

BAB II

KAJIAN LITERATUR

2.1 Landasan Teori

2.1.1 *Data Mining*

Data Mining merupakan suatu proses menggali sekumpulan data dan mengubahnya dalam bentuk informasi yang bermanfaat bagi pengguna. *Data mining* memiliki beberapa teknik yang terkenal dan sering digunakan oleh peneliti, diantaranya seperti *clustering*, *classification*, *association*, dan beberapa perkembangan teknik sesuai dengan perubahan kecenderungan data pada saat ini.

Data mining merupakan langkah dalam proses KDD yang terdiri dari penerapan analisis data dan algoritma penemuan yang menghasilkan penghitungan pola atau model tertentu melalui data. Fayyad *et al.* (1996) di dalam penelitiannya menjelaskan bahwa terdapat beberapa langkah di dalam proses KDD (*Knowledge Discovery in Database*) diantaranya secara berurutan *selection*, *preprocessing*, *transformation*, *DATA MINING*, dan *Interpretation/evaluation*.

Menurut Fayyad *et al.* (1996) di dalam buku yang ditulis Suyanto (2017), tugas-tugas *data mining* dapat dikelompokkan ke dalam enam kelompok berikut ini:

- Klasifikasi (*classification*): men-generalisasi struktur yang diketahui untuk diaplikasikan pada data-data baru. Misalkan, klasifikasi penyakit ke dalam sejumlah jenis, klasifikasi email ke dalam spam atau bukan.
- Klasterisasi (*clustering*): mengelompokkan data, yang tidak diketahui label kelasnya, ke dalam sejumlah kelompok tertentu sesuai dengan ukuran kemiripannya.

- Regresi (*regression*): menemukan suatu fungsi yang memodelkan data dengan galat (kesalahan prediksi) seminimal mungkin.
- Deteksi anomali (*anomaly detection*): mengidentifikasi data yang tidak umum, bisa berupa *outlier*, perubahan atau deviasi yang mungkin sangat penting dan perlu investigasi lebih lanjut.
- Pembelajaran aturan asosiasi (*association rule mining*) atau pemodelan kebergantungan (*dependency modeling*): mencari relasi antar variabel.
- Perangkuman (*summarization*): menyediakan representasi data yang lebih sederhana, meliputi visualisasi dan pembuatan laporan.

Selain itu, beberapa pengaplikasian *data mining* dapat digunakan dalam berbagai bidang, diantaranya adalah (Suyanto, 2017) :

1. Marketing dan Bisnis

Perusahaan selalu memiliki data-data yang dapat dimanfaatkan dalam strategi marketing dan bisnis. Seperti melakukan pemilihan *vendor* yang tepat, penggunaan strategi penjualn produk, serta penggunaan aturan dalam kartu loyalti perusahaan. Berikut ini contoh aplikasi *data mining* dalam marketing dan bisnis:

a. *Market Basket Analysis*

Analisis keranjang belanja yang memungkinkan dapat mengetahui kebiasaan belanja konsumen. MBA dikenal juga dengan *association rule* (aturan asosiasi) yaitu salah satu konsep dalam *data mining* yang berusaha menemukan asosiasi atau keterkaitan data.

b. *Recommender System*

Merupakan sistem rekomendasi dari beberapa variabel sehingga dapat dilakukan pemilihan lebih tepat seperti dalam pemilihan rekomendasi *supplier* mana yang menunjukkan performansi baik. Aplikasi ini biasanya menggunakan teknik kalsterisasi ataupun klasifikasi.

c. *Churn Prediction*

Merupakan analisis dari loyal apa tidaknya suatu pelanggan berdasarkan variabel-variabelnya. Sebagai contoh perusahaan telokomunikasi yang memiliki pelanggan hampir ratusan juta ingin melihat pelanggan apakah tetap loyal apa tidak, dengan menggunakan teknik *data mining* sehingga hal

tersebut menjadi mudah dan cepat dilakukan. Teknik yang biasa digunakan adalah teknik klasifikasi dan kalsterisasi.

d. *Fraud Detection*

Fraud detection digunakan dalam menemukan pelanggan yang mungkin melakukan kecurangan. Sejumlah data yang besar apabila dilakukan secara manual akan membutuhkan biaya dan waktu yang lama sehingga penggunaan teknik *data mining* dapat mempercepat dalam penemuan kecurangan di dalam suatu basis data pelanggan. Sistem ini dapat dibangun menggunakan teknik *anomaly detection*.

2. Sains dan Teknik

Beberapa teknik *data mining* dapat digunakan dalam dunia sains dan teknik untuk menyelesaikan permasalahan yang kompleks, seperti genetika, medis, teknik elektro, dan sebagainya.

3. Seni dan Hiburan

Data mining juga dapat diaplikasikan ke dalam seni dan hiburan, seperti menentukan lagu kesukaan yang sering kali diputar ataupun merekomendasikan jenis lagu ataupun video yang memiliki kemiripan yang sama dengan lagu atau video favorit.

2.1.2 Association Rule Mining

Aturan Asosiasi atau disebut dengan *association rule* merupakan salah satu teknik *data mining* yang berguna dalam mencari aturan asosiatif dari suatu barang. Aturan Asosiasi dibentuk dengan cara menganalisis pola data yang sering muncul (*frequent pattern*) dan dengan menggunakan parameter *support* dan *confidence* untuk mengidentifikasi hubungan yang paling penting. *Support* adalah indikasi yang menunjukkan seberapa sering suatu item muncul di dalam database. Sedangkan *Confidence* menunjukkan berapa kali pernyataan tersebut tersebut benar (Gunadi & Sensuse, 2012).

1) *Support*

Support adalah suatu ukuran yang menunjukkan seberapa besar tingkat dominasi suatu *item/item set* dari keseluruhan transaksi. Ukuran ini akan menentukan apakah suatu *item/item set* layak untuk dicari *confidece* selanjutnya.

$$Support = P(X \cap Y) = \frac{\sum \text{Transaksi yang mengandung } X \text{ dan } Y}{\text{Jumlah Transaksi}} \quad (2.1)$$

2) Confidence

Confidence adalah suatu ukuran yang menunjukkan hubungan antar 2 *item* secara *conditional* seperti seberapa sering *item* *Y* dibeli jika orang membeli *X*.

$$Confidence = P(Y/X) = \frac{P(X \cap Y)}{\sum \text{Transaksi yang mengandung } X} \quad (2.2)$$

3) Lift Ratio

Lift Ratio adalah salah satu cara yang baik untuk melihat kuat atau tidaknya aturan asosiasi yang terbentuk. *Lift Ratio* juga merupakan nilai yang menunjukkan kevalidan proses transaksi dan memberikan informasi apakah benar produk *X* dibeli bersamaan dengan produk *Y*. Nilai *Lift* mengukur kemungkinan *X* dan *Y* terjadi bersama-sama dibagi dengan kemungkinan *X* dan *Y* terjadi jika mereka adalah peristiwa independen.

$$Lift Ratio = \frac{Support(X \cap Y)}{Support(X) \cdot Support(Y)} \quad (2.3)$$

Dalam *data mining*, aturan asosiasi berguna untuk menganalisis dan memprediksi perilaku pelanggan dan berperan penting dalam analisis data keranjang belanja, pengelompokan produk, desain katalog dan tata letak toko. Terdapat beberapa algoritma yang diterapkan pada teknik *association rule*. Berdasarkan penelitian yang dilakukan oleh Dhanalakahmi & Porkodi (2017), beberapa algoritma yang pernah digunakan oleh peneliti diantaranya yaitu Algoritma *Apriori*, *Eclat*, dan *FP-growth*

2.1.3 Algoritma Apriori

Algoritma Apriori merupakan suatu algoritma yang diajukan oleh R. Agrawal dan R. Srikant pada tahun 1994 pada suatu penelitian yang berguna untuk menemukan *frequent itemset* untuk *Boolean association rule*. Nama algoritma apriori berdasarkan fakta bahwa algoritma tersebut menggunakan pengetahuan sebelumnya (*prior knowledge*) dari pencarian *frequent itemset mining* (Han et al., 2012).

Aturan asosiasi (*Association rule*) biasanya diperlukan untuk memenuhi minimum *support* yang ditentukan *user* dan minimum *confidence* yang ditentukan pengguna pada saat yang sama. Pembuatan aturan asosiasi biasanya dibagi menjadi dua langkah terpisah (Dhanalakshmi & Porkodi, 2017):

- Ambang batas minimum *support* diterapkan untuk menemukan semua *frequent itemset* di dalam database.
- Batasan minimum *confidence* diterapkan pada *frequent itemset* digunakan untuk membentuk aturan.

Di dalam Han *et al.* (2012) menuliskan ada dua langkah di dalam menggunakan Algoritma Apriori, yaitu langkah *join* dan *prune*.

1. *Join Step* : Untuk mencari suatu itemset yang baru (L_k), beberapa kandidat dari k-itemset dibentuk oleh penggabungan itemset sebelumnya (L_{k-1}). Pada proses ini setiap item dikombinasikan dengan item yang lainnya sampai tidak terbentuk kombinasi lagi.
2. *Prune Step* : Proses ini adalah proses pemangkasan itemset yang telah dibentuk sebelumnya berdasarkan dengan minimum *support* yang ditentukan. Sehingga menghasilkan suatu itemset yang sesuai dengan keinginan oleh *user* (Ariana & Asana, 2013)

Sedangkan di dalam Tampubolon *et al.* (2013) Algoritma Apriori memiliki beberapa langkah dalam pencarian aturan asosiasi. Langkah ini hampir sama dengan sebelumnya, hanya saja pendefinisian dalam langkah ini yang berbeda:

1. Pembentukan Kandidat Itemset : pemangkasan kandidat k-itemset yang subsetnya berisi k-1 item. Dimana kandidat itemset dibentuk dari kombinasi (k-1)-itemset yang berasal dari iterasi sebelumnya.
2. Perhitungan *support* dari tiap kandidat k-itemset.
3. Menetapkan pola frekuensi tinggi
4. Bila tidak didapat pola frekuensi tinggi maka seluruh proses dihentikan.

2.1.4 Algoritma FP-Growth

FP-growth merupakan salah satu alternatif algoritma yang dapat digunakan untuk menentukan himpunan data yang paling sering muncul (*frequent itemset*) dalam sebuah kumpulan data (Gunadi & Sensuse, 2012). Pada penelitian yang dilakukan oleh Gunadi

& Sensuse (2012) tersebut menuliskan bahwa terdapat 2 tahap proses yang dilakukan dalam menentukan *frequent itemset*.

1) Pembuatan *FP-Tree*

FP-Tree merupakan struktur data yang dimampatkan. *FP-Tree* dibangun dengan memetakan data transaksi ke dalam setiap lintasan tertentu dalam *FP-Tree* karena dalam setiap transaksi mungkin ada transaksi yang memiliki *item* yang sama, maka lintasannya pemampatan dengan struktur data *FP-Tree* semakin efektif.

2) Penerapan Algoritma *FP-Growth*

a. Pembangkitan *Conditional Pattern Base*

Conditioneal Pattern Base merupakan sub database yang berisi *prefix path* (lintasan *prefix*) dan *suffix pattern* (pola akhiran). Pembangkitan *Conditional Pattern Base* didapatkan melalui *FP-Tree*.

b. Pembangkitan *Conditional FP-Tree*

Pada tahap *Conditional Pattern Base* dicari *support count*, kemudian setiap *item* yang memiliki jumlah *support count* yang lebih besar sama dengan minimum *support count* s akan dibangkitkan dengan *Conditional FP-Tree*.

c. Pencarian *Frequent Itemset*

Conditional FP-Tree yang mempunyai lintasan tunggal (*single path*), maka didapatkan *frequent item set* dengan melakukan kombinasi *item* untuk setiap *Conditional FP-Tree*.

2.1.5 Proses *Knowledge Discovery in Data Mining (KDD)*

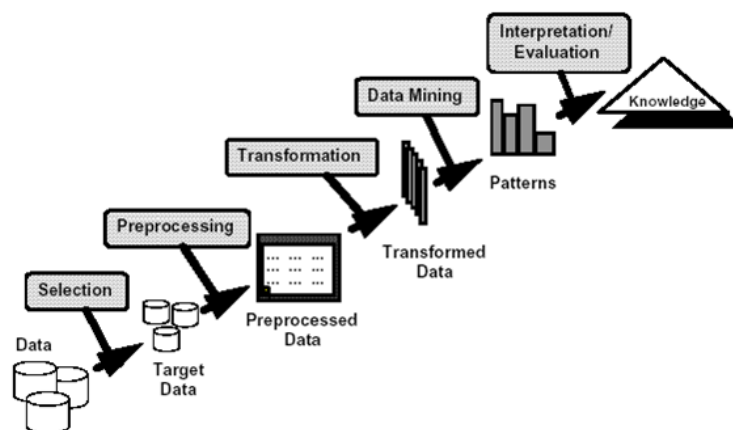
Metodologi pada penelitian ini menggunakan proses KDD (*Knowledge Discovery in Database*). Proses ini menjelaskan secara sistematis dalam mencari suatu hubungan baru di dalam *Market Basket Analysis* menggunakan beberapa tahap pengolahan data. Fayyad *et al.* (1996) di dalam penelitiannya menjelaskan bahwa terdapat beberapa langkah di dalam proses KDD (*Knowledge Discovery in Database*) diantaranya secara berurutan *selection, preprocessing, transformation, data mining, dan Interpretation/evaluation*.

a) **Selection** : Data yang didapatkan dilakukan proses pemilihan terlebih dahulu.

Dengan adanya data *selection*, proses pengolahan akan menjadi lebih baik sesuai dengan tujuan penelitian yang akan dicapai.

- b) **Preprocessing** : Merupakan beberapa proses persiapan data sebelum dilakukannya proses *data mining*. Data yang dilakukan *preprocessing* biasanya dilakukan dengan beberapa tahap seperti *cleaning, reduction, integration*.
- c) **Transformation** : Data harus dilakukan transformasi sebelum dilakukan pengolahan menggunakan *data mining*. Hal ini bertujuan untuk menyesuaikan data yang diolah berdasarkan algoritma dan *software* yang digunakan di dalam pengolahan data.
- d) **Data Mining** : Proses pengolahan data berdasarkan algoritma sesuai dengan teknik *data mining*. Algoritma yang digunakan pada penelitian yaitu Algoritma *FP-Growth* yang merupakan salah satu algoritma di dalam metode *association rule mining*. Sehingga hasil luaran dari teknik ini berupa *rule* dari data yang diolah dan parameter *support, confidence, lift ratio*.
- e) **Interpretation/evaluation** : Merupakan proses menginterpretasikan hasil *rule* yang didapatkan dari teknik *data mining*. Pada bagian ini diberikan rekomendasi strategi penjualan produk berdasarkan hasil *rule* yang didapat dari hasil pengolahan data. Selanjutnya proses evaluasi dilihat dari parameter algoritma yang digunakan (*support, confidence, lift*).

Berikut ini gambar ilustrasi dari proses *Knowledge Discovery in Database*.



Gambar 2.1 Proses *Knowledge Discovery in Database*

Sumber : Fayyad *et al.* (1996)

2.1.6 Analisis Overall Variability of Association Rule (OCVR)

Overall Variability of Association Rule (OCVR) merupakan indikator baru yang pertama kali dikenalkan oleh Papavasileiou dan Tsadiras pada tahun 2011 dimana indikator ini diterapkan pada supermarket yang berfokus dalam analisis produk keranjang belanja dengan asumsi variabilitas yang tinggi pada perilaku konsumen dalam membeli produk. Parameter variabilitas dari aturan asosiasi yang dihubungkan dengan perubahan dalam kebiasaan belanja selama beberapa waktu, dapat berkontribusi lebih lanjut untuk efisiensi analisis keranjang belanja dan strategi pemasaran yang tepat untuk mempromosikan penjualan (Papavasileiou & Tsadiras, 2011).

Hasil dari *support*, *lift*, dan *confidence* pada hasil aturan asosiasi sering kali berbeda-beda setiap periodenya. *Rules* yang terbentuk setiap bulan transaksi mengalami naik turun berdasarkan kebiasaan belanja yang tidak pasti. OCVR berperan dalam menghasilkan nilai indeks variabilitas dari perubahan-perubahan parameter *confidence* dan *lift*. Perhitungan indeks variabilitas ini mengambil konsep dari analisis statistik standar deviasi. *Variability Index CV* dihitung dengan persamaan berikut.

$$CV = \frac{S}{\bar{X}} \quad (2.4)$$

Keterangan : CV = Variability Index
 S = Standar Deviasi
 \bar{X} = Rata-rata

Indeks variabilitas ini digunakan untuk menghitung variabilitas indeks dari *lift* dan *confidence* dapat disebut sebagai *Index Variability Lift* (CVL) dan *Index Variability Confidence* (CVC). Sehingga, analisis *Overall Variability of Association Rule* (OCVR) berasal dari penggabungan antar CVL dan CVC dan dapat dirumuskan sebagai berikut.

$$OCVR = \frac{CVL+CVC}{2} \quad (2.5)$$

Hasil dari analisis OCVR menunjukkan tingkat variasi aturan asosiasi dari satu periode ke periode selanjutnya. Sehingga, hasil tersebut menjadi tambahan pertimbangan dalam menentukan *rules* apa saja yang menjadi prioritas di dalam *Market Basket Analysis* (MBA). Selain itu, *rule* dengan nilai OCVR yang tinggi yang menandakan perubahan yang tinggi setiap periodenya dapat dianalisis lebih lanjut penyebab dari perubahan tersebut dan pengambilan keputusan dalam penentuan strategi yang baik dalam menurunkan tingkat variabilitas tersebut.

Hasil dari OCVR yang berhubungan dengan perubahan dalam kebiasaan belanja selama beberapa periode waktu, dapat berkontribusi lebih lanjut dalam efisiensi dari *market basket analysis* dan strategi pemasaran yang tepat untuk mempromosikan penjualan. Strategi tersebut dapat berupa perubahan tata letak produk di rak penjualan, kebijakan promosi diskon, atau juga dapat membantu dalam sistem rekomendasi.

2.2 Penelitian Terdahulu

Metode *association rule mining* masih menjadi topik yang aktif di dalam lingkungan peneliti. Pengembangan algoritma baru dan menemukan pengetahuan adalah salah satu penelitian yang aktif dilakukan peneliti. Sandhu *et al.* (2011) menemukan metodologi baru dalam mengembangkan metode *association rule* berdasarkan asumsi *profit and quantity*. Algoritma Apriori digunakan pada penelitian untuk menghasilkan seperangkat aturan asosiasi dari database. Penelitian tersebut mengusulkan sebuah pendekatan yang efisien berdasarkan faktor bobot dan utilitas untuk penambahan data yang efektif dari *association rule* dengan utilitas yang tinggi.

Memahami pola belanja konsumen di dalam suatu supermarket dengan menggunakan *market basket analysis* banyak dilakukan beberapa peneliti. Akan tetapi, semakin berkembangnya perilaku belanja konsumen yang memiliki kecenderungan yang berubah-ubah setiap waktu dan ketidakpastian menjadi tantangan tersendiri bagi peneliti untuk mengembangkan algoritma yang sesuai agar hasil yang didapatkan efisien dan dapat diterapkan pada keadaan nyata. Seperti penelitian yang dilakukan Kaur & Kang (2016). Penelitian tersebut menjelaskan bahwa beberapa algoritma pada *association rule* yang ada hanya bekerja pada data yang statis dan tidak dapat menangkap perubahan data dengan waktu. Algoritma yang diusulkan pada penelitian ini tidak hanya data statis tetapi juga memberikan cara baru untuk memperhitungkan perubahan akun yang terjadi pada

data. Penelitian tersebut membahas teknik data mining seperti *association rule* dan menyediakan algoritma baru yang dapat membantu menentukan kebiasaan konsumen dan membantu dalam meningkatkan penjualan.

Semakin berkembangnya metode *Market Basket Analysis*, para peneliti mulai mengembangkan beberapa algoritma yang dapat menanggulangi beberapa kekurangan pada metode tersebut. Seperti metode pengembangan algoritma yang di ajukan oleh Chen *et al.* (2005). Metode *Market Basket Analysis* yang ada terdapat kemungkinan kegagalan dalam menemukan pola kebiasaan belanja konsumen jika diterapkan pada lingkungan toko yang banyak (*multiple-store environment*). Dikarenakan asumsi bahwa produk yang sedang dipertimbangkan berada di rak sepanjang waktu di semua toko dan belum tentu satu toko dengan lainnya memiliki susunan rak yang sama. Jurnal ini bertujuan untuk mengajukan metode baru yang dapat mengatasi kelemahan tersebut. Hasil dari penelitian menunjukkan bahwa metode yang diusulkan efisien secara komputasi, dan bahwa ia memiliki keunggulan dibandingkan metode tradisional ketika toko memiliki ukuran yang beragam, bauran produk berubah dengan cepat dari waktu ke waktu, jumlah toko dan periode yang lebih besar.

Selain itu, kabiasaan belanja konsumen yang bervariasi setiap waktu menjadi alasan utama pada penelitian yang dilakukan oleh Papavasileiou & Tsadiras (2011). Penelitian tersebut mengenalkan konsep variabilitas dari aturan asosiasi produk menggunakan asumsi indikator baru yang disebut *Overall Variability of Association Rule* (OCVR). Parameter variabilitas dari aturan asosiasi dihubungkan dengan perubahan kebiasaan belanja konsumen selama beberapa waktu, dapat berkontribusi dalam analisis keranjang belanja secara efisien pada penelitian.

Tabel 2.1 Perkembangan Metode *Association Rule*

Penulis	Tujuan	Metode
Kaur & Kang (2016)	Pengembangan Algoritma	Ketidakpastian dan Perubahan kebiasaan pelanggan
Sandhu <i>et al.</i> (2011)	Pengembangan Algoritma	<i>Profit</i> dan <i>Quantity</i>
Chen <i>et al.</i> (2005)	Pengembangan Algoritma	<i>Multiple store environment</i>

Papavasileiou & Tsadiras (2011)	Pengembangan Indikator	Perubahan dan variabilitas hasil aturan asosiasi
Penelitian yang diajukan (2018)	Pengaplikasian Indikator	Perubahan kebiasaan belanjaa konsumen

Dengan demikian, dari beberapa penjelasan terhadap penelitian terdahulu, penelitian ini bertujuan untuk menentukan strategi marketing yang tepat berdasarkan asumsi bahwa kebiasaan belanja konsumen yang berubah-ubah setiap waktu. Penelitian yang sesuai dengan keadaan pada perubaahn kebiasaan belanja konsumen adalah penelitian yang dilakukan oleh Papavasileiou & Tsadiras (2011). Oleh karena itu, penelitian ini menggunakan indikator *Overall Variability of Association Rule* (OCVR) yang dikenalkan oleh Papavasileiou & Tsadiras (2011) dalam menerapkan *Market Basket Analysis*. Penelitian ini diharapkan mendapatkan aturan asosiasi yang efektif dalam pengambilan keputusan pada strategi marketing. Beberapa strategi marketing yang dilakukan seperti, tata letak barang di supermarket, mengubah lokasi barang pada rak, keputusan dalam menentukan barang diskon, hingga kebijakan dalam membuat katalog.