

## BAB I PENDAHULUAN

### I.1 Latar Belakang

Kopi merupakan jenis minuman populer kedua setelah teh dan merupakan komoditi yang paling banyak diperdagangkan. Pada tahun 2023 produksi perkebunan kopi di Indonesia sebesar 4,1 ton dan perkebunan teh sebesar 73,4 ton (Badan Pusat Statistik, 2024). Indonesia sendiri termasuk sebagai 5 negara produsen dan eksportir kopi terbesar di dunia. Perkebunan kopi di Indonesia mencakup sekitar 1,27 juta hektar pada akhir tahun 2023. Berdasarkan data 2019, terdapat 933 hektar perkebunan kopi robusta dan 307 hektar perkebunan arabika, dengan sebagian besar perkebunan terletak di pulau Sumatera. Lima provinsi dengan luas perkebunan kopi terbesar berada di pulau Sumatera, sementara dataran tinggi Aceh dan Sumatera Utara sangat cocok untuk perkebunan arabika (Indonesia Investments, 2024).

Nilai pasar bisnis kedai kopi mencapai Rp. 4,8 triliun per tahun, perkiraan konsumsi domestik pada tahun 2020 sebesar 13,9%, hasil riset memperlihatkan bahwa 6 dari 10 orang menyukai kopi susu kekinian. Konsumsi kopi di Indonesia pada tahun 2019/2020 mencapai 294.000 ton, tumbuh sebesar 13,9% dari tahun sebelumnya. Pertumbuhan jumlah kedai kopi di Indonesia pun berkembang 3x lipat dari tahun 2016 hingga tahun 2019, terlebih lagi industri *coffee shop* telah mencapai gelombang ke 4, dimana gelombang ini *market size* telah bertumbuh besar dan juga semakin banyak pelanggan kopi yang tumbuh akibat naik daunnya produk kopi RTD ala kafe yang diperkenalkan oleh gerai *coffee to-go* dengan harga yang lebih terjangkau (Toffin, 2020).

Berdasarkan penelitian Snapcart yang dilakukan pada 12-15 September 2023 terhadap 4.538 responden, ditemukan bahwa 79% masyarakat Indonesia mengonsumsi kopi, dengan sebagian besar dari mereka meminumnya sekali sehari, terutama di pagi hari. Meskipun konsumsi kopi umumnya dilakukan di pagi hari, hanya kurang dari 20% yang mengonsumsinya saat sarapan atau dalam perjalanan ke tempat kerja atau sekolah. Perilaku konsumsi kopi ini tidak mengalami perubahan signifikan sejak pandemi, dan mayoritas masyarakat dari berbagai tingkat ekonomi cenderung memilih kopi dengan harga terjangkau (IDR

6.000 – IDR 20.000), yang umumnya dibeli dalam bentuk bubuk atau sachet di warung atau minimarket. Selain itu, 63% dari konsumen yang rutin mengunjungi kafe lebih sering membeli kopi pada hari kerja dibandingkan akhir pekan (Snapcart, 2023).

Penggunaan data yang didapatkan oleh *coffee shop* akan sangat berguna dalam menentukan banyak hal yang dapat berdampak pada kedainya. Beberapa informasi yang bisa didapatkan dengan mengolah data yang telah didapatkan yaitu seperti melihat performa penjualan, melihat perbandingan produk *repeat sales* dan penjualan baru, mengetahui angka permintaan produk, memperkiraan harga pasaran, dan informasi lainnya (Sudianto, 2019). Tentunya, data yang didapatkan dapat menjadi informasi yang bermanfaat dan menguntungkan bagi *coffee shop* dengan mengolahnya secara tepat.

Paket *bundling* pada dasarnya strategi pemasaran yang menggabungkan dua atau lebih produk berbeda dalam satu paket promosi, umumnya promosi seperti ini terdapat pada berbagai *event* tertentu seperti promo awal bulan Ramadhan atau promo hari natal. Strategi pemasaran seperti ini mampu mendorong konsumen untuk membeli lebih banyak produk dari satu *brand* yang sama (Ismail, 2021). Tentu saja membuat menu *bundling* tidak dapat sembarangan, harus sesuai dengan minat konsumen dan sasaran yang tepat. Data tersebut bisa didapatkan dari perilaku pembelian produk dari suatu *brand*.

Oleh karena itu, agar mendapatkan informasi yang sesuai untuk *bundling* promo suatu produk, peneliti akan melakukan analisa penjualan pada salah satu *brand* yang terdapat di kota Bandung yaitu *Authen Cafe and Space*. Analisa data penjualan ini dilakukan agar mendapatkan informasi yang sesuai mengenai *bundling* promo menu apa yang akan diterapkan pada *Authen and Space* agar mendapatkan keuntungan yang lebih banyak.

## **I.2 Rumusan Masalah**

Berdasarkan latar belakang di atas maka rumusan permasalahan penelitian ini yaitu ‘Bagaimana analisis penjualan cafe untuk menentukan *bundling menu* dengan metode Algoritma Apriori?’. Selanjutnya permasalahan tersebut diuraikan dalam sub-sub permasalahan:

1. Bagaimana penentuan menu *bundling* berdasarkan metode Algoritma Apriori?
2. Bagaimana penentuan potongan harga maksimal berdasarkan setiap menu yang digabungkan (*bundling*) dengan menggunakan pendekatan *Data Mining*?
3. Bagaimana evaluasi terhadap strategi *bundling menu* menggunakan metode LIFT?

## **I.3 Tujuan Tugas Akhir**

Berdasarkan rumusan permasalahan di atas, maka tujuan yang ingin dicapai dalam penelitian ini adalah untuk menganalisis penjualan cafe dalam menentukan *bundling menu* dengan metode Algoritma Apriori. Tujuan lain dirincikan:

1. Menentukan menu *bundling* berdasarkan metode Algoritma Apriori.
2. Menentukan potongan harga maksimal berdasarkan setiap menu yang digabungkan (*bundling*) dengan menggunakan pendekatan *Data Mining*.
3. Melakukan evaluasi terhadap strategi *bundling menu* menggunakan metode LIFT

## **I.4 Batasan dan Asumsi Penelitian**

Pada penelitian kali ini peneliti memberikan batasan untuk memfokuskan pembahasan masalah agar sesuai dengan tujuan yang dicapai. Batasan penelitian tersebut sebagai berikut:

1. Data yang diambil bersumber dari transaksi penjualan *Authen Cafe and Space* mulai dari bulan April - Agustus 2023.

2. Pengolahan data dilakukan hanya untuk mendapatkan informasi menu yang berpotensi untuk dijadikan *bundling menu* dan potongan harga maksimal yang tepat.

Adapun asumsi penelitian kali ini sebagai berikut:

1. Pendekatan algoritma yang digunakan adalah Algoritma Apriori untuk pengolahan data transaksi.
2. Persentase beban operasional untuk setiap produk adalah 20-25% dari harga jual.

### **I.5 Manfaat Tugas Akhir**

Penelitian ini diharapkan dapat memberikan manfaat untuk pihak-pihak yang berkaitan, di antaranya sebagai berikut:

1. Bagi Universitas Telkom  
Penelitian ini dapat memberikan kontribusi bagi pengembangan ilmu pengetahuan, terutama di bidang sistem informasi.
2. Bagi *Manajer Authen Cafe and Space*  
Penelitian ini bermanfaat untuk membantu *manajer* menentukan menu mana yang paling tepat untuk melakukan *bundling menu* dan seberapa besar maksimal potongan harga untuk menu yang dibundling agar dapat menarik minat konsumen lebih banyak lagi.
3. Bagi Peneliti  
Penelitian ini dapat menambah wawasan, pengetahuan, serta keterampilan dalam mengimplementasikan teori serta memperdalam teknik *data mining*, pengolahan data, dan analisis data dengan Algoritma Apriori yang didapatkan pada saat masa perkuliahan di *Telkom University*.
4. Bagi Peneliti Lain  
Penelitian ini dapat dijadikan referensi bagi peneliti bidang *Data Science*, untuk melakukan penelitian selanjutnya yang terkait pengimplementasian algoritma *data mining* khususnya apriori pada data penjualan objek penelitian.

## I.6 Sistematika Penulisan

Penelitian ini diuraikan dengan sistematika penulisan sebagai berikut:

- BAB I                   PENDAHULUAN**
- Pada bab ini berisikan latar belakang dalam penelitian yang dilakukan di *Authen Cafe and Space*, rumusan masalah, tujuan tugas akhir, batasan dan asumsi, manfaat tugas akhir, dan sistematika penulisan.
- BAB II                   LANDASAN TEORI**
- Pada bab ini berisikan penjabaran-penjabaran mengenai teori dasar dalam pembentukan penelitian ini. Terdapat penelitian terdahulu, konsep-konsep dasar dan ruang lingkup *bundling menu* yang disertai oleh teori dari para ahli. Serta penjabaran Metode *Apriori Algorithm* sebagai *tools* yang akan membantu dalam perhitungan dalam menentukan *bundling menu*.
- BAB III                 METODOLOGI PERANCANGAN**
- Pada bab ini terdapat pengembangan model konseptual, sistematika penyelesaian masalah, dan kebutuhan data.
- BAB IV                 ANALISIS DAN PERANCANGAN DATA**
- Pada bab ini dijelaskan mengenai analisis data dan tahap *pre-processing* yang dirincikan dari data selection, data cleaning, transformasi data, dan data processing.
- BAB V                  IMPLEMENTASI DAN EVALUASI HASIL**
- Pada bab ini membahas analisis kebutuhan *bundling menu* dan evaluasi hasil.
- BAB VI                 KESIMPULAN DAN SARAN**
- Pada bab ini merupakan penutup yang berisikan jawaban dari rumusan masalah dan merupakan kesimpulan akhir penelitian, serta saran sebagai bentuk rekomendasi dari peneliti.

## BAB II LANDASAN TEORI

### II.1 Penelitian Terdahulu

Penelitian terdahulu bertujuan untuk mendapatkan bahan perbandingan dan menjadi salah satu acuan dasar dalam melakukan penelitian, sehingga memperkaya teori-teori yang digunakan untuk mengkaji penelitian. Maka dalam kajian pustaka ini, pada Tabel II – 1 peneliti mencantumkan hasil-hasil penelitian terdahulu yang memiliki korelasi dengan penggunaan Algoritma Apriori sebagai berikut.

Tabel II- 1 Penelitian Terdahulu

| Peneliti                                  | Judul   | Metode  | Fungsi Tujuan   | Objek   |
|---|---|---|---|---|
| Ping-Ho Ting, Steve Pan, Shuo-Shiung Chou | <i>Finding Ideal Menu Items Assortments: An Empirical Application of Market Basket Analysis</i> | <i>Market Basket Analysis</i>                           | Mengeksplorasi variasi item menu yang didefinisikan sebagai kumpulan pasangan item menu yang sering dipesan.                            | <i>Chicken fish, Spicy Chicken, Eel Chicken, Miso salmon, Corn Soup, Salad, Onion Soup, Bread, Rice, Red Tea, Ice Tea, Orange Juice, Mango Cheese</i> |
| Wen-Yu Chiang                             | <i>Applying Data Mining for Online CRM Marketing Strategy: an Empirical Case of Coffee Shop</i> | <i>Fuzzy Clustering Algorithm and Apriori Algorithm</i> | Mengusulkan pendekatan <i>data mining</i> untuk menambang informasi pasar yang memiliki nilai tinggi untuk strategi pemasaran manajemen | <i>Coffee, Breads, Snacks, Teas (Chinese Teas), Cakes, Cookies, Milks</i>   |

| <b>Peneliti</b>   | <b>Judul</b>   | <b>Metode</b>  | <b>Fungsi Tujuan</b>   | <b>Objek</b>   |
|---|--|--|--|--|
|   | <i>Industry in Taiwan</i>  |  | hubungan pelanggan secara online.  |  |
| Melinska Ayu Febrianti, Roaida Yanti, Qurtubi, Hari Purnomo | <i>Analysis of Customers Buying Pattern Based on Sales Transaction Using Multilevel Association Rules-Market Basket Analysis</i> | <i>Multilevel Association Rules-Market Basket Analysis</i> | Merekomendasikan strategi pemasaran, seperti perbaikan tata letak dalam toko, desain planogram, dan penawaran bundling produk  | <i>Cake, Ice Cream, Instant Noodles, Sausage, Bread, Dish Soap, Powdered Drinks, Ground Coffee, These, Body Care</i> |
| Stefan Stremersch, Gerard J. Tellis                         | <i>Strategic Bundling of Products and Price: a New Synthesis for Marketing</i>   | <i>Bundling Strategies</i>                                 | Menentukan strategi <i>bundling</i> mana yang optimal dalam berbagai konteks dan juga memberikan kerangka kerja untuk memahami dan memilih <i>bundling</i> strategi. | <i>Bundling, Price Bundling, Product Bundling, Pure Bundling, Mixed Bundling</i>                                     |
| Milfa Yetri, Suardi Yakun, Elpandri                         | Data Mining untuk Analisis Pola Pemilihan Menu Pada Penang <i>Corner Cafe</i>  | Algoritma Apriori  | Menentukan pola-pola pemilihan menu antara menu yang satu dengan menu yang lainnya untuk   | Kwetiau Penang, Kentang Goreng, Tempe Goreng, Teh Manis, Teh Pahit, Teh Tarik, <i>Black Coffee, Special Drink</i>    |

| Peneliti  | Judul   | Metode               | Fungsi Tujuan   | Objek  |
|---|---|----------------------|---|--|
|   | dan Resto<br>Menggunakan<br>Algoritma<br>Apriori  |                      | dapat menjadi<br>bahan<br>pertimbangan<br>untuk<br>menentukan<br>strategi<br>penjualan.   |  |
| Marina<br>Rajagukguk,<br>Rafiqa<br>Dewi, Eka<br>Irawan, Jaya<br>Tata<br>Hardinata,<br>Irfan<br>Sudahri<br>Damanik | Implementasi<br><i>Association<br/>Rule Mining</i><br>untuk<br>Menentukan<br>Pola<br>Kombinasi<br>Makanan<br>dengan<br>Algoritma<br>Apriori | Algoritma<br>Apriori | Menentukan pola<br>kombinasi<br>makanan<br>berdasarkan<br>database untuk<br>memberikan<br>peningkatan<br>kualitas dan<br>sistem informasi<br>yang akurat. | <i>Spaghetti Black<br/>Paper, Bola<br/>Udang,<br/>Tenderloin Steak,<br/>Nasi Goreng<br/>Cabe Hijau,<br/>Siomay Ayam,<br/>Salmon Steak,<br/>Spaghetti<br/>Bolognese,<br/>French Fries</i> |

## II.2 Pemilihan Teori/Metode

Karakteristik dan keunggulan dari tiap metode menyesuaikan dengan kebutuhan penelitian. Berikut di bawah ini macam teori/metode yang sering digunakan:

### 1. *Fuzzy Clustering Algorithm*

Himpunan fuzzy adalah dasar dari teori fuzzy, menurut Zadeh pada tahun 1965. Himpunan fuzzy dapat mengambil bentuk kuantitatif untuk mewakili beberapa konsep fuzzy yang ambigu. Mengingat hal ini, Dunn (1974) menciptakan metode fuzzy c-means (FCM) berdasarkan gagasan fuzzy clustering (Ruspini, 1969). Bezdek menciptakan algoritma fuzzy c-means, yang didasarkan pada metode k-means, pada tahun yang sama.

Fungsi tujuan pengelompokan fuzzy dapat digunakan untuk memecahkan masalah dalam berbagai konteks. Pengertian pemrograman non-linier dapat digunakan pada proses pengelompokan fuzzy. Algoritma pengelompokan

standar, yang menggunakan ruang Euclidean untuk merepresentasikan vektor, terinspirasi oleh pengelompokan keras. Pengelompokan fuzzy digunakan oleh Ruspini (1969) untuk menghasilkan bidang hiper untuk vektor. Ide partisi  $c$  untuk menghilangkan kendala pertama kali dikemukakan oleh Krishnapuram dan Keller pada tahun 1993. Namun, konvergensi tidak terjadi cukup cepat.

## 2. *Multilevel Association Rule*

Menurut Triyanto, *Association Rule Mining* merupakan teknik data mining yang digunakan untuk menemukan aturan asosiatif antar sesuatu. Menurut penelitian Kim, Xu, dan Gupta, analisis keranjang pasar—juga dikenal sebagai analisis asosiasi—menjadi terkenal karena penggunaannya dalam memeriksa pola pembelian konsumen dengan mencari asosiasi dan korelasi antara berbagai item di keranjang belanja setiap pelanggan.

Salah satu jenis aturan asosiasi yang umum adalah aturan asosiasi multi-level. Saat menganalisis data dalam jumlah besar, korelasi antara kelompok kecil objek tertentu ditemukan menggunakan aturan asosiasi multi-level. Hal-hal sering kali dikategorikan secara hierarki berdasarkan kategori untuk mengatasi masalah ini. Menggunakan aturan asosiasi multi-level di setiap kategori hierarki, pencarian kombinasi asosiasi objek dilakukan secara bertahap. Akibatnya, akan lebih mudah untuk mengidentifikasi interaksi antara kedua hal tersebut.

## 3. Algoritma Apriori

Algoritma Apriori adalah salah satu algoritma yang digunakan dalam data mining untuk menemukan asosiasi atau hubungan antara item dalam dataset besar, seperti dalam analisis keranjang belanja atau rekomendasi produk. Algoritma ini digunakan untuk menemukan item-item yang sering muncul bersama dalam transaksi atau dataset, yang dapat memberikan wawasan tentang pola pembelian atau asosiasi antar-item. Berikut adalah prinsip dasar dari algoritma Apriori:

a. *Support* (Dukungan)

*Support* adalah ukuran seberapa sering suatu *itemset* (kumpulan item) muncul dalam *dataset*. Untuk suatu *itemset*, *support* dihitung sebagai jumlah transaksi yang berisi *itemset* tersebut dibagi dengan total jumlah transaksi dalam *dataset*.

b. *Confidence* (Kepercayaan)

*Confidence* mengukur seberapa sering aturan asosiasi (*rule*) yang menghubungkan dua *itemset* ditemui dalam *dataset*. Untuk suatu aturan asosiasi  $A \rightarrow B$ , *confidence* dihitung sebagai *support* dari  $A$  dan  $B$  dibagi dengan *support* dari  $A$ .

4. Uji *Lift Rasio*

Untuk mengetahui kekuatan aturan asosiasi yang telah dibuat, tahapan uji rasio lift dilakukan. Pada tahap ini, pengujian kevalidan aturan apakah produk  $A$  dan produk  $B$  dibeli bersamaan terbukti benar. Jika *Lift Ratio* lebih dari 1, maka aturan itu kuat dan dapat digunakan sebagai rekomendasi produk (Renaldi Guntur et al. 2003).

$$\text{Lift Ratio} = (\text{Confidence}(A,B)) / (\text{Benchmark Confidence}(A,B)).$$

$$\text{Benchmark Confidence} = (\text{jumlah Transaksi yang mengandung item B}) / (\text{Total Transaksi}).$$

## II. 2. 1 Alasan Pemilihan Metode

Algoritma Apriori dipilih dalam analisis asosiasi karena kemampuannya untuk menemukan asosiasi antar-item dalam dataset dengan menggunakan konsep support dan confidence, yang memungkinkan pengguna untuk mengontrol relevansi dan kepercayaan dalam aturan asosiasi. Selain itu, Apriori dapat mengelola dataset besar dan memiliki berbagai implementasi terbuka yang tersedia. Kelebihan lainnya termasuk interpretasi yang mudah dari hasilnya dan rekam jejak yang baik dalam berbagai aplikasi bisnis. Namun, pemilihan algoritma Apriori harus dipertimbangkan dengan bijaksana sesuai dengan karakteristik dataset dan tujuan analisis yang diinginkan. Berikut adalah alasan pemilihan metode berdasarkan karakteristik permasalahan:

Tabel II- 2 Alasan Pemilihan Metode

| Karakteristik Permasalahan   | Metode            |                                   |                                    |
|--|-------------------|-----------------------------------|------------------------------------|
|  | Algoritma Apriori | <i>Fuzzy Clustering Algorithm</i> | <i>Multilevel Association Rule</i> |
| <b>Tujuan Utama</b><br>Menentukan aturan asosiasi dalam dataset, yaitu menemukan hubungan atau pola antara item-item dalam data transaksi.                     | ✓                 |                                   | ✓                                  |
| <b>Analisis Data</b><br>Memiliki teknik dalam analisis data yang digunakan untuk mengungkap pola atau struktur dalam dataset.                                  | ✓                 | ✓                                 | ✓                                  |
| <b>Tingkatan Level antar Hubungan</b><br>Mencari hubungan antar-item pada satu tingkat transaksi atau dataset tunggal.   | ✓                 |                                   |                                    |
| <b>Dampak pada Pengambilan Keputusan</b><br>Dapat digunakan untuk pengambilan keputusan yang lebih baik dan perencanaan strategis dalam bisnis dan organisasi. | ✓                 | ✓                                 | ✓                                  |

Dengan demikian, metode Algoritma Apriori adalah metode yang tepat digunakan dalam penelitian kali ini karena dapat menyelesaikan tujuan utama, menganalisis data, menyelesaikan permasalahan sesuai dengan kondisi saat ini yaitu pada tingkatan level antar hubungan, dan memiliki dampak pada pengambilan keputusan.

### II.3 Teori Dasar

Teori dasar adalah prinsip atau konsep fundamental yang membentuk dasar pemahaman dalam suatu bidang studi. Ini mencakup definisi, prinsip, dan model yang menjelaskan bagaimana sesuatu bekerja dan memberikan landasan untuk

pengembangan pengetahuan lebih lanjut dan aplikasi praktis. Berikut dibawah ini macam teori dasar yang digunakan:

1. *Market Basket Analysis* (MBA)

MBA pertama kali diperkenalkan oleh Agrawal, Imielinski, dan Swami (1993). Ini juga disebut sebagai analisis asosiasi atau penambangan aturan asosiasi. Penjualan silang, dukungan keputusan, analisis perilaku pelanggan, dan manajemen hubungan pelanggan adalah beberapa bidang di mana MBA telah digunakan (Tang, Chen, dan Hu 2008; Haughton dkk. 2003). Ia menggunakan data mining untuk mengidentifikasi hubungan dari data transaksi point-of-sale (POS) dengan tujuan mengidentifikasi tren pembelian klien (Chen et al. 2005; Berry dan Linoff 2004). Data asosiasi item yang dihasilkan dapat digunakan dalam inisiatif pemasaran termasuk positioning produk, penjualan silang, dan desain katalog (Auslander 2008; Berry dan Linoff 2004).

2. *Bundling*

*Bundling*, menurut Adams dan Yellen (1976, p. 475), adalah praktik 'menjual barang dalam bentuk paket'. *Bundling* dijelaskan oleh Guiltinan (1987, p. 74) sebagai 'praktik pemasaran dua atau lebih produk dan/atau jasa dalam satu paket dengan harga khusus.' Sebagaimana dinyatakan di halaman 350 buku mereka tahun 1993, Yadav dan Monroe mendefinisikannya sebagai 'penjualan dua atau lebih produk dan/atau jasa dengan harga tunggal.' Tanpa definisi standar, legalitas bundling menjadi ambigu, dan penerapannya dalam situasi dunia nyata tidak jelas.

3. Harga Pokok Penjualan

Harga dapat menghasilkan penerimaan penjualan, karena harga mempengaruhi sejumlah faktor, termasuk tingkat penjualan dan tingkat keuntungan yang dicapai perusahaan saat menetapkan harga. Jika produk dijual, pendapatan dilaporkan sebagai penjualan, dan biaya produk dianggap sebagai beban, yang dikenal sebagai harga pokok barang dagang yang dijual atau harga pokok penjualan. Haryono (2001,342) menyatakan bahwa, pertama-tama, persediaan awal ditambah dengan harga pokok barang yang dibeli sama

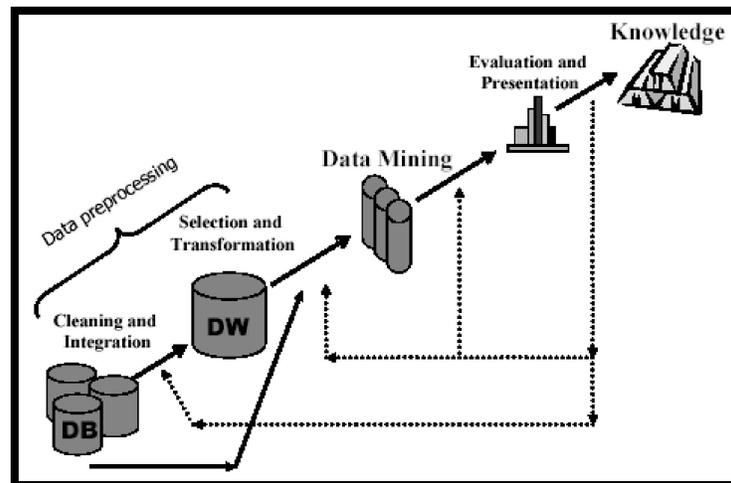
dengan harga pokok barang yang dijual, dan kedua-duanya, harga pokok barang yang tersedia untuk dijual dikurangi persediaan akhir sama dengan harga pokok penjualan. Menurut (Wiratna, 2016), harga pokok penjualan adalah biaya keseluruhan yang dikeluarkan dalam rangka mendapatkan barang yang dijual atau harga perolehan dari barang yang dijual. Perhitungan harga pokok penjualan dapat dilakukan dengan rumus sebagai berikut:

$$\text{HPP} = \text{Biaya Bahan Baku} + \text{Biaya Tenaga Kerja Langsung} + \text{Biaya Overhead Pabrik} + \text{Biaya Distribusi dan Administrasi}$$

HPP dihitung dengan membagi biaya produksi total (termasuk biaya bahan baku, tenaga kerja langsung, dan overhead pabrik) dengan jumlah unit barang yang dihasilkan. Ini memberikan HPP per unit barang atau jasa yang dihasilkan.

#### II.4 Data Mining

*Data mining* adalah proses menggali dan menganalisis data dalam jumlah yang sangat besar untuk mendapatkan sesuatu yang nyata, baru dan berguna dan akhirnya dapat ditemukan satu atau lebih pola dalam data tersebut. (Han dan Kamber, 2006). Data mining merupakan bagian integral dari knowledge discovery databases (KDD), yang merupakan keseluruhan proses transformasi data mentah menjadi model data yang menarik yang merupakan informasi yang dibutuhkan pengguna sebagai pengetahuan, seperti yang ditunjukkan pada gambar II.1.



Gambar II. 1 Proses Data Mining

Pada Gambar II.1 terdapat beberapa langkah yang ditempuh untuk melakukan data mining, di antaranya sebagai berikut:

1. *Cleaning and Integration*

Sebuah data besar pada umumnya memiliki ciri khas masing-masing dan belum diolah sehingga terjadi beberapa situasi seperti data yang *noise* sehingga menyebabkan ketidakkonsistenan dan sumber data yang terpisah dapat disatukan sebagaimana seharusnya. Oleh sebab itu, perlu dilakukan *data cleaning* agar *noise* data teratasi dan *data integration* supaya dapat menyatukan data yang terpisah.

2. *Selection and Transformation*

Setelah data telah menunjukkan kekonsistenan, maka selanjutnya adalah mensortir data yang relevan dengan kebutuhan. Maka dari itu perlu dilakukan *data selection* agar data yang dimiliki tidak mencakup data yang tidak relevan. Kemudian data yang sudah terseleksi perlu disatukan menjadi bentuk atau susunan yang tepat untuk mendapatkan ringkasan dari data tersebut, sehingga perlu dilakukan *data transformation*.

3. *Data Mining*

Mengekstraksi pengetahuan atau data yang berharga atau pola tersembunyi dari dataset yang besar dan kompleks. Pada proses ini membutuhkan metode intelijen yang digunakan untuk mengekstrak pola dari data terkait.

4. *Evaluation and Presentation*

Setelah menerapkan metode intelijen, diperlukan sebuah evaluasi dari hasil yang telah didapatkan. Evaluasi yang dilakukan dapat melibatkan penggunaan metrik seperti akurasi, presisi, *recall*, atau berbagai metrik lain tergantung dari kebutuhan dan jenis analisis yang sedang dilakukan. Lalu dari hasil analisis yang telah dilakukan perlu disajikan dengan cara yang dapat dimengerti oleh pemangku kepentingan ataupun pembaca lainnya. Penyajian tersebut dapat berupa laporan, visualisasi data, ataupun *dashboard*.

## 5. Knowledge

Informasi ataupun pengetahuan yang didapatkan dari proses *data mining* adalah *output* atau hasil akhir yang diterima oleh pemangku kepentingan ataupun pembaca lainnya. *Knowledge* ini dapat membantu pemahaman lebih tentang data, pengambilan keputusan, peramalan, atau identifikasi tren atau sebuah pola pada data tersebut.

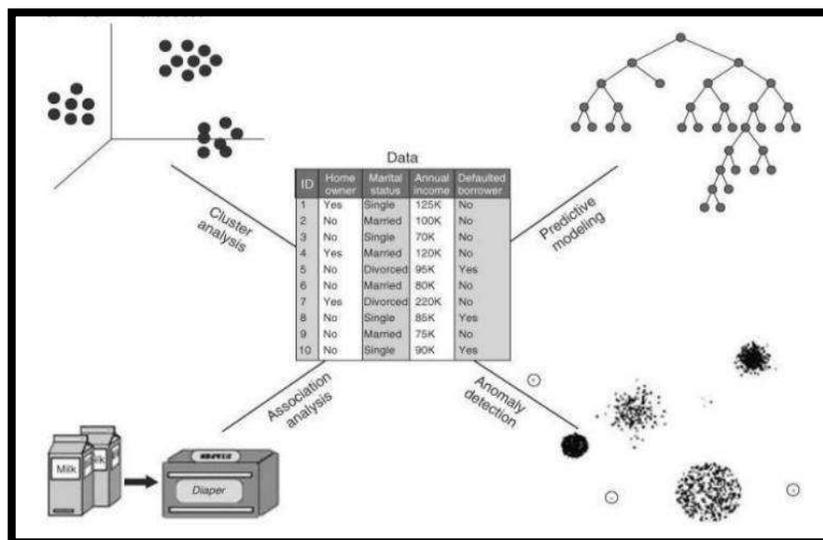
## 6. Database

*Database* adalah kumpulan data yang terstruktur yang disimpan dalam satu lokasi yang dapat diakses dan dikelola dengan mudah. Database digunakan untuk menyimpan, mengatur, dan mengelola informasi dalam format yang dapat diambil kembali dengan cepat.

## 7. Data Warehouse

*Data Warehouse* adalah sistem yang dirancang khusus untuk menyimpan dan mengelola data dalam jumlah besar dari berbagai sumber yang berbeda untuk keperluan analisis dan pengambilan keputusan. Data warehouse memfasilitasi proses data mining dengan menyediakan tempat sentral untuk data yang diolah sebelum diambil untuk analisis lebih lanjut.

Pada dasarnya terdapat empat tugas utama dalam *data mining* seperti gambar di bawah ini:



Gambar II. 2 Tugas Utama Data Mining

Adapun penjelasan dari Gambar II.2 di antaranya sebagai berikut:

### 1. Model Prediktif

Model prediktif memungkinkan untuk membangun model variabel target sebagai fungsi dari variabel penjelas. Variabel penjelas dalam hal ini adalah semua atribut yang digunakan untuk membuat prediksi, sedangkan variabel target adalah properti yang nilainya akan diprediksi. Model prediktif dibagi menjadi dua kategori, yaitu : klasifikasi yang digunakan untuk memprediksi nilai variabel target diskrit dan regresi yang digunakan untuk memprediksi nilai variabel target kontinu.

### 2. Analisis Asosiasi

Analisis asosiasi adalah penemuan aturan asosiasi yang menunjukkan pola yang sering terjadi pada data. Terdapat nilai *support* dan *confidence* yang dapat menunjukkan keandalan suatu aturan. *Support* adalah ukuran dominasi item atau set item atas transaksi secara keseluruhan. Sedangkan *confidence* adalah ukuran yang secara kondisional menyatakan hubungan antara dua unsur. Analisis asosiasi digunakan untuk menemukan aturan asosiasi yang menunjukkan kondisi nilai atribut yang sering terjadi bersama-sama dalam suatu kumpulan data.

### 3. Analisis Klaster

Tidak seperti analisis pengklasifikasi kelas data fitur yang berisi label. *Clustering* digunakan untuk menganalisis objek dari data tanpa memeriksa label kelas yang diketahui. Label kelas dimasukkan dalam data pelatihan karena tidak diketahui sebelumnya. *Clustering* adalah proses mengelompokkan satu set objek yang sangat mirip.

### 4. Deteksi Anomali / Eksploitasi Pengecualian

*Database* mungkin berisi data dari objek yang tidak cocok atau menyimpang dari model data. Objek data ini disebut pengecualian. Banyak metode data mining menghilangkan *outlier* ini. Eksplorasi *outlier* juga biasa dikenal sebagai deteksi anomali, adalah metode pendeteksian data yang tujuannya adalah untuk menemukan objek yang berbeda dari

kebanyakan objek lainnya. Anomali dapat dideteksi dengan menggunakan uji statistik yang menerapkan model distribusi atau probabilitas pada data.