

BAB 1

PENGENALAN DATA MINING

1.1 Apa Itu Data Mining?

Dalam definisi sederhana, data mining merupakan ekstraksi informasi ataupun pola yang penting dan menarik dari data yang terdapat dalam sebuah database dalam jumlah besar. Data mining juga dapat diartikan sebagai suatu istilah yang digunakan untuk menguraikan pengetahuan di dalam database. Secara umum data mining terbagi atas 2(dua) kata yaitu:

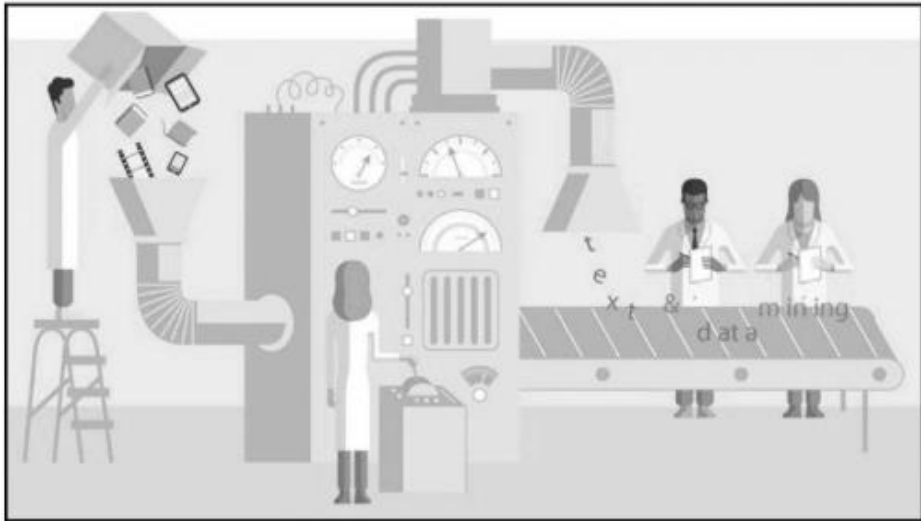
1. Data yaitu kumpulan fakta yang terekam atau sebuah entitas yang tidak memiliki arti dan selama ini terabaikan.
2. Mining yaitu proses penambangan.

BAB 1

Sehingga data mining dapat diartikan sebagai proses penambangan data yang menghasilkan sebuah output berupa pengetahuan. Berikut merupakan definisi data mining yang dikutip dari beberapa sumber, yaitu:

1. Data mining adalah analisis otomatis dari data yang berjumlah besar atau kompleks dengan tujuan untuk menemukan pola atau kecenderungan yang penting yang biasanya tidak disadari keberadaannya. (Pramudiono, 2006).
2. Data mining merupakan analisis dari peninjauan kumpulan data untuk menemukan hubungan yang tidak diduga dan meringkas data dengan cara berbeda dengan cara yang berbeda dengan sebelumnya, yang dapat dipahami dan bermanfaat bagi pemilik data. (Larose, 2005)
3. Data mining merupakan bidang dari beberapa bidang keilmuan yang menyatukan teknik dari pembelajaran mesin, pengenalan pola, statistik, database, dan visualisasi untuk penanganan permasalahan pengambilan informasi dari database yang besar (Larose, 2005)
4. Data mining adalah proses ekstraksi suatu data (sebelumnya tidak diketahui, bersifat implisit, dan dianggap tidak berguna) menjadi informasi atau pengetahuan atau pola dari data yang jumlahnya besar (Written, Ian H. Frank, 2011)
5. Menurut Gartner Group data mining adalah suatu proses menemukan hubungan yang berarti pola dan kecenderungan dengan memeriksa dalam sekumpulan besar data yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika (Larose, 2005).

BAB 1



Gambar 1.1 Ilustrasi Data Mining

Dari definisi yang telah disampaikan dapat disimpulkan bahwa data mining merupakan proses penambangan data dalam jumlah besar untuk memperoleh pengetahuan dengan menyatukan teknik dari berbagai bidang yaitu pembelajaran mesin, pengenalan pola, statistic, database, dan visualisasi informasi.

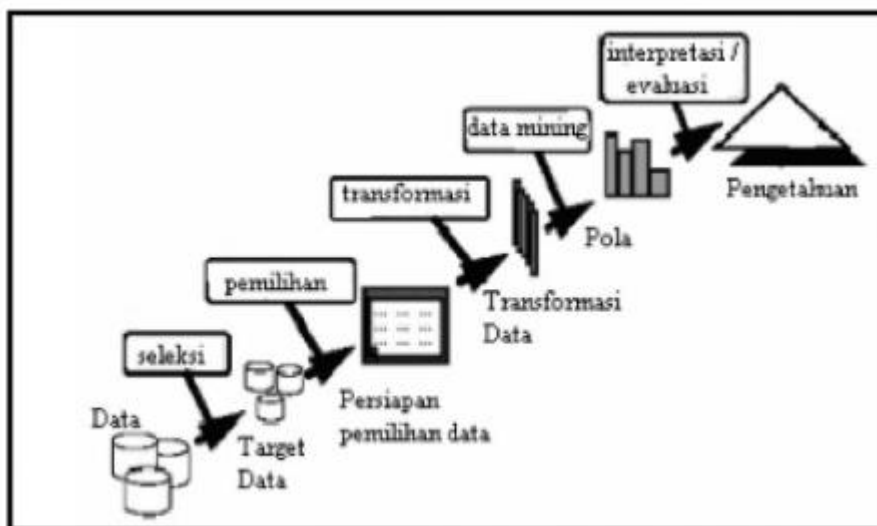
Data mining tidak hanya digunakan untuk melakukan penambangan data pada data transaksi saja. Penelitian di bidang data mining saat ini sudah merambah ke sistem database lanjut seperti *object oriented database*, *image/spatia database*, *time-series data/temporal database*, *text* (dikenal dengan nama *text mining*) dan *multimedia database*. Kemudian data mining juga memiliki karakteristik sebagai berikut:

- a. Data mining berhubungan dengan penemuan sesuatu yang tersembunyi dan pola data tertentu yang tidak diketahui sebelumnya.

BAB 1

- b. Data mining biasa menggunakan data yang sangat besar. Biasanya data tersebut digunakan untuk membuat hasil lebih dapat dipercaya.
- c. Data mining berguna untuk membuat keputusan kritis.

Dalam jurnal ilmiah data mining dikenal dengan istilah Knowledge Discovery in Database (KDD). Proses KDD secara garis besar dapat dijelaskan sebagai berikut (Fayyad, 1996).



Gambar 1.2 Proses Knowledge Discovery Database

1. Data Selection

Pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalan informasi saat KDD dimulai. Data hasil seleksi yang akan digunakan untuk proses data mining disimpan dalam suatu berkas yang terpisah dari database operasional.

BAB 1

2. *Pre-processing/Cleaning*

Sebelum proses data mining dapat dilaksanakan, perlu dilakukan proses cleaning pada data yang menjadi focus KDD. Proses cleaning mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak. Selain itu juga dilakukan proses enrichment, yaitu proses memperkaya data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan untuk KDD, seperti data atau informasi eksternal.

3. *Transformation*

Coding adalah proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses data mining. Proses *coding* dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

4. *Data Mining*

Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.

5. *Interpretation/Evaluation*

Pola informasi yang dihasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya.

BAB 1

Perlukah kita mempelajari data mining? Perlu, karena manusia banyak sekali menghasilkan data dalam jumlah besar dari berbagai bidang baik dalam bidang Bisnis, Kedokteran, Cuaca, Olahraga, Politik dan sebagainya. Dari bidang olahraga kita mengetahui berapa perolehan gol dari setiap tim dalam satu musim. Pada bidang bisnis khususnya saham, kita memperoleh data dari Bursa Efek Jakarta, kapan harga saham naik ataupun turun. Kemudian pada bidang cuaca kita memiliki data curah hujan, tingkat kelembaban, dan lain sebagainya yang diperoleh dari BMKG.

1.2 Sejarah Data Mining

Tahun 90-an merupakan awal kemunculan data mining. Data mining memang termasuk ke dalam salah satu cabang ilmu computer yang relative baru. Dan hingga saat ini orang-orang masih memperdebatkan data mining untuk ditempatkan di dalam bidang ilmu mana, karena data mining menyangkut berbagai aspek seperti database, kecerdasan buatan (*artificial intelligence*), statistic, dan sebagainya. Beberapa pihak ada yang berpendapat bahwa data mining tidak lebih dari machine learning atau analisa statistic yang berjalan di atas database saja. Namun beberapa pihak yang lain berpendapat bahawa database memiliki peranan penting dalam data mining karena data mining mengakses data yang ukurannya besar dan disini peran penting database terlihat terutama dalam proses optimisasi querynya. Data mining hadir dengan dilatarbelakangi terjadinya ledakan jumlah data yang dialami saat ini dimana banyak organisasi baik itu milik pemerintah maupun non-pemerintah yang telah mengumpulkan banyak data sekian tahun lamanya. Data tersebut hampir semuanya dimasukkan ke dalam komputer yang digunakan untuk menangani transaksi sehari-hari.

BAB 1

Bisa dibayangkan berapa transaksi perbankan yang terjadi dari sebuah bank dalam seharinya dan betapa besarnya data yang dimiliki oleh bank tersebut jika aktivitas tersebut telah berjalan beberapa tahun. Dari hal tersebut timbul sebuah pertanyaan, apakah data tersebut akan tersimpan begitu saja lalu pada akhirnya dibuang, ataukah kita dapat menggali informasi yang berguna untuk organisasi dari sekian banyak data tersebut.

Data mining melakukan eksplorasi pada basis data untuk melakukan pola-pola yang tersembunyi, mencari informasi yang mungkin saja terlupakan oleh pelaku bisnis karena hal tersebut diluar ekspektasi mereka. Sementara itu pelaku bisnis tersebut memiliki kebutuhan untuk memanfaatkan kumpulan data yang sudah dimiliki, melihat peluang tersebut para peneliti menciptakan sebuah teknologi baru yang dapat menjawab kebutuhan tersebut, yaitu data mining. Teknologi tersebut sekarang sudah banyak digunakan oleh berbagai perusahaan untuk memecahkan berbagai persoalan bisnis. Meningkatnya kebutuhan dunia bisnis yang selalu ingin memperoleh nilai tambah dari data yang telah dikumpulkan telah mendorong penerapan teknik analisis data yang berasal dari berbagai bidang seperti statistika, kecerdasan buatan, dan lain sebagainya untuk mengolah data berskala besar tersebut.

Berawal dari penerapan dalam dunia bisnis, kini data mining juga diterapkan pada bidang lain yang membutuhkan analisis data dalam skala besar seperti bioinformasi dan pertahanan negara.

1.3 Tujuan Data Mining

Penggunaan data mining tidak hanya sekedar hanya menggabungkan beberapa bidang untuk memperoleh suatu informasi saja, lebih dari itu

BAB 1

memiliki tujuan utama yaitu sebagaimana dijelaskan berikut ini. (Hoffer, Presscott, dan McFadden, 2007)

a. Explanatory

Untuk menjelaskan beberapa kondisi yang terdapat di dalam penelitian, seperti mengapa penjualan handphone meningkat di suatu negara.

b. Confirmatory

Untuk mempertegas hipotesis, seperti halnya suatu promosi yang dilakukan pada social media lebih menarik perhatian banyak orang daripada promosi yang dilakukan pada media cetak.

c. Exploratory

Untuk menganalisis data yang memiliki hubungan yang baru. Misalnya, pola apa yang cocok diterapkan untuk strategi promosi penjualan.

1.4 Fungsi Data Mining

Teknik data mining tidak hanya digunakan untuk menemukan pola namun juga dapat digunakan untuk meperediksi tren masa kini. Selain itu data mining juga memiliki keuntungan yang kompetitif termasuk di dalamnya peningkatan pendapatan, berkurangnya pengeluaran, dan meningkatnya kemampuan pasar. Data mining dibagi menjadi dua kategori utama (Han dan Kamber, 2006:21-29) yaitu:

1. Prediktif

Tujuan dari tugas prediktif adalah untuk memprediksi nilai dari atribut tertentu berdasarkan pada nilai atribut-atribut lain. Atribut yang diprediksi umumnya dikenal sebagai target atau variable tak bebas,

BAB 1

sedangkan atribut-atribut yang digunakan untuk membuat prediksi dikenal dengan istilah explanatory atau variable bebas.

2. Deskriptif

Tujuan dari tugas deskriptif adalah untuk menurunkan pola-pola (korelasi, trend, cluster, teritori, dan anomali) yang meringkas hubungan pokok dalam data. Tugas data mining deskriptif sering berupa penyelidikan dan seringkali memerlukan teknik *post-processing* untuk validasi dan penjelasan hasil.

Selain itu data mining juga memiliki beberapa fungsionalitas yaitu *Concept/Class Description: Characterization and Discrimination, Mining Frequent Patterns, Associations, and Correlations, Classification and Prediction, Cluster Analysis, Outlier analysis and Evolution analysis* (Han dan Kamber, 2006: 21-27)

Berikut ini merupakan penjelasan dari masing-masing fungsi yang telah disebutkan.

1. *Concept/Class Description: Characterization and Discrimination*

Data characterization adalah ringkasan dari semua karakteristik atau fitur dari data yang telah diperoleh dari target kelas atau fitur dari data yang telah diperoleh dari target kelas. Data yang sesuai dengan kelas yang telah ditentukan oleh pengguna biasanya dikumpulkan di dalam database. Sedangkan data discrimination adalah perbandingan antara fitur umum objek data target kelas dengan fitur umum objek dari satu atau satu set kelas lainnya. Target diambil melalui query database.

BAB 1

2. *Mining Frequent Patterns, Associations, and Correlations*

Frequent patterns adalah pola yang sering terjadi di dalam data. Frequent pattern memiliki jenis yang beragam termasuk di dalamnya pola, sekelompok item set, sub-sequence, dan sub struktur. Sebuah frequent patterns biasanya mengacu pada satu set item yang sering muncul bersama dalam sebuah kumpulan data transaksional.

Association analysis adalah pencarian aturan-aturan asosiasi yang menunjukkan kondisi-kondisi nilai atribut yang sering terjadi bersama dalam sekumpulan data. Association analysis sering digunakan untuk menganalisa *Market Basket Analysis* dan data transaksi.

3. *Classification and Prediction*

Klasifikasi adalah proses untuk menemukan model atau fungsi yang menggambarkan dan membedakan kelas data atau konsep dengan tujuan memprediksikan kelas untuk data yang tidak diketahui kelasnya. Model yang diturunkan didasarkan pada analisis dari training data. Model yang diturunkan tersebut dapat direpresentasikan dalam berbagai bentuk seperti if-then klasifikasi, decision tree, dan sebagainya.

Teknik classification bekerja dengan mengelompokkan data berdasarkan data training dan nilai atribut klasifikasi. Aturan pengelompokan tersebut akan digunakan untuk klasifikasi data baru dalam kelompok yang ada. Classification dapat direpresentasikan dalam bentuk pohon keputusan. Setiap node dalam pohon keputusan menyatakan suatu tes terhadap atribut dataset, sedangkan setiap cabang menyatakan hasil dari test tersebut. Pohon keputusan yang terbentuk dapat diterjemahkan menjadi sekumpulan aturan dalam bentuk IF condition THEN outcome. (Mewati Ayub, 2007: 7)

BAB 1

4. *Cluster Analysis*

Cluster adalah kumpulan objek data yang mirip satu sama lain dalam kelompok yang sama dan berbeda dengan objek data di kelompok lain. Sednagkan, clustering atau analisis cluster adalah proses pengelompokan satu set benda-benda fisik atau abstrak ke dalam kelas objek yang sama. Tujuannya adalah untuk menghasilkan pengelompokkan objek yang mirip satu sama lain dalam suatu kelompok. Semakin besar kemiripan objek dalam suatu cluster serta semakin besar pula perbedaan dalam suatu cluster maka kualitas analisis cluster tersebut semakin baik.

5. *Outlier Analysis*

Outlier merupakan objek data yang tidak mengikuti perilaku umum dari data. Outlier dianggap sebagai noise atau pengecualian. Analisis data outliers dapat dianggap sebagai noise atau pengecualian. Analisis data outlier dinamakan outlier mining. Teknik ini biasanya digunakan dalam fraud detection dan rare events analysis.

6. *Evolution Analysis*

Analisis evolusi data menjelaskan dan memodelkan trend dari objek yang memiliki perilaku yang berubah setiap waktu. Teknik ini dapat meliputi krakterisasi, diskriminasi, asosiasi, klarifikasi, atau clustering dari data yang berkaitan dengan waktu

1.5 Tipe Data Pada Data Mining

Secara garis besar terdapat 2 (dua) tipe data ayang harus dipahami dalam data mining yaitu:

1. Numeric merupakan tipe data yang bisa di kalkulasi

BAB 1

2. Nominal merupakan tipe data yang tidak bisa di kalkulasi baik tambah, kurang, kali, maupun bagi.

Contoh pemanfaatan tipe data dapat terlihat pada tabel 1.1 berikut.

Tabel 1.1 Tipe Data Dalam Data Mining

No	NAMA	V1	V2	V3	Ket
1	Dini	0.25	73.6	79.3	Gagal
2	Dino	3.75	98.9	87	Lulus
3	Dina	3.85	99	85	Lulus
4	Dani	0.56	60.3	65	Gagal
5	Dana	3.15	95.7	84.3	Lulus
6	Danu	0.35	52.6	56	Gagal
7	Doni	1.72	68.3	73	Gagal
8	Dono	0.75	79.4	80	Gagal

Numeric

Nominal

1.6 Perkembangan Data Mining

Perkembangan awal dari data mining dimulai pada tahun 1763 ketika Thomas Bayes mempublikasikan Teorema Bayes. Teori ini sangat penting dalam data mining, karena memungkinkan estimasi suatu kejadian berdasarkan kejadian yang telah berlangsung. Pada tahun 1805 mulai berkembang teori regresi yang mempelajari hubungan antar variable. Regresi menjadi salah satu alat penting dalam data mining.

Penggunaan computer untuk mengolah data dalam jumlah besar dimulai ketika alan turing memperkenalkan ide mesin pengolah data yang bersifat universal pada tahun 1936. Tahun 1943 Warren McCulloch dan Walters Pitts menciptakan konsep dasar jaringan syaraf tiruan. Konsep

BAB 1

jaringan syaraf tiruan berusaha meniru cara kerja otak manusia dalam mengingat pola. Pengembangan system basis data yang pesat mulai tahun 1970 memungkinkan manusia untuk menyimpan dan mengelola data berukuran besar. Perkembangan itu diikuti pula oleh perkembangan berbagai algoritma untuk pengolahan data, misalnya algoritma genetika pada tahun 1975 dan *Support Vector Machines* (SVM) pada tahun 1992.

Perkembangan pesat data mining, baik dari segi perangkat keras maupun algoritma, memungkinkan implementasi data mining dalam berbagai bidang.

Kemajuan luar biasa yang terus berlanjut pada data mining yang didorong oleh beberapa faktor, antara lain (Larose,2005) :

1. Mempunyai pertumbuhan yang cepat dalam kumpulan data.
2. Penyimpanan data, sehingga seluruh perusahaan memiliki akses ke dalam database yang andal.
3. Peningkatan akses data melalui navigasi web dan intranet
4. Adanya tekanan kompetisi bisnis untuk meingkatkan penguasaan pasar dalam globalisasi ekonomi.
5. Perkembangan teknologi perangkat lunak untuk data mining (ketersediaan teknologi)
6. Perkembangan yang hebat dalam kemampuan komputasi dan pengembangan kapasitas media penyimpanan.

Hal penting yang terkait dengan data mining adalah :

1. Data mining adalah suatu proses otomatis terhadap data yang sudah ada.
2. Data yang akan diproses berupa data yang sangat besar
3. Tujuan data mining adalah mendapatkan hubungan atau pola yang mungkin memberikan indikasi yang bermanfaat.

BAB 1

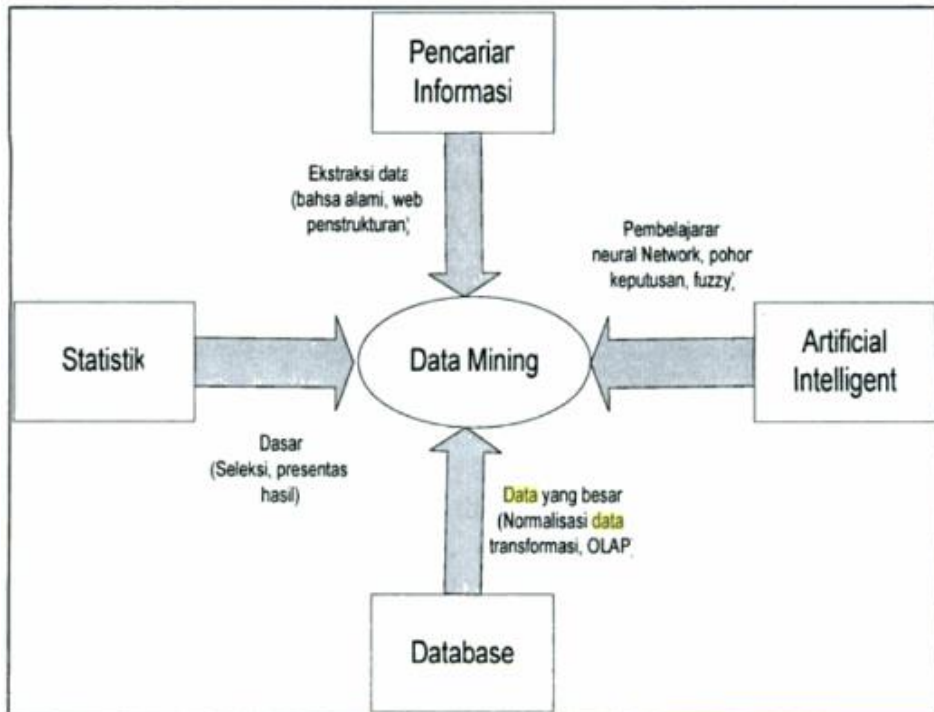
Hubungan yang dicari dalam data mining dapat berupa hubungan antara dua atau lebih dalam satu dimensi. Misalnya dalam dimensi produk dapat melihat keterkaitan pembelian suatu produk dengan produk yang lain. Selain itu, hubungan juga dapat dilihat antara dua atau lebih atribut dan dua atau lebih objek (Ponniah, 2001). Beberapa definisi awal dari data mining menyertakan fokus pada proses otomatisasi. Berry dan Linoff dalam buku *Data Mining Technique for Marketing, Sales, and Customers Support* mendefinisikan data mining sebagai suatu proses eksplorasi dan analisis secara otomatis maupun semiotomatis terhadap data dalam jumlah besar dengan tujuan menemukan pola atau aturan yang berarti (Larose, 2005).

Tiga tahun kemudian, dalam buku *mastering Data mining* memberikan definisi ulang terhadap pengertian data mining dan memberikan pernyataan bahwa jika ada yang kami sesalkan adalah frasa “secara otomatis maupun semiotomatis” karena kami merasa hal tersebut memberikan fokus lebih pada teknik otomatis dan kurang pada eksplorasi dan analisis. Hal tersebut memberikan pemahaman yang salah bahwa data mining merupakan produk yang dapat dibeli dibandingkan keilmuan yang harus dikuasai (Larose, 2005).

Dari pernyataan tersebut menegaskan bahwa dalam data mining otomatisasi tidak menggantikan campur tangan manusia. Manusia harus ikut aktif dalam setiap fase proses data mining. Kehebatan data mining yang terdapat saat ini memungkinkan terjadinya kesalahan penggunaan yang berakibat fatal. Pengguna mungkin menerapkan analisis yang tidak tepat terhadap kumpulan data dengan menggunakan pendekatan yang berbeda. Oleh karena itu, dibutuhkan pemahaman tentang statistik dan

BAB 1

struktur model matematika yang mendasari kerja perangkat lunak (Larose, 2005).



Gambar 1.3 Bidang Ilmu Data Mining

Salah satu kesulitan untuk mendefinisikan data mining adalah kenyataan bahwa data mining mewarisi banyak aspek dan teknik dari bidang-bidang ilmu yang sudah mapan terlebih dahulu. Gambar 1.3 menunjukkan bahwa data mining memiliki akar yang panjang dari bidang ilmu seperti kecerdasan buatan (artificial intelligent), machine learning, statistik, data base, dan juga information retrieval (Pramudiono, 2006).

BAB 1

1.7 Implementasi Data Mining

Data mining memiliki penerapan yang sangat luas di berbagai bidang, hal tersebut dikarenakan aktivitas yang dilakukan banyak menghasilkan data. Berikut ini merupakan contoh penerapan data mining dalam berbagai bidang.

1. Kesehatan

Pada bidang kesehatan data mining memiliki potensi yang besar untuk memperbaiki kesehatan. Analisis terhadap data praktik terbaik untuk meningkatkan perawatan serta meminimalkan biaya. Selain itu juga data mining diimplementasikan untuk diagnosis sebuah penyakit dan digunakan untuk memprediksi volume pasien dari setiap kategori. Selain itu data mining juga digunakan oleh perusahaan asuransi untuk menghindari kecurangan.

2. Analisis pasar

Analisis pasar digunakan untuk melakukan pemodelan dengan menarik kesimpulan bahwa jika seseorang membeli item tertentu, maka cenderung akan membeli item lainnya. Teknik ini memungkinkan seorang penjual memahami perilaku dari pelanggannya sehingga penataan item pada toko akan disesuaikan dengan hasil analisis tersebut. Selain itu analisis tersebut juga untuk perbandingan antara pelanggan dalam suatu kelompok demografis yang berbeda.

3. Pendidikan

Terdapat bidang baru yang muncul dalam Pendidikan, yaitu Educational Data Mining (EDM). Bidang tersebut berkaitan dengan metode pengembangan dengan menemukan pengetahuan baru yang berasal dari lingkungan Pendidikan. Tujuan dari EDM dapat

BAB 1

diidentifikasi sebagai perilaku belajar siswa di masa yang akan datang, mempelajari dampak dari dukungan pendidikan serta untuk memajukan pengetahuan ilmiah tentang pembelajaran.

4. Rekayasa Manufaktur

Dalam bidang manufaktur data mining dapat digunakan untuk menemukan proses manufaktur yang kompleks. Data mining dapat juga digunakan untuk membuat suatu rancangan system yang digunakan untuk mengekstrak hubungan antara arsitektur produk, portofolio produk, data kebutuhan pelanggan, serta perkembangan produk.

5. CRM

Customer Relationship Management merupakan bagian yang memiliki tugas untuk mengakuisisi dan mempertahankan pelanggan, serta meningkatkan loyalitas pelanggan dan menerapkan strategi yang berfokus pada pelanggan. Untuk menjaga hubungan yang benar dengan pelanggan maka pelaku bisnis harus mengumpulkan data dan menganalisis informasi. Dengan menggunakan teknologi data mining, data yang telah dikumpulkan dapat digunakan untuk analisis guna menghasilkan informasi yang bermanfaat.

6. Fraud Detection (Deteksi Penipuan/Kecurangan)

Metode tradisional yang digunakan untuk melakukan deteksi terhadap kecurangan memakan waktu dan kompleks dalam pelaksanaannya.

Data mining berperan membantu dalam menemukan pola yang berarti serta mengubah data menjadi sebuah informasi, dimana setiap informasi yang valid dan berguna merupakan pengetahuan. Sebuah system dikatakan sempurna adalah sistem yang dapat melindungi semua

BAB 1

informasi pengguna. Metode yang mendapat pengawasan yaitu pengumpulan catatan sampel, karena catatan ini tergolong curang atau tidak palsu. Kemudian dibangun sebuah model dengan menggunakan data ini dan algoritma dibuat untuk mengidentifikasi apakah rekamannya itu salah atau tidak.

7. Intrusion detection

Gangguan dapat didefinisikan sebagai setiap tindakan yang membahayakan integritas dan kerahasiaan sumber daya. Langkah yang dilakukan untuk menghindari gangguan tersebut meliputi otentikasi pengguna, meminimalkan kesalahan pemrograman, dan perlindungan terhadap informasi. Data mining dapat membantu memperbaiki deteksi intrusi dengan menambah tingkat fokus terhadap deteksi anomaly. Hal tersebut membantu analisis untuk membedakan aktivitas yang terjadi sehari-hari pada jaringan. Data mining juga dapat berguna untuk membantu mengekstrak data yang lebih relevan dengan masalah yang ada.

8. Deteksi Kebohongan

Penegakan hukum bisa dilakukan dengan menggunakan teknik data mining untuk menyelidiki kejahatan, memantau komunikasi tersangka yang dianggap sebagai teroris. Hal ini termasuk ke dalam teks mining. Proses ini berjalan untuk menemukan pola yang berarti dalam data yang biasanya berupa teks yang tidak terstruktur. Hasil pengumpulan sampel data dari penelitian sebelumnya akan dibandingkan dengan model untuk melakukan deteksi terhadap kebohongan yang dilakukan dimana model yang dibuat diciptakan sesuai dengan kebutuhan.

BAB 1

9. Segmentasi Pelanggan

Penelitian terhadap suatu pasar dapat membantu untuk melakukan segmentasi pelanggan, namun dengan penggunaan data mining hal tersebut dapat dilakukan lebih mendalam serta dapat meningkatkan efektivitas pasar. Selain itu data mining juga dapat digunakan untuk menyelaraskan pelanggan menjadi segmen yang berbeda dan dapat melakukan penentuan kebutuhan berdasarkan pelanggan. Dalam buku ini akan dilakukan simulasi dalam implementasi data mining yaitu segmentasi data pelanggan pada sebuah perusahaan yang akan dijelaskan pada bab selanjutnya.

10. Perbankan/Keuangan

Komputerisasi data dalam jumlah besar pada perbankan pelanggan mempengaruhi bertambahnya data yang diperoleh dari setiap transaksi terutama transaksi baru. Dalam hal ini data mining memiliki kontribusi untuk memecahkan masalah bisnis di bidang perbankan dan keuangan dengan menemukan pola, sebab-akibat, serta korelasi dalam informasi bisnis dan harga pasar yang tidak jelas terlihat oleh manajer dikarenakan volume data yang terlalu besar atau dihasilkan terlalu cepat. Para ahli melakukan penyaringan dan pengolahan terhadap data sehingga para manajer dapat memperoleh informasi untuk segmentasi, penargetan, perolahan, penahanan, serta pemeliharaan pelanggan yang lebih baik

11. Pengawasan Perusahaan

Pengawasan perusahaan merupakan pemantauan terhadap perilaku seseorang atau kelompok oleh perusahaan. Data yang diperoleh biasanya sering digunakan untuk tujuan pemasaran atau dijual ke perusahaan lain, namun dibagi secara reguler dengan instansi pemerintah terkait. Hal

BAB 1

tersebut dapat digunakan oleh para pelaku bisnis untuk menyesuaikan produk mereka yang diharapkan oleh pelanggan. Data tersebut juga dapat digunakan untuk tujuan pemasaran langsung, seperti iklan bertarget pada mesin pencari Google, dimana iklan ditargetkan pada pengguna dengan menganalisis riwayat pencarian mereka.

12. Analisis Riset

Data mining memiliki peran yang sangat membantu dalam pembersihan data, pra-pengolahan data serta integrasi database. Penemuan yang didapat oleh peneliti dari data yang serupa dalam database dapat membawa perubahan dalam penelitian. Identifikasi dan korelasi antar aktivitas apapun dapat diketahui. Kemudian visualisasi data yang dilakukan oleh data mining dapat memberi gambaran yang jelas tentang data.

13. Investigasi Kriminal

Proses yang bertujuan untuk mengidentifikasi karakteristik kejahatan disebut dengan kriminologi. Analisis kejahatan meliputi peninjauan dan deteksi kejahatan serta hubungannya dengan penjahat. Volume dataset kejahatan yang tinggi juga kompleksitas hubungan antara data semacam ini menjadikan kriminologi sebagai bidang yang tepat untuk menerapkan teknik data mining. Hasil laporan kejahatan berbasis teks dapat diubah menjadi file pengolahan kata dimana informasi yang diperoleh bisa digunakan untuk melakukan proses pencocokan kejahatan.

14. Bioinformatika

Pendekatan data mining dalam bioinformatika membantu untuk mengekstrak pengetahuan yang berguna dari kumpulan data yang dikumpulkan dalam biologi serta bidang ilmu kehidupan lainnya yang

BAB 1

terkait seperti kedokteran dan ilmu saraf. Pengaplikasian data mining untuk bioinformatika meliputi penemuan gen, inferensi fungsi protein, diagnosis terhadap penyakit, optimasi pengobatan penyakit, rekonstruksi jaringan interaksi protein dan gen serta prediksi lokasi sub-seluler protein.

1.8 Pengelompokan Data Mining

Berdasarkan tugas yang dapat dilakukan, data mining dibagi ke dalam beberapa kelompok, yaitu (Larose, 2005):

1. Deskripsi

Peneliti dan analis terkadang secara sederhana ingin mencoba menemukan cara untuk menggambarkan pola serta kecenderungan yang terdapat dalam data. Misal, petugas pengumpulan suara mungkin tidak dapat menemukan keterangan atau fakta bahwa siapa yang tidak cukup profesional akan sedikit didukung dalam pemilihan kepala daerah. Deskripsi dari pola dan kecenderungan sering memberikan kemungkinan penjelasan terhadap suatu pola atau kecenderungan.

2. Estimasi

Estimasi memiliki kesamaan dengan klasifikasi, kecuali pada variable target estimasi lebih ke arah numerik daripada ke arah kategori. Sebuah model dibangun dengan menggunakan record lengkap yang menyediakan nilai dari variable target sebagai nilai prediksi. Selanjutnya pada tahap peninjauan berikutnya estimasi nilai dari variable target dibuat berdasarkan nilai dari variable prediksi. Misal, akan dilakukan estimasi tekanan darah sistolik pada seorang pasien di sebuah rumah sakit berdasarkan umur pasien, jenis kelamin, indeks berat badan, dan level sodium darah. Hubungan antara tekanan darah sistolik dan nilai variable

BAB 1

prediksi dalam proses pembelajaran akan menghasilkan model estimasi. Model estimasi yang dihasilkan dapat digunakan untuk kasus baru lainnya. Contoh lain yaitu estimasi nilai indeks prestasi kumulatif mahasiswa program pascasarjana berdasarkan nilai indeks prestasi mahasiswa tersebut pada saat mengikuti program sarjana.

3. Prediksi

Prediksi memiliki karakteristik yang hamper sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan ada di masa mendatang.

Contoh prediksi dalam dunia bisnis dan penelitian adalah:

- Prediksi harga beras dalam tiga bulan yang akan datang.
- Prediksi persentase kenaikan kecelakaan lalu lintas tahun depan jika batas bawah kecepatan dinaikan.

Beberapa metode dan teknik yang digunakan dalam klasifikasi dan estimasi bisa juga digunakan (dalam keadaan yang tepat) untuk prediksi.

4. Klasifikasi

Pada klasifikasi, terdapat target variable kategori. Misal, penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu pendapatan tinggi, pendapatan sedang dan pendapatan rendah.

Contoh lain klasifikasi dalam dunia bisnis maupun penelitian adalah:

- Menentukan apakah suatu transaksi pada sebuah kartu kredit merupakan transaksi yang curang atau bukan.
- Memperkirakan apakah suatu pengajuan hipotek oleh nasabah merupakan suatu kredit yang baik atau buruk
- Mendiagnosis penyakit seorang pasien untuk mendapatkan informasi penyakit tersebut termasuk dalam kategori apa.

BAB 1

5. Pengklusteran

Pengklusteran bisa juga disebut dengan pengelompokan *record*, pengamatan, atau memperhatikan dan memebentuk kelas objek-objek yang memiliki kemiripan. Kluster adalah kumpulan *record* yang memiliki kemiripan satu dengan yang lainnya dan memiliki ketidakmiripan dengan *record* dalam kluster lain.

Pengklusteran berbeda dengan klasifikasi yaitu tidak adanya variable target dalam pengklusteran. Dalam pengklusteran tidak dilakukan proses klasifikasi, estimasi, atau memprediksi nilai dari variable target. Namun demikian, algoritma pengklusteran mencoba untuk melakukan pembagian terhadap keseluruhan data menjadi kelompok-kelompok yang memiliki kemiripan (homogen), yang mana kemiripan *record* dalam suatu kelompok akan bernilai maksimal, sedangkan kemiripan dengan record dalam kelompok lain akan bernilai minimal

Contoh pengkluteran dalam dunia bisnis maupun penelitian adalah:

- Mendapatkan kelompok-kelompok pelanggan untuk target pemsaran dari suatu produk bagi perusahaan yang tidak memiliki dana pemasaran yang besar.
- Untuk tujuan audit akuntansi, yaitu melakukan pemisahan terhadap perilaku finansial dalam baik dan mencurigakan.
- Melakukan pengklusteran terhadap ekspresi dari gen, untuk mendapatkan kemiripan perilaku dari gen dalam jumlahe besar.

6. Asosiasi

Dalam data mining, asosiasi memiliki tugas menemukan atribut yang muncul dalam satu waktu. Dalam dunia bisnis leih umum disebut dengan *Market Basket Analysis* atau analisis keranjang belanja.

BAB 1

Contoh asosiasi dalam dunia bisnis maupun penelitian adalah:

- Meneliti jumlah pelanggan dari perusahaan telekomunikasi seluler yang diharapkan untuk memberikan respons positif terhadap penawaran peningkatan layanan yang diberikan.
- Menemukan barang dalam supermarket yang dibeli secara bersamaan dan barang yang tidak pernah dibeli secara bersamaan.

1.9 Cara Kerja Data Mining

Cara kerja data mining yaitu menggali hal-hal penting yang belum diketahui sebelumnya atau memprediksi apa yang akan terjadi. Teknik yang digunakan untuk melaksanakan tugas ini disebut pemodelan. Pemodelan adalah sebuah kegiatan untuk membangun sebuah model pada situasi yang telah diketahui “jawabannya” dan kemudian menerapkannya pada situasi lain yang akan dicari jawabannya.

Data Mining untuk menentukan pola-pola dalam data. Pola yang ditemukan harus penuh arti dan pola tersebut memberikan keuntungan. Karakteristik data mining sebagai berikut :

- a. Data mining berhubungan dengan penemuan sesuatu yang tersembunyi dan pola data tertentu yang tidak diketahui sebelumnya.
- b. Data mining biasa menggunakan data yang sangat besar. Biasanya data yang besar digunakan untuk membuat hasil lebih mudah dipercaya
- c. Data mining berguna untuk membuat keputusan yang kritis, terutama strategi (Davies,2004), juga dapat digunakan untuk pengambilan keputusan dimasa depan berdasarkan informasi yang diperoleh dari data masa lalu. Tergantung pada aplikasinya, data bisa

BAB 1

berupa data mahasiswa, pasien, dll. Banyak Kasus dalam sehari-hari yang dapat diselesaikan dengan data mining diantaranya :

- 1) Memprediksi berapa jumlah mahasiswa baru di perguruan tinggi berdasarkan data pendaftar pada tahun-tahun sebelumnya.
- 2) Memprediksi nilai IPK berdasarkan nilai IP setiap semester sebelumnya.
- 3) Produk apa yang akan dibeli pelanggan secara bersamaan jika membeli produk di swalayan

Tentu masih banyak lagi contoh-contoh dalam bidang lain atau kasus lain yang kaitannya dengan panggalian data sehingga bias menghasilkan pengetahuan baru dan informasi baru yang dapat menjadi strategi dalam mengembangkan suatu bidang usaha.

1.10 Algoritma Data Mining

Proses pemecahan masalah yang dilakukan saat pengolahan data dalam data mining yang bertujuan untuk menemukan sebuah pola yang tersembunyi dalam data tidak lepas dari sebuah algoritma. Penggunaan algoritma itu sendiri disesuaikan dengan melihat informasi apa yang ingin didapat serta data yang akan diolah menggunakan data mining. Berikut ini merupakan algoritma yang populer dan sering digunakan dalam data mining.

1. Klasifikasi.

Pada data mining proses klasifikasi bertujuan untuk mengelompokkan data menjadi beberapa kelompok. Proses pengelompokan data mengacu pada data yang telah diketahui terlebih

BAB 1

dahulu kelompok atau kelasnya. Data yang belum memiliki kelompok ditentukan kelompoknya melalui proses perbandingan dengan data yang sudah diketahui kelompoknya. Berikut merupakan algoritma klasifikasi populer.

- a. *Decision tree*
- b. *Naïve bayes*
- c. *K-nearest neighbor (KNN)*
- d. Jaringan syaraf tiruan
- e. Algoritma genetika
- f. *Support vector machine (SVM)*

2. *Clustering.*

Clustering memiliki tujuan yang sama dengan klasifikasi yaitu mengelompokkan data. Namun, proses pengelompokan data pada clustering tidak menggunakan data lain yang sudah diketahui kelompoknya sebagai pembandingan. Pengelompokan clustering berlangsung otonom dengan cara membandingkan semua data yang belum memiliki kelas dan membaginya kedalam beberapa kelas berdasarkan kemiripan antara data. Beberapa algoritma clustering yang banyak digunakan adalah.

- a. *K-means*
- b. *Density-based spatial clustering of applications with noise (DBSCAN)*
- c. *Expectation-Maximization (EM)*
- d. *Fuzzy C-Means*
- e. *Hierarchical clustering*
- f. *Gaussian mixtures*

BAB 1

3. Regresi

Regresi berbeda dengan klasifikasi dan clustering yang bertujuan dalam mengelompokan data. Regresi bertujuan untuk melakukan prediksi atau peramalan. Konsep dasara regresi pada data mining diturunkan dari teori statistika. Pada dasarnya, regresi berusaha mengidentifikasi relasi antar beberapa variable terikat dengan variable bebas. Selanjutnya model matematika yang telah dihasilkan dapat digunakan untuk memperkirakan nilai dari suatu variable terikat berdasarkan nilai variable bebasnya. Berikut ini beberapa algoritma regresi yang banyak digunakan adalah.

- a. Regresi linear sederhana
- b. Regresi linear berganda

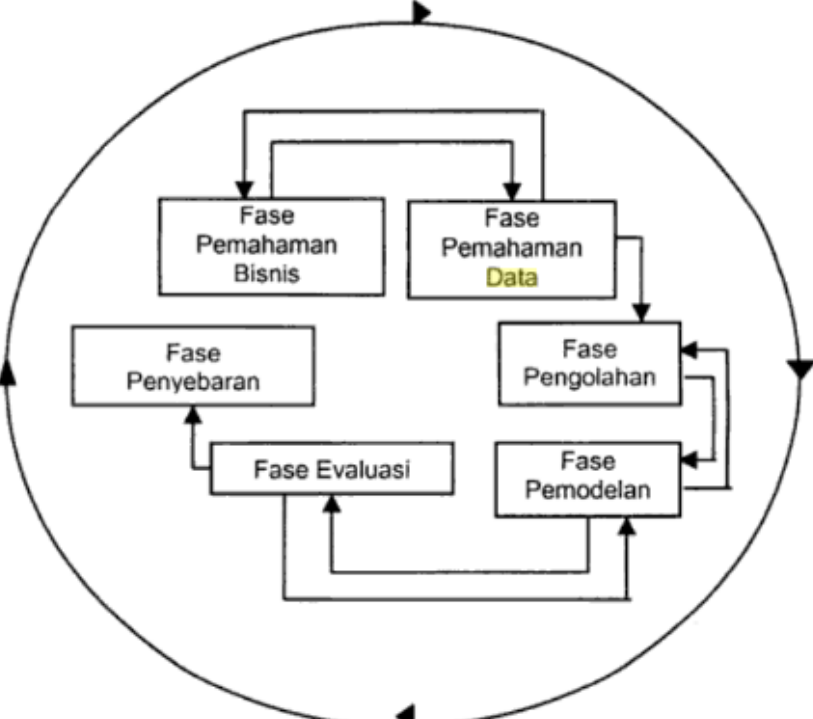
4. *Association Rule*.

Association rule merupakan metode pencarian pola relasi antar data dalam sebuah kumpulan data. Berdasarkan pola tersebut, kemunculan suatu data dapat diprediksi berdasarkan kemunculan data lainnya. Algoritma *association rule* yang populer diantaranya adalah sebagai berikut.

- a. Apriori
- b. Eclat
- c. *Frequent-pattern growth (FP-growth)*

1.11 CRISP-DM

Pada tahun 1996 analis dari beberapa industri seperti DaimlerChrysler, SPSS dan NCR mengembangkan cross-Industry Standard Process for Data Mining (CRISP-DM). CRISP DM menyediakan standar proses data mining sebagai strategi pemecahan masalah secara umum dari bisnis atau unit penelitian. Dalam CRISP-DM, sebuah proyek



BAB 1

Terdapat enam fase CRISP-DM (Larose, 2005) :

1. Fase Pemahaman Bisnis (Business Understanding Phase)
 - a. Penentuan tujuan proyek dan kebutuhan secara detail dalam lingkup bisnis atau unit penelitian secara keseluruhan.
 - b. Menerjemahkan tujuan dan batasan menjadi formula dari permasalahan data mining.
 - c. Menyiapkan strategi awal untuk mencapai tujuan
2. Fase Pemahaman Data (Data Understanding Phase)
 - a. Mengumpulkan data
 - b. Menggunakan analisis penyelidikan data untuk mengenali lebih lanjut data dan pencarian pengetahuan awal
 - c. Mengevaluasi kualitas data
 - d. Jika diinginkan, pilih sebagian kecil group data yang mungkin mengandung pola dari permasalahan
3. Fase Pengolahan Data (Data Preparation Phase)
 - a. Siapkan data awal, kumpulan data yang akan digunakan untuk keseluruhan fase berikutnya. Fase ini merupakan pekerjaan perlu dilaksanakan secara intensif
 - b. Pilih kasus dan variabel yang akan dianalisis dan sesuai analisis yang akan dilakukan.
 - c. Lakukan perubahan pada beberapa variabel jika dibutuhkan.
 - d. Siapkan data awal sehingga siap untuk ke perangkat pemodelan.
4. Fase Pemodelan (Modeling Phase)
 - a. Pilih dan aplikasikan teknik pemodelan yang sesuai
 - b. Kalibrasi aturan model untuk mengoptimalkan hasil

BAB 1

- c. Perlu diperhatikan bahwa beberapa teknik mungkin untuk digunakan pada permasalahan data mining yang sama.
 - d. Jika diperlukan, proses dapat kembali ke fase pengolahan data untuk menjadikan data ke dalam bentuk yang sesuai dengan spesifikasi kebutuhan teknik data mining tertentu.
5. Fase Evaluasi (Evaluation Phase)
- a. Mengevaluasi satu atau lebih model yang digunakan dalam fase pemodelan untuk mendapatkan kualitas dan efektivitas sebelum disebarkan untuk digunakan.
 - b. Menetapkan apakah terdapat model yang memenuhi tujuan pada fase awal
 - c. Menentukan apakah terdapat permasalahan penting dari bisnis atau penelitian yang tidak tertangani dengan baik
 - d. Mengambil keputusan berkaitan dengan penggunaan hasil dari data mining.
6. Fase Penyebaran (Deployment Phase)
- a. Menggunakan model yang dihasilkan. Terbentuknya model tidak menandakan telah terselesaikannya proyek.
 - b. Pembuatan Laporan. Penerapan proses data mining secara paralel pada departemen lain

1.12 Tools Data Mining

Tools data mining digunakan untuk menunjang proses kerja dari data mining. Tools tersebut dibuat untuk mendefinisikan dan mencapai berbagai tujuan serta untuk membantu mendapatkan informasi yang lebih

BAB 1

terperinci. Berikut ini beberapa tools atau software yang digunakan dalam data mining.

1. Rapidminer



Gambar 1.5 Logo Rapidminer

Rapidminer merupakan software yang bersifat open source. Rapidminer merupakan salahsatu solusi untuk melakukan analisis terhadap data mining, text mining dan analisis prediksi. Rapidminer menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik. Rapidminer merupakan software yang berdiri sendiri untuk analisis data dan sebagai mesin data mining yang dapat diintegrasikan pada produknya sendiri. Rapidminer ditulis dengan menggunakan bahasa pemrograman java sehingga dapat bekerja pada semua sistem operasi.

Rapidminer memiliki beberapa sifat sebagai berikut.

1. Ditulis dengan bahasa pemrograman java sehingga dapat dijalankan di berbagai sistem operasi.
2. Proses penemuan pengetahuan dimodelkan sebagai operator trees.
3. Representasi XML internal untuk memastikan format standar pertukaran data.

BAB 1

4. Bahasa scripting memungkinkan untuk eksperimen skala besar dan otomatisasi eksperimen
5. Konsep multi-layer untuk menjamin tampilan data yang efisien dan menjamin penanganan data.
6. Memiliki GUI, command line mode, dan java API yang dapat dipanggil dari program lain.

Selain itu Rapidminer juga memiliki beberapa fitur diantaranya adalah sebagai berikut.

1. Banyaknya algoritma data mining, seperti decision tree dan self-organization map.
2. Bentuk grafis yang grafis , seperti tumpeng tindih diagram histogram, tree chart dan 3D Scatter plots.
3. Banyaknya variasi plugin, spserti text plugin untuk melakukan analisis teks.
4. Menyediakan prosedur data mining dan machine learning termasuk ETL (extraction, transformation, loading), data preprocessing, visualisasi, modelling dan evaluasi.
5. Proses data mining tersusun atas operator-operator yang nestable, dideskripsikan dengan XML, dan dibuat dengan GUI.
6. Mengintegrasikan proyek data mining Weka dan statistika R.

BAB 1

2. Weka



Gambar 1.6 Logo Weka

Weka merupakan software terintegrasi yang berisi implementasi dari metode-metode data mining. Weka dikembangkan oleh Universitas Waikato, Selandia Baru menggunakan bahasa pemrograman java. Oleh karena itu, Weka merupakan singkatan dari *Waikato Environment for Knowledge Analysis*. Dengan mengadopsi konsep open source software, menjadikan Weka dapat digunakan dan dimodifikasi siapapun secara gratis.

Weka memiliki keunggulan jika dibandingkan dengan perangkat lunak data mining lainnya. Penggunaan Weka murah karena aplikasi tersebut berlisensi *GNU General Public License*, yang artinya dapat

BAB 1

digunakan secara gratis. Penggunaan bahasa pemrograman java dalam pengembangan Weka menyebabkan Weka dapat diinstal pada hampir semua sistem operasi, sepanjang sistem operasi tersebut mendukung *Java Virtual Machine*. Berbagai macam algoritma data mining, mulai dari pemrosesan awal sampai dengan pemodelan data, telah disertakan dalam Weka sehingga memudahkan pengguna dalam menganalisis data. Apabila algoritma yang akan digunakan tidak tersedia pada Weka, pengguna dapat menambahkan algoritma tersebut melalui bahasa pemrograman java. Penggunaan Weka pun tergolong mudah karena telah dibekali dengan antarmuka grafis (*Graphical User Interface*) sehingga pengguna dapat menggunakannya tanpa perlu menulis baris kode pun.

3. R



Gambar 1.7 Logo R

BAB 1

R adalah nama bahas pemrograman computer yang ditujukan secara khusus untuk menangani komuptasi statistic dan memudahkan penyajian grafik. Bahasa ini diciptakan oleh Ross Ihaka dan Robert Gentleman di Universitas Auckland, Selandia Baru. Versi pertama dirilis pada tahun 1995. Adapun nama R disematkan berdasarkan nama depan kedua penciptanya.

Bahasa R tergolong sebagai bahasa skrip, yakni bahasa yang memungkinkan perintah-perintah yang digunakan ditulis dalam skrip, berekstensi .r, yang disimpan dalam bentuk berkas teks. Selain itudengan menggunakan konsol, dimungkinkan untuk memberikan perintah secara interaktif. Dalam hal ini, begitu perintah diberikan dan tombol enter ditekan, perintah tersebut akan dieksekusi dan hasilnya akan segera terlihat. R menggunakan lisensi *GNU General Public License* sehingga dapat digunakan atau bebas dipakai oleh siapa saja selain itu bahasa R juga dapat berjalan hampir di semua sistem operasi baik itu Windows, Linux, maupun Mac OS X.

Bahasa R sangat cocok digunakan untuk menangani operasi yang melibatkan vektor dan matriks. Dengan bahasa R, operasi vektor dan matriks dapat dikerjakan dengan perintah yang sangat singkat. Dengan demikian, R dapat digunakan sebagai alternative terhadap bahasa MATLAB maupun Octave. Aplikasi utama R adalah untuk menangani komputasi statistic dan memudahkan dalam penyajian grafik. Namun , dengan paket-paket tambahan yang juga bersifat gratis, R dapat digunakan untuk menangani pengolahan citra (*image processing*), pembelajaran mesin (*machine learning*), amupun data besar (*big data*).

BAB 2

SEGMENTASI DAN *PROFILING* PELANGGAN

2.1 Segmentasi Pelanggan

Segmentasi terus menjadi konsep pemasaran yang penting juga dalam konteks relationship marketing. Meningkatkan hubungan dengan pelanggan menjadi lebih menarik dan akan menghasilkan pemahaman yang lebih baik tentang kebutuhan pelanggan. Segmentasi adalah proses membagi pelanggan menjadi beberapa klaster dengan kategori loyalitas pelanggan untuk membangun strategi pemasaran. Segmentasi pelanggan adalah salah satu langkah awal dalam membuat model bisnis.

BAB 2

Don Pepper dan Martha Roger dalam bukunya “*Managing Customer Relationship : A Strategic Framework*”. Melakukan kategori pelanggan sebagai berikut:

1. *Most Valuble Customer* (MVC), yaitu pelanggan dengan nilai paling tinggi bagi perusahaan. Merupakan pelanggan yang memberikan keuntungan terbesar bagi perusahaan.
2. *Most Groable Customer*, yaitu pelanggan yang tanpa disadari memiliki potensi besar.
3. *Below Zeros*, yaitu pelanggan yang memberikan keuntungan lebih sedikit daripada biaya untuk memberikan pelayanan.
4. *Migrators*, yaitu pelanggan yang berada pada posisi diantara *below zeros* dan *most growable customer* dan perlu dianalisis agar dapat diketahui kategori asalnya.

Segmentasi pelanggan dapat didefinisikan sebagai pembagian basis pelanggan menjadi kelompok yang berbeda dan konsisten secara internal dengan karakteristik serupa dimana memungkinkan perusahaan untuk mengembangkan strategi pemasaran yang berbeda sesuai dengan karakteristik pelanggan. Pemahaman seperti itu akan membantu perusahaan dalam mempertahankan pelanggan dan menciptakan nilai tambah bagi pelanggan itu melalui pengembangan hubungan pelanggan. Akibat dari manajemen hubungan pelanggan (Customer Relationship Management - CRM).

Segmentasi pelanggan juga mempresentasikan elemen kunci dalam identifikasi pelanggan dalam customer relationship management. (Ngai dkk, 2009). Customer relationship management berguna untuk meningkatkan hubungan dengan pelanggan, memfokuskan dalam hal

BAB 2

mengintegrasikan nilai, harapan, dan perilaku pelanggan dengan melakukan analisa data dari transaksi pelanggan (Peppard, 2000). Untuk mencapai tujuan customer relationship management , maka biasanya perusahaan memanfaatkan teknologi informasi untuk membantu perusahaan dalam mengatur hubungan pelanggan dengan suatu cara yang sistematis untuk meningkatkan loyalitas pelanggan dan meningkatkan keuntungan bisnis secara menyeluruh (Kalakota & Robinson, 1999).

Karakteristik pelanggan dapat direpresentasikan oleh beberapa kategori variabel yang terkait dengan pengelompokan, seperti berikut ini:

- Demographics: Umur, jenis kelamin, besarnya keluarga, besarnya kediaman, siklus kehidupan keluarga, pemasukan, pekerjaan atau profesi, pendidikan, kepemilikan rumah, status sosial ekonomi, agama, kewarganegaraan.
- Psychographics: kepribadian, gaya hidup, nilai-nilai, sikap.
- Behaviour: manfaat yang dicari, status pembelian, tingkat penggunaan produk, frekuensi pembelian.
- Geographic: negara, provinsi, kota, kode pos, iklim.

Skema segmentasi yang berbeda dapat dikembangkan menurut tujuan bisnis yang spesifik dari organisasi. Segmentasi umumnya digunakan melalui riset data pasar untuk mendapatkan wawasan tentang sikap pelanggan, keinginan, pandangan, preferensi, dan opini tentang perusahaan dan kompetisi. Segmentasi pelanggan berdasarkan pada riset pasar dan demografi seringkali membutuhkan pemahaman karakteristik semua pelanggan agar lebih efektif mengetahui segmen apa yang menjadi menarik pelanggan. Penggalan data dapat mengembangkan segmentasi pelanggan yang juga mengidentifikasi segmentasi pada perilaku

BAB 2

pelanggan. Selain data penelitian eksternal atau pasar, data transaksi dan pembayaran pelanggan juga dapat digunakan untuk mendapatkan wawasan tentang perilaku pelanggan. Segmentasi dengan cara tersebut, dapat mengalokasikan pelanggan untuk membentuk kelompok berdasarkan jumlah pengeluaran mereka. Hal ini dapat digunakan untuk mengidentifikasi pelanggan yang bernilai tinggi dan memprioritaskan pelayanan. Karakter dari pelanggan dijelaskan pada tabel 2.1.

Tabel 2.1 Karakteristik Pelanggan

Kelas Pelanggan	Karakteristik
Superstar	<ul style="list-style-type: none">a. Pelanggan dengan loyalti yang tinggi.b. Mempunyai nilai monetary yang tinggi.c. Mempunyai frekuensi yang tinggi.d. Mempunyai transaksi paling tinggi.
Golden Customer	<ul style="list-style-type: none">a. Mempunyai nilai monetary tertinggi yang ke dua.b. Frequency yang tinggi.c. Mempunyai rata-rata transaksi.
Typical Customer	Mempunyai rata-rata nilai monetary dan rata-rata transaksi.

BAB 2

Occasional customer	<ul style="list-style-type: none">a. Nilai monetary terendah kedua setelah dormant customerb. Nilai recency paling rendahc. Transaksi paling tinggi
Everyday shopper	<ul style="list-style-type: none">a. Memiliki peningkatan transaksib. Transaksi yang rendahc. Mempunyai nilai monetary sedang sampai dengan rendah.
Dormant customer	<ul style="list-style-type: none">a. Mempunyai frequency dan monetary yang paling rendahb. Nilai recency yang paling rendah

2.2 *Profiling Pelanggan*

Customer Profiling merupakan langkah yang dilakukan untuk memetakan dan mendalami profil pelanggan atau pelanggan dengan lebih baik. Pemetaan profil pelanggan dapat dilakukan dengan kombinasi data eksplisit (informasi mengenai pelanggan yang didapatkan dari proses pendaftaran dan survei) dan data implisit (informasi perilaku pelanggan yang didapatkan dengan pengamatan langsung).

Istilah pelanggan diartikan sebagai dua jenis pelanggan, yaitu: pelanggan individu dan pelanggan organisasi. Pelanggan individu membeli barang dan jasa untuk digunakan sendiri, maupun oleh anggota keluarga yang lain. Pelanggan individu sering disebut pelanggan akhir karena langsung digunakan oleh individunya. Sedangkan pelanggan

BAB 2

organisasi meliputi organisasi bisnis, yayasan, kantor, dan lembaga lainnya. Jenis pelanggan organisasi membeli produk dan jasa untuk menjalankan kegiatan organisasinya (tidak dikonsumsi sendiri). Perilaku pelanggan adalah tindakan yang langsung terlibat dalam mendapatkan, mengonsumsi, dan menghabiskan produk atau jasa, termasuk proses keputusan yang mendahului dan menyusuli tindakan ini.

Kebutuhan dan keinginan pelanggan selalu menjadi perhatian utama bagi pemilik usaha, yaitu dengan selalu memperhatikan perilaku pelanggannya. Oleh sebab itu, suatu perusahaan dituntut untuk selalu memperhatikan perilaku pelanggan dan menyesuaikan pengenalan produknya kepada pelanggan dengan mengadakan penyempurnaan dan perbaikan terhadap produknya serta menyesuaikan kembali kebutuhan mereka untuk saat ini maupun kebutuhan masa depan. Berikut ini merupakan definisi perilaku pelanggan menurut beberapa ahli.

1. Perilaku Pelanggan (consumer behavior) didefinisikan sebagai studi tentang unit pembelian (Buying units) dan proses pertukaran yang melibatkan perolehan, konsumsi, dan pembuangan barang, jasa, pengalaman, serta ide-ide. Proses pertukaran merupakan unsur mendasar dari perilaku pelanggan. Pertukaran terjadi antara pelanggan dengan perusahaan, disamping itu juga terjadi di antara perusahaan pada situasi pembelian industrial. Akhirnya, pertukaran juga terjadi diantara pelanggan sendiri, seperti pada saat tetangga meminjam secangkir gula atau mesin pemotong rumput (John C. Mowen and Minor, 2002).
2. Perilaku pelanggan sebagai tindakan yang langsung terlibat dalam mendapatkan, mengonsumsi, dan menghabiskan produk dan jasa,

BAB 2

termasuk proses keputusan yang mendahului dan menyusuli tindakan ini. (Engel, et al 2008).

3. Perilaku pelanggan menurut Shiffman adalah perilaku yang ditunjukkan dalam mencari, membeli, menggunakan, menilai dan menentukan produk jasa dan gagasan. Sedangkan menurut Philip perilaku pelanggan adalah Bidang ilmu perilaku pelanggan mempelajari bagaimana individu, kelompok dan organisasi memilih, memakai serta memanfaatkan barang, jasa, gagasan atau pengalaman dalam rangka memuaskan kebutuhan dan Hasrat mereka.
4. Menurut Carl McDaniel perilaku pelanggan menggambarkan bagaimana pelanggan membuat keputusan pembelian dan bagaimana mereka menggunakan serta mengatur pembelian barang atau jasa. Dari beberapa pengertian diatas disimpulkan bahwa setiap pelanggan dalam membeli produk mempunyai perilaku yang berbeda antara satu dengan yang lain.
5. Perilaku pelanggan adalah perilaku dari pelanggan akhir, individu dan rumah tangga, yang membeli barang dan jasa untuk konsumsi pribadi. Faktor-faktor yang mempengaruhi perilaku pelanggan adalah kebudayaan, sosial, pribadi, psikologis (Kotler dan Keller, 2007).

Berdasarkan pendapat para ahli tersebut, maka dapat disimpulkan bahwa perilaku pelanggan adalah tindakan yang dilakukan oleh individu, kelompok, atau organisasi yang secara langsung terlibat atau berhubungan dengan proses pengambilan keputusan yang meliputi tindakan mengevaluasi, mendapatkan, dan mengkonsumsi produk, baik barang maupun jasa yang dapat dipengaruhi lingkungan, termasuk sebelum dan sesudah proses pengambilan keputusan pembelian. Perilaku pelanggan

BAB 2

adalah dinamis, berarti bahwa perilaku seorang pelanggan, group pelanggan, ataupun masyarakat luas selalu berubah dan bergerak sepanjang waktu. Perilaku pelanggan melibatkan pertukaran. Itu merupakan hal terakhir yang ditekankan dalam definisi perilaku pelanggan, yaitu pertukaran di antara individu. Hal tersebut membuat definisi dari perilaku pelanggan tetap konsisten dengan definisi pemasaran yang sejauh ini juga menekankan pertukaran. Namun pada kenyataannya, peran pemasaran adalah untuk menciptakan pertukaran dengan pelanggan melalui formulasi dan penerapan strategi pemasaran.

2.3 Perlunya Mempelajari Perilaku Pelanggan

Kajian atau studi tentang perilaku pelanggan yang dilakukan oleh para ahli menyimpulkan bahwa mempelajari perilaku dari pelanggan itu harus dilakukan dikarenakan akan memiliki dampak yang dapat membantu para pelaku bisnis untuk melakukan hal berikut ini.

1. Merancang bauran pemasaran
2. Menetapkan segmentasi
3. Merumuskan positioning dan pembedaan produk
4. Memformulasikan analisis lingkungan bisnisnya
5. Mengembangkan riset pemasaran

Selain itu, analisis perilaku pelanggan juga memiliki peranan penting dalam merancang kebijakan publik. Bagi orang yang memiliki peranan penting pada suatu negara, kajian ini diperlukan untuk merumuskan kebijakannya dalam kerangka perlindungan pelanggan. Dengan mengetahui perilaku pelanggan mungkin dapat dimanfaatkan untuk

BAB 2

kepentingan pengembangan kemampuan seorang pelaku bisnis dalam menjalankan tugasnya.

2.4 Jenis-Jenis Pelanggan

Keputusan pelanggan untuk pembelian dan mengonsumsi suatu produk sangat dipengaruhi oleh berbagai faktor. Sebagai seorang individu, konsumsi suatu produk akan dipengaruhi oleh persepsi, proses pembelajaran dan memori, motivasi dan nilai, konsep diri, sikap, kepribadian dan gaya hidup. Sebagai pengambil keputusan, hal ini akan tergantung dari tipe keputusan (rutin atau jarang), situasi pembelian yang dihadapi, kelompok atau orang yang mempengaruhi dan menjadi acuan. Selanjutnya, kebudayaan dan subbudaya juga memiliki pengaruh kepada perilaku pelanggan. Pembahasan lengkap dari topik-topik di atas akan Saudara temukan pada modul-modul berikutnya beserta contoh-contoh untuk memudahkan saudara memahaminya.

Kata pelanggan (consumer) lebih umum menjelaskan setiap orang yang terlibat dengan suatu kegiatan, seperti yang tercantum pada definisi perilaku pelanggan di atas, yaitu mengevaluasi, memperoleh, menggunakan, dan membuang barang atau jasa. Dengan demikian, pelanggan terkait dengan hubungannya dengan perusahaan tertentu, sedangkan pelanggan tidak.

Pelanggan umum merupakan seseorang yang memiliki kebutuhan atau dorongan, melakukan pembelian, selanjutnya membuang produk dalam 3 tahap proses konsumsi (Solomon, 2002). Pelanggan memiliki beberapa peran dalam ketiga proses tersebut, yaitu berikut ini.

1. Pencetus ide (initiator).

BAB 2

2. Pembeli (Purchaser/Buyer).
3. Membayar (Payer).
4. Pengguna/pemakai (User).
5. Pemberi pengaruh (Influencer).
6. Pengambil keputusan (decision maker).
7. Pelanggan organisasi atau kelompok, di mana satu orang atau sekelompok orang akan membuat keputusan untuk organisasi

2.5 Macam-Macam Model Perilaku Pelanggan

Menurut model perilaku pelanggan yang dikemukakan oleh Henry Assael (1998) terdapat beberapa faktor yang mempengaruhi perilaku pelanggan. Dengan model perilaku pelanggan yang sederhana Henry Assael menunjukkan bahwa interaksi antara pemasar dengan pelanggan perlu dilakukan karena dapat menimbulkan adanya proses untuk merasakan dan mengevaluasi informasi merek produk, mempertimbangkan berbagai alternatif merek dapat memenuhi kebutuhan pelanggan dan pada akhirnya memutuskan merek apa yang akan dibeli pelanggan. Model perilaku pelanggan adalah suatu gambar atau kerangka yang mencerminkan atau menjelaskan tahap demi tahap yang akan dilakukan oleh pelanggan dalam memutuskan untuk melakukan keputusan pembelian.

2.6 Faktor-Faktor yang Mempengaruhi Perilaku Pelanggan

Terdapat banyak faktor yang mampu mempengaruhi perilaku pelanggan dalam melakukan sebuah pembelian. Faktor-faktor tersebut

BAB 2

berawal dari dalam diri pelanggan serta dari luar pelanggan. Keputusan pembelian dari pembeli sangat dipengaruhi oleh faktor kebudayaan, sosial, pribadi dan psikologi dari pembeli. Sebagian besar adalah faktor-faktor yang tidak dapat dikendalikan oleh pemasar, tetapi harus benar-benar diperhitungkan. Faktor-Faktor yang Mempengaruhi Perilaku Pelanggan

1. Faktor Budaya

Menurut Sumarwan (2004) budaya adalah segala nilai, pemikiran, simbol yang mempengaruhi perilaku, sikap, kepercayaan dan kebiasaan seseorang dan masyarakat. Adapun unsur-unsur budaya antara lain budaya, subbudaya dan kelas sosial.

2. Subbudaya

Setiap kebudayaan terdiri dari subbudaya-subbudaya yang lebih kecil yang memberikan identifikasi dan sosialisasi yang lebih spesifik untuk para anggotanya. Subbudaya dapat dibedakan menjadi empat jenis: kelompok nasionalisme, kelompok keagamaan, kelompok, ras, dan area geografis.

3. Kelas social

Kelas-kelas social adalah kelompok yang relative homogen dan bertahan lama dalam suatu masyarakat, yang tersusun secara hierarki dan yang keanggotaannya mempunyai nilai, minat, dan perilaku yang serupa

4. Faktor Sosial

- Kelompok Referensi

Kelompok referensi seseorang terdiri dari seluruh kelompok yang mempunyai pengaruh langsung maupun tidak langsung terhadap sikap atau perilaku seseorang. Beberapa diantaranya adalah

BAB 2

kelompok primer yaitu kelompok yang memiliki interaksi yang cukup berkesinambungan seperti, keluarga, teman, tetangga, dan teman sejawat. Kelompok sekunder merupakan kelompok yang cenderung lebih resmi dan yang mana interaksi yang terjadi kurang berkesinambungan. Kelompok yang seseorang ingin menjadi anggotanya disebut kelompok aspirasi. Kemudian kelompok diasosiatif merupakan sebuah kelompok yang nilai atau perilakunya tidak disukai oleh individu.

- Keluarga

Pengaruh Keluarga yaitu keluarga memberikan pengaruh yang besar dalam perilaku pembelian. Para pelaku pasar telah memeriksa peran dan pengaruh suami, istri, dan anak dalam pembelian produk yang berbeda. Anak-anak sebagai contoh, memberikan pengaruh yang besar dalam keputusan yang melibatkan restoran fast food. Faktor sosial terdiri dari kelompok acuan, keluarga, peran dan status (Setiadi, 2003). Dalam kehidupan pembeli keluarga dibedakan menjadi dua yaitu.

- 1) Keluarga orientasi

Merupakan orangtua seseorang. Dari orangtualah seseorang mendapatkan pandangan tentang agama, politik, ekonomi, dan merasakan ambisi pribadi nilai atau harga diri dan cinta.

- 2) Keluarga prokreasi

Merupakan pasangan hidup serta anak-anak dari seseorang, keluarga ini merupakan organisasi pembeli dari seorang pelanggan yang paling penting dalam suatu masyarakat dan telah diteliti secara intensif.

BAB 2

- Peran dan Status

Setiap orang umumnya berpartisipasi dalam kelompok selama hidupnya, baik itu keluarga, organisasi ataupun klub. Posisi seseorang dalam setiap kelompok dapat diidentifikasi dalam peran dan status.

5. Faktor Pribadi

Keputusan pembelian juga dipengaruhi oleh karakteristik pribadi, antara lain sebagai berikut

- 1) Umur dan tahapan dalam siklus hidup

Konsumsi yang dilakukan oleh seseorang juga dipengaruhi oleh tahapan siklus hidup keluarga. Beberapa penelitian juga pernah melakukan identifikasi terhadap tahapan-tahapan dalam siklus hidup psikologis. Orang-orang dewasa biasanya mengalami perubahan atau transformasi tertentu pada saat mereka menjalani hidupnya.

- 2) Pekerjaan

Para pelaku bisnis mengidentifikasi kelompok-kelompok pekerja yang memiliki minat di atas rata-rata terhadap produk dan jasa tertentu.

- 3) Kepribadian dan konsep diri

Dalam hal ini yang dimaksud kepribadian adalah karakteristik psikologis yang berbeda dan setiap orang yang memandang responsnya terhadap lingkungan yang relative konsisten. Kepribadian merupakan suatu unsur yang sangat berguna dalam hal analisis perilaku pelanggan. Bila keragaman dari kepribadian dapat diklasifikasikan dan

BAB 2

memiliki korelasi yang kuat antara jenis-jenis kepribadian tersebut dan berbagai pilihan [produk atau merek.

4) Situasi ekonomi

Keadaan atau situasi ekonomi yang dimaksud dari seseorang terdiri dari pendapatan yang dapat dibelanjakan, tabungan dan hartanya, kemampuan untuk meminjam dan sikap terhadap mengeluarkan lawan menabung.

5) Gaya hidup.

Merupakan pola hidup seseorang di dunia yang diekspresikan dengan kegiatan, minat dan pendapat seseorang. Gaya hidup menggambarkan seseorang secara utuh yang berinteraksi dengan lingkungannya. Gaya hidup juga dapat mencerminkan sesuatu di balik kelas sosial seseorang

6. Faktor Psikologis

Pilihan pembelian seseorang dipengaruhi oleh empat faktor psikologis utama yaitu

1) Motivasi

Manusia memiliki beberapa kebutuhan diantaranya bersifat biogenik, kebutuhan biogenik yaitu kebutuhan yang timbul dari suatu keadaan biologis tertentu, seperti rasa lapar, haus, resah tidak nyaman. Adapun kebutuhan lain bersifat psikogenik, yaitu kebutuhan yang timbul dari keadaan psikologis tertentu, seperti kebutuhan untuk diakui oleh orang lain ataupun sebagainya. Untuk mencukupi kebutuhan

BAB 2

tersebut diperlukan lah sebuah motivasi dimana hal tersebut dapat mendorong seseorang untuk melakukan sesuatu.

2) Persepsi

Persepsi dapat diartikan sebagai proses dimana seseorang memilih, mengorganisasikan, mengartikan masukan informasi untuk menciptakan suatu gambaran yang berarti dari dunia ini.

3) Pembelajaran

Pembelajaran biasa juga disebut dengan proses belajar dapat menjelaskan perubahan dalam perilaku seseorang yang timbul diakibatkan dari pengalaman.

4) Keyakinan dan sikap.

Keyakinan dan sikap merupakan salahsatu faktor yang penting, karena keyakinan merupakan sebuah gagasan deskriptif yang dimiliki seseorang terhadap sesuatu.

Keputusan pembelian terhadap sebuah produk yang dilakukan oleh seseorang memiliki keterkaitan yang cukup rumit dan dipengaruhi oleh berbagai faktor yang telah disebutkan sebelumnya. Namun faktor-faktor tersebut sangat berguna dalam hal mengidentifikasi pembeli atau pelanggan yang mungkin memiliki minat terhadap suatu produk.