BAB II
TINJAUAN PUSTAKA

Bab ini menjelaskan tentang dasar teori yang digunakan dalam penyusunan tugas akhir ini. Dijelaskan pengertian tentang manajemen pemasaran, *data mining* beserta macam-macamnya, selain itu juga dijelaskan tentang *data warehouse*, serta *MBA*.

2.1 Manajemen Pemasaran

Menurut Kotler (2005) dalam Lestari (2009) manajemen pemasaran sebagai seni dan ilmu untuk memilih pasar sasaran serta mendapatkan, mempertahankan, dan menambah jumlah pelanggan melalui penciptaan, penyampaian, dan pengkomunikasian nilai pelanggan yang unggul. Pemasaran adalah proses sosial yang dengan proses itu individu dan kelompok mendapatkan apa yang mereka butuhkan dan inginkan dengan menciptakan, menawarkan, dan secara bebas mempertukarkan produk dan jasa yang bernilai dengan pihak lain. Pemasar adalah seorang yang mencari tanggapan (perhatian, pembelian, pemberian suara, sumbangan) dari pihak lain yang disebut calon pelanggan. Tugas pemasar adalah menyusun program atau rencana pemasaran untuk mencapai tujuan yang diinginkan perusahaan.

Strategi pemasaran adalah rencana yang menyeluruh, terpadu, dan menyatu di bidang pemasaran, yang memberikan panduan tentang kegiatan yang akan dijalankan untuk dapat tercapainya tujuan pemasaran suatu perusahaan (Assauri, 2004 dalam Lestari, 2009). Strategi pemasaran yang paling utama adalah *segmentation*, *targeting*, dan *positioning* (Kotler, 2005 dalam Lestari, 2009).

1. Segmentasi

Strategi segmentasi merupakan bagian yang paling penting dalam menentukan kesuksesan suatu perusahaan. Segmentasi adalah proses membagi pasar yang bersifat heterogen ke dalam beberapa segmen sehingga masing-masing segmen cenderung bersifat homogen dalam segala aspek. Variabel segmentasi utama yang digunakan untuk pasar konsumen adalah geografi, demografi, psikografi, perilaku dan individual.

2. *Targeting*

Menurut Kotler (2005) dalam Lestari (2009) penentuan pasar sasaran atau *targeting* penting untuk dilakukan setelah perusahaan menetapkan segmen pasar yang memiliki peluang paling besar. Perusahaan-perusahaan akan berhasil secara gemilang apabila mereka cermat dalam memilih pasar sasaran dan mempersiapkan program

pemasaran yang dirancang khusus untuk masing- masing pasar tersebut. *Targeting* merupakan pemilihan satu atau lebih segmen yang dianggap paling potensial dan menguntungkan, sekaligus mengembangkan produk dan program pemasaran untuk segmen-segmen yang dipilih. Perusahaan perlu mengevaluasi sebelum melakukan pemilihan segmen pasar yang berbeda, dengan memperhatikan daya tarik segmen secara keseluruhan serta tujuan dan sumberdaya perusahaan. Perusahaan dapat mempertimbangkan lima pola pemilihan pasar sasaran yaitu konsentrasi segmen tunggal, spesialisasi selektif, spesialisasi produk, spesialisasi pasar, dan cakupan ke seluruh pasar.

3. *Positioning*

Positioning adalah strategi yang berusaha menciptakan diferensiasi yang unik dalam benak pelanggan sasaran sehingga terbentuk citra merek atau produk yang lebih unggul dibandingkan merek produk pesaing. Hasil akhir penetapan posisi adalah keberhasilan penciptaan proporsi nilai yaitu alasan yang meyakinkan pelanggan untuk membeli produk perusahaan (Kotler, 2005 dalam Lestari, 2009).

2.2 Sasaran Pemasaran

Kotler (2005) dalam Lestari (2009) mengemukakan sasaran pemasaran adalah seperangkat alat pemasaran yang digunakan perusahaan untuk mencapai tujuan pemasarannya dalam pasar sasaran. sasaran ini terdiri dari empat unsur, yang biasa disebut empat P (4P), yaitu produk (*product*), harga (*price*), tempat/ distribusi (*place*), dan promosi (*promotion*).

1. Produk (*product*)

Produk merupakan alat bauran pemasaran yang paling mendasar dan pokok. Produk menjadi penawaran berwujud produsen kepada pasar, yang mencakup kualitas, rancangan, bentuk, merek, dan kemasan produk. Strategi produk dikonsentrasikan pada pengembangan produk yang tepat bagi pasar yang ingin dituju.

2. Harga (*price*)

Strategi ini meliputi pemilihan metode penetapan harga produk, memodifikasi harga yang sudah ada serta menetapkan dan menanggap perubahan harga. Tujuan dari strategi ini adalah untuk mempertahankan pangsa pasar, mencapai keuntungan yang maksimum, memperoleh pendapatan maksimum, dan mencapai pertumbuhan penjualan yang tinggi. Produsen harus menentukan apa yang ingin dicapai dari produk yang dipasarkannya, sebelum penetapan harga tersebut dilakukan, dengan

mempertimbangkan faktor pelanggan, pesaing, dan biaya produksi.

3. Tempat/distribusi (*place*)

Kegiatan ini bertujuan mengantarkan produk ke konsumen dan menyediakan bagi pelanggan sasaran. Produsen harus mengerti berbagai jenis pengecer, pedagang grosir, perusahaan distribusi fisik dan bagaimana mereka membuat suatu keputusan.

4. Promosi (*promotion*)

Promosi meliputi semua kegiatan yang dilakukan perusahaan untuk mengkomunikasikan dan mempromosikan produknya kepada pasar. Perusahaan harus membuat komunikasi dan promosi yang terdiri dari iklan, promosi penjualan, hubungan masyarakat serta pemasaran langsung maupun *online*. Promosi penting bagi produsen dalam meningkatkan nilai penjualan dan pertumbuhan produk.

2.3 Pengertian Penjualan

Menurut Assauri (2004) dalam Lestari (2009) kegiatan penjualan merupakan kegiatan pelengkap dari pembelian untuk memungkinkan terjadinya transaksi. Kegiatan pembelian dan penjualan merupakan satu kesatuan agar terlaksananya transfer hak atau transaksi. Oleh karena itu, kegiatan penjualan seperti halnya kegiatan pembelian, terdiri dari serangkaian kegiatan yang meliputi penciptaan permintaan, menemukan si pembeli, negosiasi harga, dan syarat-syarat pembayaran. Penjual harus menentukan kebijaksanaan dan prosedur yang akan diikuti untuk memungkinkan dilaksanakannya rencana penjualan yang telah ditetapkan.

2.4 Struktur Dasar Bisnis Ritel (Usaha Eceran)

Menurut Sopiah dan Syihabudhin (2008) dalam Lestari (2009) berlanjutnya pertumbuhan ekonomi dan meningkatnya rata-rata pendapatan yang bisa dibelanjakan memperbesar permintaan atau akan memperbanyak toko eceran yang lebih khusus dan spesifik. Pemasaran adalah kegiatan memasarkan barang atau jasa secara umum kepada masyarakat dan secara khusus kepada pembeli potensial. Kegiatan perdagangan besar dan perdagangan eceran sangatlah penting dalam proses penyaluran barang dan jasa.

Ada dua kepentingan mengapa produk perlu ditempatkan agar konsumen bisa dengan mudah memperolehnya.

1. Kepentingan pertama adalah kepentingan produsen, yakni kepentingan terhadap produknya agar ditempatkan di posisi yang layak dengan maksud agar konsumen terstimulasi untuk membelinya.

2. Kepentingan kedua adalah kepentingan pengecer yang berkeinginan agar konsumen terstimulasi untuk membeli produk yang ditawarkan oleh berbagai produsen

Perdagangan eceran memegang peranan yang sangat penting, baik ditinjau dari sudut produsen, pedagang eceran dipandang sebagai pihak yang ahli dalam bidang penjualan produk perusahaannya. Pengecer ujung tombak perusahaan yang sangat menentukan laku atau tidaknya produk perusahaan. Sementara jika dipandang dari sudut konsumen, pedagang eceran juga memiliki peranan yang sangat penting. Pedagang eceran bertindak sebagai agen yang membeli, mengumpulkan, dan menyediakan barang atau jasa untuk memenuhi kebutuhan atau keperluan pihak konsumen.

2.4.1 Perilaku Konsumen dalam *Retailing*

Menurut Sopiah dan Syihabudhin (2008) dalam Lestari (2009) khusus dalam pembelian ritel terdapat pola perilaku tertentu pada konsumen, pola perilaku tersebut terbagi kedalam tujuh kategori, dimana masing- masing kategori bisa berubah urutannya. Pada dasarnya setiap manusia berbeda, perilakunya pun berbeda walaupun perilaku tersebut relatif sama. Pola perilaku tersebut sebagai berikut:

1. Perilaku sebelum pembelian (*Pre-purchase*)

   Pada tahap sebelum pembelian, ada dua tahapan perilaku konsumen, yaitu:

   a. Konsumen mencari informasi sebanyak-banyaknya dengan membaca Koran, majalah, buletin, mendengarkan siaran radio, melihat TV, mencari informasi dari teman, orang tua, pramuniaga, dan lain-lain.
   b. Perilaku pada tahap kedua, konsumen berusaha mendapatkan uang, mengambil uang tunai di bank, mengambil uang di ATM, meminjam teman, menggunakan kartu kredit, atau menggunakan kartu debet bank tertentu.

2. Perilaku pembelian (*Purchase*)

   Ada lima tahap perilaku konsumen yaitu:

   a. *Store contact*

      Pada tahap ini, konsumen memilih lokasi belanja yang dirasa cocok, memasuki toko, dan memilih produk yang dicari.

   b. *Product contact*

      Pada tahap ini, konsumen memilih dan menemukan produk yang dicari, lalu membawanya ke kasir untuk melakukan pembayaran.

   c. *Transaction*

Pada tahap ini, konsumen membayar barang yang telah di pilih di kasir dan membawa produk untuk dikonsumsi.

d. *Consumption*

Pada tahap ini, konsumen menggunakan produk, membuang sisanya jika ada, dan melakukan pembelian ulang jika konsumen merasa puas.

e. *Communication*

Di tahap terakhir ini, konsumen memberikan informasi tentang produk yang baru dikonsumsi tersebut kepada orang lain, mengisi kartu garansi (jika ada) dan memberikan informasi tentang produk, harga, pelayanan, dan lain-lain kepada peritel.

2.5 *Data Warehouse*

Data warehouse adalah sebuah sistem yang mengambil dan menggabungkan data secara periodik dari sistem sumber data ke penyimpanan data bentuk dimensional atau normal (Rainardi, 2008 dalam Huda, 2010). *Data warehouse* merupakan penyimpanan data yang berorientasi objek, terintegrasi, mempunyai *variant* waktu, dan menyimpan data dalam bentuk *nonvolatile* sebagai pendukung manejemen dalam proses pengambilan keputusan (Han, 2006 dalam Huda, 2010).

Data warehouse menyatukan dan menggabungkan data dalam bentuk multidimensi. Pembangunan *data warehouse* meliputi pembersihan data, penyatuan data dan transformasi data dan dapat dilihat sebagai praproses yang penting untuk digunakan dalam *data mining*. Selain itu *data warehouse* mendukung *On-line Analitycal Processing* (OLAP), sebuah kakas yang digunakan untuk menganalisis secara interaktif dari bentuk multidimensi yang mempunyai data yang rinci. Sehingga dapat memfasilitasi secara efektif *data generalization* dan *data mining*.

Banyak metode-metode *data mining* yang lain seperti asosiasi, klasifikasi, prediksi, dan *clustering*, dapat diintegrasikan dengan operasi OLAP untuk meningkatkan proses *mining* yang interaktif dari beberapa level dari abstraksi. Oleh karena itu *data warehouse* menjadi platform yang penting untuk data analisis dan OLAP untuk dapat menyediakan platform yang efektif untuk proses *data mining*.

Empat karakteristik dari *data warehouse* meliputi:

1. *Subject oriented*: sebuah *data warehouse* disusun dalam subjek utama, seperti pelanggan, suplier, produk, dan sales. Meskipun *data warehouse* terkonsentrasi pada operasi harian dan proses transaksi dalam perusahaan, *data warehouse*

fokus pada pemodelan dan analisis data untuk pembuat keputusan. Oleh karena itu *data warehouse* mempunyai karakter menyediakan secara singkat dan sederhana gambaran seputar subjek lebih detail yang dibuat dari data luar yang tidak berguna dalam proses pendukung keputusan.

2. *Integrated*: *Data warehouse* biasanya dibangun dari bermacam-macam sumber yang berbeda, seperti *database* relasional, *flat files*, dan *on-line transaction records*. Pembersihan dan penyatuan data diterapkan untuk menjamin konsistensi dalam penamaan, struktur kode, ukuran atribut, dan yang lainnya.
3. *Time Variant*: data disimpan untuk menyajikan informasi dari sudut pandang masa lampau (misal 5 – 10 tahun yang lalu). Setiap struktur kunci dalam *data warehouse* mempunyai elemen waktu baik secara implisit maupun eksplisit
4. *Nonvolatile*: sebuah *data warehouse* secara fisik selalu disimpan terpisah dari data aplikasi operasional. Penyimpanan yang terpisah ini, *data warehouse* tidak memerlukan proses transaksi, *recovery* dan mekanisme pengendalian konkurensi. Biasanya hanya membutuhkan dua operasi dalam akses data yaitu *initial load of data* dan *access of data*

Dari pengertian tersebut, sebuah *data warehouse* merupakan penyimpanan data tetap sebagai implementasi fisik dari pendukung keputusan model data. *Data warehouse* juga biasanya dilihat sebagai arsitektur, pembangunan dan penyatuan data dari bermacam macam sumber data yang berbeda untuk mendukung struktur dan atau *query* tertentu, laporan analisis, dan pembuatan keputusan (Han, 2006 dalam Huda, 2010).

2.6 Pengertian *Data Mining*

Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan di dalam database. *Data mining* adalah proses yang menggunakan teknik *statistic*, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar (Turban dkk, 2005 dalam kusrini dkk, 2009).

Selain itu juga *data mining* adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual (Pramudiono, 2006 dalam kusrini dkk, 2009).

Data mining, sering juga disebut sebagai *knowledge discovery in database*

(KDD). KDD adalah kegiatan yang meliputi pengumpulan, pemakaian data, historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar (Santoso, 2007).

Data mining adalah kegiatan menemukan pola yang menarik dari data dalam jumlah besar, data dapat disimpan dalam *database*, *data warehouse*, atau penyimpanan informasi lainnya. *Data mining* berkaitan dengan bidang ilmu – ilmu lain, seperti *database system, data warehousing*, statistik, *machine learning*, *information retrieval*, dan komputasi tingkat tinggi. Selain itu, *data mining* didukung oleh ilmu lain seperti *neural networ*k, pengenalan pola, *spatial data analysis*, *image database*, *signal processing* (Han, 2006 dalam Huda, 2010). *Data mining* didefinisikan sebagai proses menemukan pola-pola dalam data. Proses ini otomatis atau seringnya semiotomatis. Pola yang ditemukan harus penuh arti dan pola tersebut memberikan keuntungan, biasanya keuntungan secara ekonomi. Data yang dibutuhkan dalam jumlah besar (Witten, 2005 dalam Huda, 2010).

Karakteristik *data mining* sebagai berikut:

1. *Data mining* berhubungan dengan penemuan sesuatu yang tersembunyi dan pola data tertentu yang tidak diketahui sebelumnya.
2. *Data mining* biasa menggunakan data yang sangat besar. Biasanya data yang besar digunakan untuk membuat hasil lebih dipercaya.
3. *Data mining* berguna untuk membuat keputusan yang kritis, terutama dalam strategi (Davies, 2004 dalam Huda, 2010).

Berdasarkan beberapa pengertian tersebut dapat ditarik kesimpulan bahwa *data mining* adalah suatu teknik menggali informasi berharga yang terpendam atau tersembunyi pada suatu koleksi data (*database*) yang sangat besar sehingga ditemukan suatu pola yang menarik yang sebelumnya tidak diketahui. Kata 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*, statistik dan *database*. Beberapa metode yang sering disebut- sebut dalam literatur *data mining* antara lain *clustering, classification, association rules mining, neural network, genetic algorithm* dan lain-lain (Pramudiono, 2003).

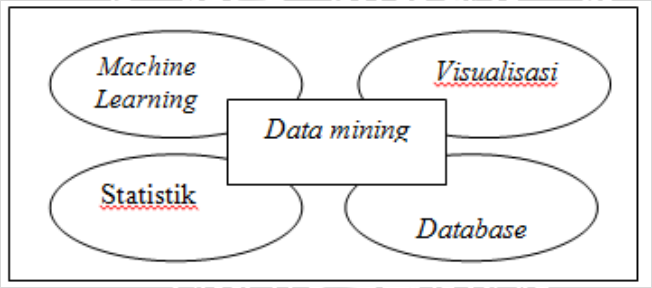
2.7 Pengenalan Pola, *Data mining*, dan *Machine Learning*

Pengenalan pola adalah suatu disiplin ilmu yang mempelajari cara-cara

mengklasifikasikan obyek ke beberapa kelas atau kategori dan mengenali kecenderungan data. Tergantung pada aplikasinya, obyek-obyek ini bisa berupa pasien, mahasiswa, pemohon kredit, *image* atau *signal* atau pengukuran lain yang perlu diklasifikasikan atau dicari fungsi regresinya (Santoso, 2007).

Data mining, sering juga disebut *knowledge discovery in database* (KDD), adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar. Keluaran dari *data mining* ini bisa dipakai untuk memperbaiki pengambilan keputusan di masa depan. Sehingga istilah *pattern recognition* jarang digunakan karena termasuk bagian dari *data mining* (Santoso, 2007).

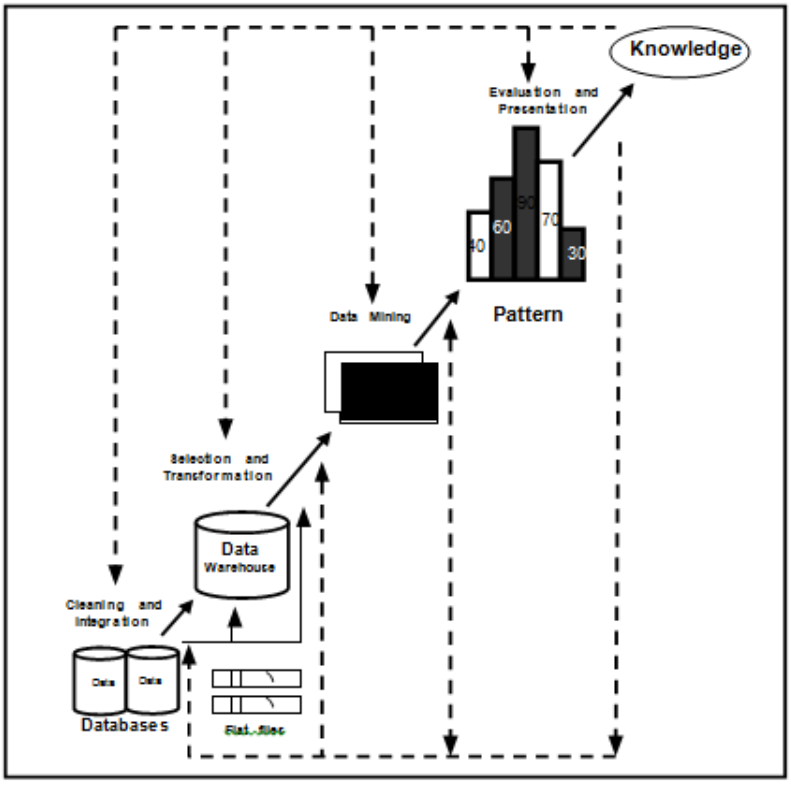
Machine Learning adalah suatu area dalam *artificial intelligence* atau kecerdasan buatan yang berhubungan dengan pengembangan teknik-teknik yang bisa diprogramkan dan belajar dari data masa lalu. Pengenalan pola, *data mining* dan *machine learning* sering dipakai untuk menyebut sesuatu yang sama. Bidang ini bersinggungan dengan ilmu probabilitas dan statistik kadang juga optimasi. *Machine learning* menjadi alat analisis dalam *data mining*. Bagaimana bidang-bidang ini berhubungan bisa dilihat dalam gambar 2.1.



Gambar 2.1 *Data mining* merupakan irisan dari berbagai disiplin (Santoso, 2007)

2.8 Tahap-Tahap *Data Mining*

Sebagai suatu rangkaian proses, *data mining* dapat dibagi menjadi beberapa tahap yang diilustrasikan di Gambar 2.2. Tahap-tahap tersebut bersifat interaktif, pemakai terlibat langsung atau dengan perantaraan *knowledge base*.



Gambar 2.2 Tahap-tahap *data mining*
(Han, 2006 dalam Huda, 2010).

Tahap-tahap *data mining* yaitu :

1. Pembersihan data (*data cleaning*)

   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. Integrasi 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-aribut yang mengidentifikasikan 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 korelasi 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 faktor kecenderungan orang membeli dalam kasus *MBA*, tidak perlu mengambil nama pelanggan, cukup dengan id pelanggan saja.

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 data kategorikal. Karenanya data berupa angka numerik yang berlanjut perlu dibagibagi 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 diperoleh tidak sesuai hipotesa ada beberapa alternatif yang dapat diambil seperti menjadikannya umpan balik untuk memperbaiki proses *data mining*, mencoba metode *data mining* lain yang lebih sesuai, atau menerima hasil ini sebagai suatu hasil yang di luar dugaan yang mungkin bermanfaat.

7. Presentasi pengetahuan (*knowledge presentation*)

   Merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna. Tahap terakhir dari proses *data mining* adalah bagaimana memformulasikan keputusan

atau aksi dari hasil analisis yang didapat. Ada kalanya hal ini harus melibatkan orang-orang yang tidak memahami *data mining*. Karenanya presentasi hasil *data mining* dalam bentuk pengetahuan yang bisa dipahami semua orang adalah satu tahapan yang diperlukan dalam proses *data mining*. Dalam presentasi ini, visualisasi juga bisa membantu mengkomunikasikan hasil *data mining* (Han, 2006 dalam Huda, 2010).

2.9 Metode *Data mining*

Dengan definisi *data mining* yang luas, ada banyak jenis metode analisis yang dapat digolongkan dalam *data mining*.

2.9.1 *Association rules*

Association rules (aturan asosiasi) atau *affinity analysis* (analisis afinitas) berkenaan dengan studi tentang "apa bersama apa". Sebagai contoh dapat berupa berupa studi transaksi di supermarket, misalnya seseorang yang membeli susu bayi juga membeli sabun mandi. Pada kasus ini berarti susu bayi bersama dengan sabun mandi. Karena awalnya berasal dari studi tentang *database* transaksi pelanggan untuk menentukan kebiasaan suatu produk dibeli bersama produk apa, maka aturan asosiasi juga sering dinamakan *MBA*.

Aturan asosiasi ingin memberikan informasi tersebut dalam bentuk hubungan "if-then" atau "jika-maka". Aturan ini dihitung dari data yang sifatnya probabilistik (Santoso, 2007).

Analisis asosiasi dikenal juga sebagai salah satu metode *data mining* yang menjadi dasar dari berbagai metode *data mining* lainnya. Khususnya salah satu tahap dari analisis asosiasi yang disebut analisis pola frekuensi tinggi (*frequent pattern mining*) menarik perhatian banyak peneliti untuk menghasilkan algoritma yang efisien. Penting tidaknya suatu aturan assosiatif dapat diketahui dengan dua parameter, *support* (nilai penunjang) yaitu prosentase kombinasi item tersebut dalam *database* dan *confidence* (nilai kepastian) yaitu kuatnya hubungan antar item dalam aturan assosiatif. Analisis asosiasi didefinisikan suatu proses untuk menemukan semua aturan assosiatif yang memenuhi syarat minimum untuk support (*minimum support*) dan syarat minimum untuk confidence (*minimum confidence*) (Pramudiono, 2007 dalam Huda, 2010).

Ada beberapa algoritma yang sudah dikembangkan mengenai aturan asosiasi, namun ada satu algoritma klasik yang sering dipakai yaitu algoritma *apriori*. Ide dasar

dari algoritma ini adalah dengan mengembangkan *frequent itemset*. Dengan menggunakan satu item dan secara rekursif mengembangkan *frequent itemset* dengan dua item, tiga item dan seterusnya hingga *frequent itemset* dengan semua ukuran.

Untuk mengembangkan *frequent set* dengan dua item, dapat menggunakan *frequent set item*. Alasannya adalah bila set satu item tidak melebihi *support minimum*, maka sembarang ukuran itemset yang lebih besar tidak akan melebihi *support minimum* tersebut. Secara umum, mengembangkan set dengan fc-item menggunakan frequent set dengan k – 1 item yang dikembangkan dalam langkah sebelumnya. Setiap langkah memerlukan sekali pemeriksaan ke seluruh isi *database*.

Dalam asosiasi terdapat istilah *antecedent* dan *consequent, antecedent* untuk mewakili bagian "jika" dan *consequent* untuk mewakili bagian "maka". Dalam analisis ini, *antecedent* dan *consequent* adalah sekelompok item yang tidak punya hubungan secara bersama (Santoso, 2007).

Dari jumlah besar aturan yang mungkin dikembangkan, perlu memiliki aturan-aturan yang cukup kuat tingkat ketergantungan antar item dalam *antecedent* dan *consequent*. Untuk mengukur kekuatan aturan asosiasi ini, digunakan ukuran *support* dan *confidence*. *Support* adalah rasio antara jumlah transaksi yang memuat *antecedent* dan *consequent* dengan jumlah transaksi. *Confidence* adalah rasio antara jumlah transaksi yang meliputi semua item dalam *antecedent* dan *consequent* dengan jumlah transaksi yang meliputi semua item dalam *antecedent.*

$$S = \frac{\Sigma\,(Ta+Tc)}{\Sigma\,(T)} \qquad (2\text{-}1)$$

Sumber: Santoso, 2007

Keterangan:

S = *Support*

$\Sigma\,(Ta+Tc)$ = jumlah transaksi yang mengandung *antecedent* dan *consequent*

$\Sigma\,(T)$ = jumlah transaksi

$$C = \frac{\Sigma\,(Ta+Tc)}{\Sigma\,(Ta)} \qquad (2\text{-}2)$$

Sumber: Santoso, 2007

Keterangan:

C = *Confidence*

$\Sigma\,(Ta+Tc)$ = jumlah transaksi yang mengandung *antecedent* dan *consequenct*

$\Sigma\,(Ta)$ = jumlah transaksi ang mengandung *antecedent*

Langkah pertama algoritma *apriori* adalah, *support* dari setiap item dihitung dengan men-*scan database*. Setelah *support* dari setiap item didapat, item yang memiliki *support* lebih besar dari *minimum support* dipilih sebagai pola frekuensi tinggi dengan panjang 1 atau sering disingkat 1-itemset. Singkatan k-itemset berarti satu set yang terdiri dari k item.

Iterasi kedua menghasilkan 2-itemset yang tiap set-nya memiliki dua item. Pertama dibuat kandidat 2-itemset dari kombinasi semua 1-itemset. Lalu untuk tiap kandidat 2-itemset ini dihitung *support*-nya dengan men-scan *database*. *Support* artinya jumlah transaksi dalam *database* yang mengandung kedua item dalam kandidat 2-itemset. Setelah *support* dari semua kandidat 2-itemset didapatkan, kandidat 2-itemset yang memenuhi syarat *minimum support* dapat ditetapkan sebagai 2-itemset yang juga merupakan pola frekuensi tinggi dengan panjang 2. (Pramudiono, 2007 dalam Huda, 2010).

Untuk selanjutnya iterasi ke-k dapat dibagi lagi menjadi beberapa bagian:

1. Pembentukan kandidat itemset

   Kandidat k-itemset dibentuk dari kombinasi (k-1)-itemset yang didapat dari iterasi sebelumnya. Satu ciri 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 men-scan *database* untuk menghitung jumlah transaksi yang memuat semua item di dalam kandidat k-itemset tersebut. Ini adalah juga ciri dari algoritma *apriori* yaitu diperlukan penghitungan dengan scan 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*. Kemudian dihitung *confidence* masing-masing kombinasi item. Iterasi berhenti ketika semua item telah dihitung sampai tidak ada kombinasi item lagi. (Pramudiono, 2007 dalam Huda, 2010)

   Secara ringkas algoritma *apriori* sebagai berikut:

   Create L1 = set of supported itemsets of cardinality one

   Set k to 2

while (Lk−1 _= ∅) {

Create Ck from Lk−1

Prune all the itemsets in Ck that

are not supported, to create Lk

Increase k by 1

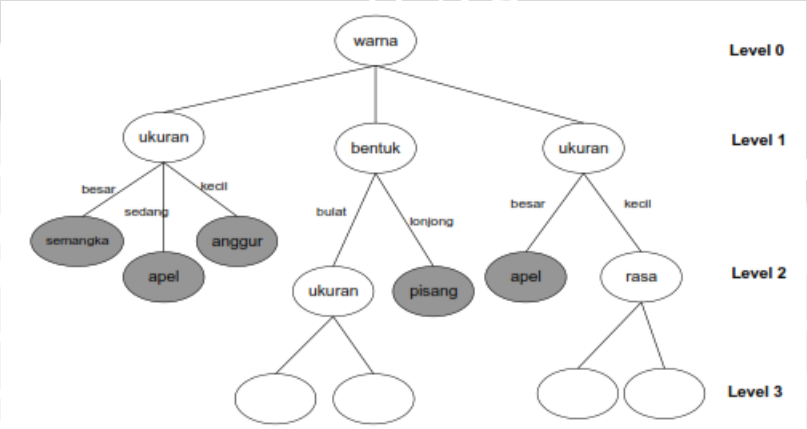
}

The set of all supported itemsets is L1 ∪ L2 ∪ · · · ∪ Lk

Selain algoritma *apriori*, terdapat juga algoritma lain seperti *FP-Grwoth*. Perbedaan algoritma *apriori* dengan *FP-Growth* pada banyaknya *scan database.* Algoritma *apriori* melakukan *scan database* setiap kali iterasi sedangkan algoritma *FP-Growth* hanya melakukan sekali di awal (Bramer, 2007 dalam Huda, 2010).

2.9.2 *Decision Tree*

Dalam *decision tree* tidak menggunakan *vector* jarak untuk mengklasifikasikan obyek. Seringkali data observasi mempunyai atribut-atribut yang bernilai nominal. Seperti yang diilustrasikan pada gambar 2.6, misalkan obyeknya adalah sekumpulan buah-buahan yang bisa dibedakan berdasarkan atribut bentuk, warna, ukuran dan rasa. Bentuk, warna, ukuran dan rasa adalah besaran nominal, yaitu bersifat kategoris dan tiap nilai tidak bisa dijumlahkan atau dikurangkan. Dalam atribut warna ada beberapa nilai yang mungkin yaitu hijau, kuning, merah. Dalam atribut ukuran ada nilai besar, sedang dan kecil. Dengan nilai-nilai atribut ini, kemudian dibuat *decision tree* untuk menentukan suatu obyek termasuk jenis buah apa jika nilai tiap-tiap atribut diberikan (Santoso, 2007).



Gambar 2.6 *Decision Tree*
(Santoso, 2007)

Ada beberapa macam algoritma *decision tree* diantaranya CART dan C4.5. Beberapa isu utama dalam *decision tree* yang menjadi perhatian yaitu seberapa detail dalam mengembangkan *decision tree*, bagaimana mengatasi atribut yang bernilai *continues*, memilih ukuran yang cocok untuk penentuan atribut, menangani data training yang mempunyai data yang atributnya tidak mempunyai nilai, memperbaiki efisiensi perhitungan (Santoso, 2007).

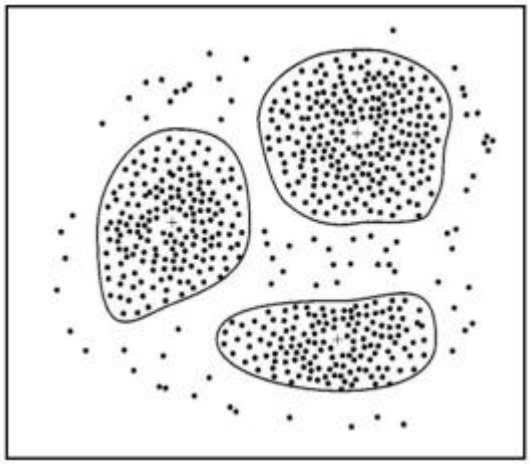
Decision tree sesuai digunakan untuk kasus-kasus yang keluarannya bernilai diskrit. Walaupun banyak variasi model *decision tree* dengan tingkat kemampuan dan syarat yang berbeda, pada umumnya beberapa ciri yang cocok untuk diterapkannya *decision tree* adalah sebagai berikut:

1. Data dinyatakan dengan pasangan atribut dan nilainya
2. Label/keluaran data biasanya bernilai diskrit
3. Data mempunyai *missing value* (nilai dari suatu atribut tidak diketahui Dengan cara ini akan mudah mengelompokkan obyek ke dalam beberapa kelompok. Untuk membuat *decision tree* perlu memperhatikan hal-hal berikut ini :
   a) Atribut mana yang akan dipilih untuk pemisahan obyek
   b) Urutan atribut mana yang akan dipilih terlebih dahulu
   c) Struktur *tree*
   d) Kriteria pemberhentian
   e) Pruning

(Santoso, 2007)

2.9.3 *Clustering*

Clustering termasuk metode yang sudah cukup dikenal dan banyak dipakai dalam *data mining*. Sampai sekarang para ilmuwan dalam bidang *data mining* masih melakukan berbagai usaha untuk melakukan perbaikan model *clustering* karena metode yang dikembangkan sekarang masih bersifat *heuristic*. Usaha-usaha untuk menghitung jumlah *cluster* yang optimal dan pengklasteran yang paling baik masih terus dilakukan. Dengan demikian menggunakan metode yang sekarang, tidak bisa menjamin hasil pengklasteran sudah merupakan hasil yang optimal. Namun, hasil yang dicapai biasanya sudah cukup bagus dari segi praktis.



Gambar 2.7 *Clustering*
(Santoso, 2007)

Tujuan utama dari metode *clustering* adalah pengelompokan sejumlah data/obyek ke dalam *cluster* (*group*) sehingga dalam setiap *cluster* akan berisi data yang semirip mungkin seperti diilustrasikan pada gambar 2.7. Dalam *clustering* metode ini berusaha untuk menempatkan obyek yang mirip (jaraknya dekat) dalam satu klaster dan membuat jarak antar klaster sejauh mungkin. Ini berarti obyek dalam satu *cluster* sangat mirip satu sama lain dan berbeda dengan obyek dalam *cluster-cluster* yang lain. Dalam metode ini tidak diketahui sebelumnya berapa jumlah *cluster* dan bagaimana pengelompokannya (Santoso, 2007).

2.10 *Market Basket Analysis* (*MBA*)

Fungsi *association rules* seringkali disebut dengan analisis keranjang belanja atau *Market Basket Analysis* (*MBA*) yang digunakan untuk menemukan relasi atau korelasi diantara himpunan *item-item.* Analisis keranjang belanja adalah analisis dari kebiasaan membeli *customer* dengan mencari asosiasi dan korelasi antara *item-item* berbeda yang diletakan *customer* dalam keranjang belanjanya.

Menurut Megaputer (2007) dalam Lestari (2009), mengemukakan *MBA* merupakan salah satu tipe analisis data yang paling sering digunakan dalam dunia pemasaran. Tujuan dari *MBA* adalah untuk menentukan produk-produk (jasa) apa saja yang paling sering dibeli atau digunakan sekaligus oleh para pelanggan. Istilah analisis ini sendiri datang dari kejadian yang sudah sangat umum terjadi di dalam *supermarket,* yaitu ketika para pelanggan memasukan semua barang yang mereka beli ke dalam yang umumnya telah disediakan oleh pihak *supermarket* itu sendiri.

Menurut David Olson dan Yong Shi (2008) dalam Lestari (2009), *MBA* mengacu

pada berbagai metodologi yang mempelajari komposisi keranjang belanja yang terdiri atas produk-produk yang dibeli pada satu kejadian belanja. Teknik ini telah dikembangkan dalam berbagai operasi pasar swalayan. Data keranjang belanja dalam bentuknya yang paling mentah adalah daftar transaksi pembelian oleh pelanggan, yang mengindikasikan hanya barang-barang yang di beli bersamaan. Data ini menantang karena beberapa hal:

1. Jumlah datanya besar (biasanya jutaan transaksi pertahun)
2. Pembagiaannya luas (setiap keranjang belanja hanya mengandung sebagian kecil dari benda-benda yang dijual)
3. Heterogenitas (orang-orang dengan selera yang berbeda cenderung sekelompok barang yang spesifik)

Menurut Albion Research (2007) dalam Lestari (2009), mengemukakan Informasi mengenai produk- produk yang biasanya dibeli secara bersama-sama oleh para pelanggan dapat memberikan wawasan tersendiri bagi para pengelola toko atau swalayan untuk menaikkan laba bisnisnya. Misalnya, seorang manajer toko bisa saja memanfaatkan informasi seperti ini untuk menempatkan produk-produk yang umumnya dibeli bersama-sama ke dalam sebuah area yang berdekatan jaraknya. Hal ini bertujuan untuk memudahkan mayoritas pembeli agar semakin senang membeli beberapa produk berlainan sekaligus. Informasi ini juga dapat digunakan untuk menyampaikan ide mendasar dari *MBA*, dimulai dengan melihat gambar keranjang belanja yang berisi bermacam-macam barang-barang yang dibeli oleh seseorang di sebuah *supermarket.*

Keranjang ini berisi bermacam-macam barang-barang seperti *orange juice*, pisang, *soft drink*, pembersih jendela, dan *detergent*. Sebuah keranjang memberitahukan kepada kita tentang apa saja yang dibeli oleh seorang konsumen dalam satu waktu. Sebuah daftar belanjaan yang lengkap yang diperoleh dari semua konsumen memberikan kita informasi yang sangat banyak, dan ini dapat menjelaskan barang-barang apa saja yang paling penting dari bisnis penjualan yaitu jenis barang dan kapan konsumen membeli. 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- konsumen untuk menyediakan tanda atau informasi yaitu siapa mereka dan mengapa mereka melakukan pembelian tersebut.

MBA juga menyediakan pengertian tentang barang dagangan dengan

memberitahukan kepada kita produk-produk mana yang memungkinkan untuk dibeli secara bersamaan dan produk mana yang lebih disetujui untuk di promosikan. Informasi ini dapat digunakan dalam menentukan tata letak rak toko, menetapkan produk-produk mana yang akan diletakan dalam tempat yang khusus, untuk mengidentifikasi ketika persoalannya berpasangan atau kupon*/issue coupons*, dan lain sebagainya. *MBA* menawarkan sekumpulan permasalahan bisnis yang berkaitan untuk mengetahui *point of sale* dari data transaksi. Teknik yang umum digunakan adalah *assosiation rules.*

Fungsi ini paling banyak digunakan untuk menganalisa data dalam rangka keperluan strategi pemasaran, *desain catalog*, dan proses pembuatan keputusan bisnis. Tipe *association rules* dapat dinyatakan dengan contoh: 70 persen dari orang-orang yang membeli mie, *juice*, dan saus akan membeli juga roti tawar. Aturan asosiasi meng*capture item* atau kejadian dalam data berukuran besar yang berisi data transaksi.

2.11 Penelitian yang Relevan

Meiwati dan Mustikasari (2006) melakukan penelitian mengenai aplikasi *data mining* menggunakan aturan asosiasi dengan metode *apriori* untuk analisis keranjang pasar pada data transaksi penjualan apotek. Dalam penelitian ini dilakukan teknik untuk menemukan pola berupa produk-produk yang sering dibeli bersamaan atau cenderung muncul bersama dalam sebuah transaksi. Penelitian ini bertujuan untuk membuat sebuah aplikasi *data mining* menggunakan aturan asosiasi dengan metode *apriori* sebagai teknik analisis keranjang pasarnya. Data yang diambil dalam penelitian ini adalah data transaksi penjualan disuatu apotek di Perumnas 1. Hasil dari aturan asosiasi yang didapat yaitu berupa kombinasi dari jenis obat yang sering dibeli oleh konsumen. Dari hasil tersebut diharapkan dapat membantu manajemen apotek untuk merancang strategi pemasaran obat di apoteknya. Aplikasi ini dibuat menggunakan perangkat lunak *Java* dan didukung oleh media penyimpanan *database Microsoft Access*.

Budhi dan Soedjianto (2007) melakukan penelitian mengenai aplikasi *data mining MBA* pada table data absensi elektronik untuk mendeteksi kecurangan absensi (*check-lock*) karyawan di perusahaan**.** Dalam penelitian ini Peneliti berkeinginan untuk mengatasi permasalahan absensi tersebut dengan memanfaatkan metode *data mining*, khususnya metode *MBA*, untuk mendeteksi kecurangan ini. Perangkat lunak yang dibuat ini akan mentranformasikan data absensi pegawai menggunakan metode *MaxDiff*

Histogram menjadi format *compact transaction* yang selanjutnya akan diproses menggunakan Algoritma *Pincer Search* menjadi *frequent itemset.* Pada akhirnya dari data *frequent itemset* ini didapat *association rule* pegawai untuk disajikan kepada pengguna, yaitu bagian *Human Resource Development* (HRD) perusahaan. Dari hasil pengujian dapat diketahui bahwa metode *Data mining MBA* dapat dimanfaatkan untuk menggali *pattern* kebiasaan absensi (*check-lock*) pegawai sebuah perusahaan. Dari sini kemungkinan terjadinya kecurangan saat melakukan absensi masuk / pulang dapat dideteksi.

Lestari (2009) melakukan penelitian mengenai analisis keranjang belanja pada data transaksi penjualan (studi kasus toserba yogya banjar). Dalam penelitian ini dilakukan pengolahan data transaksi tesebut yang dapat digunakan sebagai sumber informasi penting dalam mempertahankan keberadaan usaha tersebut. Penelitian ini bertujuan mempelajari sebuah aplikasi *MBA* yang berkaitan dengan data transaksi penjualan yang menggunakan teknik assosiatif, mengidentifikasi beberapa tipe dari kaidah asosiasi (*association rules)* yang berkaitan dengan data transaksi penjualan yaitu nilai-nilai *support* dan *confidence* serta mengetahui implikasi manajerial dari *MBA.* Penelitian ini dilaksanakan di Toserba Yogya Banjar, data yang digunakan adalah data sekunder yang bersumber dari dokumen perusahaan berupa data transaksi. Penelitian diolah dengan menggunakan *software Microsoft Office Excel* 2007 dan *software Minitab 14.* Data dianalisis dengan menggunakan algoritma *apriori* yang menghasilkan aturan asosiasi, dengan pola "*if-then*".

Tabel 2.1 Perbandingan Penelitian Terdahulu yang Relevan

Peneliti	Karakterstik Penelitian		
	Topik Penelitian	Penggunaan *software*	Hasil
Meiwati dan Mustikasari (2006)	aplikasi *data mining* menggunakan aturan asosiasi dengan metode *apriori* untuk analisis keranjang pasar pada data transaksi penjualan apotek	*JDK1.6.0*, *NetBeans IDE 6.8* dan *Microsoft Office Access*	membantu manajemen apotek untuk merancang strategi pemasaran obat di apoteknya.
Budhi dan Soedjianto (2007)	aplikasi *data mining MBA* pada tabel data absensi elektronik untuk mendeteksi kecurangan absensi (*check-lock*) karyawan di perusahaan	*MaxDiff Histogram*	menggali pattern kebiasaan absensi (check-lock) pegawai sebuah perusahaan.
Lestari (2009)	analisis keranjang belanja pada data transaksi penjualan (studi kasus toserba yogya banjar)	*Microsoft Office Excel* 2007 dan *Minitab 14*	menentukan kecenderungan pelanggan dalam membeli barang-barang secara bersamaan
Penelitian ini (2013)	perancangan aplikasi *data mining* dengan algoritma *apriori* untuk analisis keranjang belanja pada data transaksi penjualan (studi kasus di swalayan kpri universitas brawijaya)	*Visual Basic 6.0* dan *Microsoft Office Access* 2007	menghasilkan suatu sistem pendukung keputusan untuk suatu pasar swalayan dimana sistem ini dapat membantu manager dalam menentukan strategi pemasaran dan juga masalah peletakan barang.