

**PENERAPAN METODE ASSOCIATION RULE MENGGUNAKAN  
ALGORITMA FP-GROWTH DAN ECLAT PADA POLA KONSUMSI  
MASYARAKAT PERKOTAAN DI INDONESIA**

**(Skripsi)**

**Oleh**  
**ALDIANSYAH**



**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS LAMPUNG  
BANDAR LAMPUNG  
2022**

## **ABSTRACT**

### **APPLICATION OF ASSOCIATION RULE METHOD USING FP-GROWTH ALGORITHM AND ECLAT ON CONSUMPTION PATTERNS OF URBAN COMMUNITY IN INDONESIA**

**By**

**Aldiansyah**

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. 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, Confidence

## **ABSTRAK**

### **PENERAPAN METODE ASSOCIATION RULE MENGGUNAKAN ALGORITMA *FP-GROWTH* DAN *ECLAT* PADA POLA KONSUMSI MASYARAKAT PERKOTAAN DI INDONESIA**

**Oleh**

**Aldiansyah**

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, mampu mempertahankan daya konsumsi masyarakat. Dalam penelitian ini akan dilakukan mencarian pola konsumsi masyarakat perkotaan di Indonesia menggunakan dua algoritma, untuk mendapatkan algortima 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. Pola terbaik yang dihasilkan adalah masyarakat yang mengonsumsi makanan berdasarkan faktor agama adalah masyarakat yang berusia diantara 30 dan 50 tahun.

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

**PENERAPAN METODE ASSOCIATION RULE MENGGUNAKAN  
ALGORITMA FP-GROWTH DAN ECLAT PADA POLA KONSUMSI  
MASYARAKAT PERKOTAAN DI INDONESIA**

**Oleh  
ALDIANSYAH  
1917031044**

**Skripsi**

Sebagai Salah Satu Syarat untuk Mencapai Gelar  
**SARJANA MATEMATIKA**

Pada  
Jurusan Matematika  
Fakultas Matematika dan Ilmu Pengetahuan Alam  
Universitas Lampung



**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS LAMPUNG  
BANDAR LAMPUNG  
2022**

Judul Skripsi

: **PENERAPAN METODE ASSOCIATION  
RULE MENGGUNAKAN ALGORITMA  
FPGROWTH DAN ECLAT PADA POLA  
KONSUMSI MASYARAKAT PERKOTAAN  
DI INDONESIA**

Nama Mahasiswa

: **Aldiansyah**

Nomor Pokok Mahasiswa

: **1917031044**

Jurusan

: **Matematika**

Fakultas

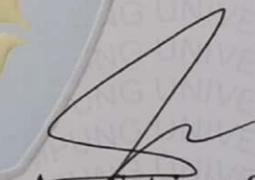
: **Matematika dan Ilmu Pengetahuan Alam**



*[Handwritten signature of Drs. Eri Setiawan, M.Si.]*

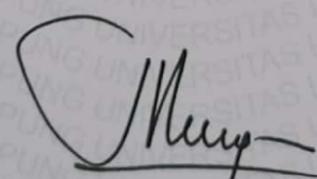
**Drs. Eri Setiawan, M.Si.**

NIP. 19581101 198803 1 002



**Agus Sutrisno, S.Si., M.Si.**  
NIP. 19700831 199983 1 002

2. Ketua Jurusan Matematika



*[Handwritten signature of Dr. Aang Nuryaman, S.Si., M.Si.]*

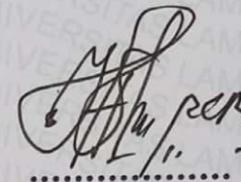
**Dr. Aang Nuryaman, S.Si., M.Si.**  
NIP. 19740316 200501 1 001

**MENGESAHKAN**

1. Tim Penguji

Ketua

: Drs. Eri Setiawan, M.Si.



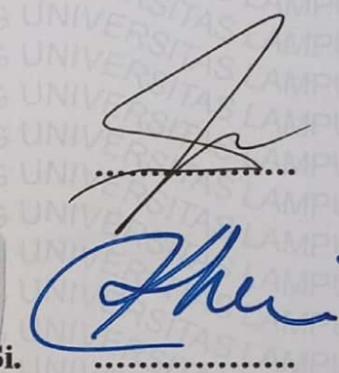
Sekretaris

: Agus Sutrisno, S.Si., M.Si.

Penguji

Bukan Pembimbing

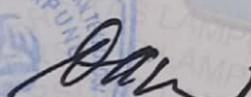
: Dr. Khoirin Nisa, S.Si., M.Si.



2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam

Dr. Eng. Suripto Dwi Yuwono, S.Si., M. T.

NIP. 19740705 200003 1 001



Tanggal Lulus Ujian Skripsi: **20 Desember 2022**

## **PERNYATAAN SKRIPSI MAHASISWA**

Saya yang bertanda tangan di bawah ini:

Nama : Aldiansyah  
Nomor Pokok Mahasiswa : 1917031044  
Jurusan : Matematika  
Judul Skripsi : **PENERAPAN METODE ASSOCIATION RULE MENGGUNAKAN ALGORITMA FP-GROWTH DAN ECLAT PADA POLA KONSUMSI MASYARAKAT PERKOTAAN DI INDONESIA**

Dengan ini menyatakan bahwa skripsi ini adalah hasil pekerjaan saya sendiri dan semua tulisan yang tertuang dalam skripsi ini telah mengikuti kaidah karya penulisan ilmiah Universitas Lampung.

Bandar Lampung, 20 Desember 2022

Penulis



**Aldiansyah**  
**NPM. 1917031044**

## **RIWAYAT HIDUP**

Penulis bernama lengkap Aldiansyah lahir di Srikaton, Semaka, Tanggamus pada 23 November 2000. Penulis merupakan anak kedua dari dua bersaudara pasangan Bapak Surya Efendi dan Ibu Jumilah.

Penulis menempuh pendidikan Sekolah Dasar di SDN 1 Srikaton pada tahun 2007 sampai dengan 2013. Penulis melanjutkan pendidikan Sekolah Menengah Pertama di SMPN 1 Semaka pada tahun 2013 sampai dengan 2016. Kemudian menempuh pendidikan Sekolah Menengah Atas di SMAN 1 Gadingrejo pada tahun 2016 sampai dengan 2019. Pada tahun 2019 penulis melanjutkan pendidikan Strata Satu (S1) di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam (FMIPA) Universitas Lampung melalui jalur SBMPTN.

Selama menjadi mahasiswa, penulis aktif dalam organisasi Natural FMIPA Unila sebagai Anggota Bidang Sirkulasi dan Periklanan (Ruslan) 2019 dan Kepala Biro Bidang Sirkulasi dan Periklanan (Ruslan) pada periode 2020. Pada periode 2021, penulis menjadi anggota luar biasa Koperasi Mahasiswa (Kopma) Unila.

Pada bulan Juli sampai dengan Agustus 2021 penulis mengikuti kegiatan Kredensial Mikro Mahasiswa Indonesia (KMMI) yaitu di Medan, Universitas Prima Indonesia. Pada bulan November 2021, penulis mengikuti relawan Warga Bantu Warga di RS. Jawa Timur. Pada bulan Januari sampai Februari, sebagai bentuk aplikasi bidang ilmu kepada masyarakat, penulis melaksanakan kegiatan Kuliah Kerja Nyata (KKN) di Kacapura, Kecamatan Semaka, Kabupaten Tanggamus. Pada bulan Juni sampai Juli penulis melaksanakan Kerja Praktik (KP) di Badan Pusat Statistik (BPS) Kabupaten Tanggamus.

## **KATA INSPIRASI**

“Berfikirlah positif, tidak peduli seberapa keras kehidupanmu”  
(Ali bin Abi Thalib)

“Dan barang siapa yang bertakwa kepada Allah, niscaya Allah menjadikan baginya kemudahan dalam urusannya”  
(Q.S. At-Talaq: 4)

“Allah tidak akan membebani seseorang melainkan sesuai dengan kesanggupannya”  
(Q.S. Al-Baqarah: 286)

“Janganlah menunda urusanmu berlarut-larut, setiap detik ketertinggalanmu menjadikan keterdahuluan bagi yang lain”  
(Aldiansyah)

## **PERSEMBAHAN**

*Bismillahirrahmanirrahim  
Alhamdulillahirobbil'alamin*

Dengan segala kerendahan hati mengucapkan syukur kepada Allah SWT  
Atas segala ridhonya sehingga penulis dapat menyelesaikan skripsi ini.

Saya persembahkan karya ini kepada :

### ***Kedua Orang Tua***

Terima kasih atas segala kasih sayang, pengorbanan, doa, pengetahuan, dan nasehat. Terima kasih yang tiada terhingga telah mendukung segala bentuk keinginan penulis. Terima kasih telah mengajarkan banyak hal luar biasa, dan banyak pelajaran yang berharga sehingga menjadikan penulis menjadi seseorang yang kuat .

### ***Dosen Pembimbing dan Pembahas***

Terimakasih kepada dosen pembimbing dan pembahas yang sudah sangat membantu, memberikan motivasi, serta memberikan arahan dan ilmu yang berharga.

### ***Sahabat-sahabatku***

Terimakasih atas semua keceriaan dan semangat yang telah diberikan.

***Almamater Tercinta, Universitas Lampung***

## SANWACANA

Puji syukur kehadirat Allah SWT, atas segala rahmat dan karunianya sehingga penulis dapat menyelesaikan skripsi dengan judul "**Penerapan Metode Association Rule Menggunakan Algoritma FP-Growth dan Eclat Pada Pola Konsumsi Masyarakat Perkotaan di Indonesia**". Skripsi ini disusun sebagai salah satu syarat memperoleh gelar Sarjana Matematika (S. Mat) pada Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.

Dalam kesempatan ini penulis ingin mengucapkan terimakasih kepada :

1. Bapak Drs. Eri Setiawan, M.Si., selaku pembimbing utama atas kesediaan waktu, pemikiran dalam memberikan evaluasi, arahan, dan saran yang membangun dalam proses penyusunan skripsi ini.
2. Bapak Agus Sutrisno, S.Si., M.Si., selaku pembimbing kedua atas kesediaan waktu, arahan dan saran yang membangun dalam proses penulisan skripsi ini.
3. Ibu Dr. Khoirin Nisa, S.Si., M.Si., selaku dosen pembahas atas kesediaan waktu, saran, dan masukan yang membangun selama proses penyusunan skripsi ini.
4. Ibu Dra. Dorrah Aziz, M.Si., selaku dosen pembimbing akademik yang telah memberikan bimbingan, motivasi, dan nasehat selama penulis menjalankan perkuliahan.
5. Bapak Dr. Aang Nuryaman, S.Si., M.Si., selaku Ketua Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
6. Bapak Dr. Eng. Suripto Dwi Yuwono, M.T., selaku dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
7. Seluruh Dosen, Staf, dan Civitas Akademika Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.

8. Bapak dan Mamak tersayang, Yayuk, beserta keluarga besar yang selalu memberikan semangat dukungan, dan do'a kepada penulis.
9. Untuk keluarga besar berencana, Maji, Sinta, Dini, Puja, Wiranto, dan Fazri, terimakasih untuk segala motivasi, dukungan, kebersamaan, kehangatan, dan semangat positif dalam menjalani perkuliahan dan selama proses penyusunan skripsi ini.
10. Untuk sahabat seperjuangan dari Semaka, Galih, Didik, Wilda, Egi, Nurul, Ridho dan Dinda atas segala doa dukungan, semangat, dan menemani penulis dari SD hingga saat ini.
11. Teman-teman pimpinan Natural 2020. Terimakasih atas segala pengalaman dan kebersamaan selama ini.
12. Teman-teman KKN Kacapura, untuk segala kebersamaanya selama ini.
13. Teman-teman Matematika 2019, dan Abang Yunda yang telah membantu selama perkuliahan selama ini.
14. Seluruh pihak terkait yang membantu dalam menyelesaikan skripsi ini yang tidak dapat penulis sebut satu persatu.

Penulis menyadari skripsi ini masih jauh dari kata sempurna dan masih terdapat banyak kekurangan. Oleh sebab itu, saran dan kritikan yang membangun senantiasa penulis harapkan. Dan semoga skripsi ini dapat bermanfaat bagi pihak yang memerlukan. Aamiin.

Bandar Lampung, 20 Desember 2022

Penulis

**Aldiansyah**

NPM. 1917031044

## **DAFTAR ISI**

Halaman

<b>DAFTAR TABEL</b> .....	xv
<b>DAFTAR GAMBAR</b> .....	xvi
<b>I. PENDAHULUAN</b> .....	1
1.1 Latar Belakang dan Masalah .....	1
1.2 Tujuan Penelitian.....	3
1.3 Manfaat Penelitian.....	3
<b>II. TINJAUAN PUSTAKA</b> .....	4
2.1 Uji Validitas dan Reliabilitas .....	4
2.1.1 Uji Validitas .....	4
2.1.2 Uji Reliabilitas .....	5
2.2 Data Mining.....	5
2.3 <i>Machine Learning</i> .....	8
2.4 <i>Unsupervised Learning</i> .....	8
2.5 Kaidah Asosiasi ( <i>Association Rule</i> ) .....	9
2.5.1 Parameter-Parameter <i>Association Rule</i> .....	9
2.5.2 Ukuran Kepentingan <i>Association Rule</i> .....	15
2.5.3 Kriteria Kebaikan Parameter <i>Association Rule</i> .....	15
2.6 <i>FP-Growth (Frequent Pattern Growth)</i> .....	17
2.6.1 <i>FP-Tree (Frequent Pattern Tree)</i> .....	18
2.7 <i>ECLAT (Equivalence Class Transformation)</i> .....	20
2.8 Evaluasi Hasil <i>Association Rule</i> .....	21
<b>III. METODOLOGI PENELITIAN</b> .....	23
3.1 Waktu dan Tempat Penelitian .....	23
3.2 Data Penelitian .....	23
3.3 Metode Penelitian.....	23

<b>IV. HASIL DAN PEMBAHASAN.....</b>	<b>26</b>
4.1 Uji Validitas dan Reliabilitas .....	26
4.1.1 Uji Validitas .....	27
4.1.2 Uji Reliabilitas .....	29
4.2 Proses Data Mining .....	29
4.2.1 Seleksi Data.....	30
4.2.2 Data <i>Preprocessing</i> .....	33
4.2.3 Transformasi Data.....	33
4.3 Pencarian Frequent Itemset .....	35
4.4 Assosiation Rule Mining Dengan <i>FP-Growth</i> .....	36
4.4.1 Input Data.....	37
4.4.2 Penentuan Minimum <i>Support</i> dan Minimum <i>Confidence</i> .....	38
4.5 Assosiation Rule Mining dengan <i>Eclat</i> .....	41
4.5.1 Input Data.....	42
4.5.2 Penentuan Minimum <i>Support</i> dan Minimum <i>Confidence</i> .....	42
4.6 Hasil Association Rule dengan <i>FP-Growth</i> dan <i>Eclat</i> .....	44
4.7 Evaluasi Association Rule <i>FP-Growth</i> dan <i>Eclat</i> .....	45
4.8 Pembentukan Pola Association Rule <i>Eclat</i> .....	49
4.9 Pola Pandangan Konsumsi Masyarakat Perkotaan di Indonesia.....	49
<b>V. KESIMPULAN .....</b>	<b>52</b>
<b>DAFTAR PUSTAKA .....</b>	<b>53</b>
<b>LAMPIRAN</b>	

## DAFTAR TABEL

Tabel	Halaman
1. Nilai Kebaikan <i>Support</i> .....	16
2. Nilai Kebaikan <i>Confidence</i> .....	16
3. Nilai Kebaikan <i>Liftt</i> .....	16
4. Format Perbedan Data Horizontal dan Vertikal .....	20
5. Indikator Data.....	26
6. Data untuk Diuji Validitas dan Reliabilitas.....	27
7. Uji Validitas 710 Sampel .....	28
8. Data Awal Pandangan Konsumsi Masyarakat Perkotaan di Indonesia.....	30
9. Data Seleksi Pandangan Konsumsi Masyarakat Perkotaan di Indonesia.....	31
10. Deskriptif Data .....	32
11. Pengecekan <i>Missing Data</i> .....	33
12. Data Keterangan Kode Umur .....	34
13. Data Hasil Transformation Untuk <i>FP-Growth</i> .....	34
14. Data Hasil Transformasi Untuk <i>Eclat</i> .....	35
15. Frekuensi Kemunculan <i>Support</i> Setiap Item.....	36
16. Hasil Pengujian 1 .....	38
17. Hasil Pengujian 2 .....	38
18. Hasil Pengujian 3 .....	39
19. Hasil Pengujian 5 .....	39
20. Hasil Pengujian 6 .....	40
21. Hasil Pengujian 8 .....	40
22. Hasil Pola Data dengan <i>Conclution Faktor</i> .....	41
23. Hasil Pengujian Parameter <i>Eclat</i> .....	42
24. Hasil Pola Data dengan <i>Conclution Faktor</i> .....	43
25. Hasil Uji Coba Parameter <i>FP-Growth</i> .....	44
26. Hasil Uji Coba Parameter <i>Eclat</i> .....	44
27. Evaluasi <i>Association Rule</i> .....	45
28. Perbandingan Nilai RI Antar Algoritma .....	48
29. Hasil Pembentukan <i>Rule Eclat</i> .....	49
30. Hasil Pola Faktor Pandangan Konsumsi Masyarakat Perkotaan di Indonesia	50

## **DAFTAR GAMBAR**

Gambar	Halaman
1. Proses Data Mining .....	6
2. Kontruksi <i>FP-Tree</i> .....	19
3. Flowchart Algoritma <i>FP-Growth</i> dan <i>Eclat</i> .....	25
4. Operator Discrete dan Numeric to Binomial.....	37
5. Operator <i>FP-Growth</i> dan <i>Create Assosiation Rule</i> .....	37
6. Perbandingan Banyaknya <i>Rule</i> .....	46
7. Perbandingan Waktu dalam Mengeluarkan <i>Output</i> .....	47
8. Plot <i>Rule Relationship Between Items</i> .....	51

## I. PENDAHULUAN

### 1.1 Latar Belakang dan Masalah

*Unsupervised learning* merupakan salah satu algoritma dalam *machine learning* yang digunakan untuk menarik kesimpulan dari data set yang terdiri dari input data *non labeled response*. Salah satu algoritma yang digunakan adalah *association rule*, dimana metode ini digunakan untuk menentukan faktor yang paling mempengaruhi suatu hal berupa *item set* yang dapat menentukan pola *item set* pada suatu data.

Salah satu pola *item set* yang bisa dilihat adalah mengenai pola konsumsi masyarakat. Pola konsumsi masyarakat di Indonesia merupakan salah satu tolak ukur dalam menentukan indeks perekonomian masyarakat Indonesia. Banyak faktor yang menunjang terjadinya perubahan dalam konsumsi masyarakat Indonesia. Menurut BPS Indonesia (2022), perekonomian di Indonesia pada Triwulan II 2022 mengalami pertumbuhan impresif sebesar 5,44%. Dengan melihat pertumbuhan ekonomi Indonesia yang naik, maka daya konsumsi masyarakat Indonesia juga naik. Untuk mempertahankan kenaikan ekonomi ini, maka industri makanan dan minuman perlu melakukan adaptasi agar ketahanan pangan dan *sustainability* lebih terasa. Dengan melihat faktor-faktor yang dapat menentukan pola konsumsi masyarakat Indonesia, 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, salah satunya adalah *Frequent Pattern Growth* atau dinamakan *FP-Growth*. Algoritma ini membangkitkan *frequent itemset* dengan menelusuri *FP-tree* secara *bottom-up* menggunakan strategi *divide-and conquer* dalam memilah masalah menjadi lebih kecil dengan mengkonstruksi *conditional FP-tree* dari *global FP-tree* (Karyawati & Winarko, 2011).

Selain algoritma *FP-Growth*, terdapat algoritma lain yang dapat digunakan untuk menentukan pola data dari sekumpulan *itemset* yaitu *Equivalence Class Transformation* atau *Eclat*. Algoritma ini pada dasarnya melakukan pencarian secara *Depth First Search (DFS)* dengan tata letak vertikal dan jika *database* berbentuk horizontal terlebih dahulu dikonversi menjadi vertikal (Wijaya, dkk., 2020).

Berdasarkan kedua algoritma dari *association rule* tersebut, keduanya bisa digunakan untuk menentukan pola yang sering muncul dalam sebuah *itemset* data, maka sangat menarik untuk mengkaji perbandingan hasil analisis algoritma yang memiliki hasil lebih efisien dan akurat dalam menentukan prediksi pola data.

Berbagai penelitian yang mengkaji perbandingan algoritma *FP-Growth* dan algoritma *Eclat* pernah dilakukan sebelumnya. Penelitian Wijaya, dkk. (2020), melakukan perbandingan kedua algoritma ini pada data transaksi minimarket dihasilkan bahwa sedikit *rule* yang lebih efisien atau aturan yang kuat dan waktu proses yang lebih cepat pada algoritma *Eclat*. Sudarsono, dkk. (2019) melakukan perbandingan algoritma *FP-Growth* dan *Eclat* pada penjualan barang minimarket 212 *mart* diperoleh algoritma *FP-Growth* lebih banyak menghasilkan *rule* dalam pola pembelian dibandingkan dengan algoritma *Eclat*. Srinadh (2022), melakukan evaluasi antara algoritma *apriori*, *FP-Growth* dan *Eclat* dengan menggunakan data transaksi diperoleh bahwa *Eclat* lebih baik dari ke tiga algoritma dalam menghasilkan *rule* dengan menghabiskan lebih sedikit waktu.

Dari penelitian-penelitian yang sudah dilakukan, data yang digunakan merupakan data transaksi barang. Oleh karena itu, penulis 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.

## 1.2 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah:

1. Menerapkan dan mengevaluasi algoritma *FP-Growth* dan *Eclat* dilihat dari nilai kebaikan parameter, efisiensi waktu dan banyaknya *rule*.
2. Mendapatkan pola *item set* untuk variabel yang paling mempengaruhi pola konsumsi masyarakat perkotaan di Indonesia.

## 1.3 Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah:

1. Menambah pengetahuan dalam penerapan algoritma *FP-Grwoth* dan *Eclat*.
2. Mendapatkan algoritma dengan parameter terbaik yang dapat digunakan untuk menentukan pola data sehingga mampu digunakan untuk prediksi data serta dijadikan referensi untuk penelitian selanjutnya.
3. Mengetahui faktor apa saja yang mempengaruhi pola konsumsi masyarakat perkotaan di Indonesia.

## **II. TINJAUAN PUSTAKA**

### **2.1 Uji Validitas dan Reliabilitas**

#### **2.1.1 Uji Validitas**

Validitas adalah suatu ukuran yang menunjukkan tingkat-tingkat kevalidan suatu kuisioner. Suatu kuesioner yang kurang valid berarti validitasnya rendah. Rumus yang digunakan adalah yang dikemukakan oleh Pearson yang dikenal dengan korelasi Pearson (Arikunto, 2006) sebagai berikut:

$$r_{xy} = \frac{N \sum XY - (\sum X)(\sum Y)}{\sqrt{\{N \sum X^2 - (\sum X)^2\}\{N \sum Y^2 - (\sum Y)^2\}}} \quad (2.1)$$

dengan,

$r_{xy}$  : Koefisien korelasi Pearson

$N$  : Jumlah subjek uji coba

$\sum X$  : Jumlah skor butir

$\sum Y$  : Jumlah total

Selanjutnya angka korelasi yang diperoleh dibandingkan dengan angka kritis tabel korelasi nilai  $r_{tabel}$ . Apabila  $r_{hitung}$  nilainya diatas angka taraf nyata 5% maka pernyataan tersebut valid dan sebaliknya apabila  $r_{hitung}$  nilainya dibawah taraf nyata 5% maka pernyataan tidak valid.

### 2.1.2 Uji Reliabilitas

Reliabilitas menunjukkan bahwa suatu kuesioner cukup dapat dipercaya untuk digunakan sebagai alat pengumpul data yang sudah baik (Arikunto, 2006). Dalam pengujian untuk mencari reliabilitas indikator yang skornya bukan 1 dan 0.

$$r_{xy} = \left( \frac{k}{k-1} \right) \left( 1 - \frac{\sum \sigma b^2}{\sigma_t^2} \right) \quad (2.2)$$

dengan,

- $r_{xy}$  : Reliabilitas kuesioner
- $k$  : Banyaknya butir pertanyaan
- $\sum \sigma b^2$  : Jumlah varian butir
- $\sigma_t^2$  : Varian total

Apabila nilai  $r_{xy}$  ini dikonsultasikan dengan nilai  $r$  Pearson, dapat diketahui bahwa lebih kecil dari  $r_{tabel}$  yang ada. Dengan demikian dapat disimpulkan bahwa kuesioner tersebut tidak reliabel. Dalam pengujian reliabilitas peneliti menggunakan uji cronbach's alpha, nilai lebih dari 0.6 maka butir atau pertanyaan dapat dikatakan reliabel.

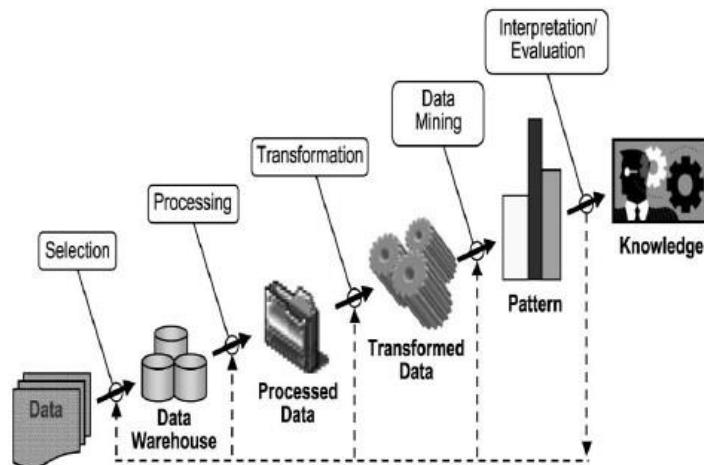
## 2.2 Data Mining

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.

Menurut Ramadhanti, dkk. (2020), data mining merupakan analisis dataset untuk menemukan hubungan yang tak terduga dan meringkas dataset tersebut dengan cara yang berbeda. Data mining adalah proses penentuan, pengekplorasian dan pemodelan data untuk menemukan pola atau hubungan yang jelas dan berguna dalam proses analisis data (Bellazzi & Zupan, 2008).

Data mining merupakan irisan dari *machine learning*, *artificial intelligence*, pengenalan pola, statistik, dan sistem basis data. Karena irisan tersebut melahirkan data mining sebagai metode baru yang mampu memberikan solusi bagi pengelolahan data dengan ukuran besar, data dengan dimensi yang tinggi, data heterogen.

Menurut Amri (2022), data mining adalah bagian integral dari penemuan pengetahuan dalam database yang merupakan proses dengan urutan sebagai berikut.



Gambar 1. Proses Data Mining

## 1. Seleksi Data

Seleksi data merupakan pengambilan data yang berhubungan dengan analisis basis data dimana hanya data yang dapat dipakai untuk analisis saja yang akan diambil dari database.

## 2. Praproses Data

Dalam praproses data terdapat dua tahap yaitu:

### a. Pembersihan Data

Pembersihan data merupakan proses untuk menghilangkan *noise* dan inkonsistensi data.

b. Data *integration*

Proses data *integration* merupakan proses dimana beberapa data *sources* akan dikombinasikan.

3. Transformasi Data

Transformasi data merupakan proses dalam pemilihan data yang telah dipilih. Data akan ditransformasikan ke bentuk yang lebih terstruktur untuk mempermudah proses data mining. Normalisasi data merupakan salah satu cara transformasi data.

Menurut Lewis (2017), terdapat beberapa teknik normalisasi data, yaitu:

a. *Min-max Normalization*

Teknik ini digunakan untuk mengatasi perbedaan nilai yang cukup besar antar dataset dengan mengubah nilai data actual menjadi nilai skala (0,1) tanpa mengubah informasi yang ada.

$$x' = \frac{(x - x_{min})}{(x_{max} - x_{min})} \quad (2.3)$$

dengan,

$x'$  : Data hasil normalisasi

$x$  : Data asli

$x_{min}$  : Nilai minimum dari  $x$

$x_{max}$  : Nilai maksimum dari  $x$

4. Data mining

Data mining merupakan proses utama data mining dimana teknik data mining diterapkan untuk menghasilkan pola data.

5. *Pattern evaluation*

Pola informasi yang dihasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak berkepentingan.

### 2.3 *Machine Learning*

*Machine learning* merupakan cabang ilmu bagian dari *artificial intelligent* (AI) yang bekerja pada data besar untuk menghasilkan pola tertentu. *Machine learning* fokus pada pengembangan sistem yang mampu belajar sendiri untuk memutuskan sesuatu tanpa harus berulang kali diprogram oleh manusia (Retnoningsih & Pramudita, 2020).

*Machine learning* menggunakan teori statistik yang dapat memproses algoritma pembelajaran dengan jumlah informasi yang banyak dan menggunakan teori statistika dalam membangun model matematis untuk membuat kesimpulan dari sampel *input* yang tersedia.

Algoritma *machine learning* digunakan untuk mengekstrak model yang didapat dari pengolahan data mentah menjadi data yang dapat digunakan untuk tugas dan tujuan tertentu. Secara garis besar algoritma *machine learning* dibagi menjadi tiga jenis, yaitu *supervised learning*, *unsupervised learning* dan *reinforcement learning* (Supangat & Joyonegoro, 2019).

### 2.4 *Unsupervised Learning*

*Unsupervised learning* merupakan proses pembelajaran yang tidak terawasi dimana tidak memerlukan target *output*. Dalam *unsupervised learning* algoritma tidak membutuhkan data berlabel. Dalam hal ini tidak ada pengawas dan hanya input data yang tersedia (Arhami & Nasir, 2020).

Algoritma yang digunakan dalam *unsupervised learning* adalah *clustering* dan *association rule*. Kedua algoritma ini tidak membutuhkan label sehingga algoritma ini lebih leluasa mencari pola yang mungkin sebelumnya diketahui.

## 2.5 Kaidah Asosiasi (*Association Rule*)

*Association rule* merupakan pencarian hubungan dari item-item pada data transaksi sehingga tercipta aturan yang dapat digunakan untuk pengambilan keputusan. *Association rule* bertujuan mencari pola yang sering muncul pada banyak transaksi, dimana setiap transaksi terdiri dari beberapa item (Zhang, *et al.*, 2003).

Kaidah asosiasi dikenal dengan nama analisis keranjang pasar (*Market-Basket Analysis*) karena dapat menganalisis keranjang para konsumen yang melakukan transaksi. Suatu transaksi  $T$  dikatakan 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  $A \cap B = \emptyset$ .

Misalkan  $I = \{i_1, i_2, \dots, \dots, i_d\}$  sebagai semua item set dalam keranjang belanja dan  $T = \{t_1, t_2, \dots, \dots, t_N\}$  adalah semua set dalam transaksi. Yang mana transaksi  $t_1$  merupakan subset dari item yang dipilih dari  $I$ . Dalam analisis asosiasi, sebuah itemset adalah kumpulan dari 0 atau lebih item. Jika sebuah item dikatakan  $k$  item, maka ini dinamakan  $k$ -itemset.

### 2.5.1 Parameter-Parameter *Association Rule*

Dalam *association rule* ada tiga parameter penting yang berfungsi untuk pembentukan *rule* yaitu *support*, *confidence* dan *lift*. Dari parameter-parameter tersebut, memiliki kriteria-kriteria tertentu untuk menentukan mana *rule* yang paling baik digunakan. Sehingga dihasilkan aturan asosiasi yang kuat (*strong rules*).

#### a. *Support*

Menurut Thanathamathee & Sawangarreerak (2022), *support* merupakan ukuran yang menunjukkan sering munculnya suatu item dalam semua transaksi. *Support* juga menunjukkan persentase kumpulan data yang muncul dalam database, yang mana dapat dituliskan dalam rumus berikut:

1. Untuk 1-item:

$$Support(A) = P(A) = \frac{frekuensi(A)}{Jumlah total transaksi dalam data} \quad (2.4)$$

2. Untuk 2-item:

$$\begin{aligned} Support(A \rightarrow B) &= P(A \cap B) \\ &= \frac{frekuensi(A \cap B)}{Jumlah total transaksi dalam data} \end{aligned} \quad (2.5)$$

3. Untuk 3-item:

$$\begin{aligned} Support(A, B \rightarrow C) &= P(A \cap B \cap C) \\ &= \frac{frekuensi(A, B \cap C)}{Jumlah total transaksi dalam data} \end{aligned} \quad (2.6)$$

4. Untuk n-item

$$\begin{aligned} Support(A \cap B \cap C \cap \dots \cap n) &= \cap_A^n x \\ &= \frac{frekuensi(A, B, C, \dots \cap n)}{Jumlah total transaksi dalam data} \end{aligned} \quad (2.7)$$

dengan,

- $A \cap B$  : Jumlah transaksi item A dan B
- $A \cap B \cap C$  : Jumlah transaksi item A,B dan C
- $A \cap B \cap C \cap \dots \cap n$  : Jumlah transaksi item A,B,C dan seterusnya hingga item n
- $x$  : Item transaksi

Dalam menentukan *support* ada istilah *support count*, yaitu angka dari sebuah transaksi yang mana dimasukan ke itemset yang lebih spesifik.

Dalam matematika *support count* ditulis  $\sigma(X)$ , untuk itemset X dapat dituliskan:

$$\sigma(X) = |\{t_i | X \subseteq t_i, t_i \in T\}| \quad (2.8)$$

b. *Confidence*

*Confidence* merupakan nilai presentase kuatnya hubungan antar item dalam kaidah asosiasi. Sebuah *assiation rule* dengan *confidence* sama atau lebih besar dari minimum *confidence* dapat dikatakan sebagai *valid asosiation rule* dimana *confidence* dapat dituliskan sebagai berikut (Thanathamathee & Sawanggarerak, 2022):

$$\begin{aligned} Confidence(A \rightarrow B) &= P(B|A) \\ &= \frac{frekuensi(A \cap B)}{Jumlah transaksi yang mengandung A} \end{aligned} \quad (2.9)$$

$$\begin{aligned} Confidence(A, B \rightarrow C) &= P(C|A, B) \\ &= \frac{frekuensi(A, B \cap C)}{Jumlah transaksi yang mengandung A dan B} \end{aligned} \quad (2.10)$$

$$\begin{aligned} Confidence(A, B, C \dots n \rightarrow x) &= P(x|A, B, C \dots n) \\ &= \frac{frekuensi(A, B, C \dots n \cap x)}{Jumlah transaksi yang mengandung A, B, C \dots n} \end{aligned} \quad (2.11)$$

dengan,

$x$  dan  $n$  : Item transaksi

$P(B|A)$  dinamakan *probability ratio (PR)*, kondisi peluang asosiasi dijelaskan dalam derajat  $p(B|A)$  relative terhadap  $p(B)$  dengan:

$$PR(B|A) = \begin{cases} \frac{p(B|A) - p(B)}{1 - p(B)} & \text{jika } p(B|A) \geq p(B), p(B) \neq 1 \\ \frac{p(B|A) - p(B)}{1 - p(B)}, \text{jika } p(B) > p(B|A), p(B) \neq 0 & \end{cases}$$

Karena  $P(B|A) = \frac{p(A \cup B)}{p(B)}$ , maka

$$PR(B|A) = \begin{cases} \frac{p(A \cup B) - p(A)p(B)}{p(A)(1 - p(B))} & \text{jika } p(A \cup B) \geq p(A)p(B) \\ & p(A)(1 - p(B)) \neq 0 \\ \frac{p(A \cup B) - p(A)p(B)}{p(A)p(B)} & \text{jika } p(A \cup B) < p(A)p(B) \\ & p(A)p(B) \neq 0 \end{cases}$$

Atau

$$R(B|A) = \begin{cases} \text{jika } \frac{\text{supp}(A \cup B) - \text{supp}(A)\text{supp}(B)}{\text{supp}(A)(1 - \text{supp}(B))} \\ \quad \quad \quad \text{supp}(A \cup B) \geq \text{supp}(A)\text{supp}(B) \\ \quad \quad \quad \text{supp}(A)(1 - \text{supp}(B)) \neq 0 \\ \text{jika } \frac{\text{supp}(A \cup B) - \text{supp}(A)\text{supp}(B)}{\text{supp}(A)\text{supp}(B)} \\ \quad \quad \quad \text{supp}(A \cup B) < \text{supp}(A)\text{supp}(B) \\ \quad \quad \quad \text{supp}(A)\text{supp}(B) \neq 0 \end{cases}$$

### c. Lift Ratio

*Lift* adalah suatu ukuran (parameter) untuk mengetahui kekuatan atau kebaikan aturan asosiasi yang telah terbentuk dari nilai *support* dan *confidence*. *Lift ratio* mengukur seberapa besar kemungkinan barang *A* dibeli ketika barang *A* dibeli, sambil mengendalikannya seberapa terkenalnya barang *B* (Arhami & Nasir, 2020).

*Lift* dapat ditulis sebagai berikut:

$$\text{Lift}(A \rightarrow B) = \frac{P(A \cup B)}{P(A)P(B)} = \frac{\text{Support}(A, B)}{\text{support}(A) * \text{support}(B)} \quad (2.12)$$

Nilai *lift ratio* digunakan sebagai penentu valid tidaknya suatu aturan asosiasi. Ada tiga kemungkinan yang akan dihasilkan saat menghitung *lift ratio* yaitu:

1. Jika hasil dari perhitungan *lift ratio*  $< 1$ , maka itemset *A* dan itemset *B* memiliki korelasi negatif atau jika konsumen membeli barang *A* maka tidak akan membeli barang *B*.
2. Jika *lift ratio*  $> 1$ , maka itemset *A* dan *B* memiliki korelasi positif, jika konsumen membeli *A* maka akan membeli *B*.
3. Jika nilai *lift ratio*  $= 1$ , maka itemset *A* dan item set *B* independent satu sama lain, jika konsumen membeli *A* maka belum tentu membeli barang *B*.

Menurut Zhang & Zhang (2007), misalkan  $I$  adalah set dari itemset di dalam database  $D, A, B \subseteq I$  dalam itemset,  $A \cap B = \emptyset$ ,  $p(A) \neq 0$  dan  $p(B) \neq 0$ . Minimal support ( $\text{minsupp}$ ) dan minimal confidence ( $\text{minconf}$ ) telah ditentukan. Maka  $A \rightarrow B$  adalah valid jika:

1.  $\text{supp}(A \cup B) \geq \text{minsupp}$ ,
2.  $\text{conf}(A \rightarrow B) = \frac{\text{supp}(A \cup B)}{\text{supp}(A)} \geq \text{minconf}$ ,

dimana  $\text{conf}(A \rightarrow B)$  menyatakan kepercayaan dari aturan  $A \rightarrow B$ .

Dalam *association rule* terdapat dua ketetapan pengukuran dalam parameter yaitu:

1.  $\text{support}(A \rightarrow B) \approx \text{support}(A) \times \text{support}(B)$

Dengan menggunakan konsep peluang maka  $\text{support}(A \cup B) = p(A \cup B)$  dan  $\text{confidance}(A \rightarrow B) = p(A|B) = \frac{p(A \cup B)}{p(A)}$ . Sehingga dapat dituliskan sebagai berikut:

$$p(A \cup B) \approx p(A)p(B)$$

2.  $\text{Interest}(A, B) = \frac{p(A \cup B)}{p(A)p(B)}$

Dengan ekstensi yang jelas ke lebih dari dua set, rumus diatas digunakan sebagai *interest* dari  $A \rightarrow B$ , dengan salah satu pengukuran utama ketidakpastian dari aturan asosiasi. Dalam hal ini, semakin jauh nilainya dari 1, semakin besar ketergantungan atau untuk  $1 > \text{mininterest} > 0$ , jika,

$$\left| \frac{p(A \cup B)}{p(A)p(B)} - 1 \right| \geq \text{mininterest}$$

Maka  $A \rightarrow B$  adalah aturan yang menarik.

Bukti:

$$\left| \frac{p(A \cup B)}{p(A)p(B)} - 1 \right| \geq \text{mininterest}$$

Atau

$$\left| \frac{\text{supp}(A \cup B)}{\text{supp}(A)\text{supp}(B)} - 1 \right| \geq \text{mininterest}$$

Disederhanakan menjadi

$$\frac{|supp(A \cup B) - supp(A)suppB|}{supp(A)supp(B)} \geq \frac{mininterest}{supp(A)supp(B)}$$

Atau

$$\left| \frac{supp(A \cup B)}{supp(A)supp(B)} - 1 \right| \geq \frac{mininterest}{supp(A)supp(B)}$$

Karena  $supp(A)supp(B) \leq 1$ , maka

$$\frac{mininterest}{supp(A)supp(B)} \geq \frac{mininterest}{1} = mininterest$$

Maka,

$$\left| \frac{supp(A \cup B)}{supp(A)supp(B)} - 1 \right| \geq mininterest$$

Sehingga,

$$\left| \frac{p(A \cup B)}{p(A)p(B)} - 1 \right| \geq mininterest$$

Menurut Zhang & Zhang (2007), aturan yang menarik atau  $interest(A, B)$  dibagi menjadi 3 bagian, yaitu:

1. Jika  $\frac{p(A \cup B)}{p(A)p(B)} = 1$  maka  $p(A \cup B) = p(A)p(B)$  atau  $B$  dan  $A$  adalah *independent*.
2. Jika  $\frac{p(A \cup B)}{p(A)p(B)} > 1$ , atau  $p(A \cup B) > p(A)p(B)$  maka  $B$  adalah positif *dependent* dalam  $A$ .
3. Jika  $\frac{p(A \cup B)}{p(A)p(B)} < 1$  atau  $p(A \cup B) < p(A)p(B)$  maka  $B$  adalah negatif *dependent* dalam  $A$ , atau  $B$  adalah positif *dependent* dalam  $A$ .

Sehingga untuk *rule* yang *interest* maka untuk  $1 > mininterest > 0$  jika:

a.  $\frac{p(A \cup B)}{p(A)p(B)} - 1 \geq mininterest.$

Maka  $A \rightarrow B$  adalah aturan yang *interest*.

b.  $-\left(\frac{p(A \cup B)}{p(A)p(B)}\right) - 1 \geq mininterest$

Maka  $A \rightarrow -B$  adalah aturan yang *interest*.

Misalkan  $I$  adalah set dari itemset dalam database  $TD$ ,  $A, B \subseteq I$  dalam itemset,  $A \cap B = \emptyset$ ,  $p(A) \neq \emptyset$  dan  $p(B) \neq \emptyset$ . Dengan  $\text{minsupp}$ ,  $\text{minconf}$  dan  $\text{mininterest} > 0$  telah ditentukan, maka  $A \rightarrow B$  dapat dikatakan aturan valid jika:

1.  $p(A \cup B) \geq \text{minsupp}$ .
2.  $p(B|A) \geq \text{minconf}$ .
3.  $|p(A \cup B) - p(A)p(B)| \geq \text{mininterest}$ .

### 2.5.2 Ukuran Kepentingan Association Rule

Menurut Zhang & Zhang (2007), untuk aturan asosiasi  $A \rightarrow B$ , misal  $p(A)$ ,  $p(B)$  dan  $p(A \cup B)$  menunjukkan fraksi baris yang memenuhi kondisi  $A, B, AB$ , masing-masing dan  $p(B|A) = \frac{p(A \cup B)}{p(A)}$ , maka didefinisikan ukuran kepentingan dengan menggunakan rumus  $J\text{-measure}$  sebagai berikut:

Untuk satu sisi:

$$I^J(A, B) = p(A)p(B|A) \log_2 \frac{p(B|A)}{p(B)}$$

Untuk mengukur ketertarikan karena  $p(B|A) = \frac{p(A \cup B)}{p(A)}$  maka,

$$I^J(A, B) = p(A \cup B) \log_2 \frac{p(A \cup B)}{p(A)p(B)}$$

Untuk dua sisi:

$$J(A, B) = p(A) \left[ p(B|A) \log_2 \frac{p(B|A)}{p(B)} + (1 - p(B|A)) \log_2 \frac{1 - p(B|A)}{1 - p(B)} \right]$$

### 2.5.3 Kriteria Kebaikan Parameter Association Rule

Menurut Yetri, dkk. (2018), semakin besar nilai *minimum support* dan nilai *minimum confidence* yang digunakan, maka semakin sedikit waktu yang

dibutuhkan untuk menghasilkan rekomendasi yang diberikan. Sedangkan itu nilai *lift* yang baik adalah lebih dari 1 yang menunjukkan ukuran keakuratan *rule* yang dibuat. Berikut merupakan kriteria kebaikan parameter dalam *association rule*.

### 1. Kriteria Kebaikan Support

Tabel 1. Nilai Kebaikan *Support*

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

### 2. Kriteria Kebaikan *Confidence*

Tabel 2. Nilai Kebaikan *Confidence*

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

### 3. Kriteria Kebaikan *Lift*

Tabel 3. Nilai Kebaikan *Lift*

<b>Lift</b>	<b>Kriteria</b>
<i>lift ratio</i> < 1	Korelasi negatif atau <i>rule</i> tidak valid
<i>lift ratio</i> > 1	Korelasi positif atau <i>rule</i> valid
<i>lift ratio</i> = 1	<i>Rule</i> tidak dapat dipastikan

## 2.6 FP-Growth (*Frequent Pattern Growth*)

*Frequent pattern growth (FP-Growth)* adalah algoritma alternatif yang digunakan untuk menentukan himpunan data yang paling sering muncul (*frequent itemset*) dalam sebuah *dataset*. Algoritma *FP-Growth* menggunakan konsep kontruksi pohon yang disebut *FP-Tree* guna memperoleh *Frequent itemset*. Algoritme *FP-Growth* memberikan cara singkat untuk menghitung itemset yang sering dilakukan dengan meringkas catatan transaksi menggunakan struktur data grafik khusus yang disebut *FP-Tree*.

*FP-Growth* pertama-tama menghasilkan *FP-Tree* dan menggunakan pohon yang sering muncul untuk menghasilkan itemset yang sering. Efisiensi dari algoritma *FP-Growth* tergantung pada seberapa banyak pengurangan yang dapat dicapai dalam menghasilkan *FP-Tree* (Chauhan & Kaur, 2017).

Menurut Wijaya, dkk. (2020), metode *FP-Growth* dibagi menjadi 3 tahapan yaitu:

1. Tahap pembagkitan *conditional pattern base*, tahap ini merupakan tahap yang berisi *path* dan *suffix pattern* yang diperoleh dari *FP-Tree* yang telah dibentuk.
2. Tahap pembangkitn *conditional FP-Tree*. Pada tahap ini nilai *support* dari setiap item pada tahap 1 dijumlahkan, kemudian setiap item memiliki jumlah yang lebih besar atau sama dengan nilai minimum *support* yang akan dibangkitkan.
3. Tahap pencarian *frequent itemset*, jika tahap 2 adalah lintasan tunggal , maka bisa didapat *frequent pattern* dari kombinasi item dan jika bukan lintasan tunggal, maka lakukan pembangkitan secara rekursif.

Menurut Ramadhanti, dkk. (2020), langkah untuk menentukan aturan asosiasi menggunakan algoritme *FP-Growth* adalah sebagai berikut:

1. Menggunakan *minimum support* dan *minimum confidence*.
2. Menghitung nilai *support* pada setiap *item* dengan rumus berikut.

$$Support(A \rightarrow B) = \frac{frekuensi(A \cap B)}{Jumlah total transaksi dalam data base} \quad (2.15)$$

3. Hapus itemset yang nilai *support*  $\leq$  *minimum support*.
4. Membangun *FP-Tree* yang berisi akar, *label item*, *support count*, dan panah penghubung.
5. Menentukan *frequent itemset* dengan *FP-Tree* yang telah dibentuk yang merupakan lintasan tunggal.
6. Menghitung nilai *confidence* pada *frequent itemset* yang dihasilkan dengan rumus berikut.

$$Confidence(A \rightarrow B) = \frac{Support(A, B)}{Support(A)} \quad (2.16)$$

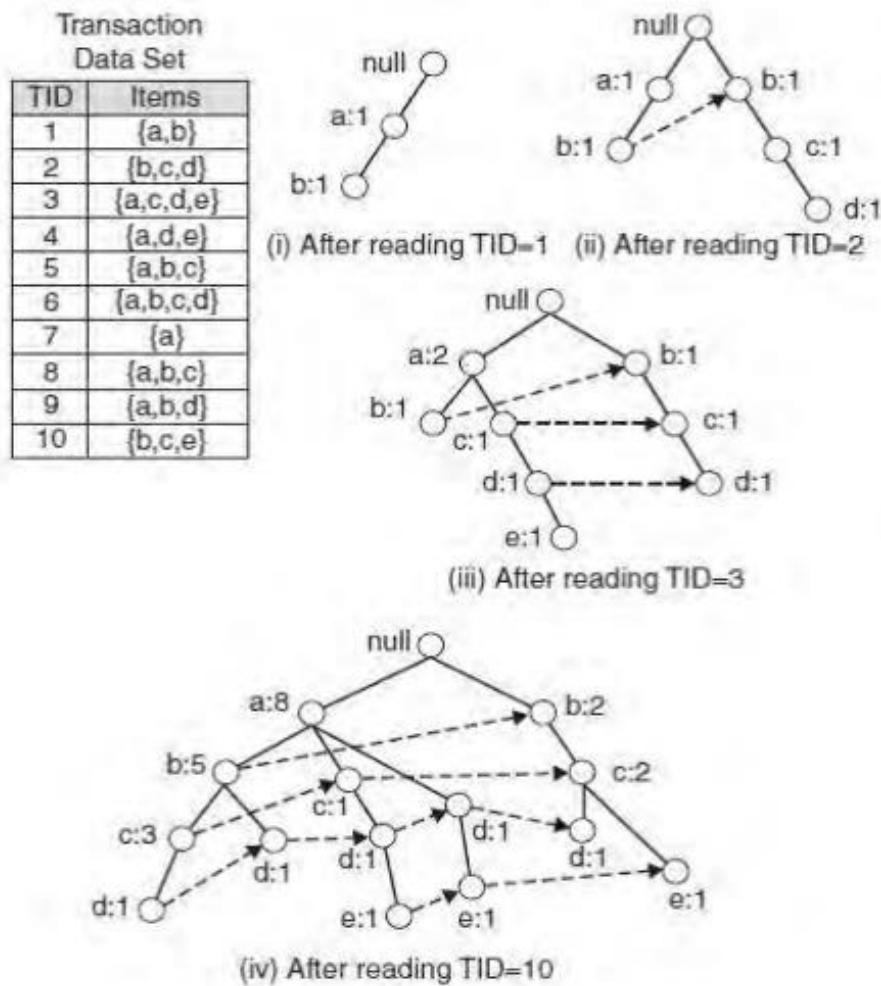
7. Menghitung nilai *lift* pada masing-masing *itemset* menggunakan rumus berikut.

$$Lift(A \rightarrow B) = \frac{Confidance(A \rightarrow B)}{Support(B)} \quad (2.17)$$

Nilai *confidence* yang terlalu tinggi dapat mengabaikan nilai *support* dari *itemset* yang muncul dalam aturan asosiasi yang telah terbentuk. Hal tersebut dapat diatasi dengan menghitung nilai *lift* pada setiap *itemset* aturan asosiasi. Jika nilai *lift*  $\geq 1$  maka menyatakan terdapat manfaat pada aturan asosiasi yang terbentuk.

### 2.6.1 FP-Tree (*Frequent Pattern Tree*)

*FP-Ttree* merupakan cara mencari itemset dengan meringkas catatan transaksi menggunakan struktur data grafik khusus kontruksi pohon. Tahapan ini merupakan tahapan yang telah dibatasi dengan menggunakan *support count* yang telah ditentukan, kemudian dibangun *tree* (Lestari, 2015).

Gambar 2. Kontruksi *FP-Tree*

Menurut Khan, dkk. (2017), langkah-langkah dalam membuat *FP-Tree* adalah:

1. Data dipindai untuk menghasilkan nilai *support* pada setiap item. Item kemudian diurutkan dari urutan menurun yang paling sering.
2. Setelah membaca transaksi pertama (a,b), *node* a dan b dibuat, kemudian jalur dibentuk dari root → a → b untuk mewakili transaksi di *tree*.
3. Setelah itu ketika transaksi kedua dilalui (b,c,d) dibuat *node* baru untuk mewakilinya.
4. Kemudian transaksi ketiga (a,c,d,e) yang memiliki awalan yang sama yaitu a. Selama pendahuluan sama maka tetap tumpeng tindih.
5. Proses serupa dilanjutkan dan dimasukan kedalam *FP-Tree*.

## 2.7 ECLAT (*Equivalence Class Transformation*)

*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 kebentuk vertikal (Evadini, 2022).

Berbeda dengan algoritma *FP-Growth*, algoritma *Eclat* menggunakan data yang ditransformasikan dengan bentuk vertikal. Dimana untuk kolom pertama adalah item dan kolom kedua adalah *TID\_set*. Berikut perbedaan format data horizontal dan vertikal.

Tabel 4. Format Perbedan Data Horizontal dan Vertikal

<b>Horizontal</b>		<b>Vertikal</b>	
<b>TID</b>	<b>Item set</b>	<b>Item</b>	<b>TID_set</b>
T01	I1,I2,I3	I1	T01,T02,T05
T02	I1,I3	I2	T01,T04,T06
T03	I3	I3	T01,T04,T05
T04	I2,I4	I4	T04,T05
T05	I1,I4		

Menurut Wijaya, dkk. (2020), dalam melakukan analisis pada algoritma *Eclat* ada 3 tahapan yaitu:

1. Inisialisasi

Inisialisasi dilakukan dengan cara mengkontruksi perhitungan global untuk *frequent 2-itemset*.

2. Tahap transformasi

Tahap transformasi digunakan untuk membagi *frequent 2-itemset* dan dipartisi ke dalam prosessor, selanjutnya melakukan transformasi vertikal database.

3. Tahap *asynchronous*

Merupakan tahap pembentukan *frequent k-itemset* sesuai dengan kebutuhan atau tujuan yang akan dicapai.

Mernurut Budiono (2015), berikut ini merupakan ilustrasi proses *mining frequent itemset* dengan menggunakan algoritma *Eclat*:

1. Transformasi data kedalam bentuk format vertikal.
2. Menentukan nilai minimum *support*.
3. Mnenetukan nilai minimum *confidence*.
4. Cari data kandidat k-itemset kemudian hitung masing-masing nilai *support*.
5. Eliminasi nilai *support* yang tidak memenuhi minimum *support*.
6. Kombinasikan semua itemset berdasarkan TID\_set yang sama pada data untuk mendapatkan k-itemset.
7. Ulangi langkah ke 4 sampai ke 6 sampai tidak menghasilkan itemset baru.
8. Mencari aturan asosiasi dengan menentukan nilai *confidence* nya.
9. Eliminasi data yang tidak memenuhi nilai minimum *confidence*.

## 2.8 Evaluasi Hasil Association Rule

Menurut Niu & Ji (2014), indeks evaluasi menyeluruh merupakan kombinasi antara kriteria evaluasi objektif dan evaluasi subjektif. Indikator evaluasi objektif ada tiga, yaitu: *support*, *confidence* dan *lift*. Indikator evaluasi subyektif terbagi menjadi tiga, yaitu: *novelty*, *simplicity* dan tingkat minat pengguna.

Karena nilai setiap indeks tidak sama, maka indeks evaluasi komprehensif didefinisikan sebagai tertimbang rata-rata dari indikator evaluasi ini. Indek evaluasi komprehensif dinamakan *RI* dengan rumus:

$$RI = S^{w_1} \times C^{w_2} \times lift^{w_3} \times W_i^{w_4} \times USI^{w_5} \times \left(\frac{1}{CN}\right)^{w_6} \quad (2.13)$$

Persamaan 2.13 disederhanakan menjadi,

$$RI = k_1 \times S + k_2 \times C + k_3 \times lift + k_4 \times W_i + k_3 \times USI + k_6 \times Wi \quad (2.14)$$

dimana,

$W_i$	: Nilai dari <i>novelty</i>
$USI$	: Tingkat minat pengguna
$CN$	: Nilai <i>simplicity</i>
$S$	: <i>Support</i>
$C$	: <i>Confidence</i>
$w_1 \geq 0, (i = 1, 2, \dots, 6) \sum_{i=1}^6 w_1 = 1$	

### **III. METODOLOGI PENELITIAN**

#### **3.1 Waktu dan Tempat Penelitian**

Penelitian ini dilaksanakan pada semester ganjil tahun ajaran 2022/2023 bertempat di Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung.

#### **3.2 Data 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 Seda, dkk. (2020), dengan judul *Dataset on The Cultural Dimension Of Urban Society Food Consumption in Indonesia* dari Jurusan Sosiologi, Universitas Indonesia. Data sebanyak 710 data beserta 22 atribut yang diperoleh dari 5 kota besar di Indonesia yaitu Jakarta, Bandung, Surabaya, Makasar dan Bali.

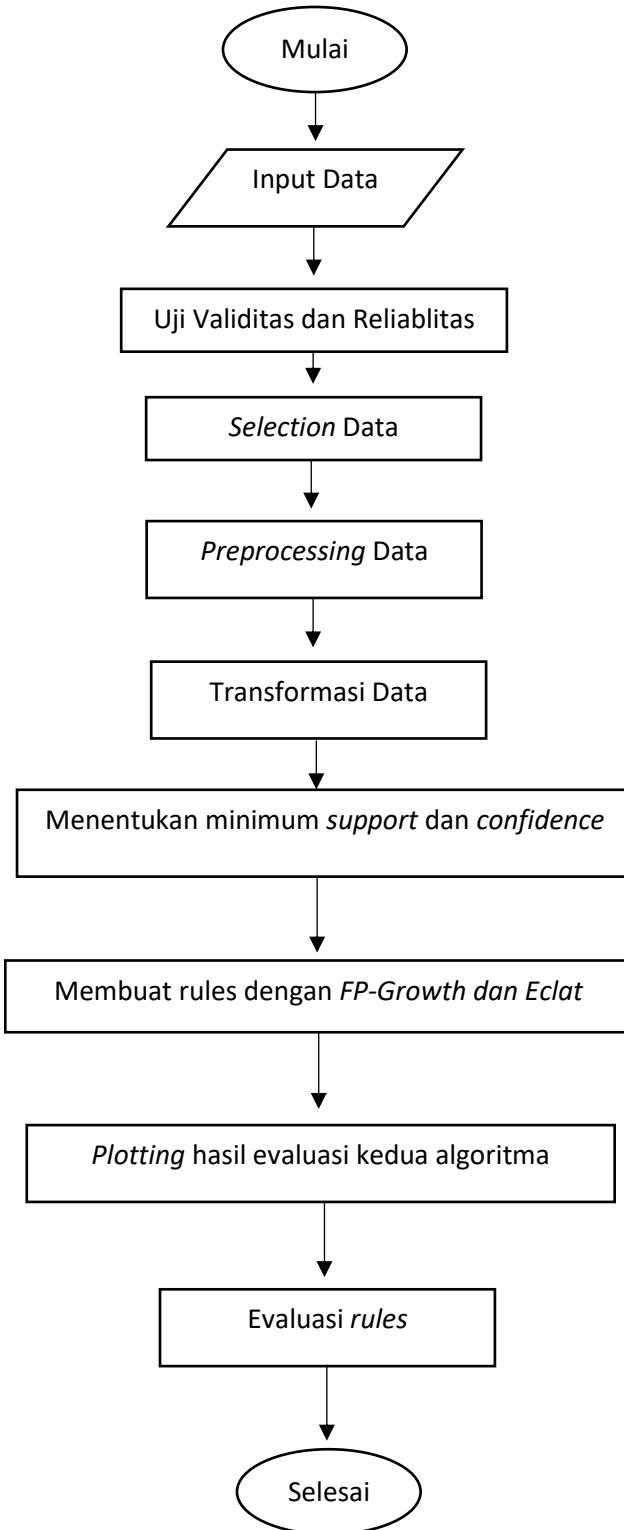
#### **3.3 Metode Penelitian**

Pada penelitian ini akan menentukan algoritma mana yang terbaik antara *FP-Growth* dan *Eclat* untuk menentukan pola konsumsi masyarakat perkotaan di Indonesia dengan bantuan *software RapidMiner* dan *software R*.

Berikut adalah tahapan-tahapan yang dilakukan dalam penelitian:

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:
  - 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:
  - 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. Evaluasi antara algoritma *FP-Growth* dan *Eclat*.
7. Menentukan aturan *association rule* untuk menentukan pola yang sering terjadi pada pola konsumsi masyarakat perkotaan di Indonesia.

Dibawah ini merupakan *flowchart* dari kedua metode.



Gambar 3. Flowchart Algoritma *FP-Growth* dan *Eclat*.

## V. KESIMPULAN

Berdasarkan hasil dari pembahasan mengenai algoritma *FP-Growth* dan *Eclat* dapat disimpulkan sebagai berikut.

1. 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.
2. Berdasarkan 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.
3. Hasil prediksi pola konsumsi masyarakat perkotaan di Indonesia dengan menggunakan algoritma *Eclat* terbaik 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**

- 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.
- Arhami, M., & Nasir, M. 2020. *Data Mining Algoritma dan Implementasi*. Penerbit Andi, Yogyakarta.
- Arikunto. 2006. *Prosedur Penelitian Suatu Pendekatan Praktik*. Rineka Citra, Jakarta.
- Bellazzi, R. & Zupan, B. 2008. Predictive Data Mining In Clinical Medicine. *International Journal of medical informatic.* **2**(3): 81-97.
- BPS Indonesia. 2022. Ekonomi Indonesia. <http://www.bps.go.id/>. Diakses pada tanggal 10 Oktober 2022.
- Budiono, A. 2015. Implementasi Data Mining Pada Penjualan Produk DI PT.Focus Gaya Graha Menggunakan Metode Association Rule. *Komputa.* **1**(1): 1-8.
- Chauhan, R. & Kaur, H. 2017. Predictive Analytics and Data Mining. *Advances in Secure Computing, Internet Services, and Applications.* **3**(2): 73-88.
- Evadini, S. 2022. Analisis Faktor Resiko Kematian Dengan Penyakit Komorbide COVID-19 Menggunakan Algoritma Eclat. *Jurnal Informasi dan Teknologi.* **4**(1): 52-57.

Karyawati., E & Winarko, E. 2011. Class Association Rule Pada Metode Associative Classification. *IJCCS*. **5**(3): 17-24.

Khan, M.A., Solaiman, K.M., & Pritom, T.H. 2017. Market Basket Analysis For Improving The Effective Of Marketing and Sales Using Apriori, FP Growth and Eclat Algoritm. Brac University.

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.

Lewis, N.D. 2017. *Neural Network For Time Series Forecasting With R*. Pearson Education, New York.

Niu, X., & Ji, X. 2014. Evaluation Method For Asssociaiton Rules in Spatial Kmowledge Base. *ISPRS Annals of the Photogrammetry*. **11**(4): 53-

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 KNPM, Surakarta.

Retnoningsih, E., & Pramudita, R. 2020. Mengenal Machine Learning Dengan Teknik Supervised Learning dan Unsupervised Learning Menggunakan Pyhton. *Bina Insansi ICT Journal*. **7**(2): 156-165.

Seda, F.S., Setyawati, L., Tirta, T., & Nobel, K. 2020. Dataset on The Cultural Dimension of Urban Society Food Consumption in Indonesia. *Elsivier Inc*. **14**(7): 1-9.

Srinadh, V. 2022. Evaluation Of Apriori, FP Growth And Eclat Associaation Rule Miningalgorithms. *International Journal Of Kesehatan Science*. **6**(2): 7475-7485.

Sudarsono., Wijaya, A., & Andri. 2019. Perbandingan Algoritma Eclat Dan Fp-Growth Pada Penjualan Barang (Studi Kasus: Minimarket 212 Mart Veteran Utama). *Bina Darma Conference on Computer Science*. **1**(1): 208-217.

Supangat., & Joyonegoro, M.R. 2020. Perbandingan Akurasi Algoritma Machine Learning Untuk Prediksi Pendaftar Mahasiswa Baru Di Universitas XYZ. *Jurnal Turnitin*. 1(1): 2-9.

Thanathamathee, P., & Sawangarreerak, S. 2022. Discovering Future Earning Pattern Through FP-Growth and Eclat Algorithms With Optimal Discretization. *Emerging Science Journal*. 6(6): 12328-1345.

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 Informasi dan Sistem Informasi*. 7(2): 364-373.

Yetri, M., Yakun, S., & Elpandri. 2018. Data Mining Untuk Analisis Pola Pemilihan Menu Pada Penag Corner Café Dan Resto Menggunakan Algoritma Apriori. *Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD*. 1(2): 114-123.

Zhang, C., & Zhang, S. 2007. *Association Rule Mining*. Springer, Australia.

Zhang, S., Zhang, C., & Yang, Q. 2003. Data Preparation For Data Mining. *Applied Artificial Intelligence*. 2(1): 375-381.