

**“PENERAPAN ALGORITMA *ECLAT* DAN *FP- GROWTH* SEBAGAI  
DASAR PERTIMBANGAN PENENTUAN PAKET PROMO DI TOKO SPC  
MART”**

**Tugas Akhir**

**Diajukan untuk Memenuhi Sebagian Persyaratan Program Strata I Program  
Studi Teknologi Informasi**



**Oleh:**

**I Putu Kevin Ari Narayana**

**1905551030**

**Reviewer:**

**I Made Andre Dwi Winama Putra**

**1905551003**

**PROGRAM STUDI TEKNOLOGI INFORMASI**

**FAKULTAS TEKNIK**

**UNIVERSITAS UDAYANA**

**2022**

## **BAB I**

### **PENDAHULUAN**

Bab I Pendahuluan menjelaskan proses dalam mengidentifikasi masalah penelitian. Pembahasan dalam bab ini diantaranya adalah latar belakang masalah, perumusan masalah, tujuan penelitian, dan manfaat kegiatan penelitian terhadap solusi dari permasalahan yang diangkat, batasan masalah serta sistematika penulisan dalam penyusunan laporan Tugas Akhir.

#### **1.1 Latar Belakang**

Badan Pusat Statistik menyebutkan bahwa *Covid-19* menyebabkan terjadinya penurunan ekonomi di Indonesia sebesar 2,07% dan sebanyak 87,5% UMKM di Indonesia terdampak oleh pandemi ini. Penurunan ekonomi ini menimbulkan persaingan bisnis yang semakin ketat dan mengharuskan para pemilik bisnis untuk berpikir kritis agar kepuasan pelanggan tetap terjaga serta penjualan produk tetap meningkat. Kemampuan untuk memperkirakan volume penjualan secara akurat sangat penting guna kelangsungan hidup bisnis. Peningkatan penjualan produk juga dapat dilakukan dengan strategi promosi yang efisien dan efektif.

Pengertian promosi menurut Kotler & Keller (2009) "*promotion means activities that communicate the merits of the product and persuade target customers to buy it*", yang artinya promosi merupakan kegiatan yang mengkomunikasikan manfaat dari sebuah produk dan membujuk target konsumen untuk membeli produk tersebut. Menurut Putri & Safri (2015) Promosi merupakan salah satu komponen dari bauran pemasaran (*marketing mix*) memiliki peranan penting dalam mengkomunikasikan suatu produk, dan dapat menciptakan preferensi konsumen atau calon konsumen mengenai keefektifan dan keefisienan dari bauran promosi (*promotion mix*) yang digunakan. Bauran promosi yang digunakan oleh setiap perusahaan tidak sama, hal ini tergantung dari kondisi perusahaan.

Promosi yang tepat akan memengaruhi keberlanjutan jangka panjang dari penjualan. Dengan mengetahui pola pembelian pelanggan, manajemen dapat menentukan strategi penjualan terkait dengan promosi, prediksi persediaan produk, dan kombinasi produk yang akan dijual.

Toko SPC MART adalah toko di Sidakarya yang menjual kebutuhan sehari-hari. Masalah yang dihadapi oleh pemilik toko adalah pencatatan persediaan barang dan penjualan yang masih dilakukan secara semi-manual, serta belum ada analisis terhadap data penjualan secara menyeluruh. Hal ini memberikan beberapa masalah, salah satunya adalah pemberian promo bagi pelanggan. Karena proses bisnis dilakukan secara manual, maka pemilik toko kesulitan dalam menganalisis barang apa saja yang laku terjual dengan cepat dan barang apa saja yang dibeli bersamaan.

Proses pemberian promo masih dilakukan dengan cara berdiskusi dengan pihak internal. Perkiraan pembelian barang valid dan pemberian promo tidak disarankan hanya berdasarkan intuisi, maka diperlukan suatu data yang mendukung pengambilan keputusan. Meningkatnya jumlah pelanggan dan aktivitas pembelian akan menghasilkan sejumlah besar data historis yang berharga. Data historis transaksi yang telah terakumulasi ini dapat dimanfaatkan untuk membantu dalam menemukan hubungan pembelian antar berbagai barang. Data yang dimiliki akan membantu mempelajari pelanggan dan mencari kombinasi barang yang sering dibeli bersamaan. Kemudian barang-barang tersebut ditawarkan sebagai rekomendasi kepada pelanggan.

Rekomendasi Promo ini perlu diterapkan pada data historis penjualan dari SPC Mart. Pembuatan produk rekomendasi yang paling populer menggunakan algoritma *Association Rule*. *Association rules* adalah salah satu task data mining deskriptif yang bertujuan untuk menemukan aturan asosiasi antara item-item data. Langkah utama yang perlu dalam *association rules* adalah mengetahui seberapa sering kombinasi item muncul dalam database, yang disebut sebagai *frequent patterns* (Han, 2011)

Penelitian ini dilakukan menggunakan algoritma *Eclat* dan *FP-Growth* untuk mencari pola hubungan antar item. Berdasarkan hasil penelitian sebelumnya, algoritma *FP-Growth* mampu menggali informasi pada data dan dapat menampilkan hubungan antar atributnya. Algoritma *FP-Growth* ini sudah banyak diterapkan dalam berbagai penelitian sebelumnya, digunakan untuk memprediksi pola pembelian obat pada apotek, pola belanja pelanggan pada toko buku dan minimarket.

Secara singkat, Algoritma *Frequent Pattern Growth (FP-Growth)* adalah salah satu alternatif algoritma yang dapat digunakan untuk menentukan himpunan data yang paling sering muncul (*frequent itemset*) dalam sebuah kumpulan data. Dalam Algoritma *Frequent Pattern Growth (FP-Growth)* menyimpan informasi mengenai frequent itemset dalam bentuk struktur prefix-tree atau sering disebut FP-Tree. Tidak melakukan candidate generation dalam proses pencarian frequent itemset sehingga dapat mengurangi scan database secara berulang dalam proses mining dan dapat berlangsung lebih cepat. (Jiawei, dkk: 2012)

Algoritma *ECLAT* merupakan algoritma yang berfungsi hampir serupa dengan *Fp-Growth*. *ECLAT* algoritma untuk menemukan itemset yang sering muncul. Pada dasarnya algoritma *ECLAT* melakukan pencarian secara depth-first search pada database dengan tata letak vertikal. Algoritma *ECLAT* ini masih jarang diterapkan dalam berbagai penelitian sebelumnya, oleh sebab itu ada baiknya apabila dijadikan perbandingan dengan algoritma *FP-Growth* yang kerap digunakan untuk penelitian untuk menentukan prediksi pola pembelian.

Hasil dari penelitian yang berupa daftar barang yang dapat dijadikan

promo paket diskon menggunakan algoritma *ECLAT* dan *Fp-Growth* diharapkan bisa memberikan dampak terhadap volume penjualan dari *SPC Mart*. Selain itu, laporan penjualan berupa *dashboard* yang dibuat menggunakan *tableau public* diharapkan dapat membantu meningkatkan pemahaman pemilik *SPC Mart* terhadap keadaan penjualan berdasarkan data yang dimiliki.

### **1.2 Rumusan Masalah**

Masalah yang dapat dirumuskan dari latar belakang yang telah diuraikan adalah sebagai berikut.

1. Bagaimana merancang promosi berupa paket barang yang baik untuk meningkatkan efisiensi promosi sesuai dengan transaksi pelanggan.
2. Bagaimana perbandingan algoritma *ECLAT* dan *FP-Growth* dalam menerapkan association rule untuk menentukan Itemset dari transaksi yang ada.

### **1.3 Tujuan Penelitian**

Berdasarkan permasalahan yang telah dirumuskan pada rumusan masalah, terdapat tujuan yang ingin dicapai adalah.

1. Menghasilkan suatu laporan *descriptive analytics* berupa *dashboard* untuk membantu analisis terhadap data penjualan yang dimiliki oleh *SPC Mart* sehingga dapat digunakan pihak tersebut untuk memahami keadaan penjualan.
2. Memberikan suatu rekomendasi bundle item berdasarkan algoritma *ECLAT* dan *Fp-Growth* yang dapat dijadikan sebagai sarana untuk melakukan promosi.
3. Mengetahui bagaimana pembuatan pelaporan berupa visualisasi data untuk mengetahui penjualan dari *SPC Mart*

#### 1.4 Manfaat Penelitian

Manfaat yang diharapkan dalam pembuatan rancang bangun *sistem* adalah sebagai berikut. Bagi SPC Mart, laporan beserta rekomendasi bundle untuk promosi yang dibuat diharapkan dapat membantu memberikan solusi pemanfaatan teknologi informasi dalam menyimpan, mengelola, dan menganalisa lebih dalam suatu data yang dimiliki.

#### 1.5 Batasan Masalah

Penulisan tugas akhir ini memiliki beberapa batasan masalah antara lain sebagai berikut.

1. *Data Pre-processing* untuk mengubah format data sesuai dengan format yang diperlukan untuk implementasi algoritma *ECLAT* dan *FP-Growth* dilakukan dengan *Jupyter Notebook*.
2. Dalam implementasi algoritma *ECLAT* dan *FP-Growth*, penelitian ini dibantu dengan instrumen penelitian berupa *Rapidminer*
3. Visualisasi eksplorasi *dataset* yang ditampilkan menggunakan *Jupyter Notebook* / *Tableau Student* adalah visualisasi berupa jumlah barang terjual terbanyak per bulan dan secara keseluruhan rentang waktu yang dimiliki *dataset*, rata-rata jumlah transaksi perbulan dan daftar rekomendasi item berdasarkan algoritma *Eclat* dan *Fp-Growth*.
4. Perbandingan dari algoritma *ECLAT* dan *FP-Growth* akan dibandingkan berdasarkan hasil *rules* yang dihasilkan beserta lama proses eksekusi algoritma yang diperlukan untuk memberi output berupa *rules*.

### **1.6 Sistematika Penulisan**

Sistematika penulisan dari penelitian Tugas Akhir ini dapat dijelaskan sebagai berikut.

## **BAB I PENDAHULUAN**

Bab I berisikan pembahasan gambaran umum lewat tulisan yang membahas mengenai latar belakang, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan mengenai sistematika penulisan yang digunakan dalam penulisan laporan.

## **BAB II TINJAUAN PUSTAKA**

Bab II berisikan pembahasan mengenai beberapa teori-teori yang mendasari dalam penelitian dan penulisan tugas akhir ini. Teori-teori tersebut merupakan teori-teori pendukung dalam pembuatan laporan tugas akhir ini.

## **BAB III METODOLOGI DAN PERANCANGAN SISTEM**

Bab III berisikan pembahasan mengenai metode dan perancangan sistem, meliputi tempat dan waktu penelitian, sumber data, jenis data, metode pengumpulan data, metode pengolahan data, analisis yang digunakan pada penelitian, metode dan metodologi penelitian serta perancangan sistem *dashboard*.

## **BAB IV PEMBAHASAN DAN ANALISIS**

Bab IV berisikan pembahasan mengenai pembuatan *Dashboard* untuk *descriptive analytics* sederhana, analisa untuk menentukan *bundle* atau paket barang promosi menggunakan algoritma *ECLAT* dan *Fp-Growth* dan tahap perbandingan algoritma *ECLAT* dan *Fp-Growth*.

## **BAB V PENUTUP**

Bab V berisikan penutup dari penelitian yang telah dilakukan berupa kesimpulan dan saran. Kesimpulan diambil berdasarkan pada analisis objektif yang dilakukan dalam penelitian ini. Saran dicantumkan apabila terdapat jalan keluar untuk mengatasi masalah atau kelemahan yang ada, saran yang diberikan tidak terlepas dari ruang lingkup penelitian.



## BAB II

### TINJAUAN PUSTAKA

Bab II pada laporan membahas mengenai tinjauan Pustaka yang berisikan landasan teori secara umum, *State of the Art*, dan teori-teori penunjang dalam penelitian ini.

#### 2.1 State of The Art

*State of the Art* pada penelitian ini dipelajari dari artikel atau *paper* yang berkaitan dengan penelitian ini. Analisa mengambil acuan dan penelitian-penelitian yang sudah ada. Tujuannya adalah agar pengetahuan yang sudah ada dapat dikembangkan. Pengembangan dilakukan dengan menyimak saran perbaikan dari artikel atau *paper* mengenai data mining menggunakan *Association Rules* atau sejenis.

Penelitian yang menjadi bahan literatur dalam penelitian tugas akhir ini adalah penelitian mengenai *Data Mining Menggunakan Algoritma Apriori untuk Rekomendasi Produk bagi Pelanggan*. Hasil penelitian menunjukkan bahwa implementasi algoritma *Apriori* menggunakan aplikasi Weka menghasilkan 10 aturan dengan nilai *minimum support* 85% dan nilai *confidence* terendah sebesar 90%. Penelitian ini menyarankan pembaca untuk mencoba menggunakan algoritma lain seperti *FP-Growth* atau algoritma *hash-based* (Riszky & Sadikin, 2019). Penelitian mengenai penentuan pola yang sering muncul untuk penjualan pupuk menggunakan algoritma *FP-Growth*. Penelitian ini menggunakan data transaksi penjualan pupuk pada tanggal 02 Januari 2017–11 Januari 2017 dan aplikasi *RapidMiner* sebagai alat bantu penelitian. Hasil dari penelitian ini menyatakan bahwa algoritma ini menghasilkan 2 *rule* dengan nilai *support* tertinggi yaitu 11% dan nilai *confidence* tertinggi sebesar 92,3%. Penelitian ini menyarankan pembaca untuk menggunakan aplikasi lain sebagai alat bantu dan mencoba algoritma yang lain sebagai perbandingan (Firman, 2019).

**Commented [IMADW1]:** Ini coba dipastiin ya, kalau menurutku di italic semua soalnya untuk memperjelas materi yang diangkat.

**Commented [IMADW2]:** Menurutku lebih baik menggunakan sampai kalau tanggal penuh yang disebutkan. Kecuali kalau bahasanya 02-07 januari 2017

Penelitian berikutnya adalah penelitian **tentnagn** implementasi algoritma *Apriori* dan *FP-Growth* untuk menentukan persediaan barang. Penelitian ini didasari oleh permintaan konsumen yang terus meningkat secara otomatis juga harus diimbangi dengan teknologi untuk proses penjualan dan pelaporan hasil penjualannya. Hasil laporan penjualan hanya menjadi pemberitahuan untuk pihak manajemen dan tidak dilihat sebagai sumber data untuk pengembangan diwaktu yang akan datang, padahal hasil penjualan dapat digunakan sebagai tren untuk memprediksi permintaan barang oleh konsumen kedepannya. Hasil penelitian ini menyatakan secara keseluruhan dari data sampel penjualan diperoleh 152 *rule* yang terdiri dari 24 *rule* asosiasi yang memenuhi *support* dengan ambang batas 60% dan 108 *rule* yang memenuhi *confidence* 90% namun tidak ada menginformasikan mengenai perbandingan kedua algoritma yang digunakan (Junaidi, 2019). *Paper* yang berjudul Analisa Pola Frekuensi Keranjang Belanja Dengan Perbandingan Algoritma *Fp-Growth* (*Frequent Pattern Growth*) Dan *Eclat*. Penelitian dilakukan dengan menggunakan metode *Agile Software Development* (ASD) berjangka pendek, yang memerlukan adaptasi cepat dan pengembangan terhadap perubahan apapun. Penelitian dilakukan dengan mengumpulkan data dan studi awal atau kepustakaan demi memahami penerapan data mining tentang teori-teori yang berkaitan dengan data mining. Penelitian ini menyatakan *ECLAT association rules* yang memiliki waktu pencarian frequent item set lebih cepat di banding dengan Algoritma *Apriori* dengan teknik pendekatan BDF (*Breath deep First*) dengan melakukan *scan* pada *database* sebanyak 3(tiga) kali yang mentransformasikan tabel yang dapat mengurangi proses *input* dan *output* (Wijaya et al., 2020).

Commented [IMADW3]: typo

Penelitian berjudul Implementasi Algoritma **Apriori** dan Algoritma ECLAT pada AHASS Akmal Jaya Motor menggunakan metode CRISP-DM. CRISP-DM adalah suatu standarisasi pemrosesan data mining yang telah dikembangkan dimana data yang ada akan melewati setiap fase terstruktur dan terdefinisi dengan jelas dan efisien. Hasil dari penelitian ini menciptakan dua *rules* asosiasi dengan ketentuan *support* sebesar 0.1 dan *confidence* sebesar 0.5(Sulastri et al., 2017). Penelitian mengenai Asosiasi Data Mining Menggunakan Algoritma **Apriori dan FP-Growth** sebagai dasar pertimbangan penentuan paket sembako

Commented [IMADW4]: ini tadi nama algoritmanya diitalic tapi yang ini ngga

Commented [IMADW5]: ini sama

menggunakan metode yang sama yaitu CRISP-DM. Data yang digunakan adalah data dari 1805 transaksi Koperasi Karyawan Texmarco. Algoritma *Apriori* berhasil diimplementasikan untuk rekomendasi paket sembako. Hasilnya adalah, dengan *minimum support* 0,06 dan *confidence* 0,01 didapatkan 8 *rule* dengan waktu 0,03s dan dengan tingkat akurasi sebesar 35%. Kemudian Algoritma *FP-Growth* juga berhasil diimplementasikan. Perbandingan kedua algoritma yang digunakan dilakukan dengan perhitungan akurasi didasarkan pada jumlah *support* dari masing-masing algoritma (Maulidiya & Jananto, 2020).

## **2.2 Knowledge Discovery in Database (KDD)**

*Knowledge Discovery in Database (KDD)* adalah kegiatan yang meliputi pengumpulan, pemakaian data, historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar (Santosa, 2007). Langkah-langkah dalam *KDD* sebagai berikut (Fayyad *et al.*, 1996):

### **2.2.1 Data Selection**

Data Selection adalah proses untuk pemilihan data yang akan digunakan. Data yang ada pada *database* yang sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari *database*.

### **2.2.2 Preprocessing**

*Preprocessing* merupakan tahap mempersiapkan data, dilakukan dengan pembersihan data (*data cleaning*). *Data cleaning* mencakup antara lain membuang duplikasi data, memeriksa data yang tidak relevan, memeriksa ada tidaknya *missing data*, dan memperbaiki kesalahan pada data, seperti salah ketik. Membersihkan data yang tidak relevan dapat meningkatkan dan mengefisienkan teknik *data mining* yang digunakan.

### **2.2.3 Transformation**

Data diubah kedalam format yang sesuai untuk diproses dalam *data mining* dan sesuai dengan *software data mining* yang digunakan. Beberapa metode *data mining* membutuhkan format data yang khusus sebelum bisa diaplikasikan

(Huda, 2010). Beberapa *software data mining* hanya dapat membaca data dengan format yang khusus seperti \*.csv, \*.txt atau \*.arff. Oleh karena itu, apabila data yang didapatkan misalnya berupa data dengan format \*.xls, maka harus diubah ke \*.csv, \*.txt, atau \*.arff.

Commented [IMADW6]: ini spacingnya dibenerin

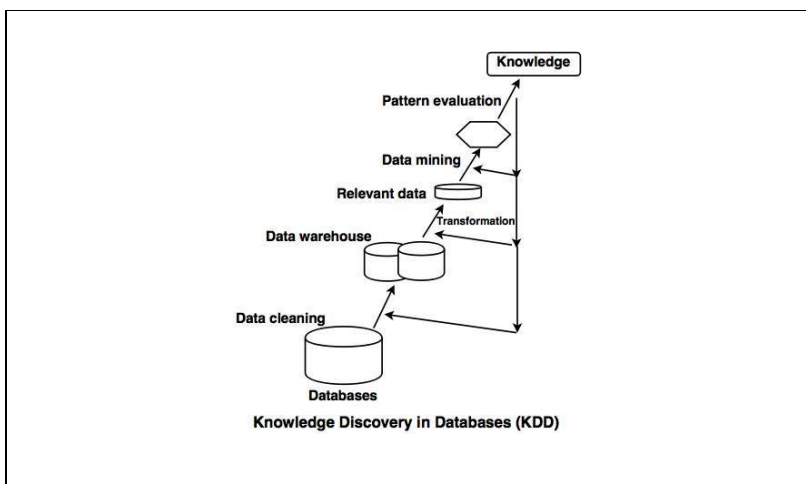
#### 2.2.4 Data Mining

Proses *data mining* yaitu proses mencari pola atau informasi menarik yang tersembunyi dalam data dengan menggunakan teknik dan algoritma tertentu. Data mining biasanya dibantu dengan algoritma dan tools tertentu.

Commented [IMADW7]: ini font sizenya

#### 2.2.5 Interpretasion/Evaluation

Menginterpretasikan pola-pola atau *patterns* menarik yang ditemukan ke dalam bentuk *knowledge* yang lebih mudah dimengerti oleh pihak yang berkepentingan. Berikut merupakan gambaran dari langkah-langkah dari *Knowledge Discovery in Database (KDD)*.



Gambar 2. 1 Gambaran Tahapan dalam Knowledge Discovery in Databases (KDD)

### 2.3. Data Mining

Menurut David Hand, Heikki Mannila, dan Padhraic Smyth dari MIT *Data Mining* adalah analisa terhadap data (biasanya data yang besar) untuk menemukan hubungan yang jelas serta menyimpulkan yang belum diketahui sebelumnya dan berguna bagi pemilik data tersebut. (Bramer, 2007).

Menurut Gartner Group, *data mining* adalah proses menemukan hubungan baru yang mempunyai arti, pola dan kebiasaan dengan memilah-milah Sebagian besar data yang disimpan dalam media penyimpanan dengan menggunakan teknologi pengenalan pola misalnya teknik *statistic* dan matematika.

*Data mining* adalah proses yang menggunakan teknik *statistic*, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar. (Pang-Ning Tan, 2005)

Dari ketiga definisi tersebut dapat diperoleh suatu kesimpulan bahwa data mining merupakan suatu proses yang melibatkan teknologi terhadap data yang sudah ada dan biasanya berukuran cukup besar dengan tujuan untuk mendapatkan hubungan atau pola yang mungkin memberikan indikasi yang bermanfaat.

Menurut Larose dalam bukunya yang berjudul “*Discovering Knowledge in Data: An Introduction to Data Mining*”, *data mining* dibagi menjadi beberapa kelompok berdasarkan tugas/pekerjaan yang dapat dilakukan, yaitu.

#### 2.3.1 Deskripsi (Descriptive)

Terkadang peneliti dan analisis secara sederhana ingin mencoba mencari cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data. Deskripsi dari pola kecenderungan sering memberikan kemungkinan penjelasan untuk suatu pola atau kecenderungan. Metode *data mining* yang satu ini bertujuan untuk memahami lebih dalam mengenai data yang masuk dalam pengamatan. Hasil akhirnya adalah mengetahui perilaku dari data itu sendiri.

Commented [IMADW8]: ini ngga kapital harusnya kan ?

Commented [IMADW9]: statistika aja pakai indonesia. Atau kurang italic

Commented [IMADW10]: ini spacingnya coba diformat lagi

### 2.3.2 Klasifikasi (Classification)

Klasifikasi *data mining* adalah sebuah proses menemukan definisi kesamaan karakteristik dalam suatu kelompok atau kelas (*class*). Klasifikasi *data mining* menjadi salah satu metode yang paling umum untuk digunakan. Metode ini dilakukan bertujuan untuk memperkirakan kelas dari suatu objek yang labelnya belum diketahui. Dalam klasifikasi, terdapat target variabel kategori. Sebagai contoh, penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu pendapatan tinggi, pendapatan sedang, dan pendapatan rendah.

### 2.3.3 Regression

*Regression* mirip dengan klasifikasi data mining. Pembedanya adalah, regresi merupakan metode yang bertujuan untuk mencari pola nilai numerik, bukan kelas. Hasil dari metode regression adalah sebuah fungsi sebagai penentu hasil yang didasarkan nilai dari *input*. Model Regresi dibangun menggunakan baris data (*record*) lengkap yang menyediakan nilai dari variabel target sebagai nilai prediksi. Selanjutnya, pada peninjauan berikutnya estimasi nilai dari variabel target dibuat berdasarkan nilai variabel prediksi. Secara umum, regresi digunakan untuk mengetahui pengaruh antara dua atau banyak variabel.

### 2.3.4 Prediksi (Forecasting)

*Forecasting* data mining adalah metode yang digunakan untuk memprediksi nilai yang akan dicapai pada satu periode. Menggunakan teknik ini membuat *noise* data dan nilai pada periode sebelumnya dijadikan dasar bahan prediksi. *Forecasting* hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam *forecasting* nilai dari hasil akan ada di masa mendatang. Beberapa metode dan teknik yang digunakan dalam klasifikasi dan estimasi dapat pula digunakan untuk *forecasting* dalam kondisi tertentu.

### 2.3.5 Pengklasteran (Clusterring)

Klaster adalah kumpulan *record* yang memiliki kemiripan satu dengan yang lainnya dan memiliki ketidakmiripan *record* dalam klaster yang lain.

Commented [IMADW11]: italic

*Clustering data mining* juga merupakan nama lain untuk metode *segmentation*. Tujuan dari segmentasi pada metode *data mining* adalah mengelompokkan suatu *class* ke dalam beberapa segmen berdasarkan atribut yang ditentukan. Penentuan atribut harus sesuai kesamaan yang dimiliki beberapa *class* tadi. Berbeda dengan klasifikasi, pada pengklasteran tidak ada variabel target. Pengklasteran tidak melakukan klasifikasi, mengestimasi, atau memprediksi nilai dari variabel target, akan tetapi, algoritma pengklasteran mencoba untuk melakukan pembagian terhadap keseluruhan data menjadi kelompok-kelompok yang memiliki kemiripan (homogen).

Commented [IMADW12]: italic

#### 2.3.6 Asosiasi (Association)

*Association Data Mining* adalah suatu metode untuk menemukan atribut yang muncul dalam satu waktu. Asosiasi sering digunakan untuk melakukan *Market Basket Analysis* dengan tujuan untuk mengidentifikasi produk yang sering dibeli bersamaan oleh pelanggan. Salah satu implementasi dari asosiasi adalah *market basket analysis* untuk menentukan suatu promo *bundling* yang efektif untuk pelanggan, sebagaimana yang akan dibahas dalam penelitian ini.

Commented [IMADW13]: italic

Commented [IMADW14]: ini italic dan kenapa dibedakan kapitalisasinya sama yang sebelumnya

#### 2.4. Aturan Asosiasi (Association Rules)

*Association rules* adalah salah satu *task data mining* deskriptif yang bertujuan untuk menemukan aturan asosiatif antara item-item data. Langkah utama yang perlu dalam *association rules* adalah mengetahui seberapa sering kombinasi item muncul dalam *database*, yang disebut sebagai *frequent patterns*. Pramudiono menyatakan bahwa penting tidaknya suatu aturan asosiatif dapat diketahui dengan dua parameter, *support* yaitu persentase kombinasi item dalam *database* dan *confidence* yaitu kuatnya hubungan antar item dalam aturan asosiatif (Aprianti et al., 2017). Bentuk umum dari association rule adalah *Antecedent* → *Consequent*. Contohnya adalah sebuah transaksi pembelian barang di dalam sebuah minimarket didapat bentuk *association rule* roti → selai. Artinya adalah pelanggan yang membeli roti ada kemungkinan pelanggan tersebut juga akan membeli selai, dimana

Commented [IMADW15]: italic

Commented [IMADW16]: italic

tidak ada batasan dalam jumlah item pada bagian *antecedent* ataupun *consequent* dalam sebuah *rule*.

terdapat suatu *interestingness measure* (ukuran kepercayaan) dalam menentukan suatu association rules. *interestingness measure* adalah salah satu parameter yang digunakan untuk perbandingan dalam data mining. Umumnya ada dua ukuran sebagai berikut.

#### 2.4.1 Support

Support adalah suatu ukuran yang menunjukkan seberapa besar tingkat dominasi suatu *item/itemset* dari keseluruhan transaksi. Ukuran ini menentukan apakah suatu *item/itemset* layak untuk dicari *confidence*-nya. Berikut merupakan persamaan dasar untuk mencari Support.

$$\text{Support} = \frac{\text{Jumlah transaksi A}}{\text{Jumlah transaksi}}$$

Sesuai dengan definisinya menunjukan seberapa tingkat dominasi item dari keseluruhan transaksi, maka persamaannya adalah jumlah transaksi yang mengandung *item A* dibagi dengan jumlah transaksi secara keseluruhan.

#### 2.4.2 Confidence

*Confidence* adalah parameter yang digunakan untuk mengetahui berapa kali pernyataan hubungan antara 2 *item* secara *conditional* tersebut benar. *Confidence* biasanya baru akan dihitung apabila nilai *support* sudah memenuhi kondisi. Berikut merupakan persamaan sederhana untuk mencari *confidence*.

$$\text{Confidence} = \frac{\text{Jumlah trans A + B}}{\text{Jumlah Trans A}}$$



Persamaan untuk mencari confidence adalah dengan cara membagi jumlah transaksi yang mengandung item A dan B dengan jumlah transaksi A. Contoh kasus untuk **support** dan **confidence** adalah sebagai berikut.

Commented [IMADW17]: italic

Commented [IMADW18]: italic

**Tabel 2. 1** Transaksi Awal

No	Barang
1	Gula, kopi
2	Permen, Marie, Gula
3	Sabun, Permen, Sikat Gigi
4	Permen, Snack, Marie
5	Air Mineral, Permen, Roti
6	Roti, Gula, Kopi, Kacang Kulit\
7	Roti, Permen
8	Pasta Gigi, Sikat Gigi
9	Roti, Permen, Snack
10	Snack, Permen, Marie
11	Susu, Kopi, Gula
12	Susu, Marie, Roti
13	Pasta Gigi, Sikat Gigi, Kopi
14	Permen, Snack
15	Kacang Kulit, Kopi, Gula
16	Gula Kopi, Snack
17	Snack, Susu, Gula
18	Kopi, Gula, Roti
19	Pasta Gigi, Gula, Susu
20	Snack, Kacang Kulit

Selanjutnya menentukan nilai **support** dengan menghitung jumlah transaksi yang mengandung item A dibagi dengan jumlah transaksi yang ada, maka didapat:

Commented [IMADW19]: italic

**Tabel 2. 2** Pencarian Nilai Support

No	Jenis Barang	Jml	Sup
1.	Gula, Kopi	5	25 %
2.	Gula,Permen	1	5 %
3.	Gula, Susu	4	20 %
4.	Permen. Marie	3	5 %
5.	Permen, Sikat gigi	1	5 %
6.	Permen, Kacang Kulit	1	5 %
7.	Kopi, Susu	1	5 %
8.	Marie, Susu	1	5 %
9.	Sikat gigi, pasta gigi	2	10 %
10.	Pasta gigi, susu	1	5 %

Setelah mendapat nilai **support** maka langkah selanjutnya adalah menentukan minimal nilai **support**. Misalkan *minimum* nilai **support** adalah 10% maka

Commented [IMADW20]: tab

selanjutnya adalah mencari nilai *confidence* dengan mencari banyaknya kemunculan item A dan B pada satu nota, maka didapat:

**Tabel 2. 3** Pencarian Nilai Confidence

no	barang	conf
1.	Gula, Kopi	5/8 = 62 %
2.	Gula, Susu	4/8 = 50 %
3.	Sikat gigi, Pasta Gigi	2/3 = 67 %

Nilai *confidence* sudah didapatkan, maka perbandingan itemset dapat dilakukan. Gula → Kopi menjadi *rules* yang memiliki nilai *support* terbesar dan *confidence* diatas 50% sehingga memungkinkan untuk diterapkan.

Beberapa perdebatan masih kerap terjadi mengenai dua parameter pengukuran yaitu *Support dan Confidence* walaupun memang sudah biasa digunakan dalam metode asosiasi. Sekitar 20 parameter pengukuran sempat diajukan melalui literatur, namun belum ada yang cocok untuk digunakan di segala kondisi sehingga untuk saat ini masih menggunakan parameter *support, confidence dan lift* (Merceron & Yacef, 2019).

## 2.5 Market Basket Analysis (MBA)

Menurut Alfagra et al, 2018 *Market Basket Analysis* merupakan salah satu proses analisis keranjang belanja dalam menentukan strategi pemasaran untuk memenuhi produk yang akan dibeli secara bersamaan oleh konsumen. *Data mining* dan model asosiasi khususnya dapat digunakan untuk mengidentifikasi produk-produk terkait yang biasanya dibeli bersamaan. MBA ini sendiri dapat digunakan dalam memahami kebiasaan konsumen dalam membeli produk yang dijual melalui transaksi belanja dengan analisis *asociation rule*. MBA bekerja dengan cara mengidentifikasi hubungan pada sekumpulan produk, item, maupun kategori. Produk-produk yang dibeli konsumen sering kali memiliki hubungan sehingga membentuk suatu hubungan jika-maka yang disebut dengan *rule if-then*.

Commented [IMADW21]: tab

Commented [IMADW22]: italic

Commented [IMADW23]: italic

Commented [IMADW24]: italic

## 2.6 Algoritma Equivalence Class Transformation (ECLAT)

Algoritma ECLAT (*Equivalence Class Transformation*) adalah suatu algoritma yang digunakan untuk melakukan *itemset mining*. Pencarian itemset untuk menemukan pola yang sering terjadi pada data seperti jika konsumen membeli roti dan juga membeli mentega. Jenis pola ini disebut aturan asosiasi dan digunakan di banyak domain aplikasi.

Menurut Dan Toomey (2014:42) Algoritma *eclat* digunakan untuk *frequent itemset mining*. Algoritma ini menggunakan persimpangan dalam data untuk menghitung dukungan kandidat. Dalam hal ini, kami mencari pola perilaku yang sama, dibandingkan dengan mencari penyimpangan. Kejadian yang sering terjadi bersamaan, seperti item keranjang belanja. Kandidat *frequent* kemudian diuji untuk mengkonfirmasi pola dalam dataset. *Eclat* dapat digunakan dalam pemrograman R dengan fungsi *eclat* dalam paket *arules*.

Commented [IMADW25]: italic

Proses pembentukan *itemset* pada algoritma ECLAT dimulai dengan mengubah bentuk transaksi, jika *item* pada transaksi berbentuk horizontal maka diubah menjadi bentuk vertikal dengan menggabungkan TID List pada transaksi yang memiliki *item* yang sama. Berikut merupakan contoh penerapan ECLAT.

Tabel 2. 4 Transaksi Awal

TID	Item
1	a, b
2	b, c, d
3	b, c, d, e
4	a, d, e
5	a, b, c

Tabel 2. 5 Transaksi Vertikal

TID	Item
a	(1,4,5)
b	(1,2,3,5)
c	(2,3,5)
d	(2,3,4)
e	(3,4)

Selanjutnya dilakukan penyilangan TID List dari kedua (k-1) subset. Model penyilangan dapat dilakukan dari atas kebawah. Contoh penyilangan dapat dilihat pada Tabel 2.6.

**Tabel 2. 6** Hasil Penyilangan 2-itemset

Itemset	TID List
(a,b)	(1,5)
(a,c)	(5)
(a,d)	(4)
(a,e)	(4)
(b,c)	(2,3,5)
(b,d)	(2,3)
(b,e)	(3)
(c,d)	(2,3)
(c,e)	(3)
(d,e)	(3,4)

Pada **Tabel 2.6** dapat dilihat bahwa *subset* item a disilangkan dengan *subset* item b, hasil penyilangan (a,b) didapat dari operasi konjungsi antara anggota *subset* a dan *subset* b, proses penyilangan terus dilakukan sampai tidak ada lagi *itemset* yang tersisa. Selanjutnya ditentukan *minimum support (minsupp)* dari setiap k-itemset. *Itemset* yang memiliki nilai kurang dari nilai *minsupp* akan dihilangkan. Sebagai contoh, ditentukan bahwa *minsupp* dari transaksi adalah 2, maka hasil transaksi terdapat pada Tabel 2.7.

**Tabel 2. 7** Hasil frequent 2-itemset

TID	Item
(a,b)	(1,5)
(b,c)	(2,3,5)
(b,d)	(2,3)
(c,d)	(2,3)
(d,e)	(3,4)

Pada Tabel 2.7 dapat dilihat bahwa semua transaksi memiliki jumlah *TIDList* kurang dari nilai *minsupp* telah dihilangkan. Sehingga didapat hasil *frequent* dari *2-itemset*. Selanjutnya apabila dilakukan penyilangan terhadap *3-itemset* dan seterusnya, penyilangan dilakukan dengan langkah yang sama dengan penyilangan sebelumnya dari atas kebawah. Setelah hasil dari *frequent itemset* ditemukan maka dilakukan perhitungan terhadap nilai *support* dan nilai *confidence* dari masing-masing *itemset* untuk menentukan ukuran seberapa valid *association rule* yang ditemukan.

Tabel 2. 8 Nilai Support dan Confidence Itemset

Itemset	Support (%)	Confidence (%)
(a→b)	0,4	0,6667
(b→c)	0,6	0,75
(b→d)	0,4	0,5
(c→d)	0,4	0,6667
(c→d)	0,4	0,6667

Pada Tabel 2.8 merupakan nilai *support* dan *confidence* dari *frequent 2-itemset*. Pada aturan asosiasi (a→b) diketahui bahwa nilai *support* atau nilai transaksi yang mengandung item a dan b dari total transaksi adalah 0,4%, sedangkan nilai *confidence* semua transaksi yang mengandung a dan b dari semua transaksi yang mengandung a dari aturan asosiasi (a→b) adalah 0,66667%. Pada Tabel 2.8 juga dapat dilihat bahwa aturan asosiasi yang mempunyai nilai *support* dan *confidence* tertinggi adalah (b→c).

2.7     **Algoritma Frequent Pattern Growth (FP-Growth)**

Algoritma *FP-growth* adalah suatu algoritma yang mengkodekan data menggunakan struktur data yang disebut *FP-tree* dan mengekstrak *frequent itemset* dari *FP-tree* tersebut. *FP-tree* adalah representasi terkompresi dari data keranjang belanja. *FP-tree* dibangun dengan membaca *dataset* transaksi per satu transaksi dan melakukan *mapping item* tiap transaksi tersebut ke sebuah *path* pada *FP-tree*

Commented [IMADW26]: italic

Commented [IMADW27]: italic

Commented [IMADW28]: italic

Commented [IMADW29]: italic

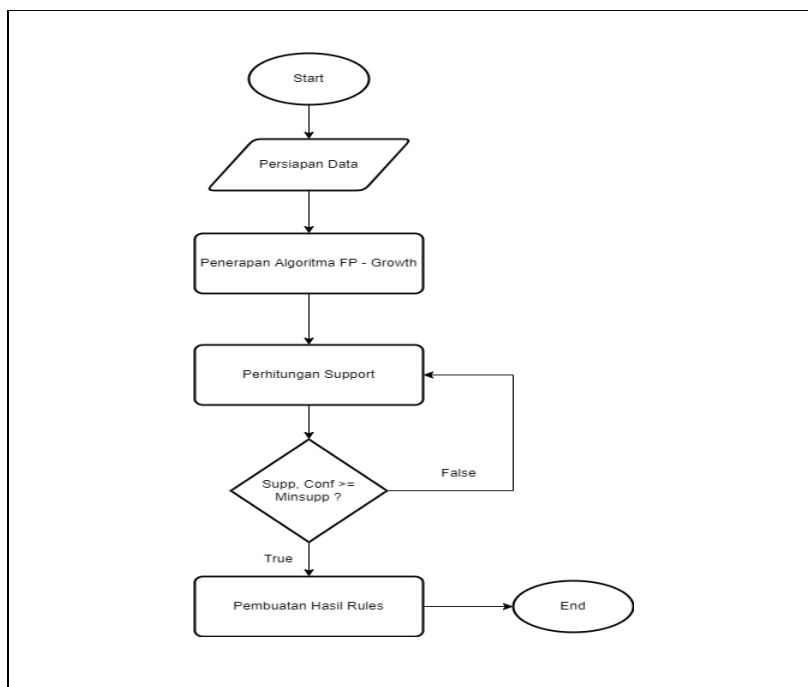
(Indrawan et al., 2018). Karakteristik algoritma *Fp-Growth* adalah struktur dari data digunakan adalah *tree* atau disebut dengan *Fp-tree*. Dengan menggunakan *Fp-Tree*, algoritma ini dapat langsung mengekstrak frequent itemset dari *Fp-Tree*. Semakin banyak data transaksi yang memiliki item yang sama, maka proses pemanfaatan dengan struktur data *FP-tree* semakin efektif. Kelebihan dari *FP-tree* adalah hanya memerlukan dua kali pemindaian data. Adapun gambaran tahapan metode *FP-Growth* adalah sebagai berikut.

Commented [IMADW30]: italic

Commented [IMADW31]: italic

Commented [IMADW32]: italic

Commented [IMADW33]: italic



**Gambar 2. 2** Flowchart Tahapan Algoritma FP-Growth

**Gambar 2.2** merupakan diagram alur yang digunakan dalam menerapkan FP-Growth, antara lain melakukan persiapan data dengan memindai data transaksi penjualan memilah atribut dan mengilangkan data yang bersifat redundansi. Proses selanjutnya melakukan perhitungan *Support* dan *confidence* untuk menghasilkan aturan dari algoritma FP-Growth.

## 2.8 Lift Ratio

Lift Ratio adalah suatu parameter yang digunakan untuk mengetahui kekuatan pengaruh aturan asosiasi yang telah terbentuk dari nilai support dan confidence. Nilai lift ratio biasanya digunakan sebagai penentu apakah aturan asosiasi valid atau tidak valid. Nilai lift ratio dapat diperoleh dengan rumus pada gambar berikut.

Commented [IMADW34]: italic

$$Lift = \frac{Support}{Support(X) * Support(Y)}$$

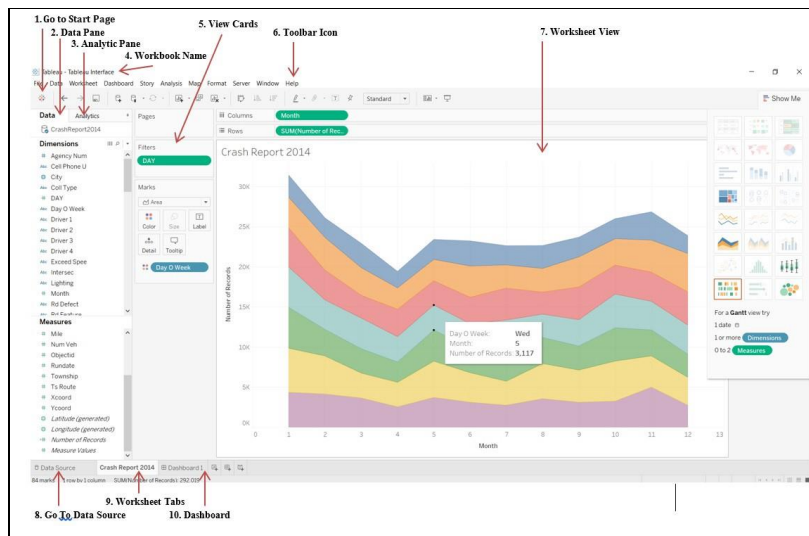
Gambar 2. 3 Rumus Perhitungan Lift Ratio

Pada Gambar 2.3 dapat diketahui bahwa cara mengetahui nilai *lift ratio* adalah dengan membagi nilai *support* dari suatu *rule* dengan Gabungan dari *support itemset* A dengan *Itemset* B. Korelasi akan dinyatakan negatif apabila hasil dari perhitungan *lift ratio* < 1 atau dengan kata lain, jika konsumen membeli barang A maka ia tidak akan membeli barang B. *Itemset* A dan *itemset* B dinyatakan memiliki korelasi positif apabila *lift ratio* > 1 yang berarti, jika konsumen membeli barang A maka ia juga akan membeli barang B. *Itemset* A dan *itemset* B dinyatakan independen satu sama lain apabila *lift ratio* = 1, yang berarti jika konsumen membeli barang A maka ia belum tentu membeli barang B.

## 2.9 Tableau

Tableau adalah sebuah *tools* yang dapat mempermudah pembuatan analisis visual interaktif dalam bentuk *dashboard*. Tableau banyak digunakan untuk memvisualisasikan data dan menyederhanakan pola analisis *business intelligence* (BI) karena Tableau menyediakan *dashboard* dan *scorecards*, *ad hoc analysis and queries*, pemrosesan analitik *online*, penemuan data, pencarian BI,

integrasi *spreadsheet*, dan lainnya. Berikut merupakan tampilan dari Tableau Public.



Gambar 2. 4 Tampilan Sederhana Tableau Public

Gambar 2.4 merupakan tampilan sederhana dari *tableau public*. *Tabelau Public* adalah versi yang dapat diakses secara gratis dari *Tableau Desktop*. *Tableau Public* juga memungkinkan pengguna untuk membagikan *dashboard* yang telah dibuat sehingga memungkinkan pengguna lainnya untuk mengaksesnya.

## 2.10 Jupyter Notebook

**Jupyter Notebook** merupakan singkatan dari tiga bahasa pemrograman, yakni *Julia* (Ju), *Python* (Py), dan *R*. Jupyter Notebook adalah sebuah aplikasi *web* gratis yang paling banyak dipakai oleh *data scientist*. Aplikasi ini dipakai untuk membuat dan membagikan dokumen yang memiliki kode, hasil hitungan, visualisasi, dan teks melalui *google collaboratory*. Ketiga bahasa pemrograman pada Jupyter Notebook sendiri adalah sesuatu yang penting bagi seorang *data scientist*.

**Sederhananya**, Jupyter Notebook berfungsi membantu *data scientist* dalam membuat narasi komputasi. Salah satunya adalah melakukan *data pre-*

**Commented [IMADW35]:** ini kalau aku sih di italic juga saat ini, coba pastiin lagi

**Commented [IMADW36]:** tab



*processing*. Aplikasi ini juga memudahkan *data scientist* melakukan kolaborasi dengan sesama *data scientist*, *data researchers*, maupun *data engineers* lainnya.

**“PENERAPAN ALGORITMA *ECLAT* DAN *FP- GROWTH* SEBAGAI  
DASAR PERTIMBANGAN PENENTUAN PAKET PROMO DI TOKO SPC  
MART”**

**Tugas Akhir**

**Diajukan untuk Memenuhi Sebagian Persyaratan Program Strata I Program  
Studi Teknologi Informasi**



**Oleh:**

**I Putu Kevin Ari Narayana**

**1905551030**

**Reviewer:**

**Ellyas Immanuel Sinaga**

**(1905551055)**

**PROGRAM STUDI TEKNOLOGI INFORMASI**

**FAKULTAS TEKNIK**

**UNIVERSITAS UDAYANA**

**2022**

## **BAB I**

### **PENDAHULUAN**

Bab I Pendahuluan menjelaskan proses dalam mengidentifikasi masalah penelitian. Pembahasan dalam bab ini diantaranya adalah latar belakang masalah, perumusan masalah, tujuan penelitian, dan manfaat kegiatan penelitian terhadap solusi dari permasalahan yang diangkat, batasan masalah serta sistematika penulisan dalam penyusunan laporan Tugas Akhir.

#### **1.1 Latar Belakang**

Badan Pusat Statistik menyebutkan bahwa *Covid-19* menyebabkan terjadinya penurunan ekonomi di Indonesia sebesar 2,07% dan sebanyak 87,5% UMKM di Indonesia terdampak oleh pandemi ini. Penurunan ekonomi ini menimbulkan persaingan bisnis yang semakin ketat dan mengharuskan para pemilik bisnis untuk berpikir kritis agar kepuasan pelanggan tetap terjaga serta penjualan produk tetap meningkat. Kemampuan untuk memperkirakan volume penjualan secara akurat sangat penting guna kelangsungan hidup bisnis. Peningkatan penjualan produk juga dapat dilakukan dengan strategi promosi yang efisien dan efektif.

Pengertian promosi menurut Kotler & Keller (2009) "*promotion means activities that communicate the merits of the product and persuade target customers to buy it*", yang artinya promosi merupakan kegiatan yang mengkomunikasikan manfaat dari sebuah produk dan membujuk target konsumen untuk membeli produk tersebut. Menurut Putri & Safri (2015) Promosi merupakan salah satu komponen dari bauran pemasaran (*marketing mix*) memiliki peranan penting dalam mengkomunikasikan suatu produk, dan dapat menciptakan preferensi konsumen atau calon konsumen mengenai keefektifan dan keefisienan dari bauran promosi (*promotion mix*) yang digunakan. Bauran promosi yang digunakan oleh setiap perusahaan tidak sama, hal ini tergantung dari kondisi perusahaan.

Promosi yang tepat akan memengaruhi keberlanjutan jangka panjang dari penjualan. Dengan mengetahui pola pembelian pelanggan, manajemen dapat menentukan strategi penjualan terkait dengan promosi, prediksi persediaan produk, dan kombinasi produk yang akan dijual.

Toko SPC MART adalah toko di Sidakarya yang menjual kebutuhan sehari-hari. Masalah yang dihadapi oleh pemilik toko adalah pencatatan persediaan barang dan penjualan yang masih dilakukan secara semi-manual, serta belum ada analisis terhadap data penjualan secara menyeluruh. Hal ini memberikan beberapa masalah, salah satunya adalah pemberian promo bagi pelanggan. Karena proses bisnis dilakukan secara manual, maka pemilik toko kesulitan dalam menganalisis barang apa saja yang laku terjual dengan cepat dan barang apa saja yang dibeli bersamaan.

Proses pemberian promo masih dilakukan dengan cara berdiskusi dengan pihak internal. Perkiraan pembelian barang valid dan pemberian promo tidak disarankan hanya berdasarkan intuisi, maka diperlukan suatu data yang mendukung pengambilan keputusan. Meningkatnya jumlah pelanggan dan aktivitas pembelian akan menghasilkan sejumlah besar data historis yang berharga. Data historis transaksi yang telah terakumulasi ini dapat dimanfaatkan untuk membantu dalam menemukan hubungan pembelian antar berbagai barang. Data yang dimiliki akan membantu mempelajari pelanggan dan mencari kombinasi barang yang sering dibeli bersamaan. Kemudian barang-barang tersebut ditawarkan sebagai rekomendasi kepada pelanggan.

Rekomendasi Promo ini perlu diterapkan pada data historis penjualan dari SPC Mart. Pembuatan produk rekomendasi yang paling populer menggunakan algoritma *Association Rule*. *Association rules* adalah salah satu task data mining deskriptif yang bertujuan untuk menemukan aturan asosiasi antara item-item data. Langkah utama yang perlu dalam *association rules* adalah mengetahui seberapa sering kombinasi item muncul dalam database, yang disebut sebagai *frequent patterns* (Han, 2011)

Penelitian ini dilakukan menggunakan algoritma *Eclat* dan *FP-Growth* untuk mencari pola hubungan antar item. Berdasarkan hasil penelitian sebelumnya, algoritma *FP-Growth* mampu menggali informasi pada data dan dapat menampilkan hubungan antar atributnya. Algoritma *FP-Growth* ini sudah banyak diterapkan dalam berbagai penelitian sebelumnya, digunakan untuk memprediksi pola pembelian obat pada apotek, pola belanja pelanggan pada toko buku dan minimarket.

Secara singkat, Algoritma *Frequent Pattern Growth (FP-Growth)* adalah salah satu alternatif algoritma yang dapat digunakan untuk menentukan himpunan data yang paling sering muncul (*frequent itemset*) dalam sebuah kumpulan data. Dalam Algoritma *Frequent Pattern Growth (FP-Growth)* menyimpan informasi mengenai frequent itemset dalam bentuk struktur prefix-tree atau sering disebut FP-Tree. Tidak melakukan candidate generation dalam proses pencarian frequent itemset sehingga dapat mengurangi scan database secara berulang dalam proses mining dan dapat berlangsung lebih cepat. (Jiawei, dkk: 2012)

Algoritma *ECLAT* merupakan algoritma yang berfungsi hampir serupa dengan *Fp-Growth*. *ECLAT* algoritma untuk menemukan itemset yang sering muncul. Pada dasarnya algoritma *ECLAT* melakukan pencarian secara depth-first search pada database dengan tata letak vertikal. Algoritma *ECLAT* ini masih jarang diterapkan dalam berbagai penelitian sebelumnya, oleh sebab itu ada baiknya apabila dijadikan perbandingan dengan algoritma *FP-Growth* yang kerap digunakan untuk penelitian untuk menentukan prediksi pola pembelian.

Hasil dari penelitian yang berupa daftar barang yang dapat dijadikan

promo paket diskon menggunakan algoritma *ECLAT* dan *Fp-Growth* diharapkan bisa memberikan dampak terhadap volume penjualan dari *SPC Mart*. Selain itu, laporan penjualan berupa *dashboard* yang dibuat menggunakan *tableau public* diharapkan dapat membantu meningkatkan pemahaman pemilik *SPC Mart* terhadap keadaan penjualan berdasarkan data yang dimiliki.

### **1.2 Rumusan Masalah**

Masalah yang dapat dirumuskan dari latar belakang yang telah diuraikan adalah sebagai berikut.

1. Bagaimana merancang promosi berupa paket barang yang baik untuk meningkatkan efisiensi promosi sesuai dengan transaksi pelanggan.
2. Bagaimana perbandingan algoritma *ECLAT* dan *FP-Growth* dalam menerapkan association rule untuk menentukan Itemset dari transaksi yang ada.

### **1.3 Tujuan Penelitian**

Berdasarkan permasalahan yang telah dirumuskan pada rumusan masalah, terdapat tujuan yang ingin dicapai adalah.

1. Menghasilkan suatu laporan *descriptive analytics* berupa *dashboard* untuk membantu analisis terhadap data penjualan yang dimiliki oleh *SPC Mart* sehingga dapat digunakan pihak tersebut untuk memahami keadaan penjualan.
2. Memberikan suatu rekomendasi bundle item berdasarkan algoritma *ECLAT* dan *Fp-Growth* yang dapat dijadikan sebagai sarana untuk melakukan promosi.
3. Mengetahui bagaimana pembuatan pelaporan berupa visualisasi data untuk mengetahui penjualan dari *SPC Mart*

#### 1.4 Manfaat Penelitian

Manfaat yang diharapkan dalam pembuatan rancang bangun *sistem* adalah sebagai berikut. Bagi SPC Mart, laporan beserta rekomendasi bundle untuk promosi yang dibuat diharapkan dapat membantu memberikan solusi pemanfaatan teknologi informasi dalam menyimpan, mengelola, dan menganalisa lebih dalam suatu data yang dimiliki.

#### 1.5 Batasan Masalah

Penulisan tugas akhir ini memiliki beberapa batasan masalah antara lain sebagai berikut.

1. *Data Pre-processing* untuk mengubah format data sesuai dengan format yang diperlukan untuk implementasi algoritma *ECLAT* dan *FP-Growth* dilakukan dengan *Jupyter Notebook*.
2. Dalam implementasi algoritma *ECLAT* dan *FP-Growth*, penelitian ini dibantu dengan instrumen penelitian berupa *Rapidminer*
3. Visualisasi eksplorasi *dataset* yang ditampilkan menggunakan *Jupyter Notebook* / *Tableau Student* adalah visualisasi berupa jumlah barang terjual terbanyak per bulan dan secara keseluruhan rentang waktu yang dimiliki *dataset*, rata-rata jumlah transaksi perbulan dan daftar rekomendasi item berdasarkan algoritma *Eclat* dan *Fp-Growth*.
4. Perbandingan dari algoritma *ECLAT* dan *FP-Growth* akan dibandingkan berdasarkan hasil *rules* yang dihasilkan beserta lama proses eksekusi algoritma yang diperlukan untuk memberi output berupa *rules*.

### **1.6 Sistematika Penulisan**

Sistematika penulisan dari penelitian Tugas Akhir ini dapat dijelaskan sebagai berikut.

## **BAB I PENDAHULUAN**

Bab I berisikan pembahasan gambaran umum lewat tulisan yang membahas mengenai latar belakang, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan mengenai sistematika penulisan yang digunakan dalam penulisan laporan.

## **BAB II TINJAUAN PUSTAKA**

Bab II berisikan pembahasan mengenai beberapa teori-teori yang mendasari dalam penelitian dan penulisan tugas akhir ini. Teori-teori tersebut merupakan teori-teori pendukung dalam pembuatan laporan tugas akhir ini.

## **BAB III METODOLOGI DAN PERANCANGAN SISTEM**

Bab III berisikan pembahasan mengenai metode dan perancangan sistem, meliputi tempat dan waktu penelitian, sumber data, jenis data, metode pengumpulan data, metode pengolahan data, analisis yang digunakan pada penelitian, metode dan metodologi penelitian serta perancangan sistem *dashboard*.



## **BAB IV PEMBAHASAN DAN ANALISIS**

Bab IV berisikan pembahasan mengenai pembuatan *Dashboard* untuk *descriptive analytics* sederhana, analisa untuk menentukan *bundle* atau paket barang promosi menggunakan algoritma *ECLAT* dan *Fp-Growth* dan tahap perbandingan algoritma *ECLAT* dan *Fp-Growth*.

## **BAB V PENUTUP**

Bab V berisikan penutup dari penelitian yang telah dilakukan berupa kesimpulan dan saran. Kesimpulan diambil berdasarkan pada analisis objektif yang dilakukan dalam penelitian ini. Saran dicantumkan apabila terdapat jalan keluar untuk mengatasi masalah atau kelemahan yang ada, saran yang diberikan tidak terlepas dari ruang lingkup penelitian.

## **BAB II**

### **TINJAUAN PUSTAKA**

Bab II pada laporan membahas mengenai tinjauan Pustaka yang berisikan landasan teori secara umum, *State of the Art*, dan teori-teori penunjang dalam penelitian ini.

#### **2.1 State of The Art**

*State of the Art* pada penelitian ini dipelajari dari artikel atau *paper* yang berkaitan dengan penelitian ini. Analisa mengambil acuan dan penelitian-penelitian yang sudah ada. Tujuannya adalah agar pengetahuan yang sudah ada dapat dikembangkan. Pengembangan dilakukan dengan menyimak saran perbaikan dari artikel atau *paper* mengenai data mining menggunakan *Association Rules* atau sejenis.

Penelitian yang menjadi bahan literatur dalam penelitian tugas akhir ini adalah penelitian mengenai *Data Mining* Menggunakan Algoritma *Apriori* untuk Rekomendasi Produk bagi Pelanggan. Hasil penelitian menunjukkan bahwa implementasi algoritma *Apriori* menggunakan aplikasi Weka menghasilkan 10 aturan dengan nilai *minimum support* 85% dan nilai *confidence* terendah sebesar 90%. Penelitian ini menyarankan pembaca untuk mencoba menggunakan algoritma lain seperti *FP-Growth* atau algoritma *hash-based* (Riszky & Sadikin, 2019). Penelitian mengenai penentuan pola yang sering muncul untuk penjualan pupuk menggunakan algoritma *FP-Growth*. Penelitian ini menggunakan data transaksi penjualan pupuk pada tanggal 02 Januari 2017–11 Januari 2017 dan aplikasi *RapidMiner* sebagai alat bantu penelitian. Hasil dari penelitian ini menyatakan bahwa algoritma ini menghasilkan 2 *rule* dengan nilai *support* tertinggi yaitu 11% dan nilai *confidence* tertinggi sebesar 92,3%. Penelitian ini menyarankan pembaca untuk menggunakan aplikasi lain sebagai alat bantu dan mencoba algoritma yang lain sebagai perbandingan (Firman, 2019).

Penelitian berikutnya adalah penelitian tentang implementasi algoritma *Apriori* dan *FP-Growth* untuk menentukan persediaan barang. Penelitian ini didasari oleh permintaan konsumen yang terus meningkat secara otomatis juga harus diimbangi dengan teknologi untuk proses penjualan dan pelaporan hasil penjualannya. Hasil laporan penjualan hanya menjadi pemberitahuan untuk pihak manajemen dan tidak dilihat sebagai sumber data untuk pengembangan di waktu yang akan datang, padahal hasil penjualan dapat digunakan sebagai tren untuk memprediksi permintaan barang oleh konsumen kedepannya. Hasil penelitian ini menyatakan secara keseluruhan dari data sampel penjualan diperoleh 152 *rule* yang terdiri dari 24 *rule* asosiasi yang memenuhi *support* dengan ambang batas 60% dan 108 *rule* yang memenuhi *confidence* 90% namun tidak ada menginformasikan mengenai perbandingan kedua algoritma yang digunakan (Junaidi, 2019). *Paper* yang berjudul Analisa Pola Frekuensi Keranjang Belanja Dengan Perbandingan Algoritma *Fp-Growth* (*Frequent Pattern Growth*) Dan *Eclat*. Penelitian dilakukan dengan menggunakan metode *Agile Software Development* (ASD) berjangka pendek, yang memerlukan adaptasi cepat dan pengembangan terhadap perubahan apapun. Penelitian dilakukan dengan mengumpulkan data dan studi awal atau kepustakaan demi memahami penerapan data mining tentang teori-teori yang berkaitan dengan data mining. Penelitian ini menyatakan *ECLAT association rules* yang memiliki waktu pencarian frequent item set lebih cepat dibanding dengan Algoritma *Apriori* dengan teknik pendekatan BDF (*Breath deep First*) dengan melakukan *scan* pada *database* sebanyak 3(tiga) kali yang mentransformasikan tabel yang dapat mengurangi proses *input* dan *output* (Wijaya et al., 2020).

Commented [EIS1]: Remove space after paragraph

Penelitian berjudul Implementasi Algoritma *Apriori* dan Algoritma *ECLAT* pada AHASS Akmal Jaya Motor menggunakan metode *CRISP-DM*. *CRISP-DM* adalah suatu standarisasi pemrosesan data mining yang telah dikembangkan dimana data yang ada akan melewati setiap fase terstruktur dan terdefinisi dengan jelas dan efisien. Hasil dari penelitian ini menciptakan dua *rules* asosiasi dengan ketentuan *support* sebesar 0.1 dan *confidence* sebesar 0.5 (Sulastris et al., 2017). Penelitian mengenai Asosiasi Data Mining Menggunakan Algoritma *Apriori* dan *FP-Growth* sebagai dasar pertimbangan penentuan paket sembako

menggunakan metode yang sama yaitu CRISP-DM. Data yang digunakan adalah data dari 1805 transaksi Koperasi Karyawan Texmarco. Algoritma *Apriori* berhasil diimplementasikan untuk rekomendasi paket sembako. Hasilnya adalah, dengan *minimum support* 0,06 dan *confidence* 0,01 didapatkan 8 *rule* dengan waktu 0,03s dan dengan tingkat akurasi sebesar 35%. Kemudian Algoritma *FP-Growth* juga berhasil diimplementasikan. Perbandingan kedua algoritma yang digunakan dilakukan dengan perhitungan akurasi didasarkan pada jumlah *support* dari masing-masing algoritma (Maulidiya & Jananto, 2020).

## **2.2 Knowledge Discovery in Database (KDD)**

*Knowledge Discovery in Database (KDD)* adalah kegiatan yang meliputi pengumpulan, pemakaian data, historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar (Santosa, 2007). Langkah-langkah dalam *KDD* sebagai berikut (Fayyad *et al.*, 1996):

### **2.2.1 Data Selection**

Data Selection adalah proses untuk pemilihan data yang akan digunakan. Data yang ada pada *database* yang sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari *database*.

### **2.2.2 Preprocessing**

*Preprocessing* merupakan tahap mempersiapkan data, dilakukan dengan pembersihan data (*data cleaning*). *Data cleaning* mencakup antara lain membuang duplikasi data, memeriksa data yang tidak relevan, memeriksa ada tidaknya *missing data*, dan memperbaiki kesalahan pada data, seperti salah ketik. Membersihkan data yang tidak relevan dapat meningkatkan dan mengefisienkan teknik *data mining* yang digunakan.

### **2.2.3 Transformation**

Data diubah kedalam format yang sesuai untuk diproses dalam *data mining* dan sesuai dengan *software data mining* yang digunakan. Beberapa metode *data mining* membutuhkan format data yang khusus sebelum bisa diaplikasikan

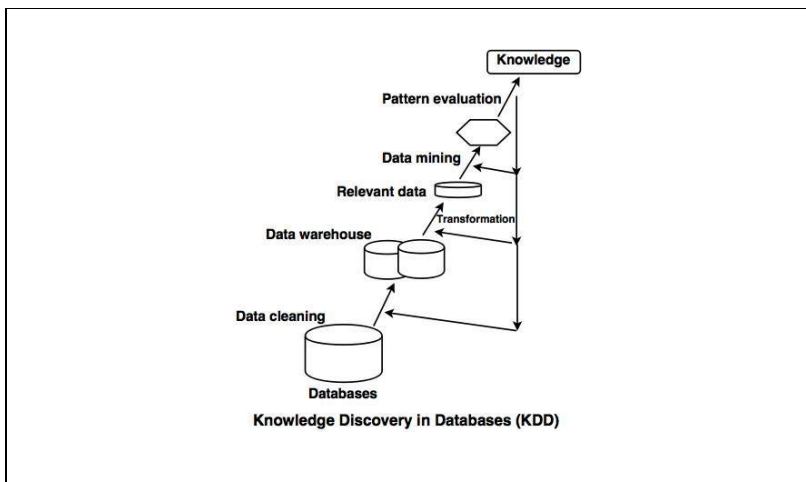
(Huda, 2010). Beberapa *software data mining* hanya dapat membaca data dengan format yang khusus seperti \*.csv, \*.txt atau \*.arff. Oleh karena itu, apabila data yang didapatkan misalnya berupa data dengan format \*.xls, maka harus diubah ke \*.csv, \*.txt, atau \*.arff.

#### 2.2.4 Data Mining

Proses *data mining* yaitu proses mencari pola atau informasi menarik yang tersembunyi dalam data dengan menggunakan teknik dan algoritma tertentu. Data mining biasanya dibantu dengan algoritma dan tools tertentu.

#### 2.2.5 Interpretation/Evaluation

Menginterpretasikan pola-pola atau *patterns* menarik yang ditemukan ke dalam bentuk *knowledge* yang lebih mudah dimengerti oleh pihak yang berkepentingan. Berikut merupakan gambaran dari langkah-langkah dari *Knowledge Discovery in Database (KDD)*.



**Gambar 2. 1** Gambaran Tahapan dalam Knowledge Discovery in Databases (KDD)

**Commented [EIS2]:** Jangan ada spasi, “Gambar 2.1”.  
Aplikasikan ke caption lainnya juga

### 2.3. Data Mining

Menurut David Hand, Heikki Mannila, dan Padhraic Smyth dari MIT *Data Mining* adalah analisa terhadap data (biasanya data yang besar) untuk menemukan hubungan yang jelas serta menyimpulkan yang belum diketahui sebelumnya dan berguna bagi pemilik data tersebut. (Bramer, 2007).

Commented [EIS3]: Remove space after paragraph

Menurut Gartner Group, *data mining* adalah proses menemukan hubungan baru yang mempunyai arti, pola dan kebiasaan dengan memilah-milah Sebagian besar data yang disimpan dalam media penyimpanan dengan menggunakan teknologi pengenalan pola misalnya teknik statistic dan matematika.

Commented [EIS4]: sama

*Data mining* adalah proses yang menggunakan teknik *statistic*, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar. (Pang-Ning Tan, 2005)

Commented [EIS5]: sama

Dari ketiga definisi tersebut dapat diperoleh suatu kesimpulan bahwa data mining merupakan suatu proses yang melibatkan teknologi terhadap data yang sudah ada dan biasanya berukuran cukup besar dengan tujuan untuk mendapatkan hubungan atau pola yang mungkin memberikan indikasi yang bermanfaat.

Commented [EIS6]: sama

Menurut Larose dalam bukunya yang berjudul “*Discovering Knowledge in Data: An Introduction to Data Mining*”, *data mining* dibagi menjadi beberapa kelompok berdasarkan tugas/pekerjaan yang dapat dilakukan, yaitu.

#### 2.3.1 Deskripsi (Descriptive)

Terkadang peneliti dan analisis secara sederhana ingin mencoba mencari cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data. Deskripsi dari pola kecenderungan sering memberikan kemungkinan penjelasan untuk suatu pola atau kecenderungan. Metode *data mining* yang satu ini bertujuan untuk memahami lebih dalam mengenai data yang masuk dalam pengamatan. Hasil akhirnya adalah mengetahui perilaku dari data itu sendiri.

### 2.3.2 Klasifikasi (Classification)

Klasifikasi *data mining* adalah sebuah proses menemukan definisi kesamaan karakteristik dalam suatu kelompok atau kelas (*class*). Klasifikasi *data mining* menjadi salah satu metode yang paling umum untuk digunakan. Metode ini dilakukan bertujuan untuk memperkirakan kelas dari suatu objek yang labelnya belum diketahui. Dalam klasifikasi, terdapat target variabel kategori. Sebagai contoh, penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu pendapatan tinggi, pendapatan sedang, dan pendapatan rendah.

### 2.3.3 Regression

*Regression* mirip dengan klasifikasi *data mining*. Pembedanya adalah, regresi merupakan metode yang bertujuan untuk mencari pola nilai numerik, bukan kelas. Hasil dari metode regression adalah sebuah fungsi sebagai penentu hasil yang didasarkan nilai dari *input*. Model *Regresi* dibangun menggunakan baris data (*record*) lengkap yang menyediakan nilai dari variabel target sebagai nilai prediksi. Selanjutnya, pada peninjauan berikutnya estimasi nilai dari variabel target dibuat berdasarkan nilai variabel prediksi. Secara umum, regresi digunakan untuk mengetahui pengaruh antara dua atau banyak variabel.

Commented [EIS7]: italic

Commented [EIS8]: tidak perlu italic

### 2.3.4 Prediksi (Forecasting)

*Forecasting* data mining adalah metode yang digunakan untuk memprediksi nilai yang akan dicapai pada satu periode. Menggunakan teknik ini membuat *noise* data dan nilai pada periode sebelumnya dijadikan dasar bahan prediksi. *Forecasting* hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam *forecasting* nilai dari hasil akan ada di masa mendatang. Beberapa metode dan teknik yang digunakan dalam klasifikasi dan estimasi dapat pula digunakan untuk *forecasting* dalam kondisi tertentu.

### 2.3.5 Pengklasteran (Clusterring)

Klaster adalah kumpulan record yang memiliki kemiripan satu dengan yang lainnya dan memiliki ketidakmiripan record dalam klaster yang lain.

*Clustering data mining* juga merupakan nama lain untuk metode *segmentation*. Tujuan dari segmentasi pada metode *data mining* adalah mengelompokkan suatu *class* ke dalam beberapa segmen berdasarkan atribut yang ditentukan. Penentuan atribut harus sesuai kesamaan yang dimiliki beberapa *class* tadi. Berbeda dengan klasifikasi, pada pengklasteran tidak ada variabel target. Pengklasteran tidak melakukan klasifikasi, mengestimasi, atau memprediksi nilai dari variabel target, akan tetapi, algoritma pengklasteran mencoba untuk melakukan pembagian terhadap keseluruhan data menjadi kelompok-kelompok yang memiliki kemiripan (homogen).

#### 2.3.6 Asosiasi (Association)

*Association Data Mining* adalah suatu metode untuk menemukan atribut yang muncul dalam satu waktu. Asosiasi sering digunakan untuk melakukan Market Basket Analysis dengan tujuan untuk mengidentifikasi produk yang sering dibeli bersamaan oleh pelanggan. Salah satu implementasi dari asosiasi adalah market basket analysis untuk menentukan suatu promo *bundling* yang efektif untuk pelanggan, sebagaimana yang akan dibahas dalam penelitian ini.

#### 2.4. Aturan Asosiasi (Association Rules)

*Association rules* adalah salah satu *task data mining* deskriptif yang bertujuan untuk menemukan aturan asosiatif antara item-item data. Langkah utama yang perlu dalam *association rules* adalah mengetahui seberapa sering kombinasi item muncul dalam database, yang disebut sebagai *frequent patterns*. Pramudiono menyatakan bahwa penting tidaknya suatu aturan asosiatif dapat diketahui dengan dua parameter, *support* yaitu persentase kombinasi item dalam database dan *confidence* yaitu kuatnya hubungan antar item dalam aturan asosiatif (Aprianti et al., 2017). Bentuk umum dari association rule adalah *Antecedent* → *Consequent*. Contohnya adalah sebuah transaksi pembelian barang di dalam sebuah minimarket didapat bentuk *association rule* roti → selai. Artinya adalah pelanggan yang membeli roti ada kemungkinan pelanggan tersebut juga akan membeli selai, dimana



tidak ada batasan dalam jumlah item pada bagian *antecedent* ataupun *consequent* dalam sebuah *rule*.

terdapat suatu *interestingness measure* (ukuran kepercayaan) dalam menentukan suatu association rules. *interestingness measure* adalah salah satu parameter yang digunakan untuk perbandingan dalam data mining. Umumnya ada dua ukuran sebagai berikut.

#### 2.4.1 Support

Support adalah suatu ukuran yang menunjukkan seberapa besar tingkat dominasi suatu *item/itemset* dari keseluruhan transaksi. Ukuran ini menentukan apakah suatu *item/itemset* layak untuk dicari *confidence*-nya. Berikut merupakan persamaan dasar untuk mencari Support.

$$\text{Support} = \frac{\text{Jumlah transaksi A}}{\text{Jumlah transaksi}}$$

Sesuai dengan definisinya menunjukan seberapa tingkat dominasi item dari keseluruhan transaksi, maka persamaannya adalah jumlah transaksi yang mengandung *item A* dibagi dengan jumlah transaksi secara keseluruhan.

#### 2.4.2 Confidence

*Confidence* adalah parameter yang digunakan untuk mengetahui berapa kali pernyataan hubungan antara 2 *item* secara *conditional* tersebut benar. *Confidence* biasanya baru akan dihitung apabila nilai *support* sudah memenuhi kondisi. Berikut merupakan persamaan sederhana untuk mencari *confidence*.

$$\text{Confidence} = \frac{\text{Jumlah trans A + B}}{\text{Jumlah Trans A}}$$

Persamaan untuk mencari confidence adalah dengan cara membagi jumlah transaksi yang mengandung item A dan B dengan jumlah transaksi A. Contoh kasus untuk support dan confidence adalah sebagai berikut.

**Tabel 2. 1** Transaksi Awal

No	Barang
1	Gula, kopi
2	Permen, Marie, Gula
3	Sabun, Permen, Sikat Gigi
4	Permen, Snack, Marie
5	Air Mineral, Permen, Roti
6	Roti, Gula, Kopi, Kacang Kulit\
7	Roti, Permen
8	Pasta Gigi, Sikat Gigi
9	Roti, Permen, Snack
10	Snack, Permen, Marie
11	Susu, Kopi, Gula
12	Susu, Marie, Roti
13	Pasta Gigi, Sikat Gigi, Kopi
14	Permen, Snack
15	Kacang Kulit, Kopi, Gula
16	Gula Kopi, Snack
17	Snack, Susu, Gula
18	Kopi, Gula, Roti
19	Pasta Gigi, Gula, Susu
20	Snack, Kacang Kulit

**Commented [EIS9]:** sebaiknya dibuatkan tabel asli

Selanjutnya menentukan nilai support dengan menghitung jumlah transaksi yang mengandung item A dibagi dengan jumlah transaksi yang ada, maka didapat:

**Tabel 2. 2** Pencarian Nilai Support

No	Jenis Barang	Jml	Sup
1.	Gula, Kopi	5	25 %
2.	Gula,Permen	1	5 %
3.	Gula, Susu	4	20 %
4.	Permen. Marie	3	5 %
5.	Permen, Sikat gigi	1	5 %
6.	Permen, Kacang Kulit	1	5 %
7.	Kopi, Susu	1	5 %
8.	Marie, Susu	1	5 %
9.	Sikat gigi, pasta gigi	2	10 %
10.	Pasta gigi, susu	1	5 %

Setelah mendapat nilai *support* maka langkah selanjutnya adalah menentukan minimal nilai *support*. Misalkan *minimum* nilai *support* adalah 10% maka

selanjutnya adalah mencari nilai *confidence* dengan mencari banyaknya kemunculan item A dan B pada satu nota, maka didapat:

**Tabel 2. 3** Pencarian Nilai Confidence

no	barang	conf
1.	Gula, Kopi	5/8 = 62 %
2.	Gula, Susu	4/8 = 50 %
3.	Sikat gigi, Pasta Gigi	2/3 = 67 %

Nilai *confidence* sudah didapatkan, maka perbandingan itemset dapat dilakukan. Gula → Kopi menjadi *rules* yang memiliki nilai *support* terbesar dan *confidence* diatas 50% sehingga memungkinkan untuk diterapkan.

**Commented [EIS10]:** remove space after paragraf. Periksa paragraf di bawah lainnya

Beberapa perdebatan masih kerap terjadi mengenai dua parameter pengukuran yaitu Support dan Confidence walaupun memang sudah biasa digunakan dalam metode asosiasi. Sekitar 20 parameter pengukuran sempat diajukan melalui literatur, namun belum ada yang cocok untuk digunakan di segala kondisi sehingga untuk saat ini masih menggunakan parameter support, confidence dan lift (Merceron & Yacef, 2019).

## 2.5 Market Basket Analysis (MBA)

Menurut Alfagra et al, 2018 *Market Basket Analysis* merupakan salah satu proses analisis keranjang belanja dalam menentukan strategi pemasaran untuk memenuhi produk yang akan dibeli secara bersamaan oleh konsumen. *Data mining* dan model asosiasi khususnya dapat digunakan untuk mengidentifikasi produk-produk terkait yang biasanya dibeli bersamaan. MBA ini sendiri dapat digunakan dalam memahami kebiasaan konsumen dalam membeli produk yang dijual melalui transaksi belanja dengan analisis *asociation rule*. MBA bekerja dengan cara mengidentifikasi hubungan pada sekumpulan produk, item, maupun kategori. Produk-produk yang dibeli konsumen sering kali memiliki hubungan sehingga membentuk suatu hubungan jika-maka yang disebut dengan *rule if-then*.

## 2.6 Algoritma Equivalence Class Transformation (ECLAT)

Algoritma ECLAT (*Equivalence Class Transformation*) adalah suatu algoritma yang digunakan untuk melakukan *itemset mining*. Pencarian itemset untuk menemukan pola yang sering terjadi pada data seperti jika konsumen membeli roti dan juga membeli mentega. Jenis pola ini disebut aturan asosiasi dan digunakan di banyak domain aplikasi.

Menurut Dan Toomey (2014:42) Algoritma *eclat* digunakan untuk frequent *itemset mining*. Algoritma ini menggunakan persimpangan dalam data untuk menghitung dukungan kandidat. Dalam hal ini, kami mencari pola perilaku yang sama, dibandingkan dengan mencari penyimpangan. Kejadian yang sering terjadi bersamaan, seperti item keranjang belanja. Kandidat *frequent* kemudian diuji untuk mengkonfirmasi pola dalam dataset. *Eclat* dapat digunakan dalam pemrograman R dengan fungsi *eclat* dalam paket *arules*.

Proses pembentukan *itemset* pada algoritma ECLAT dimulai dengan mengubah bentuk transaksi, jika *item* pada transaksi berbentuk horizontal maka diubah menjadi bentuk vertikal dengan menggabungkan TID List pada transaksi yang memiliki *item* yang sama. Berikut merupakan contoh penerapan ECLAT.

**Tabel 2. 4** Transaksi Awal

TID	Item
1	a, b
2	b, c, d
3	b, c, d, e
4	a, d, e
5	a, b, c

**Tabel 2. 5** Transaksi Vertikal

TID	Item
a	(1,4,5)
b	(1,2,3,5)
c	(2,3,5)
d	(2,3,4)
e	(3,4)

Selanjutnya dilakukan penyilangan TID List dari kedua (k-1) subset. Model penyilangan dapat dilakukan dari atas kebawah. Contoh penyilangan dapat dilihat pada Tabel 2.6.

**Tabel 2. 6** Hasil Penyilangan 2-itemset

Itemset	TID List
(a,b)	(1,5)
(a,c)	(5)
(a,d)	(4)
(a,e)	(4)
(b,c)	(2,3,5)
(b,d)	(2,3)
(b,e)	(3)
(c,d)	(2,3)
(c,e)	(3)
(d,e)	(3,4)

**Commented [EIS12]:** caption di pojok kanan atas tabel, aplikasikan ke tabel-tabel lainnya di bawah

**Commented [EIS11]:** size 10

Pada **Tabel 2.6** dapat dilihat bahwa *subset* item a disilangkan dengan *subset* item b, hasil penyilangan (a,b) didapat dari operasi konjungsi antara anggota *subset* a dan *subset* b, proses penyilangan terus dilakukan sampai tidak ada lagi *itemset* yang tersisa. Selanjutnya ditentukan *minimum support (minsupp)* dari setiap k-itemset. *Itemset* yang memiliki nilai kurang dari nilai *minsupp* akan dihilangkan. Sebagai contoh, ditentukan bahwa *minsupp* dari transaksi adalah 2, maka hasil transaksi terdapat pada Tabel 2.7.

**Tabel 2. 7** Hasil frequent 2-itemset

TID	Item
(a,b)	(1,5)
(b,c)	(2,3,5)
(b,d)	(2,3)
(c,d)	(2,3)
(d,e)	(3,4)

Pada Tabel 2.7 dapat dilihat bahwa semua transaksi memiliki jumlah *TIDList* kurang dari nilai *minsupp* telah dihilangkan. Sehingga didapat hasil *frequent* dari 2-*itemset*. Selanjutnya apabila dilakukan penyilangan terhadap 3-*itemset* dan seterusnya, penyilangan dilakukan dengan langkah yang sama dengan penyilangan sebelumnya dari atas kebawah. Setelah hasil dari *frequent itemset* ditemukan maka dilakukan perhitungan terhadap nilai *support* dan nilai *confidence* dari masing-masing *itemset* untuk menentukan ukuran seberapa valid *association rule* yang ditemukan.

**Tabel 2. 8** Nilai Support dan Confidence Itemset

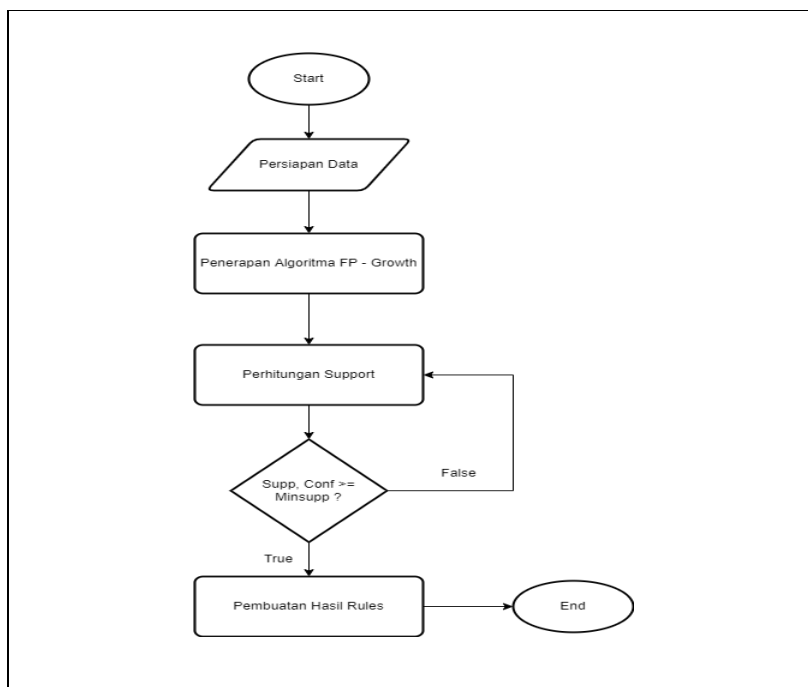
Itemset	Support (%)	Confidence (%)
(a→b)	0,4	0,6667
(b→c)	0,6	0,75
(b→d)	0,4	0,5
(c→d)	0,4	0,6667
(c→d)	0,4	0,6667

Pada **Tabel 2.8** merupakan nilai support dan confidence dari frequent 2-*itemset*. Pada aturan asosiasi (a→b) diketahui bahwa nilai support atau nilai transaksi yang mengandung item a dan b dari total transaksi adalah 0,4%, sedangkan nilai confidence semua transaksi yang mengandung a dan b dari semua transaksi yang mengandung a dari aturan asosiasi (a→b) adalah 0,66667%. Pada **Tabel 2.8** juga dapat dilihat bahwa aturan asosiasi yang mempunyai nilai support dan confidence tertinggi adalah (b→c).

## 2.7 Algoritma Frequent Pattern Growth (FP-Growth)

Algoritma *FP-growth* adalah suatu algoritma yang mengkodekan data menggunakan struktur data yang disebut *FP-tree* dan mengekstrak *frequent itemset* dari *FP-tree* tersebut. *FP-tree* adalah representasi terkompresi dari data keranjang belanja. *FP-tree* dibangun dengan membaca *dataset* transaksi per satu transaksi dan melakukan *mapping item* tiap transaksi tersebut ke sebuah *path* pada *FP-tree*

(Indrawan et al., 2018). Karakteristik algoritma *Fp-Growth* adalah struktur dari data digunakan adalah *tree* atau disebut dengan *Fp-tree*. Dengan menggunakan *Fp-Tree*, algoritma ini dapat langsung mengekstrak frequent itemset dari *Fp-Tree*. Semakin banyak data transaksi yang memiliki item yang sama, maka proses pemanpatan dengan struktur data FP-tree semakin efektif. Kelebihan dari FP-tree adalah hanya memerlukan dua kali pemindaian data. Adapun gambaran tahapan metode FP-Growth adalah sebagai berikut.



**Gambar 2. 2** Flowchart Tahapan Algoritma FP-Growth

**Gambar 2.2** merupakan diagram alur yang digunakan dalam menerapkan FP-Growth, antara lain melakukan persiapan data dengan memindai data transaksi penjualan memilah atribut dan mengilangkan data yang bersifat redundansi. Proses selanjutnya melakukan perhitungan *Support* dan *confidence* untuk menghasilkan aturan dari algoritma FP-Growth.

## 2.8 Lift Ratio

Lift Ratio adalah suatu parameter yang digunakan untuk mengetahui kekuatan pengaruh aturan asosiasi yang telah terbentuk dari nilai support dan confidence. Nilai lift ratio biasanya digunakan sebagai penentu apakah aturan asosiasi valid atau tidak valid. Nilai lift ratio dapat diperoleh dengan rumus pada gambar berikut.

$$Lift = \frac{Support}{Support(X) * Support(Y)}$$

Gambar 2. 3 Rumus Perhitungan Lift Ratio

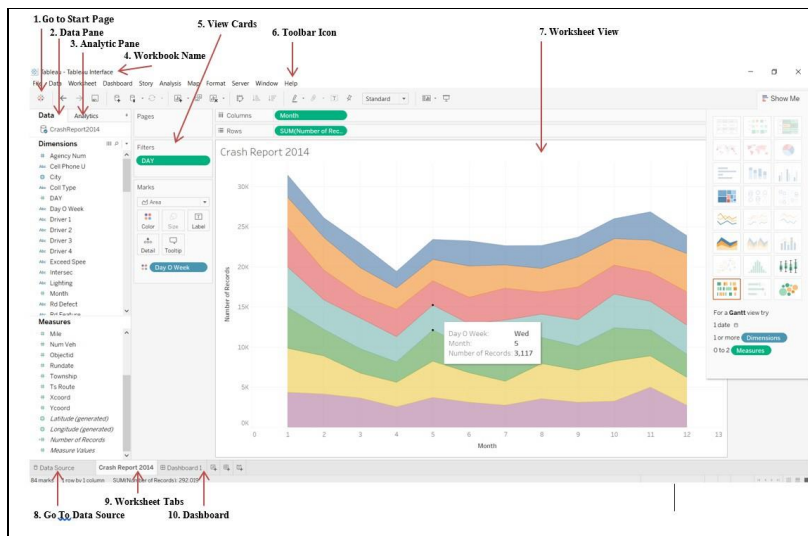
Pada Gambar 2.3 dapat diketahui bahwa cara mengetahui nilai *lift ratio* adalah dengan membagi nilai *support* dari suatu *rule* dengan Gabungan dari *support itemset* A dengan *Itemset* B. Korelasi akan dinyatakan negatif apabila hasil dari perhitungan *lift ratio* < 1 atau dengan kata lain, jika konsumen membeli barang A maka ia tidak akan membeli barang B. *Itemset* A dan *itemset* B dinyatakan memiliki korelasi positif apabila *lift ratio* > 1 yang berarti, jika konsumen membeli barang A maka ia juga akan membeli barang B. *Itemset* A dan *itemset* B dinyatakan independen satu sama lain apabila *lift ratio* = 1, yang berarti jika konsumen membeli barang A maka ia belum tentu membeli barang B.

## 2.9 Tableau

Tableau adalah sebuah *tools* yang dapat mempermudah pembuatan analisis visual interaktif dalam bentuk *dashboard*. Tableau banyak digunakan untuk memvisualisasikan data dan menyederhanakan pola analisis *business intelligence* (BI) karena Tableau menyediakan *dashboard* dan *scorecards*, *ad hoc analysis and queries*, pemrosesan analitik *online*, penemuan data, pencarian BI,



integrasi *spreadsheet*, dan lainnya. Berikut merupakan tampilan dari Tableau Public.



Gambar 2. 4 Tampilan Sederhana Tableau Public

Gambar 2.4 merupakan tampilan sederhana dari *tableau public*. *Tableau Public* adalah versi yang dapat diakses secara gratis dari *Tableau Desktop*. *Tableau Public* juga memungkinkan pengguna untuk membagikan *dashboard* yang telah dibuat sehingga memungkinkan pengguna lainnya untuk mengaksesnya.

Commented [EIS13]: penjelasan Tableau lebih detail

## 2.10 Jupyter Notebook

Jupyter Notebook merupakan singkatan dari tiga bahasa pemrograman, yakni *Julia* (Ju), *Python* (Py), dan *R*. Jupyter Notebook adalah sebuah aplikasi *web* gratis yang paling banyak dipakai oleh *data scientist*. Aplikasi ini dipakai untuk membuat dan membagikan dokumen yang memiliki kode, hasil hitungan, visualisasi, dan teks melalui *google collaboratory*. Ketiga bahasa pemrograman pada Jupyter Notebook sendiri adalah sesuatu yang penting bagi seorang *data scientist*.

Sederhananya, Jupyter Notebook berfungsi membantu *data scientist* dalam membuat narasi komputasi. Salah satunya adalah melakukan *data*

*pre-processing*. Aplikasi ini juga memudahkan *data scientist* melakukan kolaborasi dengan sesama *data scientist*, *data researchers*, maupun *data engineers* lainnya.

**Commented [EIS14]:** Penjelasan Jupyter lebih lengkap