

## **BAB II**

### **LANDASAN TEORI**

#### **2.1 Percetakan**

Pada dasarnya percetakan merupakan sebuah industri yang diperuntukkan untuk memproduksi tulisan maupun gambar dalam jumlah banyak. Media yang digunakanpun berbagai macam, namun yang sering kita jumpai yakni dalam media kertas, plastik, PVC dan sebagainya. Percetakan (*printing*) merupakan salah satu format media primer yang sulit dicari penggantinya selama lebih dari 500 tahun. *Printing* bahkan dianggap sebagai salah satu kebutuhan primer manusia (Watkins, 2012).

Berbagai macam kebutuhan sehari – hari dihasilkan dari industri percetakan, seperti buku, brosur, kotak kemasan, banner dan lain – lain. Usaha percetakan di Indonesia sendiri sudah berkembang pesat, mulai dari usaha rumahan hingga perusahaan. Persainganpun sudah biasa terjadi dalam dunia usaha terutama usaha dalam bidang jasa percetakan ini. Untuk itu diperlukannya strategi yang baik dalam menghadapi persaingan di dunia bisnis ini.

#### **2.2 Data Mining**

##### **2.2.1 Definisi Data Mining**

*Data Mining* adalah serangkaian proses untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu basis data. Informasi yang dihasilkan diperoleh dengan cara mengekstraksi dan mengenali pola yang penting atau menarik dari data yang terdapat pada basis data. *Data mining* terutama digunakan untuk mencari pengetahuan yang terdapat dalam basis data yang besar sehingga sering disebut *Knowledge Discover Database (KDD)*. (Vulandari, 2017)

Hubungan yang dicari dalam *data mining* dapat berupa hubungan antara dua atau lebih dalam suatu dimensi. Misalnya dalam dimensi produk kita dapat melihat keterkaitan pembelian suatu produk dengan produk yang lain. Selain itu, hubungan

juga dapat dilihat antara dua atau lebih atribut dan dua atau lebih objek. (Ponniah, 2001)

### 2.2.2 Pengelompokan *Data Mining*

*Data Mining* dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu (Larose, 2005):

1. Deskripsi

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.

2. Estimasi

Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih ke arah numerik dari pada ke arah kategori. Model dibangun menggunakan *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.

3. Prediksi

Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan ada di masa mendatang. Beberapa metode dan teknik yang digunakan dalam klasifikasi dan estimasi dapat pula digunakan (untuk keadaan yang tepat) untuk prediksi.

4. Klasifikasi

Dalam klasifikasi, terdapat target variabel kategori. Sebagai contoh, penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu pendapatan tinggi, pendapatan sedang, dan pendapatan rendah.

5. Pengklusteran

Pengklusteran merupakan pengelompokan *record*, pengamatan, atau memperhatikan dan membentuk kelas obyek-obyek yang memiliki kemiripan. Kluster adalah kumpulan *record* yang memiliki kemiripan satu dengan yang lainnya dan memiliki ketidakmiripan *record – record* dalam

kluster yang lain. Pengklusteran berbeda dengan klasifikasi yaitu tidak adanya variabel target dalam pengklusteran. Pengklusteran tidak mencoba untuk melakukan klasifikasi, mengestimasi, atau memprediksi nilai dari variabel target. Akan tetapi, algoritma pengklusteran mencoba untuk melakukan pembagian terhadap keseluruhan data menjadi kelompok-kelompok yang memiliki kemiripan (homogen), yang mana kemiripan *record* dalam satu kelompok akan bernilai maksimal, sedangkan kemiripan dengan *record* dalam kelompok lain akan bernilai minimal.

#### 6. Asosiasi

Tugas asosiasi dalam *data mining* adalah untuk menemukan atribut yang muncul dalam satu waktu. Dalam dunia bisnis lebih umum disebut analisis keranjang belanja.

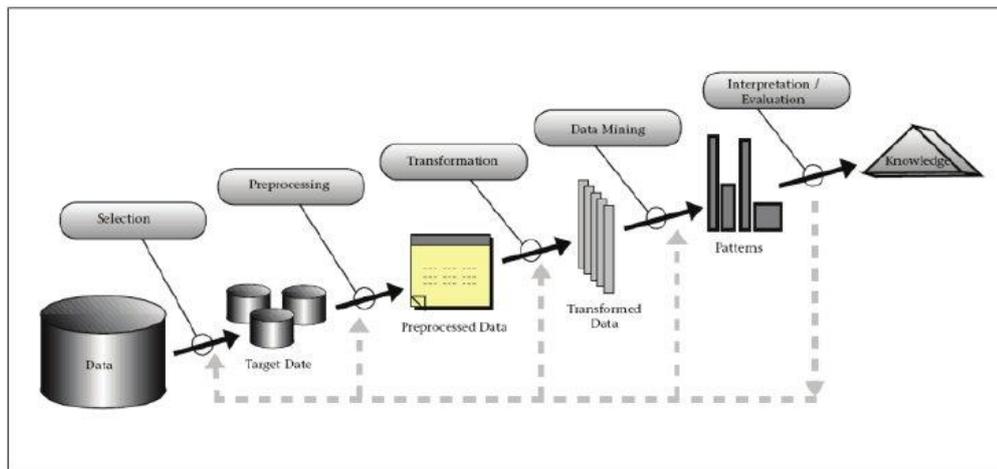
### 2.3 Proses Tahapan *Data Mining*

*Data mining* merupakan salah satu dari rangkaian *Knowledge Discovery in Database* (KDD). KDD berhubungan dengan teknik integrasi dan penemuan ilmiah, interpretasi dan visualisasi dari pola – pola sejumlah data. Serangkaian proses tersebut memiliki tahap sebagai berikut (Tan, 2004):

1. Pembersihan data (untuk membuang data yang tidak konsisten dan *noise*).
2. Integrasi data (penggabungan data dari beberapa sumber).
3. Transformasi data (data diubah menjadi bentuk yang sesuai untuk di-*mining*).
4. Aplikasi teknik *data mining*, proses ekstraksi pola dari data yang ada.
5. Evaluasi pola yang ditemukan (proses interpretasi pola menjadi pengetahuan yang dapat digunakan untuk mendukung pengambilan keputusan)
6. Presentasi pengetahuan (dengan teknik visualisasi).

Tahap ini merupakan bagian dari proses pencarian pengetahuan yang mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesa yang ada sebelumnya. Langkah terakhir KDD adalah mempresentasikan pengetahuan dalam bentuk yang mudah dipahami pengguna.

Gambar 2.1 menunjukkan proses atau tahapan KDD.



**Gambar 2.1** Tahapan *Knowledge Discovery in Database* (KDD)

([https://www.researchgate.net/profile/Anselmo\\_Chaves\\_Neto/publication/221787546/figure/fig2/AS:305508767551493@1449850374653/Steps-in-the-KDD-process-Fayyad-et-al-1996\\_W840.jpg](https://www.researchgate.net/profile/Anselmo_Chaves_Neto/publication/221787546/figure/fig2/AS:305508767551493@1449850374653/Steps-in-the-KDD-process-Fayyad-et-al-1996_W840.jpg))

#### 2.4 Association Rule

Aturan asosiasi adalah pernyataan implikasi bentuk  $X \rightarrow Y$ , dimana  $X$  dan  $Y$  adalah *itemset* yang lepas (*disjoint*), yang memenuhi persyaratan  $X \cap Y = \{\}$ . (Prasetyo, 2012).

Aturan asosiasi berguna untuk menemukan hubungan penting antar *item* dalam setiap transaksi. Hubungan tersebut menandakan kuat tidaknya suatu aturan dalam asosiasi. *Association Rule* atau aturan asosiasi digunakan untuk mengidentifikasi *item – item* produk yang kemungkinan dibeli secara bersamaan.

Kekuatan aturan asosiasi dapat diukur dengan *support* dan *confidence*. *Support* digunakan untuk menentukan seberapa banyak aturan dapat diterapkan pada set data, sedangkan *confidence* digunakan untuk menentukan seberapa sering *item* didalam  $Y$  muncul dalam transaksi yang berisi  $X$ . (Prasetyo, 2012)

Jika *itemset* pada setiap transaksi tidak sering muncul (*infrequent*), maka kandidat yang tidak sesuai dengan nilai  $support \geq \text{minsup}$  tersebut harus segera dipangkas tanpa harus menghitung *confidence*-nya. Strategi umum digunakan oleh banyak algoritma penggalian aturan asosiasi adalah memecahkan masalah ke dalam dua pekerjaan utama, yaitu:

1. *Frequent itemset generation*

Tujuannya adalah mencari semua itemset yang memenuhi ambang batas minsup. *Itemset* itu disebut *itemset* frekuen (*Itemset* yang sering muncul).

2. *Rules Generation*

Tujuannya adalah mengekstrak aturan dengan *confidence* tinggi dari *itemset* frekuen yang ditemukan dalam langkah sebelumnya. Aturan ini kemudian disebut aturan yang kuat (*strong rule*). (Prasetyo, 2012)

## 2.5 Algoritma Apriori

Algoritma apriori termasuk jenis aturan asosiasi pada data mining. Selain apriori, yang termasuk pada golongan ini adalah metode *generalized rule induction* dan *algoritma hash based*. Aturan yang menyatakan asosiasi antara beberapa atribut sering disebut *affinity analysis* atau *market basket analysis*. (Vulandari, 2017)

Algoritma Apriori menggunakan pendekatan level yang baik dalam pembangkitan asosiasi, dimana setiap level berhubungan dengan sejumlah item yang dimiliki konsekuensi aturan. Awalnya, aturan dengan *confidence* yang tinggi yang mempunyai satu *item* diekstrak. Aturan ini kemudian digunakan untuk membangkitkan kandidat aturan yang baru. (Prasetyo, 2012)

Menurut (Vulandari, 2017) metode dasar analisis asosiasi terbagi menjadi dua tahap :

1. Analisa pola frekuensi tinggi

Tahap ini mencari kombinasi *item* yang memenuhi syarat *minimum* dari nilai *support* dalam *database*. Nilai *support* sebuah *item* diperoleh dengan rumus berikut :

$$\text{support}(A) = \frac{\text{Jumlah transaksi mengandung } A}{\text{Total transaksi}} \dots\dots\dots (2.1)$$

Sedangkan nilai *support* dari 2 *item* diperoleh dari rumus berikut :

$$\text{support}(A \cap B) = \frac{\text{Jumlah transaksi mengandung } A \text{ dan } B}{\text{Total transaksi}} \dots\dots\dots (2.2)$$

## 2. Pembentukan aturan asosiatif

Setelah semua pola frekuensi tinggi ditemukan, barulah dicari aturan asosiatif yang memenuhi syarat *minimum* untuk *confidence* dengan menghitung *confidence* aturan asosiatif  $A \rightarrow B$ . Nilai *confidence* dari aturan  $A \rightarrow B$  diperoleh dari rumus berikut :

$$\text{confidence} = P(A \cap B) = \frac{\text{Jumlah transaksi mengandung A dan B}}{\text{Jumlah transaksi mengandung A}} \dots (2.3)$$

Sifat *support* merupakan ukuran yang sangat penting dalam analisis asosiasi karena aturan yang sangat lemah nilai *support*-nya berarti asosiasi yang sangat jarang terjadi dalam set data. Aturan yang sangat rendah *support*-nya juga dipandang sebagai hal yang tidak menarik dalam perspektif bisnis karena memberikan keuntungan untuk mempromosikan *item* dimana pelanggan jarang membeli barang tersebut bersama – sama. Karena alasan itulah *support* biasanya digunakan sebagai basis untuk menghilangkan aturan yang tidak menarik (tidak dibutuhkan).

Sementara, *confidence* digunakan untuk mengukur keandalan dari inferensi yang dibuat oleh aturan. Untuk aturan  $X \rightarrow Y$ , nilai *confidence* yang tinggi menandakan banyaknya  $Y$  yang muncul dalam transaksi yang berisi  $X$ . *Confidence* juga memberikan cara untuk menemukan aturan asosiasi secara efisien. (Prasetyo, 2012)

### 2.6 Lift Ratio

*Lift ratio* adalah parameter penting selain *support* dan *confidence* dalam *association rule*. *Lift ratio* mengukur seberapa penting *rule* yang telah terbentuk berdasarkan nilai *support* dan *confidence*. *Lift ratio* merupakan nilai yang menunjukkan kevalidan proses transaksi dan memberikan informasi apakah benar produk A dibeli bersamaan dengan produk B.

*Lift* atau *improvement ratio* dapat dihitung dengan rumus :

$$\text{Lift} (A, B) = \frac{P(A \cup B)}{P(A)P(B)} \dots \dots \dots (2.4)$$

$Lift(A, B)$  = Korelasi antara A dan B

$P(A \cup B)$  = Jumlah kemunculan antara A dan B dibagi dengan total transaksi

$P(A)P(B)$  = Jumlah kemunculan A dikali dengan jumlah kemunculan B pada total transaksi.

Sebuah transaksi dikatakan valid apabila mempunyai nilai "*lift*" lebih dari 1, yang artinya dalam transaksi tersebut, produk A dan B benar - benar dibeli secara bersamaan.

## 2.7 Penelitian Sebelumnya

Beberapa riset yang telah dilakukan berkaitan dengan kasus asosiasi yang menggunakan metode apriori antara lain :

Penelitian yang berjudul "*Implementasi Data Mining pada Penentuan Merk Parfum yang banyak Terjual dengan Algoritma Apriori (Studi Kasus: Copa Gabana Parfum Cabang Sentolang Gresik)*" Oleh Maharani Indah Safitri. Copa Gabana Parfum adalah *House of Perfume* yang menjual berbagai macam aroma parfum dengan kualitas premium. Copa Gabana Parfum Cabang Sentolang selalu ramai pengunjung, tidak heran jika banyak merk parfum yang banyak terjual. Sebenarnya, ada banyak nama toko lain yang membuka usaha penjualan parfum. Kondisi seperti ini yang menjadikan banyaknya persaingan dalam dunia bisnis. Sistem yang akan dibuat dapat membantu dalam pengambilan suatu keputusan, salah satunya adalah tata letak barang seharusnya dapat diatur secara optimal sesuai pembelian merk parfum yang banyak terjual, sehingga dapat memudahkan pencarian merk parfum, peningkatan penjualan dan pengembangan strategi pemasaran. Buku penjualan tersimpan banyak data maupun informasi-informasi penting seperti: informasi tentang merk parfum apa saja yang banyak terjual. Teknik Data Mining dengan menggunakan metode algoritma Apriori (*Assosiation Rule*) memanfaatkan data transaksi pada buku penjualan. Sistem ini dapat digunakan sebagai solusi yang tepat untuk meningkatkan kecepatan proses pembentukan *frequent itemset*, Pembentukan 3 *itemset* dari 2 *itemset*, Perhitungan nilai *support* dan *confidence* serta pembentukan kaidah asosiasi dari 2 *itemset* maupun 3 *itemset*. Berdasarkan tabel hasil analisa data transaksi penjualan pada

tahun 2015/2016, terdapat *rule* yang mempunyai tingkat *support* tertinggi pada masing – masing periode dari tiga kali pengujian dengan *rule 3 itemset* adalah {Lovely KW 1 , Bubble Gum} => {Fantasy} sedangkan dari *rule 2 itemset* tertinggi adalah {Heavenly Kiss} => {Antonio Banderas} yang memiliki korelasi atau hubungan yang positif dengan nilai korelasi  $\geq 1$ .

Penelitian selanjutnya oleh Naziyatun Ni'mah dengan judul “*Penerapan Algoritma Apriori Dalam Menentukan Pola Pembelian Produk Bayi Yang Terjual Bersama Di Toko Mamiko GKB*”. Toko mamiko merupakan sarana atau tempat penyedia barang kebutuhan bayi sehari – hari mulai dari perlengkapan mandi, pakaian sampai kebutuhan lainnya. Banyaknya persaingan di dunia bisnis mengharuskan para pengembang untuk menentukan suatu strategi yang dapat meningkatkan penjualan khususnya penjualan produk bayi dengan memaksimalkan pelayanan kepada konsumen. Maka dari itu pemilik toko perlu sebuah sistem dalam penjualan agar lebih mudah untuk mendapatkan informasi tentang pola pembelian konsumen tersebut ke dalam beberapa aturan asosiasi, melalui teknik data mining menggunakan metode Apriori (*Assosiation Rule*). Dengan menggunakan data transaksi penjualan tersebut maka sistem ini dapat digunakan sebagai solusi yang tepat untuk meningkatkan kecepatan proses pembentukan *frequent itemset*, Pembentukan 3 *itemset* dari 2 *itemset*, perhitungan nilai *support* dan *confidence* serta pembentukan kaidah asosiasi dari 2-*itemset* dan 3-*itemset*. Berdasarkan tabel hasil analisa data transaksi penjualan pada tahun 2015/2016, terdapat *rule* yang mempunyai tingkat *support* tertinggi pada masing-masing periode dari tiga kali pengujian dengan *rule 3 itemset* adalah {bendera cair straw kotak, bendera cair straw botol} → {bendera cair coklat kotak} sedangkan *rule 2 itemset* tertinggi adalah {happy nappy pant M20} → {sgm soya 3 van 400} dan {bendera cair straw kotak} → {bendera cair coklat kotak} yang memiliki korelasi atau hubungan yang positif dengan nilai korelasi  $\geq 1$ .