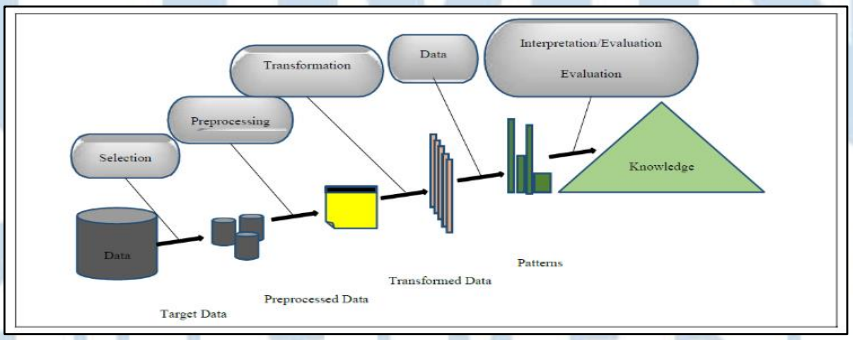
BAB 2
LANDASAN TEORI

2.1 Data Mining

Data mining adalah proses ekstraksi data dari sejumlah data yang besar menggunakan teknik statistik dan matematika sampai memanfaatkan teknik kecerdasan buatan modern [18]. Data mining merupakan proses logis untuk menemukan informasi bermanfaat. Setelah informasi serta pola ditemukan, komponen tersebut dapat digunakan sebagai instrumen pendukung saat mengambil keputusan ketika melakukan pengembangan bisnis [19]. Instrumen pada data mining memiliki kemampuan untuk memberikan jawaban terhadap berbagai pertanyaan yang berkaitan dengan permasalahan bisnis yang sulit untuk dipecahkan secara konvensional. Data mining juga dapat berfungsi untuk melakukan ramalan terhadap tren masa depan sehingga memungkinkan pemilik bisnis untuk membuat keputusan yang efektif, proaktif, dan dinamis. Pengolahan data dengan mempergunakan teknik data mining juga dapat menghasilkan pengetahuan yang sepadan dengan keinginan [20]. Data mining merupakan proses penggalian dan pemanfaatan data untuk mengidentifikasi pola atau hubungan dalam kumpulan data yang berskala besar. Hasil dari proses data mining dapat digunakan untuk mengevaluasi keputusan yang akan diambil di masa mendatang [21].

2.1.1 Tahapan data mining



Gambar 2.1 Proses tahapan data mining[12]

Pada Gambar 2.1, menjelaskan tentang tahapan data mining yaitu:

1. *Selection*

Data akan mengalami seleksi, termasuk pemilihan atribut yang akan digunakan pada tahap berikutnya. Data yang terpilih dari kumpulan data operasional akan disimpan dalam arsip terpisah untuk proses data mining.

2. *Pre-processing*

Setelah seleksi data, langkah *Cleaning* dilakukan pada data terpilih. Tahap ini mencakup pemrosesan awal dan eliminasi data yang tidak konsisten. Hal ini bertujuan untuk menghapus duplikasi dan meningkatkan kualitas data sebelum masuk ke proses data mining.

3. *Transformation*

Proses transformasi melibatkan perubahan format data terpilih ke dalam bentuk yang sesuai untuk analisis lebih lanjut.

4. Data Mining

Data Mining merupakan proses pencarian suatu informasi pada data dengan menggunakan metode tertentu. Pemilihan metode atau algoritma memiliki peranan penting karena berdampak pada tujuan dan jalannya proses data mining.

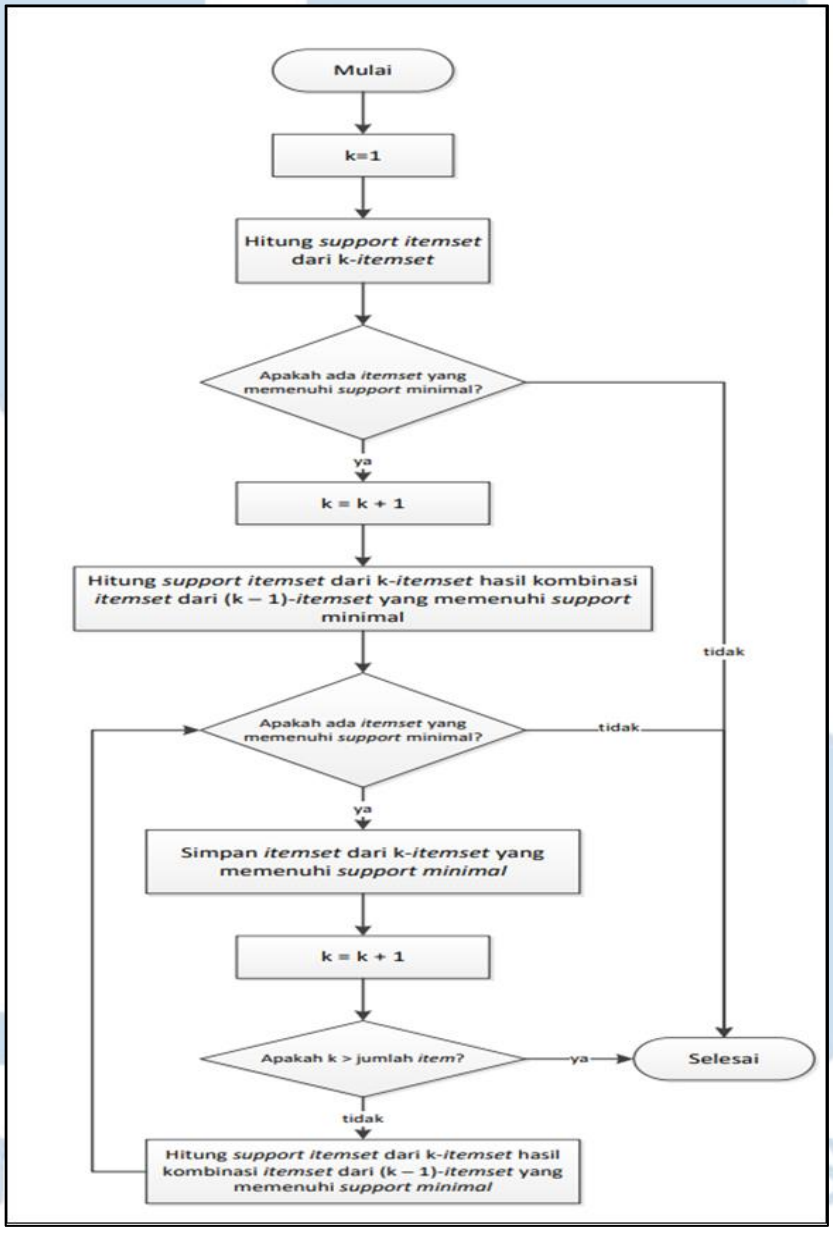
5. *Interpretation/Evaluation*

Informasi yang dihasilkan dari data mining diidentifikasi. Lalu, informasi yang ditemukan ditampilkan dalam bentuk hasil yang diinginkan.

2.2 Algoritma Apriori

Pada data mining algoritma apriori mencakup dalam peraturan asosiasi. Aturan yang mengungkapkan asosiasi diantara berbagai atribut ini seringkali dikatakan sebagai *affinity analysis* ataupun analisis keranjang pasar [22]. Aturan asosiasi ataupun *association rule* mining merujuk pada teknik data mining yang digunakan untuk menemukan peraturan gabungan item. Suatu tahapan menganalisis asosiasi yang menarik perhatian banyak periset dalam melakukan

pengembangan algoritma secara efisien yakni analisis pola frekuensi tinggi. Kriteria standar yang dapat mengetahui penting ataupun tidak peraturan asosiasi yakni merupakan *nilai support* dan *nilai confidence*. Nilai *Support* merupakan persentase dari kombinasi item tersebut dari database, sementara nilai *confidence* berarti kuatnya suatu relasi antar-item pada peraturan asosiasi [23]. Untuk alur cara kerja algoritma Apriori dapat dilihat pada Gambar 2.2.



Gambar 2.2 Flowchart Apriori [24]

Untuk menghitung nilai *support* pada suatu item didapatkan melalui penggunaan Rumus 2.1 sebagai berikut.

$$Support(A) = \frac{\sum \ transaksi\ mengandung\ A}{total\ transaksi} x100\% \quad (2.1)$$

Sedangkan untuk menghitung nilai *support* pada 2 (dua) item didapatkan melalui penggunaan Rumus 2.2 sebagai berikut.

$$Support = \frac{\sum \ transaksi\ mengandung\ A\ dan\ B}{total\ transaksi} x100\% \quad (2.2)$$

Dan untuk menghitung nilai *confidence* bisa didapatkan melalui penggunaan Rumus 2.3 sebagai berikut.

$$Confidence = \frac{\sum \ transaksi\ mengandung\ A\ dan\ B}{\sum \ transaksi\ mengandung\ A} x100\% \quad (2.3)$$

Untuk menentukan peraturan asosiasi yang ingin ditentukan sehingga harus dilakukan pengurutan mengacu pada *support* x *confidence*, serta mengambil peraturan sejumlah n aturan yang mempunyai hasil paling besar.

2.3 Product Bundling

Product Bundling adalah suatu strategi pemasaran yang telah menjadi fokus kajian dalam literatur penelitian dan juga diterapkan secara luas dalam industri *e-commerce* [25]. Dengan pertambahan jumlah produk dan kemungkinan kombinasi *bundling* yang sangat besar, diperlukan pendekatan algoritmik untuk menentukan paket produk yang dapat meningkatkan angka penjualan dan identifikasi strategi bundling yang paling menguntungkan. Penelitian sebelumnya telah mengusulkan berbagai pendekatan untuk merancang product *bundling* [25]. Perilaku konsumen yang diperoleh dari survei data belum dapat dianggap sebagai sumber data analisis yang kuat untuk memahami pola hubungan antara produk dan konsumen. Oleh karena itu, data transaksi dianggap sebagai sumber yang lebih dapat diandalkan untuk memprediksi perilaku pembelian konsumen [25].

2.4 Lift Ratio

Lift ratio adalah nilai yang menunjukkan validitas dari proses transaksi item A dan item B dibeli secara bersamaan. Rumus *lift ratio* dapat dilihat pada Rumus 2.4 [26].

$$Nilai\ lift = \frac{Support\ (A \cap B)}{Support\ (A)\ x\ Support\ (B)} \quad (2.4)$$

Sebuah transaksi dapat dianggap valid, jika memiliki nilai *lift ratio* lebih dari satu. Hal ini menunjukkan bahwa dalam transaksi yang ditentukan, *item* A dan *item* B benar-benar dibeli bersamaan [26].

2.5 White Box Testing

White box testing adalah metode pengujian suatu aplikasi atau *software* di mana pemeriksaan dilakukan secara detail terhadap struktur internal atau kode dari aplikasi tersebut. Tujuan utama dari *White Box Testing* adalah untuk memastikan bahwa setiap jalur eksekusi diuji dan setiap pernyataan kode dievaluasi. Jika model menghasilkan hasil yang tidak sesuai dengan yang diinginkan, maka akan dilakukan proses kompilasi dan pemeriksaan ulang terhadap kode-kode tersebut hingga sesuai dengan yang diharapkan [27]. Terdapat beberapa metode untuk melaksanakan pengujian *white box*, dan salah satunya adalah melalui *Unit Testing*. *Unit Testing* adalah pengujian yang setiap modul atau blok kode dilakukan pengujian terus menerus selama proses pengembangan dilakukan, pengujian ini umum dilakukan oleh *developer* yang menulis kode [28]. Salah satu teknik dalam *Unit Testing* adalah *Path Coverage*. *Path Coverage* merupakan pengujian yang bertujuan untuk melakukan validasi setiap alur program telah dijalankan setidaknya satu kali [28].

Kelebihan dari *White Box Testing* adalah sebagai berikut [29].

1. *Logic Error*

   Logic Error adalah kesalahan disaat menggunakan *syntax if* dan *loop*. Pada penelitian ini dilakukan untuk menemukan dan mengidentifikasi kondisi yang dianggap tidak tepat, serta untuk menentukan kapan proses *loop* seharusnya dihentikan.

2. Ketidaksesuaian Asumsi

Menampilkan dan melacak beberapa asumsi yang dianggap tidak sesuai dengan kondisi yang telah ditentukan, kemudian melakukan analisa dan dilakukan perbaikan.

3. *Typing Error*

Mencari dan mendeteksi kode dari program yang dianggap bersifat *case sensitive*

Ada beberapa teknik dalam Unit Testing salah satunya merupakan *Path Coverage*. *Path Coverage* adalah teknik uji coba untuk memastikan bahwa alur program dapat dilalui setidaknya satu kali [30].

2.6 USE Questionnaire

USE Questionnaire adalah suatu paket kuesioner yang bisa dimanfaatkan untuk menilai kemudahan penggunaan (*usability*) dari suatu *website*. Secara keseluruhan, konsep *usability* merujuk pada tingkat kepuasan pengguna dan sejauh mana kemampuan pengguna untuk mepelajari serta menggunakan suatu aplikasi agar dapat mencapai tujuan mereka [31]. USE Questionnaire terdiri dari 30 pertanyaan yang dibagi menjadi empat parameter, di mana setiap pertanyaan mencerminkan evaluasi ketika pengguna menggunakan aplikasi [32]. Parameter tersebut meliputi kegunaan (*usefulness*), kemudahan penggunaan (*ease of use*), kemudahan pembelajaran (*ease of learning*), *dan* kepuasan (*satisfaction*). [33]. Selanjutnya, rincian lengkap mengenai paket kuesioner yang akan digunakan dapat ditentukan menggunakan Tabel 2.1 [34].

Tabel 2.1 Kriteria pengukuran USE Questionnaire

No.	**Kriteria**
	Usefulness
U1	*It helps me be more effective.*
	Lanjut pada halaman berikutnya

Tabel 2.1 Kriteria pengukuran USE Questionnaire (lanjutan)

No	Kriteria
	Usefulness
U2	*It helps me be more productive.*
U3	*It is useful.*
U4	*It gives me more control over the activities in my life.*
U5	*It make the things I want to accomplish easier to get done.*
U6	*It saves me time when I use it.*
U7	*It meets me needs.*
U8	*It does everything I would expect it to do.*
	Ease of Use
EU1	*It is easy to use.*
EU2	*It is simple to use.*
EU3	*It is user friendly.*
EU4	*It requires the fewest steps possible to accomplish what I want to do with it.*
EU5	*It is flexible*
EU6	*Using it is effortless.*
EU7	*I can use it without written instructions.*
EU8	*I don't notice any inconsistencies as I use it.*
EU9	*Both occasional and regular users would like it.*
EU10	*I can recover from mistakes quickly and easily.*
EU11	*I can use it successfully every time.*
	Lanjut pada halaman selanjutnya

Tabel 2.1 Kriteria pengukuran USE Questionnaire (lanjutan)

	Ease of Learning
EL1	*I learned to use it quickly.*
EL2	*I easily remember how to use it.*
EL3	*It is easy to learn to use it.*
EL4	*It is quickly became skillful with it.*
	Satisfaction
S1	*I am satisfied with it.*
S2	*I would recommend it to a friend.*
S3	*It is fun to use.*
S4	*It works the way I want it to work.*
S5	*It is wonderful.*
S6	*I feel I need to have it.*
S7	*It is pleasant to use.*

2.7 Skala Likert

Pada penelitian ini, skala pengukuran yang digunakan adalah skala Likert. Skala Likert merupakan metode pengukuran sikap di mana responden diminta untuk menunjukkan tingkat persetujuan atau ketidaksetujuan terhadap setiap pernyataan atau pertanyaan yang disajikan [35]. Setiap pernyataan atau pertanyaan tersebut terkait dengan jawaban yang berupa dukungan sikap, yang dinyatakan dengan kata-kata seperti Sangat Setuju (SS), Sangat Tidak Setuju (STS), dan lainnya tergantung pada indikator penelitian yang digunakan seperti pada Tabel 2.2 dan Tabel 2.3 merupakan nilai yang digunakan untuk menghitung Skala Likert dan kelayakan suatu sistem.

Tabel 2.2 Skor Skala Likert [35]

Skor	Jawaban
1	Sangat tidak setuju
2	Tidak setuju
3	Netral
4	Setuju
5	Sangat setuju

Tabel 2.3 Kategori Kelayakan [36]

Angka (%)	Klarifikasi
<21	Sangat Tidak Layak
21-40	Tidak Layak
41-60	Cukup
61-80	Layak
81-100	Sangat Layak

Persentase skor pada pertanyaan atau pernyataan kuesioner dapat dihitung menggunakan Rumus 2.5 [36].

$$Persentase\ Skor = \frac{(SS*5)+(SS*4)+(N*3)+(TS*2)+(STS*1)}{5*(Jumlah\ Responden)} * 100\% \quad (2.5)$$

UNIVERSITAS MULTIMEDIA NUSANTARA