

BAB II

LANDASAN TEORI

2.1. Tinjauan Pustaka

Sebagai bahan referensi peneliti melakukan analisis terhadap beberapa penelitian sebelumnya yang berkaitan dengan topik penelitian yang dilakukan. Berikut merupakan hasil dari analisis penelitian sebelumnya:

- a. Pada penelitian yang dilakukan oleh Ulya Anisatur Rosidah dan Hardian Oktavianto (2018) yang berjudul “Pencarian Pola Asosiasi Keluhan Pasien Menggunakan Teknik *Association Rule Mining*” menyimpulkan bahwa algoritma apriori dapat digunakan dalam mencari *rule* asosiasi dengan menggunakan prinsip apriori dalam mencari *frequent itemset*. Selain algoritma apriori terdapat algoritma *FP-Growth* yang menggunakan konsep *tree* dalam mencari *frequent itemset*. Pada dasarnya kedua algoritma tersebut menghasilkan *output* yang sama, perbedaannya terletak pada jumlah *rule* asosiasi (Rosyidah & Oktavianto, 2018).
- b. Pada penelitian yang dilakukan oleh Mohamad Fauzy, Kemas Rahmat Saleh W, dan Ibnu Asror (2015) yang berjudul “Penerapan Metode *Association Rule* Menggunakan Algoritma Apriori pada Simulasi Prediksi Hujan Wilayah Kota Bandung” menyimpulkan bahwa metode *association rule mining* menggunakan algoritma apriori dapat digunakan dalam sistem simulasi prediksi hujan dengan *rule* yang dihasilkan memiliki nilai *lift ratio* lebih dari 1,00 (Fauzy, Saleh, & Asror, 2015).
- c. Pada penelitian yang dilakukan Adie Wahyudi Oktavia Gama, I Ketut Gede Darma Putra, dan I Putu Agung Bayupati (2016) yang berjudul “Implementasi Algoritma Apriori Untuk Menemukan *Frequent Itemset* Dalam Keranjang Belanja” menyimpulkan bahwa algoritma apriori cocok digunakan dalam metode asosiasi yang diimplementasikan untuk mencari *frequent itemset* yang tersembunyi dalam keranjang belanja (Gama, Putra, & Bayupati, 2016).

- d. Pada penelitian yang dilakukan oleh Amrin (2017) dengan judul “Data Mining Dengan Algoritma Apriori untuk Penentuan Aturan Asosiasi Pola Pembelian Pupuk” disimpulkan bahwa akurasi pembentukan *rule* dipengaruhi oleh ambang batas nilai *support* dan *confidence*. Semakin tinggi *threshold* nilai *support* dan *confidence* maka semakin akurat *rule* yang didapatkan (Amrin, 2017).

2.2. Landasan Teori

2.2.1. Data Mining

Secara umum, *data mining* didefinisikan sebagai suatu proses mengelola data dalam jumlah besar untuk menemukan suatu informasi yang dapat digunakan dalam pemecahan suatu masalah. *Data mining* atau dapat disebut juga *Knowledge Discovery in Database* (KDD) merupakan suatu kegiatan yang meliputi proses pengumpulan data, pemakaian data historis dengan tujuan menggali informasi, pengetahuan, keteraturan, ataupun pola dalam suatu data dalam jumlah besar (Buulolo, 2020). *Data mining* menemukan pola yang tersembunyi dan informasi prediktif yang sebelumnya belum pernah diketahui dengan cara memindai basis data yang ada.

Adapun tahapan dalam *data mining* yang disebut CRISP-DM (*Cross-Industry Standard Process for Data Mining*) yang dapat dilihat pada gambar 2.1 (North, 2012). Beberapa langkah dalam CRISP-DM adalah sebagai berikut.

1. *Business Understanding*

Tahap ini merupakan tahap pemahaman terhadap proses bisnis yang ada dalam organisasi, pemahaman terhadap sistem yang berjalan serta *requirement*-nya untuk menyelesaikan masalah yang terjadi.

2. *Data Understanding*

Tahap ini merupakan tahap pengumpulan data dan pemahaman terhadap data sebelum dilakukan persiapan analisa data.

3. *Data Preparation*

Pada tahap ini data yang sudah dikumpulkan akan di seleksi dan diubah menjadi bentuk sesuai model yang akan digunakan selanjutnya.

4. *Modelling*

Pada permodelan data dilakukan analisa data yang sudah disiapkan dengan menerapkan algoritma atau metode yang telah ditentukan sesuai dengan kebutuhan yang diinginkan dan sebagai representasi pemecahan masalah.

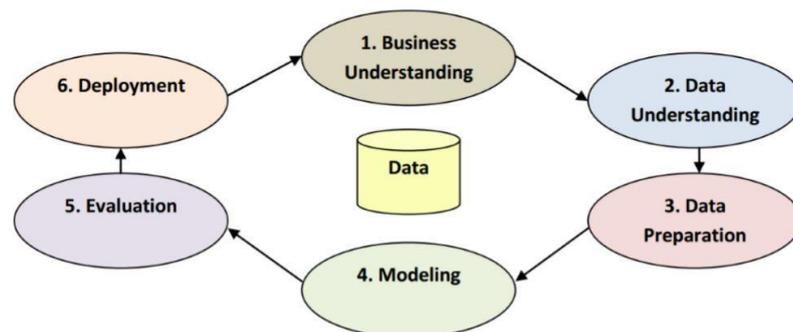
5. *Evaluation*

Setelah permodelan data dilanjutkan dengan evaluasi model data yang telah dibuat untuk mengetahui apakah sudah sesuai kebutuhan dan memberikan solusi terhadap masalah.

6. *Deployment*

Hasil dari model yang telah dievaluasi akan diterapkan dan dijadikan bentuk yang dapat diolah kembali.

CRISP-DM



Gambar 2.1 CRISP-DM

Berdasarkan fungsionalitas yang dapat dilakukan, *data mining* terbagi menjadi beberapa kelompok (Kusrini & Luthfi, 2009), yaitu:

1. *Description* (Deskripsi)

Seorang peneliti dan analis terkadang ingin mencoba menganalisis data untuk menggambarkan pola maupun kecenderungan yang terdapat dalam data. Deskripsi dari pola dan kecenderungan akan memberikan penjelasan terkait pola atau kecenderungan tersebut.

2. *Estimation* (Estimasi)

Pada estimasi terdapat variabel target estimasi yang berupa numerik. Nilai dari variabel target didapat berdasarkan model yang dibangun menggunakan data lengkap yang kemudian dijadikan nilai prediksi. Nilai variabel prediksi akan dipakai untuk membuat estimasi nilai dari variabel target.

3. *Prediction* (Prediksi)

Prediksi memiliki metode dan teknik hampir sama dengan klasifikasi dan estimasi. Pada beberapa kondisi, metode dan teknik klasifikasi dan estimasi dapat digunakan pula pada prediksi. Perbedaannya, prediksi nilai dari hasil akan terlihat di masa mendatang.

4. *Classification* (Klasifikasi)

Klasifikasi memiliki target variabel berupa kategori. Klasifikasi bekerja dengan cara mendefinisikan suatu kelompok data dengan menganalisis data yang telah diklasifikasikan berdasarkan aturan tertentu.

5. *Clustering* (Pengkusteran)

Pengkusteran merupakan aktivitas yang meliputi pengamatan, pengelompokan data, atau pembentukan kelas objek-objek yang memiliki ciri yang sama. Pengkusteran berbeda dengan klasifikasi. Pada pengkusteran tidak ada variabel target seperti pada klasifikasi, estimasi, dan prediksi.

6. *Association* (Asosiasi)

Asosiasi merupakan suatu kejadian terkait dengan peristiwa tunggal. Dalam *data mining*, tugas asosiasi adalah menemukan atribut yang timbul dalam satu waktu. Asosiasi dalam dunia bisnis juga sering disebut analisi keranjang belanja.

6.1.1. *Association Rule*

Association rule adalah salah satu metode dalam *data mining* yang digunakan untuk menentukan hubungan antar *item* dalam suatu kumpulan data (*dataset*) yang telah ditentukan (Witten & Frank, 2005). Menurut (Han, 2009), *association rule* merupakan suatu teknik untuk menemukan hubungan antar *item* dalam suatu *dataset* tertentu. *Association rule* sering digunakan pada toko untuk menganalisis pola pembelian konsumen, untuk menemukan *item* apa saja yang sering dibeli bersamaan. Penerapan *association rule* tersebut dapat membantu pemilik toko dalam mengatur penempatan barang, membuat promosi, dan mengatur persediaan barang.

Pembentukan *association rule* ditentukan dengan dua parameter, yaitu *support* dan *confidence* (Noranita & Bahtiar, 2010). *Support* (nilai penunjang) merupakan besar persentase kemunculan kombinasi *item*, sedangkan *confidence* (nilai kepastian) adalah ukuran kuat tidaknya hubungan antar *item* dalam aturan asosiatif. Aturan asosiasi harus memenuhi nilai minimum *support* dan minimum *confidence* (Pramudiono, 2003).

Adapun metodologi dalam pembentukan *association rule* terbagi menjadi dua tahap (Kusrini & Luthfi, 2009), yaitu:

1. Analisis Pola Frekuensi Tinggi

Pada tahap ini akan dicari kombinasi *item* yang memenuhi syarat minimum *support* atau *threshold*. Nilai *support* merupakan besaran persentase satu *item* atau kombinasi *item* yang ada dalam

database. Nilai *support* sebuah *item* didapatkan menggunakan rumus berikut.

$$Support(X) = \frac{\Sigma \text{Transaksi untuk } X}{\Sigma \text{Transaksi}} \dots\dots\dots (2.1)$$

Dalam artian bahwa nilai *support* didapatkan dari jumlah transaksi yang mengandung *item* X dibagi dengan total transaksi yang ada. Untuk nilai *support* dari dua *item* dapat diperoleh menggunakan rumus berikut.

$$Support(X \cap Y) = \frac{\Sigma \text{Transaksi untuk } X \& Y}{\Sigma \text{Transaksi}} \dots\dots\dots (2.2)$$

Dalam hal ini menjelaskan bahwa nilai *support* kombinasi dua *item* diperoleh dengan cara membagi jumlah transaksi yang mengandung *item* X dan Y dengan total seluruh transaksi.

2. Pembentukan Aturan Asosiatif

Setelah pola frekuensi tinggi didapatkan, dilanjutkan dengan mencari aturan asosiasi yang memenuhi syarat minimum *confidence*. Setiap kombinasi *item* akan dibentuk aturan asosiasi yang memungkinkan untuk kemudian dihitung nilai *confidence*-nya. Nilai *confidence* dapat diperoleh dengan rumus berikut.

$$Conf(X \rightarrow Y) = \frac{Support(X \cap Y)}{Support X} \dots\dots\dots (2.3)$$

Rumus diatas menjelaskan bahwa nilai *confidence* untuk X maka Y didapatkan dengan membagi nilai *support* X dan Y dengan nilai *support* X. Aturan asosiasi dapat diambil dengan syarat melebihi minimum *confidence*.

6.1.2. Algoritma Apriori

Algoritma apriori merupakan algoritma yang diusulkan pada tahun 1994 oleh Agrawal dan Srikant guna menentukan *frequent itemsets* untuk aturan 9 asosiasi Boolean. Algoritma ini juga tergolong pada 12 data mining dalam jenis aturan asosiasi.

Menurut (Rizkiana, Nurmalasari, Arisawati, & Hartini, 2019), algoritma apriori merupakan algoritma pengambilan data dengan aturan asosiatif untuk menentukan hubungan asosiatif suatu kombinasi *item*. Algoritma apriori menggunakan pendekatan *level – wise search*, yang mana *k-itemset* digunakan dalam menentukan $(k+1)$ *itemset* (Aprianti, Hafidz, & Rizani, 2017).

Tahapan algoritma apriori terbagi menjadi beberapa tahap (Buulolo, 2020), yaitu:

1. Pembentukan kandidat *itemset*.

Kandidat *k-itemset* dibentuk berdasarkan kombinasi $(k-1)$ -*itemset* yang diperoleh dari iterasi sebelumnya. Salah satu cara kerja algoritma apriori adalah dengan memangkas kandidat *k-itemset* yang subsetnya berisi $k-1$ item tidak termasuk dalam pola frekuensi tinggi dengan panjang $k-1$.

2. Perhitungan support dari tiap kandidat *k-itemset*

Setiap kandidat *k-itemset* dihitung nilai *support*-nya. Nilai *support* dari tiap kandidat *k-itemset* didapat dengan membagi jumlah transaksi yang memuat kandidat *itemset* dengan jumlah seluruh transaksi.

3. Tetapkan pola frekuensi tinggi

Pola frekuensi tinggi yang memuat *k-item* atau *k-itemset* ditentukan berdasarkan kandidat *k-itemset* yang nilai *support*-nya lebih besar dari kandidat *k-itemset* yang *support*-nya lebih besar dari minimum *support*.

4. Jika sudah tidak ada pola frekuensi tinggi baru maka seluruh proses dihentikan.

Aturan atau pola yang terbentuk menggunakan algoritma apriori akan diuji kekuatannya menggunakan *lift ratio*. *Lift ratio* bertujuan mencari nilai yang mengukur seberapa besar kekuatan aturan asosiatif yang terbentuk. Rumus yang digunakan untuk mencari nilai *lift ratio* adalah sebagai berikut.

$$\text{Lift Ratio}(X \rightarrow Y) = \frac{\text{Confidence}(X \rightarrow Y)}{\text{Bernchmark Confidence}} \dots\dots\dots (2.4)$$

Bernchmark confidence didapatkan dengan menghitung nilai *support consequent*-nya.

$$\text{Berchmark Confidence} = \frac{\Sigma \text{Transaksi untuk } Y}{\Sigma \text{Transaksi}} \dots\dots\dots (2.5)$$

