salah satu tolak ukur dalam menentukan indeks perekonomian masyarakat Indonesia dengan melihat pola konsumsi masyarakat. Perekonomian di Indonesia pada Triwulan II 2022 mengalami pertumbuhan impresif sebesar 5,44% [2]. Dengan melihat faktor-faktor yang mempengaruhi pola konsumsi masyarakat Indonesia ternyata mampu mempertahankan daya konsumsi masyarakat. Untuk melihat faktor-faktor tersebut dapat dianalisis dengan melihat pola-pola data yang terbentuk, yaitu dengan menggunakan analisis asosiasi.

Dalam analisis asosiasi digunakan sebuah algoritma yang dapat menentukan pola data yang sering muncul yaitu *Frequent Pattern Growth*. *Frequent Pattern Growth* yaitu analisis membangkitkan *frequent itemset* dengan menelusuri *FP-tree* secara bottom-up menggunakan strategi *divide-and conquer* dalam memecah masalah menjadi lebih kecil dengan mengkonstruksi *conditional FP-tree* dari *global FP-tree* [3]. Dan *Eclat* yaitu analisis pola data dilakukan pencarian secara *Depth First Search* (DFS) dengan tata letak vertikal dan jika database berbentuk horizontal terlebih dahulu dikonversi menjadi vertikal [4].

Dalam penelitiannya Srinadh (2022), melakukan evaluasi antara algoritma apriori, *FP-Growth* dan *Eclat* dengan menggunakan data transaksi diperoleh bahwa *Eclat* dan *FP-Growth* lebih baik dari apriori dalam menghasilkan rule dengan menghabiskan lebih sedikit waktu [5]. Berdasarkan uraian tersebut, ingin menerapkan dan menganalisis algoritma *FP-Growth* dan *Eclat* dengan menggunakan data bukan transaksi yaitu data pola konsumsi masyarakat perkotaan di Indonesia, sehingga hasil yang diperoleh adalah algoritma terbaik untuk menentukan pola konsumsi masyarakat perkotaan di Indonesia.

2. Metode Penelitian

Data yang digunakan dalam penelitian ini merupakan data sekunder berupa data pola konsumsi masyarakat perkotaan di Indonesia yang diperoleh dari *Kaggle Repository* pada artikel [6]. Penelitian ini dilakukan untuk mendapatkan *rule* terbaik untuk mendapatkan pola konsumsi masyarakat perkotaan di Indonesia menggunakan *software R* dan *RapidMiner*. Adapun langkah-langkah yang dilakukan dalam algoritma *FP-Growth* dan *Eclat* adalah sebagai berikut [7]:

1. Melakukan input data pandangan konsumsi masyarakat perkotaan di Indonesia dari jurnal data *in brief* ke dalam Excel.
2. Melakukan uji validasi dan reliabilitas untuk melihat kebaikan dan kevalidan data.
3. Melakukan *selection* data yang tidak diperlukan untuk di analisis.
4. Tahapan Penelitian algoritma *FP-Growth* dengan *RapidMiner* antara lain [8]:
 - a. Melakukan transformasi data agar dapat dibaca oleh *RapidMiner*.
 - b. Penginputan data pandangan konsumsi masyarakat perkotaan di Indonesia ke *RapidMiner*.
 - c. Memasukan *tools-tools* dalam analisis *FP-Growth*.
 - d. Menentukan *minimum support* dan *minimum confidence*.
 - e. Menentukan *conclusion* yang diinginkan dengan memenuhi nilai *support* dan *confidence*.

- f. Memperoleh aturan asosiasi yang akan digunakan untuk menentukan pola konsumsi masyarakat perkotaan di Indonesia.
5. Tahapan penelitian algoritma *Eclat* dengan *software R* antara lain [9]:
 - a. Melakukan transformasi data agar dapat dibaca oleh *R*.
 - b. Memasukan script algoritma *Eclat*.
 - c. Menentukan *minimum support* dan *minimum confidence*.
 - d. Menentukan *conclusion* yang diinginkan dengan memenuhi nilai *support* dan *confidence*.
 - e. Memperoleh aturan asosiasi yang akan digunakan untuk menentukan pola konsumsi masyarakat perkotaan di Indonesia.
6. Membandingkan hasil algoritma *FP-Growth* dan *Eclat*.
7. Menentukan jumlah *association rule* untuk menentukan pola yang sering terjadi pada pola konsumsi masyarakat perkotaan di Indonesia [10].

Kaidah asosiasi dikenal dengan nama analisis keranjang pasar (*Market-Basket Analysis*) karena dapat menganalisis keranjang para konsumen yang melakukan transaksi. Suatu transaksi T terdiri dari A jika dan hanya jika $A \subseteq T$ sehingga kaidah asosiasi berbentuk $A \rightarrow B$, di mana $A \subseteq I, B \subseteq I$ dan $A \cap B = \emptyset$ [11]. Dalam *association rule* ada tiga parameter penting yang berfungsi untuk pembentukan *rule* yaitu:

1. *Support*

$$Support(A \rightarrow B) = P(A \cap B) = \frac{\text{frekuensi}(A \cap B)}{\text{Jumlah total transaksi dalam data}} \quad (1)$$

2. *Confidence*

$$Confidence(A \rightarrow B) = P(B|A) = \frac{\text{frekuensi}(A \cap B)}{\text{Jumlah transaksi yang mengandung } A} \quad (2)$$

3. *Lift*

$$Lift(A \rightarrow B) = \frac{P(A \cup B)}{P(A)P(B)} = \frac{Support(A, B)}{support(A) * support(B)} \quad (3)$$

Frequent pattern growth (FP-Growth) adalah algoritma alternatif yang digunakan untuk menentukan himpunan data yang paling sering muncul dalam sebuah dataset. *FP-Growth* pertama-tama menghasilkan *FP-Tree* dan menggunakan pohon yang sering muncul untuk menghasilkan itemset yang sering [12].

Eclat merupakan algoritma asosiasi untuk menentukan *itemset* yang paling sering muncul dengan melakukan pencarian pada *database* secara vertikal, jika *database* berbentuk horizontal maka harus diubah terlebih dahulu ke bentuk vertikal [13]. Pemilihan algoritma terbaik dengan melihat kebaikan parameter *association rule* dapat dilihat pada Tabel 1:

Tabel 1. Kriteria Kebaikan Parameter

<i>Support</i>	<i>Confidence</i>	Kriteria
$0 < support < 0.5$	$0 < confidence < 0.5$	Rule cukup baik
0.5	0.5	Rule baik
$0.5 < support < 1$	$0.5 < confidence < 1$	Rule sangat baik

Evaluasi hasil *rule* dilihat menggunakan indeks evaluasi menyeluruh yang merupakan kombinasi antara kriteria evaluasi objektif dan evaluasi subjektif. Untuk melakukan evaluasi tersebut dapat dilihat indikator dari masing-masing evaluasi. Indikator evaluasi objektif ada tiga, yaitu: *support*, *confidence* dan *lift* [14]. Indikator

evaluasi subyektif terbagi menjadi tiga, yaitu: *novelty*, *simplicity* dan tingkat minat pengguna [15]. Indeks evaluasi komprehensif dinamakan *RI* dengan rumus:

$$RI = S^{w_1} \times C^{w_2} \times lift^{w_3} \times W_i^{w_4} \times USI^{w_5} \times \left(\frac{1}{CN}\right)^{w_6} \quad (4)$$

dimana,

- $w_1, w_2, w_3, w_4, w_5, w_6$: Nilai dari *novelty*
- USI* : Tingkat minat pengguna
- CN* : Nilai *simplicity*
- S* : *Support*
- C* : *Confidence*

dengan,

$$W_i \geq 0, (i = 1, 2, \dots, 6) \sum_{i=1}^6 w_i = 1$$

3. Hasil dan Pembahasan

3.1 Uji Validitas dan Reliabilitas

Validitas adalah suatu ukuran yang menunjukkan tingkat-tingkat kevalidan suatu kuisioner. Hasil uji validitas data Ekonomi Indonesia [2], diperoleh pada Tabel 2:

Tabel 2. Hasil Uji Validitas

Variabel	(r_{hitung})	r_{tabel}	Hasil Uji
X1	0.487357	0.073586	Valid
X2	0.399570	0.073586	Valid
X3	0.503305	0.073586	Valid
X4	0.321919	0.073586	Valid
X5	0.536244	0.073586	Valid
X6	0.553812	0.073586	Valid
X7	0.473457	0.073586	Valid
X8	0.458985	0.073586	Valid
X9	0.452880	0.073586	Valid
X10	0.483938	0.073586	Valid
Y1	0.421326	0.073586	Valid
Y2	0.905326	0.073586	Valid
Y3	0.897296	0.073586	Valid

Setelah melakukan uji validitas maka akan dilakukan uji reliabilitas dengan bantuan software R, diperoleh hasil berikut ini.

Tabel 3. Hasil Uji Reliabilitas

Uji	Nilai
Alpha Cronbach	0.7136665

Berdasarkan Tabel 3, hasil uji reliabilitas dengan ukuran sampel 710 diperoleh nilai *cronbach's alpha* sebesar 0.7136665 yang lebih besar di banding 0.6. Jadi dapat disimpulkan bahwa butir-butir pertanyaan dalam data dapat dikatakan reliabilitas atau handal.

3.2 Deskriptif Data

Tabel 4. Deskriptif Data

Variabel	Label	Presentase	Variabel	Label	Presentase
Jenis	Perempuan	42.1	Kelas	Bawah	34.5
Kelamin	Laki laki	57.9	Sosial	Menengah	36.2
Agama	Islam	60.1		Atas	29.3
	Protestan	14.2	Pekerjaan	Pekerja Swasta	28.5
	Katolik	15.8		Pengabd Negara	4.2
	Hindu	9.3		Pekerja Buruh	28.9
Suku	Budha	0.6		Guru	2.8
	Jawa	40.7		Pengusaha	27.2
	Batak	13.7		Pekerja Mandiri	8.5
	Bugis	8.6	Pendidikan	SD	3.4
	Sunda	10.3		SMP	5.2
	Madura	7.2		SMA	25.5
	Minang	8.2		Diploma	22.0
	Bali	3.5			

3.3 Pencarian Frequent Itemset

Langkah pertama dalam menentukan *association rule* adalah mencari *frequent itemset* yang berisi nilai *support* untuk setiap item. Berikut Tabel 5 hasil dari pencarian *frequent itemset*:

Tabel 5. Frekuensi Kemunculan Support Pada Setiap Item

Item	Frekuensi	Support	Item	Frekuensi	Support
Perempuan	411	58%	Pekerja swasta	205	29%
Laki-laki	299	42%	Pengusaha	193	27%
B1	137	19%	Pegawai pemerintah	202	28%
B2	363	51%	Guru	20	2.9%
B3	210	30%	TNI/POLRI	30	4.2%
Islam	427	60%	Kelas bawah	246	35%
Protestant	112	16%	Kelas menengah	256	37%
Katolik	101	14%	Kelas atas	208	30%
Hindu	66	9.3%	Kesehatan	267	38%
Budha	5	0.7%	Harga	208	30%
Jawa	289	41%	Agama	585	82%
Batak	51	7.1%	Tempat	332	47%
Sunda	97	14%	Tradisional	281	40%
Bugis	25	3.5%	SMP	37	5.2%
Madura	58	8.2%	SMA	181	25%
Bali	73	10%	Diploma	212	30%
Minang	61	8.6%	S1	100	14%
SD	24	3.4%	S2	156	22%

Berdasarkan Tabel 5 diperoleh nilai *support* terbesar dari 36 atribut adalah item faktor agama sebesar 82% atau nilai *support* sebesar 0.82%. Sedangkan nilai *support* terkecil adalah item budha hanya 0.4%.

3.4 Rule FP-Growth

Nilai *support*, *confidence*, *rule* dan waktu proses diperoleh menggunakan *RapidMiner* yang disajikan pada Tabel 6 berikut:

Tabel 6. Hasil Uji Parameter FP-Growth

Support	Confidence	Banyak rule	Waktu
0.1	0.5	114	01.85 s
0.2	0.5	78	01.38 s
0.3	0.5	79	01.05 s
0.1	0.6	31	01.44 s
0.2	0.6	18	01.25 s
0.1	0.7	2	01.31

3.5 Rule Eclat

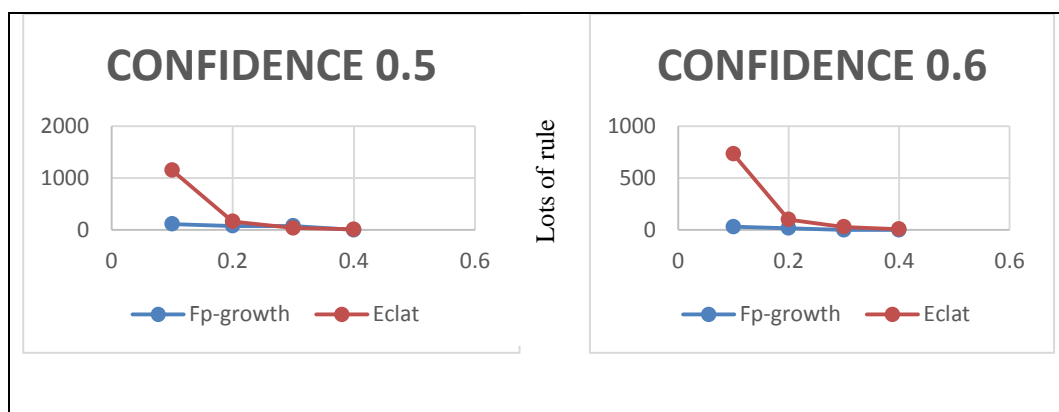
Dengan bantuan *Software R* dihasilkan nilai *support*, *confidence*, *rule* dan waktu proses pada Tabel 7 berikut:

Tabel 7. Hasil Uji Parameter FP-Growth

Support	Confidence	Banyak rule	Waktu	Support	Confidence	Banyak rule	Waktu
0.1	0.5	1151	0.02 s	0.3	0.7	26	0.00 s
0.2	0.5	163	0.08 s	0.4	0.7	8	0.01 s
0.3	0.5	38	0.00 s	0.1	0.8	431	0.02 s
0.4	0.5	12	0.01 s	0.2	0.8	68	0.08 s
0.1	0.6	732	0.02 s	0.3	0.8	23	0.00 s
0.2	0.6	101	0.08 s	0.4	0.8	8	0.01 s
0.3	0.6	29	0.00 s	0.1	0.9	48	0.02 s
0.4	0.6	8	0.01 s	0.2	0.9	2	0.08 s
0.1	0.7	581	0.02 s	0.3	0.9	1	0.00 s
0.2	0.7	83	0.08 s	0.1	1.0	15	1.02

3.6 Evaluasi Association Rule FP-Growth dan Eclat

Evaluasi hasil dari association rule bertujuan untuk melihat apakah hasil prediksi *rule* yang diperoleh mendapatkan hasil yang optimal.





Gambar 1. Perbandingan Banyaknya Rule

Dari Gambar 1 terlihat bahwa nilai semua *support* dan *confidence*, *rule* terbanyak dimiliki oleh algoritma *Eclat*.



Gambar 2. Perbandingan Waktu dalam Mengeluarkan Output

3.7 Novelty dan RI

Untuk melihat kebaikan antara parameter objektif dan subyektif dapat dilihat dari indeks evaluasi menyeluruh atau *RI*. Berikut hasil *RI* untuk masing masing algoritma dengan bantuan software R. Dengan mengambil maksimum *support* adalah diantara 0.1 sampai 0.5.

Tabel 8. Perbandingan Nilai RI Antar Algoritma

Algoritmh	RI	Novelty	Max Support
FP-growth	851.52	0.4556	0.3
Eclat	16462.24	0.4819	0.4

Dari hasil perhitungan *RI* pada Tabel 8 diperoleh bahwa algoritma *Eclat* memiliki nilai *RI* terbesar yaitu sebesar 16462.24, untuk nilai *novelty* nilai terbesar juga dimiliki oleh algoritma *Eclat* yaitu 0.4819. Begitu juga dari nilai maksimum *support* dan *confidence* terbaik adalah algoritma *Eclat*, sehingga algoritma *eclat* lebih baik dalam mengevaluasi *rule*.

3.8 Hasil Rule Eclat

Untuk mendapatkan pola yang mempengaruhi faktor pandangan konsumsi masyarakat di Indonesia, maka di tentukan *conclusion* masing masing faktor yaitu kesehatan, harga, agama, tempat, dan tradisonal. Ke 5 faktor ini akan ditentukan berdasarakan nilai *support* dan *confidence* terbesar. Berikut hasil yang diperoleh.

Tabel 9. Hasil Pola Faktor Pandangan Konsumsi Masyarakat Perkotaan di Indonesia

<i>Premise</i>	<i>Conclusion</i>	<i>Support</i>	<i>Confidence</i>	<i>Lift</i>
Islam, Kelas atas	Kesehatan	0.108	0.774	1.280
Kelas menengah	Harga	0.204	0.566	1.150
B2	Agama	0.425	0.832	1.009
Kelas atas	Tempat	0.148	0.505	1.015
B3	Tradisional	0.105	0.523	1.074

4. Kesimpulan

Berdasarkan hasil dari pembahasan mengenai algoritma *FP-Growth* dan *Eclat* dapat disimpulkan algoritma *FP-Growth* menghasilkan *rule* terbanyak sebanyak 114 *rule* dengan waktu proses terlama yaitu 1.85 detik. Untuk algoritma *Eclat* menghasilkan *rule* terbanyak sebanyak 1151 *rule* dengan waktu proses terlama yaitu 0.8 detik. Nilai *novelty* terbesar dimiliki oleh algoritma *Eclat* sebesar 0.4819 dan *RI* sebesar 16462.24.

Dari nilai kebaikan parameter dihasilkan bahwa algoritma *Eclat* merupakan algoritma terbaik yang digunakan untuk menentukan pola pandangan konsumsi masyarakat perkotaan di Indonesia dengan nilai *support* sebesar 10% , *confidence* sebesar 100%, *rule* sebanyak 15 dan waktu proses 0 detik. Hasil prediksi pola konsumsi masyarakat perkotaan di Indonesia dengan menggunakan algoritma *Eclat* yaitu masyarakat yang mengonsumsi makanan berdasarkan faktor agama adalah masyarakat yang berusia diantara 30 dan 50 tahun dengan nilai *support* 42.5% dan *confidence* 83.2%.

Daftar Pustaka

- [1] Ramadhanti, F.B., Saputro, D.R.S., & Widyaningsih, P. 2020. Penerapan Association Rule Mining-Frequent Itemset Dengan Algoritme Frequent Pattern Growth (FP-Growth) Pada Dataset Kelulusan Mahasiswa S1, hlm. 340-349. Prosiding KNPMP, Surakarta.
- [2] BPS Indonesia. 2022. Ekonomi Indonesia. <http://www.bps.go.id/> . Diakses pada tanggal 10 Oktober 2022.
- [3] Karyawati, E & Winarko, E. 2011. Class Association Rule Pada Metode Associative Classification. IJCCS. 5(3): 17-24.

- [4] Wijaya, N.K., Malik, R.F., & Nurmaini, S. 2020. Analisa Pola Frekuensi Keranjang Belanja Dengan Perbandingan Algoritma Fp-Growth (Frequent Pattern Growth) Dan Eclat Pada Minimarket. *Jurnal Teknik Informatika dan Sistem Informasi*. 7(2): 364-373.
- [5] Srinadh, V. 2022. Evaluation Of Apriori, FP Growth And Eclat Association Rule Mining algorithms. *International Journal Of Kesehatan Science*. 6(2): 7475-7485.
- [6] Seda, F.S., Setyawati, L., Tirta, T., & Nobel, K. 2020. Dataset on The Cultural Dimension of Urban Society Food Consumption in Indonesia. *International Journal Of Elsevier Inc*. 14(7): 1-9.
- [7] Sudarsono., Wijaya, A., & Andri. 2019. Perbandingan Algoritma Eclat Dan Fp Growth Pada Penjualan Barang. *Bina Darma Conference on Computer Science*. 1(1): 208-217.
- [8] Supangat., & Joyonegoro, M.R. 2020. Perbandingan Akurasi Algoritma Machine Learning Untuk Prediksi Pendaftar Mahasiswa Baru. *Jurnal Turnitin*. 1(1): 2-9.
- [9] Zhang, C., & Zhang, S. 2007. *Association Rule Mining*. Springer, Australia.
- [10] Amri, K. 2022. Penerapan Data Mining Dalam Mencari Pola Asosiasi Data Tracer Study Menggunakan Equivalence Class Transformation (Eclat). *Jurnal Nasional Komputasi dan Teknologi Informasi*. 5(3): 442-449.
- [11] Arhami, M., & Nasir, M. 2020. *Data Mining Algoritma dan Implementasi*. Penerbit Andi, Yogyakarta.
- [12] Chauhan, R. & Kaur, H. 2017. Predictive Analytics and Data Mining. *Advances in Secure Computing, Internet Services, and Applications*. 3(2): 73-88.
- [13] Evadini, S. 2022. Analisis Faktor Resiko Kematian Dengan Penyakit Komorbide COVID-19 Menggunakan Algoritma Eclat. *Jurnal Informasi dan Teknologi*. 4(1): 52-57.
- [14] Niu, X., & Ji, X. 2014. Evaluation Method For Association Rules in Spatial Knowledge Base. *ISPRS Annals of the Photogrammetry*. 11(4): 53-58.
- [15] Lestari, Y.D. 2015. Penerapan Mining Menggunakan FP-Tree dan FP-Growth Pada Data Transaksi Penjualan Obat. *Jurnal Nasional Teknologi dan Komunikasi*. 1(1): 60-64.