

BAB II

LANDASAN TEORI

2.1 Pengertian *Data Mining*

Data Mining adalah suatu istilah yang digunakan untuk menemukan pengetahuan yang tersembunyi di dalam kumpulan data. *Data Mining* merupakan proses semi otomatis yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi pengetahuan potensial dan berguna yang bermanfaat yang tersimpan di dalam *database* besar (Kusrini, dkk, 2010).

Selain definisi di atas ada definisi lain juga seperti, “*Data Mining* adalah suatu teknik menggali informasi berharga yang terpendam atau tersembunyi pada suatu koleksi (*database*) yang sangat besar sehingga ditemukan suatu pola yang menarik yang sebelumnya tidak diketahui. *Data Mining* sendiri berarti usaha untuk mendapatkan sedikit barang berharga dari sejumlah besar material dasar. Karena itu *Data Mining* sebenarnya memiliki akar yang panjang dari bidang ilmu seperti kecerdasan buatan (*artificial intelligent*), *machine learning*, *database* dan statistik” (Tri Lestari, 2009).

Berdasarkan beberapa definisi tersebut dapat ditarik kesimpulan bahwa *Data Mining* adalah suatu teknik untuk menemukan suatu informasi yang tersembunyi dalam kumpulan data yang sangat besar sehingga mampu mengidentifikasi informasi yang berguna yang sebelumnya tidak diketahui untuk membuat keputusan yang kritis, terutama dalam strategi (Hermawati, 2009).

2.2 Tahapan *Data Mining*

Terdapat enam tahapan dalam proses *Data Mining* (Dewi, 2013). Enam tahapan *Data Mining* tersebut sebagai berikut:



1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber.

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

1. Pembersihan data (*data cleanting*)

Pembersihan data merupakan proses menghilangkan *noise* dan data yang tidak konsisten atau data tidak relevan. Pada umumnya data yang diperoleh, baik dari *database* suatu perusahaan maupun hasil eksperimen, memiliki isian-isian yang tidak sempurna seperti data yang hilang, data yang tidak *valid* atau juga hanya sekedar salah ketik. Selain itu, ada juga atribut-atribut data yang tidak relevan dengan hipotesa *Data Mining* yang dimiliki. Data-data yang tidak relevan itu juga lebih baik dibuang. Pembersihan data juga akan mempengaruhi performasi dari teknik *Data Mining* karena data yang ditangani akan berkurang jumlah dan kompleksitasnya.

2. Integras data (*data integration*)

Integrasi data merupakan penggabungan data dari berbagai *database* ke dalam satu *database* baru. Tidak jarang data yang diperlukan untuk *Data Mining* tidak hanya berasal dari satu *database* tetapi juga berasal dari beberapa *database* atau file teks. Integrasi data dilakukan pada atribut-atribut yang mengidentifikasi entitas-entitas yang unik seperti atribut nama, jenis produk, nomor pelanggan, dan lainnya. Integrasi data perlu dilakukan secara cermat karena kesalahan pada integrasi data bisa menghasilkan hasil yang menyimpang dan bahkan menyesatkan pengambilan aksi nantinya. Sebagai contoh bila integrasi data berdasarkan jenis produk ternyata menggabungkan produk dari kategori yang berbeda. Maka akan didapatkan kolerasi antar produk yang sebenarnya tidak ada.

3. Seleksi data (*data selection*)

Data yang ada pada *database* sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari *database*. Sebagai contoh, sebuah kasus yang meneliti *factor* kecenderungan orang membeli dalam kasus *Market Basket Analysis*, tidak perlu mengambil nama pelanggan, cukup dengan *ID* pelanggan saja.



1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

4. Transformasi data (*data transformation*)

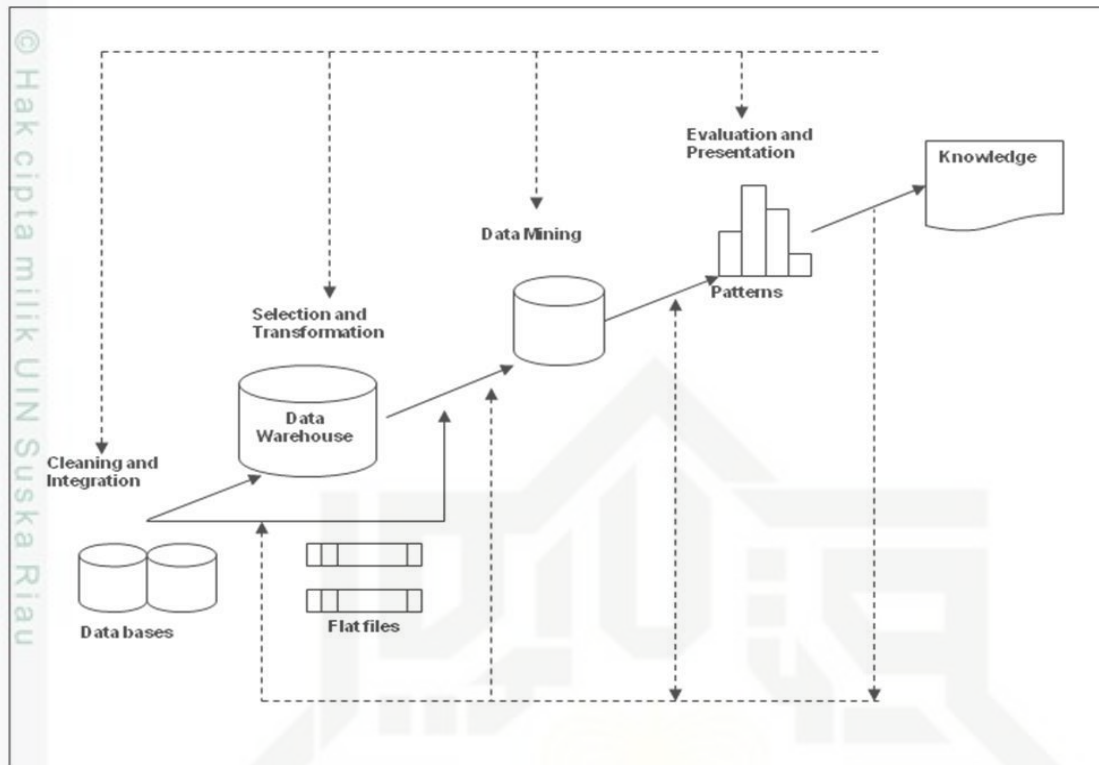
Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam *Data Mining*. Beberapa metode *Data Mining* membutuhkan format data yang khusus sebelum bisa diaplikasikan sebagai contoh beberapa metode standar seperti analisis asosiasi dan *clustering* hanya bisa menerima *input* dan kategorikal. Karena data berupa angka numerik yang perlu dibagi-bagi menjadi beberapa interval. Proses ini sering disebut transformasi data.

5. Proses *Mining*

Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data.

6. Evaluasi pola (*pattern evaluation*)

Untuk mengidentifikasi pola-pola menarik kedalam *knowledge based* yang ditemukan. Dalam tahap ini hasil dari teknik *Data Mining* berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesa yang ada memang tercapai. Bila ternyata hasil yang di peroleh tidak sesuai hipotesa ada beberapa alternatif yang dapat diambil seperti menjadikannya umpan balik untuk memperbaiki proses *Data Mining* lain yang lebih sesuai, atau menerima hasil ini sebagai suatu hasil yang di luar dugaan yang mungkin bermanfaat. Ilustrasi dari tahapan-tahapan *Data Mining* dapat dilihat pada Gambar 2.1:



Gambar 2.1 Tahapan-tahapan *Data Mining*
(Sumber: Dewi, 2013)

2.3 *Market Basket Analysis* (MBA)

Market Basket Analysis atau MBA, merupakan salah satu tipe analisis data yang paling sering digunakan dalam dunia pemasaran (Friday, dkk, 2012). Tujuan dari MBA untuk menentukan produk-produk apa saja yang paling sering dibeli atau digunakan sekaligus oleh para konsumen. Proses MBA ini adalah dengan menganalisis *buying habits* konsumen dengan menemukan asosiasi antar produk-produk yang berbeda yang diletakkan konsumen dalam *shopping basket*.

Istilah MBA sendiri datang dari kejadian yang sudah sangat umum terjadi di dalam perusahaan *retail*, yaitu ketika para konsumen memasukkan semua barang yang mereka beli ke dalam keranjang (*basket*) yang umumnya telah disediakan oleh pihak perusahaan sendiri. Informasi mengenai produk-produk yang biasanya dibeli secara bersamaan oleh para konsumen dapat memberikan “wawasan” tersendiri bagi para pengelola untuk menaikkan laba bisnisnya (Arief, 2012).

2.3.1 Defenisi MBA

MBA adalah teknik matematis yang biasa digunakan oleh marketing professional untuk menyatakan kesamaan antara produk individu atau kelompok produk. MBA berkenaan dengan sekumpulan permasalahan bisnis yang berkaitan untuk mengetahui *point of sale* dari data transaksi (Mahendra, dkk, 2012).

2.3.2 Konsep MBA



Gambar 2.2 Keranjang Belanja
(Sumber: Goldie dan Dana, 2012)

MBA merupakan salah satu contoh penerapan *Association Rule*. Untuk menyampaikan ide mendasar dari MBA, dimulai dengan melihat gambar keranjang belanja pada Gambar 2.2 yang berisi bermacam-macam barang yang dibeli oleh seseorang di sebuah supermarket. Keranjang ini berisi bermacam-macam barang seperti roti, susu, sereal, telur, mentega, gula, dan sebagainya. Sebuah keranjang yang dibeli oleh seorang konsumen memberikan informasi yang sangat banyak, dan ini dapat menjelaskan barang-barang apa saja yang paling penting dari bisnis penjualan yaitu “apa barang yang dibeli oleh konsumen dan kapan” (Goldie dan Dana, 2012).

Gambar 2.2 menjelaskan setiap konsumen membeli seperangkat barang-barang yang berbeda, dalam jumlah yang berbeda, dan dalam waktu yang berbeda. MBA menggunakan informasi apa yang dibeli oleh konsumen untuk menyediakan tanda atau informasi yaitu siapa mereka dan mengapa mereka melakukan pembelian tersebut? MBA menyediakan pengertian tentang dagangan dengan memberitahukan produk-



produk mana yang memungkinkan untuk dibeli secara bersamaan dan produk mana yang lebih disetujui untuk dipromosikan (Megaputer, 2007).

2.3.3 Tiga Level Market Basket Data (MBD)

Market Basket Data (MBD) adalah data transaksi yang menjelaskan tiga perbedaan entitas yang mendasar yaitu (Kusnawi, 2007):

1. *Customers*.
2. *Orders* atau pembelian.
3. *Items* (barang-barang).

Pengenalan konsumen setiap saat membuat mungkin untuk dikenali secara cepat, seperti frekuensi pembelian yang dilakukan oleh konsumen. Tiga level dari MBD yang penting secara cepat dapat memahami permintaan. Ada beberapa dasar pengukuran yaitu (Satia, 2007):

1. Berapa rata-rata pembelian barang yang dilakukan konsumen?
2. Berapa rata-rata barang yang khusus setiap pembelian?
3. Berapa rata-rata barang setiap pembelian?
4. Untuk barang tertentu, proposi konsumen apa yang telah membeli barang tersebut?
5. Untuk barang tertentu, berapa rata-rata dari pembelian setiap konsumen yang termasuk dalam barang tersebut?
6. Untuk barang tertentu, berapa rata-rata jumlah pembelian dalam suatu pembelian ketika barang tersebut dibeli?

2.4 Association Rule Mining

Association Rules Mining adalah suatu prosedur untuk mencari hubungan antar *item* suatu dataset yang telah ditentukan *Association Rules Mining* mencari dan menemukan hubungan antar *item* yang ada pada suatu *dataset*. Penerapan *Data Mining* dengan aturan asosiasi bertujuan menemukan informasi *item-item* yang saling berhubungan dalam bentuk aturan atau *rule*. Aturan asosiasi adalah teknik *Data Mining* untuk menemukan aturan asosiasi antar suatu kombinasi *item*. (Dr. M. Dhanabhakyan dan Dr. M. Punithavalli, 2013).



1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Metode ini dapat membantu mengenali pola-pola tertentu di dalam kumpulan data yang besar. Dalam *Association Rules*, suatu kelompok *item* dinamakan *itemset*. *support* dari *itemset* X adalah persentase transaksi di D yang mengandung X, biasa ditulis dengan $\text{supp}(X)$. Pencarian *Association Rules* dilakukan melalui dua tahap yaitu pencarian *frequent itemset* dan penyusunan *rules*. Jika *support* suatu *itemset* lebih besar atau sama dengan *minimum support* (σ), maka *itemset* tersebut dapat dikatakan sebagai *frequent itemset* atau *frequent pattern* yang tidak memenuhi dinamakan *infrequent*. *Confidence* adalah nilai ukuran seberapa besar *valid* tidaknya suatu *Association Rules*. *Confidence* suatu rule $R (X \Rightarrow Y)$ adalah proporsi dari semua transaksi yang mengandung baik X maupun Y dengan yang mengandung X, biasa ditulis sebagai $\text{conf}(R)$. Sebuah *Association Rule* dengan *confidence* sama atau lebih besar dari *minimum Confidence* (γ) dapat dikatakan sebagai *valid Association Rule* (S.O. Abdulsalam, dkk, 2014).

2.4.1 Metode Dasar *Association Rule*

Metodologi dasar analisis asosiasi terbagi menjadi dua tahap (Santoso, B, 2007 dalam Robi Yanto, dkk, 2015):

1. Analisa pola *frekuensi* tinggi.
2. Tahap ini mencari kombinasi *item* yang memenuhi syarat minimum dari nilai *support* dalam *database*, rumus analisa pola *frekuensi* untuk 1 *itemset* tertinggi dapat dilihat pada Rumus 2.1.

$$\text{Support}(A) = \frac{\text{Jumlah Transaksi Mengandung A}}{\text{Total Transaksi}} \dots\dots\dots(2.1)$$

Rumus analisa pola *frekuensi* untuk 1 *itemset* tertinggi dapat dilihat pada

Rumus 2.2.

$$\text{Support}(A,B) = \frac{\text{Jumlah transaksi kombinasi A dan B}}{\text{Total transaksi}} \times 100\% \dots\dots\dots(2.2)$$

3. Pembentukan aturan asosiatif.

Setelah semua pola *frekuensi* tinggi ditemukan, setelah itu dicari aturan asosiatif yang memenuhi syarat minimum untuk *confidence* dengan



menghitung *confidence* aturan assosiatif $A \rightarrow B$ dengan menggunakan Rumus 2.2.

$$Confidence = P(B|A) = \frac{\Sigma Transaksi mengandung A dan B}{\Sigma Transaksi mengandung A} \dots\dots\dots(2.3)$$

2.5 Pengertian *Algoritma Apriori*

Algoritma Apriori adalah suatu *algoritma* dasar yang diusulkan oleh Agrawal & Srikant pada tahun 1994 untuk penentuan *frequent itemset* untuk aturan asosiasi boolean. *Algoritma* apriori termasuk jenis aturan asosiasi pada *Data Mining*. Aturan yang menyatakan asosiasi antara beberapa atribut sering disebut *Affinity Analysis* atau MBA (Kennedi, dkk, 2011).

Untuk penerapan *Algoritma Apriori*, secara umum dibutuhkan struktur data untuk menyimpan *candidate frequent itemset* untuk suatu iterasi ke k dan untuk menyimpan *frequent itemset* yang dihasilkan. Ketika membaca tiap *item* dari seluruh transaksi, selain mendapatkan *item-item* baru juga dilakukan perhitungan nilai *support item-item* yang sudah ditemukan, sehingga untuk mendapatkan *candidate 1-itemset* beserta nilai *support*-nya cukup membutuhkan satu kali pembacaan data (Rezkiyani, 2011).

2.5.1 Kelebihan *Algoritma Apriori*

Algoritma Apriori yang akan digunakan pada penelitian ini, mempunyai beberapa kelebihan (Santoso, 2003):

1. Menggunakan format data *vertical tid-list* yang mengasosiasikan *itemset* dengan transaksi yang terjadi.
2. Menggunakan pendekatan apriori untuk mencari maksimal *frequent itemset*.
3. Membutuhkan hanya sedikit pembacaan *database*, dan meminimalkan penggunaan I atau O.

2.5.2 Tahap-Tahap *Algoritma Apriori*

Algoritma apriori dibagi menjadi beberapa tahap yang disebut narasi atau pass (Bulolo, 2013):



1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

1. Pembentukan kandidat *itemset*, kandidat *k-itemset* dibentuk dari kombinasi ($k-1$) *itemset* yang didapat dari iterasi sebelumnya. Satu cara dari *Algoritma Apriori* adalah adanya pemangkasan kandidat *k-itemset* yang *subset*-nya yang berisi $k-1$ *item* tidak termasuk dalam pola frekuensi tinggi dengan panjang $k-1$.
2. Penghitungan *support* dari tiap kandidat *k-itemset*. *Support* dari tiap kandidat *k-itemset* didapat dengan *meng-scan database* untuk menghitung jumlah transaksi yang memuat semua *item* didalam kandidat *k-itemset* tersebut. Ini adalah juga ciri dari *Algoritma Apriori* dimana diperlukan penghitungan dengan cara seluruh *database* sebanyak *k-itemset* terpanjang.
3. Tetapkan pola frekuensi tinggi. Pola frekuensi tinggi yang memuat *k-item* atau *k-itemset* ditetapkan dari kandidat *k-itemset* yang *support*-nya lebih besar dari minimum *support*.
4. Bila tidak didapat pola frekuensi tinggi baru maka seluruh proses dihentikan. Bila tidak, maka k ditambah satu dan kembali bagian satu.

2.6 Pengertian Pola Belanja Konsumen

Menurut Kamus Besar Bahasa Indonesia Pola adalah Sistem, cara kerja ataupun bentuk struktur yang pasti. Maka pola belanja konsumen bisa diartikan sebagai bentuk struktur dari kegiatan belanja konsumen yang pasti. Dari pola belanja yang dapat diprediksi, pembuat keputusan dapat membuat strategi pemasaran yang lebih efektif (A.A Gede, 2012).

2.6.1 Faktor-Faktor yang Mempengaruhi Pola Belanja Konsumen

Faktor-faktor yang mempengaruhi pola belanja konsumen atau perilaku belanja konsumen adalah sebagai berikut (Kurniasih, 2011):

1. Faktor budaya.
2. Faktor social.
3. Faktor pribadi.
4. Faktor psikologis.

Budaya merupakan salah satu penentu keinginan dan perilaku seseorang yang paling mendasar dan sesungguhnya seluruh masyarakat memiliki stratifikasi social



dimana kelas sosial menunjukkan pilihan terhadap produk dengan merek yang berbeda-beda.

Keputusan pembelian juga dipengaruhi oleh karakteristik atau ciri-ciri pribadinya, terutama yang berpengaruh adalah umur dan tahapan dalam siklus hidup pembeli, pekerjaannya, keadaan ekonominya, gaya hidupnya, pribadi dan konsep jati dirinya. Pilihan membeli seseorang juga akan dipengaruhi faktor psikologis utama, yaitu: motivasi, persepsi, proses belajar, dan kepercayaan dengan sikap. faktor-faktor yang mempengaruhi tingkah laku konsumen dapat dilihat pada Gambar 2.3.

Budaya	Sosial	Pribadi	Psikologi	PEMBELIAN
<ul style="list-style-type: none"> ✓ Kebudayaan ✓ Sub Kebudayaan ✓ Kelas sosial 	<ul style="list-style-type: none"> ✓ Kelompok acuan ✓ Keluarga ✓ Peran dan Status 	<ul style="list-style-type: none"> ✓ Umur dan Tahapan Siklus hidup ✓ Pekerjaan ✓ Situasi ekonomi ✓ Gaya hidup ✓ Kepribadian 	<ul style="list-style-type: none"> ✓ Motivasi ✓ Persepsi ✓ Pengetahuan ✓ Keyakinan ✓ Sikap 	


Gambar 2.3 Faktor-faktor yang mempengaruhi tingkah laku konsumen

(Sumber: Kurniasih, 2011)

2.7 Flowchart

Flowchart adalah suatu diagram alir yang menggunakan simbol atau tanda untuk menyelesaikan masalah. Adapun simbol-simbol yang sering digunakan dalam *flowchart* (Evi, 2014). Simbol-simbol *flowchart* dapat dilihat pada Tabel 2.1.

Tabel 2.1 Simbol-simbol *Flowchart*

Simbol <i>Flowchart</i>	Fungsi <i>Flowchart</i>
	Terminal menyatakan awal atau akhir dari suatu <i>algoritma</i> .



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.



2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Tabel 2.1 Simbol-simbol *Flowchart* (Lanjutan)

Simbol <i>Flowchart</i>	Fungsi <i>Flowchart</i>
	Menyatakan Proses.
	Proses yang terdefinisi atau sub program.
	Persiapan yang digunakan untuk memberi nilai awal suatu besaran.
	Menyatakan masukan dan keluaran (<i>input/output</i>)
	Menyatakan penyambung ke simbol lain dalam satu halaman.
	Menyatakan penyambung ke halaman lainnya.
	Menyatakan pencetakkan (dokumen) pada kertas.
	Menyatakan <i>decision</i> (keputusan) yang digunakan untuk penyeleksian kondisi di dalam program.
	Menyatakan media penyimpanan drum magnetik.
	Menyatakan input/output menggunakan disket.
	Menyatakan input/output dari kartu plong.
	Menyatakan arah alir pekerjaan (proses).



Tabel 2.1 Simbol-simbol *Flowchart* (Lanjutan)

Simbol <i>Flowchart</i>	Fungsi <i>Flowchart</i>
	Multidocument (banyak dokumen)
	Delay (penundaan atau kelambatan).

2.8 Basis Data (*Database*)

Basis data terdiri atas dua kata yaitu Basis dan Data. Basis dapat diartikan sebagai markas atau gudang, tempat bersarang. Sedangkan data adalah representasi fakta dunia nyata yang mewakili suatu objek seperti manusia (siswa, pegawai, pembeli, pelanggan), barang, hewan, peristiwa, konsep, keadaan, dan sebagainya (Fathansyah, 1999).

Basis data (*database*) adalah suatu pengorganisasian sekumpulan data yang saling terkait sehingga memudahkan aktivitas untuk memperoleh informasi. Basis data dimaksudkan untuk mengatasi problem pada sistem yang memakai pendekatan berbasis berkas. Secara lebih lengkap, tujuan basis data adalah (Abdul Kadir, 2003):

1. Kecepatan dan kemudahan (*speed*).
2. Efisiensi ruang penyimpanan (*space*).
3. Keakuratan (*accuracy*).
4. Ketersediaan (*availability*).
5. Kelengkapan (*completeness*).
6. Keamanan (*security*).
7. Kebersamaan pemakai (*sharability*).

2.9 MySQL

MySQL adalah sebuah aplikasi *Relational Database Management Server* (RDBMS) yang sangat cepat dan kokoh serta bersifat *open source*. MySQL



merupakan salah satu jenis *database server* yang banyak digunakan di dunia maya, yang menggunakan *SQL* sebagai bahasa dasar untuk mengakses *database*. *MySQL* dapat digunakan pada berbagai *platform* sistem operasi. Keunggulan *MySQL* dalam mengolah *database* adalah (Mustajib, 2016) :

1. Kecepatan.
2. Mudah digunakan.
3. *Open Source*.
4. Kapabilitas.
5. Biaya murah.
6. Keamanan.
7. Lintas *Platform*.
8. Minim "bug".

2.10 *Hypertext Pre-Processor (PHP)*

PHP singkatan dari *Hypertext Preprocessor* yang digunakan sebagai bahasa script *server-side* yang cukup handal, yang akan disatukan dengan HTML dan berada di *server*. Artinya, sintaks dan perintah yang diberikan akan sepenuhnya dijalankan di *server* sebelum dikirim ke komputer klien. Seluruh aplikasi yang berbasis *web* dapat dibuat menggunakan PHP. Salah satu kelebihan PHP adalah kemampuan untuk dapat melakukan koneksi dengan berbagai *database*, seperti *MySQL*, *Postgre SQL*, *Access*, dll. Selain itu PHP juga bersifat *open source* (Mustajib, 2016).

2.11 *Blackbox Testing*

Teknik pengujian *black-box* berfokus pada domain informasi dari perangkat lunak, dengan melakukan *test case* dengan mempartisi domain input dari suatu program dengan cara yang memberikan cakupan pengujian yang mendalam.

Blackbox Merupakan metode pengujian perangkat lunak yang dilakukan hanya mengamati hasil eksekusi data uji dan memeriksa fungsional dari perangkat lunak. Cakupan pengujian yang dilakukan pada *Blackbox* testing adalah perihal pengujian *interface* dan *form validation*. Pengujian *interface* adalah pengujian yang dilakukan secara langsung terhadap desain *interface* yang dibuat pada system. Tujuan yang



diharapkan dalam melakukan *Blackbox Testing* adalah dapat membuat desain yang sesuai dengan kebutuhan perusahaan (Mustajib, 2016).

2.12 Weka

The Waikato Environment for Knowledge Analysis (Weka) adalah rangkaian lengkap perpustakaan kelas *Java* yang mengimplementasikan banyak *state-of-the-art* pembelajaran mesin dan *algoritma Data Mining*. Weka tersedia secara bebas di *World Wide Web* dan menyertai teks baru pada dokumen *Data Mining* dan sepenuhnya menjelaskan semua *algoritma* yang dikandungnya. Aplikasi yang ditulis menggunakan *library class* pada Weka yang dapat dijalankan pada komputer manapun dengan kemampuan *browsing Web*, ini memungkinkan pengguna untuk menerapkan teknik pembelajaran mesin untuk data mereka sendiri terlepas dari *platform* komputer. (Witten, dkk, 2011). Empat tombol di atas dapat digunakan untuk menjalankan Aplikasi yang berfungsi untuk :

1. Explorer digunakan untuk menggali lebih jauh data dengan aplikasi *Weka*.
2. Experimenter digunakan untuk melakukan percobaan dengan pengujian statistic skema belajar.
3. Knowledge Flow digunakan untuk pengetahuan pendukung.
4. Simple CLI antar muka dengan menggunakan tampilan *command-line* yang memungkinkan langsung mengeksekusi perintah *weka* untuk Sistem Operasi yang tidak menyediakan secara langsung.

Weka mendukung beberapa jenis *format file* yang akan digunakan untuk inputnya, yaitu:

1. *Comma Separated Values* (CSV) merupakan *file* teks dengan pemisah tanda koma (,) yang cukup umum digunakan.
2. *Format C45* merupakan *format file* yang dapat diakses dengan menggunakan aplikasi *Weka*.
3. *Attribute-Relation File Format* (ARFF) merupakan tipe *file teks* yang berisi berbagai *instance* data yang berhubungan dengan suatu set atribut data yang dideskripsikan serta di dalam *file* tersebut.



4. *SQL Server* atau *MySql Server* dapat mengakses *database* dengan menggunakan *SQL Server* atau *MySql Server*.

2.13 Tanagra 1.4

Tanagra adalah software *Data Mining* bebas untuk tujuan akademik dan penelitian ini mengusulkan beberapa metode *Data Mining* dari analisis eksplorasi data, pembelajaran statistik, pembelajaran mesin dan daerah *database*. *Tanagra* adalah "proyek open source" karena setiap peneliti dapat mengakses ke kode sumber, dan menambahkan *algoritma* sendiri, sejauh dia setuju dan sesuai dengan lisensi distribusi perangkat lunak. Tujuan utama dari proyek *Tanagra* adalah memberikan peneliti dan mahasiswa yang mudah untuk menggunakan perangkat lunak *Data Mining*, sesuai dengan norma-norma yang hadir dari pengembangan perangkat lunak dalam domain ini (terutama dalam desain *Graphical User Interface* (GUI) dan cara untuk menggunakannya), dan memungkinkan untuk menganalisis baik data yang nyata atau sintetis. Tujuan kedua *Tanagra* adalah untuk mengusulkan kepada peneliti arsitektur yang memungkinkan mereka untuk dengan mudah menambahkan metode penambahan data mereka sendiri, untuk membandingkan kinerja mereka. *Tanagra* bertindak lebih sebagai *platform* eksperimental untuk membiarkan mereka pergi ke penting dari pekerjaan mereka, pengeluaran mereka untuk berurusan dengan bagian menyenangkan dalam programming semacam ini alat pengelolaan data (Julsam, 2009). Tujuan ketiga dan terakhir, arah pengembang pemula, terdiri dalam menyebarkan metodologi yang mungkin untuk membangun perangkat lunak semacam ini. Mereka harus mengambil keuntungan dari akses gratis untuk melihat bagaimana perangkat lunak semacam ini dibangun (Badrul, 2016).

2.14 Berkah Mart

Berkah Mart merupakan perusahaan retail yang telah berdiri sejak tahun 2014 dan beralamatkan di Jl. Suka Karya Kualu, Panam, Pekanbaru. Memiliki 6 karyawan yang dibagi menjadi dua shift kerja yaitu 07.30-15.00 dan 15.00-23.00. minimarket ini memiliki proses bisnis yang baik yaitu jumlah konsumen yang stabil dan konsumen dapat mendengarkan lagu-lagu islami ketika mereka berbelanja. Produk-



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

produk yang sering *discount* adalah bahan-bahan pokok. Seperti minyak goreng, popok bayi, dan susu bayi. Berkah Mart Memiliki Beberapa persaingan seperti Indomaret, Alfamart dan Planet Swalayan yang jaraknya tidak terlalu jauh dengan Berkah Mart.

Berdasarkan wawancara dan observasi langsung di Berkah Mart. Peneliti memperoleh informasi bahwasannya Berkah Mart ini mempunyai 8.307 *item* barang yang dijual disana, dimana jumlah transaksi yang keluar dalam sehari rata-rata 400 transaksi, dalam satu transaksi minimal terdapat 1 *item* barang dan maksimal terdapat 20 *item* barang. Tata letak sangat berpengaruh pada pembelian konsumen, dikarenakan banyaknya konsumen yang bertanya kepada karyawan letak barang yang mereka inginkan. Berkah Mart juga pernah kehabisan dan kelebihan stock sehingga *item* kadaluarsa. Informasi lebih lengkap dapat dilihat ada lampiran D. sedangkan observasi yang dilakukan di Berkah Mart memperoleh data transaksi selama 3 bulan yakni pada bulan Oktober, November, Desember 2016 yang berjumlah sekitar 12.000 transaksi berupa capturean sistem sehingga data harus diketik kembali.

2.15 Riset Terdahulu

Dibawah ini penulis paparkan beberapa penelitian terdahulu yang membahas mengenai *Data Mining* yang digunakan sebagai acuan untuk menyelesaikan tugas akhir seperti yang terlihat pada Tabel 2.2.

Tabel 2.2 Riset Terdahulu

No	Peneliti	Tahun	Judul	Hasil
1	Mustajib	2016	Implementasi <i>Association Rule Mining</i> sebagai Pendukung Keputusan Tata Letak dan Perancangan Ketersediaan Obat di Apotek Rumah Sakit Umum Daerah Temilahan.	Strategi untuk menata penempatan obat-obat digudang penyimpanan dengan menganalisa data resep obat yang jumlahnya sekitar 70 resep/hari dan total perbulan adalah 2100 resep. Hasil analisa di uji dengan Tanagra. <i>Support 10% dan Confidence 60%.</i>
4	Friday Kurniasi	2011	Analisa dan Perancangan <i>Data Mining</i> dengan metode <i>Market Basket Analysis</i> untuk Analisa Pola Belanja Konsumen pada <i>Tendencies Store</i> .	Pengolahan data transaksi dilakukan dengan menggunakan <i>Database Management System</i> yang digunakan adalah <i>MS-SQL Server 2008</i> .
5	Buulolo	2013	Implementasi <i>Algoritma Apriori</i> pada Persediaan Obat Studi Kasus : Apotek Rumah Sakit Estomihi Medan.	Hasil dari penelitian tersebut menemukan pola kombinasi paling tinggi dengan <i>Support 22%</i> dan <i>Confidence 60%</i> .

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber.

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.