

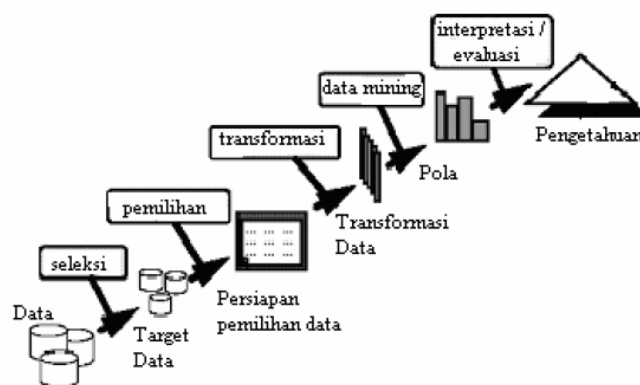
BAB II LANDASAN TEORI

2.1 Data Mining

2.1.1 Pengertian Data Mining

Kehadiran data mining dilatar belakangi dengan problema data *explosion* yang dialami akhir-akhir ini dimana banyak perusahaan telah mengumpulkan data sekian tahun lamanya (data pembelian, data penjualan, data omset penjualan, dsb). Hampir semua data tersebut dimasukkan dengan menggunakan aplikasi komputer yang digunakan untuk menangani sehari-hari yang kebanyakan adalah *OLTP*(*on line transaction processing*).

Data mining didefinisikan sebagai 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). Data mining (Connolly dan Begg, 2010) adalah suatu proses ekstraksi atau penggalian data yang belum diketahui sebelumnya, namun dapat dipahami dan berguna dari database yang besar serta digunakan untuk membuat suatu keputusan bisnis yang sangat penting. Data mining juga didefinisikan sebagai Data mining atau yang sering disebut dengan Knowledge Discovery in Database (KDD) yaitu kegiatan yang meliputi pengumpulan, pemakaian data, historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar. Tahapan data mining dalam penemuan pengetahuan (Jiawei Han, 2006) :



Gambar 2.1 Proses di dalam *Knowledge Discovery in Database*

Berikut ini adalah penjelasan dari tahapan yang ditunjukkan pada Gambar 2.1 :

1. Data cleaning, untuk membersihkan data dari noise data dan data yang tidak konsiten.
2. Data integration, megkombinasikan/mengintegrasikan beberapa sumber data.
3. Data selection, mengambil data-data yang relevan dari database untuk dianalisis.
4. Data transformation, mentransformasikan data summary ataupun operasi agregasi.
5. Data mining, merupakan proses yang esensial dimana metode digunakan untuk mengekstrak pola data yang tersembunyi.
6. Pattern evaluation, untuk mengidentifikasi pola sehingga merepresentasikan pengetahuan berdasarkan nilai-nilai yang menarik.
7. Knowledge presentation, dimana teknik representasi dan visualisai data digunakan untuk mempresentasikan pengetahuan yang diadapat kepada user

2.1.2 Metode Data Mining

Pada umumnya *data mining* dapat di kelompokkan ke dalam dua kategori yaitu: deskriptif dan prediktif. Deskriptif bertujuan untuk mencari pola yang dapat dimengerti oleh manusia yang menjelaskan karakteristik dari data. Prediktif menggunakan ciri-ciri tertentu dari data yang melakukan prediksi.

pengelompokan yang ada dalam *data mining* adalah sebagai berikut:

1. Classification

Klasifikasi (*Classification*) merupakan proses untuk menemukan sekumpulan model yang menjelaskan dan membedakan kelas-kelas data, sehingga model tersebut dapat digunakan untuk memprediksi nilai suatu kelas yang belum diketahui pada sebuah objek. Untuk mendapatkan model, kita harus melakukan analisis terhadap data latih (*training set*). Sedangkan data uji (*test set*) digunakan untuk mengetahui tingkat akurasi dari model yang dihasilkan. Klasifikasi dapat digunakan untuk memprediksi nama atau nilai kelas dari suatu obyek data.

2. *Clustering*

Pengelompokan (*Clustering*) merupakan proses untuk melakukan segmentasi. Digunakan untuk melakukan pengelompokan secara alami terhadap atribut suatu set data, termasuk kedalam supervised task. Contoh *clustering* seperti mengelompokkan dokumen berdasarkan topiknya.

3. *Assosiation*

Untuk menghasilkan sejumlah rule yang menjelaskan sejumlah data yang berhubung kuat satu dengan yang lainnya. Sebagai contoh *assosiation analysis* dapat digunakan untuk menentukan produk yang datang secara bersamaan oleh banyak pelanggan, atau bisa juga disebut dengan basket analysis.

4. *Regression*

Regression mirip dengan klasifikasi. Perbedaan utamanya adalah terletak pada atribut yang diprediksi berupa nilai yang kontinyu.

5. *Forecasting*

Prediksi (*Forecasting*) berfungsi untuk melakukan kejadian yang akan datang berdasarkan data sejarah yang ada.

6. *Sequence Analysis*

Tujuan dari metode ini adalah untuk mengenali pola dari data diskrit. Sebagai contoh adalah menemukan kelompok gen dengan tingkat ekspresi yang mirip.

7. *Deviation Analysis*

Untuk menemukan penyebab perbedaan antara data yang satu dengan data yang lain dan biasa disebut dengan *outlier detection*. Sebagai contoh adalah apakah sudah terjadi peniapaun terhadap pengguna kartu kredit dengan melihat catatan transaksi yang tersimpan dalam penggunaan basis data perusahaan kartu kredit.

2.2 Peramalan (*Forecasting*)

2.2.1 Definisi Peramalan

Peramalan pada dasarnya merupakan perkiraan suatu peristiwa di masa mendatang. Dimana situasi peramalan sangat beragam dalam horison waktu peramalan, faktor yang menentukan hasil sebenarnya, tipe pola data dan berbagai

aspek lainnya. Sebelum melakukan peramalan harus diketahui terlebih dahulu apa sebenarnya persoalan dalam pengambilan keputusan itu. Peramalan adalah pemikiran terhadap suatu besaran, misalnya untuk menentukan *omset* pada periode yang akan datang. Pada hakekatnya peramalan hanya merupakan suatu perkiraan (*guess*) dengan menggunakan teknik-teknik tertentu, maka peramalan menjadi lebih sekedar perkiraan. Peramalan dapat dikatakan perkiraan yang ilmiah (*educated guess*). Setiap pengambilan keputusan yang menyangkut keadaan di masa yang akan datang, maka pasti ada peramalan yang melandasi pengambilan keputusan. Tujuan peramalan adalah untuk meredam ketidakpastian, sehingga diperoleh suatu perkiraan yang mendekati keadaan yang sebenarnya. Jika hasil peramalan mendekati akurat, maka hal ini sangat berpengaruh besar untuk proses pengambilan keputusan pada perusahaan.

Menurut John E. Biegel:

“Peramalan adalah kegiatan memperkirakan tingkat permintaan produk yang diharapkan untuk suatu produk atau beberapa produk dalam periode waktu tertentu di masa yang akan datang”. (John E. Biegel, 1999)

Menurut Buffa:

“Peramalan atau forecasting diartikan sebagai penggunaan teknik-teknik statistik dalam bentuk gambaran masa depan berdasarkan pengolahan angka-angka historis”. (Buffa S. Elwood, 1996)

Menurut Makridakis:

“Peramalan merupakan bagian integral dari kegiatan pengambilan keputusan manajemen”. (Makridakis, 1988)

Perusahaan selalu menentukan sasaran dan tujuan, berusaha menduga faktor-faktor lingkungan, lalu memilih tindakan yang diharapkan akan menghasilkan pencapaian sasaran dan tujuan tersebut. Kebutuhan akan peramalan meningkat sejalan dengan usaha manajemen untuk mengurangi ketergantungannya pada hal-hal yang belum pasti. Peramalan menjadi lebih ilmiah sifatnya dalam menghadapi lingkungan manajemen. Karena setiap

organisasi berkaitan satu sama lain, baik buruknya ramalan dapat mempengaruhi seluruh bagian organisasi. (Makridakis, 1988)

2.2.2 Jangka waktu peramalan

Jangka waktu peramalan dapat dikelompokkan menjadi tiga kategori, yaitu (Heizer dan Render, 2005):

1. Jangka pendek (*Short Term*), peramalan untuk jangka waktu kurang dari tiga bulan.
2. Jangka menengah (*Medium Term*), peramalan untuk jangka waktu antara tiga bulan sampai tiga tahun.
3. Jangka panjang (*Long Term*), peramalan untuk jangka waktu lebih dari tiga tahun.

Untuk menghadapi penggunaan yang luas seperti itu beberapa teknik telah dikembangkan.

2.2.3 Metode Peramalan

Beberapa metode peramalan yang dapat digunakan berdasarkan sifatnya :

a. Peramalan Kualitatif

Peramalan kualitatif adalah peramalan yang didasarkan atas pendapat suatu pihak dan datanya tidak dapat direpresentasikan secara tegas menjadi suatu angka atau nilai. Hasil peramalan yang dibuat sangat bergantung pada orang yang menyusunnya. Hal ini penting karena hasil peramalan tersebut ditentukan berdasarkan pemikiran yang intuisi, pendapat dan pengetahuan serta pengalaman penyusunnya.

b. Peramalan Kuantitatif (*Statistic method*)

Peramalan kuantitatif adalah peramalan yang didasarkan atas data kuantitatif masa lalu dan dapat dibuat dalam bentuk angka (Jumingan, 2009). Peramalan kuantitatif hanya dapat digunakan apabila terdapat tiga kondisi sebagai berikut (Makridakis, 1988):

1. Informasi tentang keadaan masa lalu.
2. Informasi tersebut dapat dikuantifikasikan dalam bentuk data numerik.

3. Dapat diasumsikan bahwa beberapa aspek pola masa lalu akan terus berkelanjutan pada masa yang akan datang.

Terdapat beberapa model peramalan yang tergolong metode kuantitatif, yaitu:

- a. Model *Time series* (Deret Waktu)

Metode *Time Series* berhubungan dengan nilai-nilai suatu variabel yang diatur secara periodisasi sepanjang periode waktu dimana prakiraan permintaan diproyeksikan. Misalnya mingguan, bulanan, kwartalan, dan tahunan, tergantung keinginan dari pihak-pihak yang melakukan prakiraan permintaan ini. Metode ini semata-mata mendasarkan diri pada data dan keadaan masa lampau. Jika keadaan di masa yang akan datang cukup stabil dalam arti tidak banyak perubahan yang berarti dengan keadaan masa lampau, metode ini dapat memberikan hasil peramalan yang cukup akurat.

- b. Model *Trend Linier*

Trend Linear memiliki persamaan yang secara umum dapat dinyatakan sebagai berikut: (Riana Dwiza, 2012)

$$Y = a + bX$$

Keterangan:

Y : nilai trend pada periode tertentu

X : periode waktu

a : intersep dari persamaan trend

b : koefisien kemiringan atau gradien dari persamaan trend yang menunjukkan besarnya suatu perubahan suatu unit pada X

Ada empat metode yang bisa digunakan untuk menyusun atau menentukan trend linear, yaitu :

1. Metode Bebas (*Freehand Method*)
2. Metode Semi Rata-rata (*Semi Average Method*)

3. Metode Rata-rata Bergerak (*Moving Average Method*)

4. Metode Kuadrat Terkecil (*Least Square Method*)

2.3 Metode Kuadrat Terkecil (Least Square)

2.3.1 Pengertian Metode *Least Square*

Metode *least square* atau yang biasa disebut dengan metode kuadrat terkecil ditemukan oleh Carl F. Gauss (matematikawan dan fisikawan ternama asal Jerman, abad ke-17) ketika ia masih berumur 18 tahun, dan karyanya ini masih dipakai sampai saat ini sebagai metode yang paling baik untuk menentukan hubungan linier dari dua variabel data. Kuadrat terkecil merupakan metode yang digunakan untuk menentukan persamaan *trend* data karena metode ini menghasilkan data secara matematik. Dalam hal ini akan lebih dikhususkan untuk membahas analisis metode *least square* yang dibagi dalam dua kasus, yaitu kasus data genap dan data ganjil.

Prinsip dari metode kuadrat terkecil adalah meminimumkan jumlah kuadrat penyimpangannya (selisih) nilai variabel bebasnya (Y_i) dengan nilai trend / ramalan (Y') atau $\sum(Y_i - Y')^2$ diminimumkan.

Dengan bantuan kalkulus yaitu deviasi partial, $\sum(Y_i - Y')^2$ diminimumkan maka akan diperoleh dua buah persamaan normal sebagai berikut:

$$\sum Y_i = n \cdot a + b \cdot \sum X_i$$

$$\sum X_i Y_i = a \cdot \sum X_i + b \cdot \sum X_i^2$$

Dengan menyelesaikan kedua persamaan normal ini secara simultan, maka nilai a dan b dari persamaan trend $Y' = a + b X$ yang dicari dapat dihitung. Agar perhitungan menjadi lebih sederhana pemberian kode pada nilai X (tahun) diupayakan sedemikian rupa sehingga $\sum X_i = 0$, dengan begitu persamaan normal di atas dapat disederhanakan menjadi:

$$a = \frac{\sum Y_i}{n} \dots\dots\dots(2.1)$$

$$b = \frac{\sum X_i Y_i}{\sum X_i^2} \dots\dots\dots(2.2)$$

Setelah nilai a dan b dihitung dengan rumus di atas maka persamaan nilai trend liniernya dapat disusun sebagai berikut:

$$Y = a + bX \dots\dots\dots (2.3)$$

Y = nilai trend pada periode tertentu

a = intersep yaitu besarnya nilai Y bila nilai $X = 0$

b = slope garis trend, yaitu perubahan variabel Y untuk setiap perubahan satu unit variabel X

X = periode waktu

Untuk membuat nilai $\sum Xi = 0$ tergantung dari jumlah data tahunnya yaitu genap dan ganjil, pedomannya sebagai berikut: (Budiasih Yanti, 2012)

- (1.) Bila jumlah data tahun tidak habis dibagi dua yaitu ganjil maka dipakai skala $x = 1$ tahun. Maka tahun dasar diletakkan pada tahun yang ditengah, misalnya sebagai berikut:

Tabel 2.1 Skala X data ganjil

Tahun	2004	2005	2006	2007	2008	2009	2010
Skala x	-3	-2	-1	0	1	2	3

- (2.) Bila jumlah data tahun habis dibagi dua yaitu genap maka dipakai skala $x = 1/2$ tahun. Maka tahun dasar diletakkan pada tahun yang ditengah, misalnya sebagai berikut:

Tabel 2.2 Skala X data genap

Tahun	2004	2005	2006	2007	2008	2009	2010	2011
Skala x	-7	-5	-3	-1	1	3	5	7

Contoh soal :

1. Contoh perhitungan untuk kasus data ganjil pada penjualan tahun 2010 sampai 2014. Kemudian tentukan persamaan trendnya menurut metode kuadrat terkecil dan proyeksikan jumlah penjualan pada periode 2015.

Tabel 2.3 Data dengan jumlah n ganjil

No.	Tahun	Penjualan(Y)
1	2010	130
2	2011	145
3	2012	150
4	2013	165
5	2014	170

Penyelesaian :

- a) Jumlah data pada **Tabel 2.3** merupakan data ganjil, maka menggunakan skala x data ganjil seperti pada **Tabel 2.1**. Berikut adalah penjelasannya :

Tabel 2.4 Contoh penggunaan x

No.	Tahun	Penjualan(Y)	(X)	X^2	XY
1	2010	130	-2	4	-260
2	2011	145	-1	1	-145
3	2012	150	0	0	0
4	2013	165	1	1	165
5	2014	170	2	4	340
	Jumlah(Σ)	760	0	10	100

- b) Menghitung penjualan tahun 2015 menggunakan metode *Least Square* dengan persamaan $y = a+bx$.

Dimana untuk mencari nilai a dan b adalah:

$$a = \frac{\sum Y_i}{n}$$

$$= 760/5 = 152$$

$$b = \frac{\sum X_i Y_i}{\sum X_i^2}$$

$$= 100/10 = 10$$

$$\begin{aligned}
 Y &= a + bX \\
 &= 152 + (10 \times 3) \\
 &= 152 + 30 = 182
 \end{aligned}$$

Sehingga diperoleh hasil peramalan penjualan di tahun 2015 adalah 182.

2. Contoh perhitungan untuk kasus data genap pada penjualan tahun 2010 sampai 2015. Kemudian tentukan persamaan trendnya menurut metode kuadrat terkecil dan proyeksikan jumlah penjualan pada periode 2016.

Tabel 2.5 Data dengan jumlah n genap

No.	Tahun	Penjualan(Y)
1	2010	130
2	2011	145
3	2012	150
4	2013	165
5	2014	170
6	2015	185

Penyelesaian :

- a) Jumlah data pada **Tabel 2.5** merupakan data ganjil, maka menggunakan skala x data genap seperti pada **Tabel 2.2**. Berikut adalah penjelasannya :

Tabel 2.6 Contoh penggunaan x

No.	Tahun	Penjualan(Y)	(X)	X ²	XY
1	2010	130	-5	25	-650
2	2011	145	-3	9	-435
3	2012	150	-1	1	-150
4	2013	165	1	1	165
5	2014	170	3	9	510
6	2015	185	5	25	925
	Jumlah(Σ)	945	0	70	365

- b) Menghitung penjualan tahun 2015 menggunakan metode *Least Square* dengan persamaan $y = a+bx$.

Dimana untuk mencari nilai a dan b adalah:

$$\begin{aligned} a &= \frac{\sum Y_i}{n} \\ &= 945/6 \\ &= 157,5 \end{aligned}$$

$$\begin{aligned} b &= \frac{\sum X_i Y_i}{\sum X_i^2} \\ &= 365/70 \\ &= 5,21 \end{aligned}$$

$$\begin{aligned} Y &= a+bX \\ &= 157,5+(5,21 \times 7) \\ &= 157,5+36,47 \\ &= 193,4 \end{aligned}$$

Sehingga diperoleh hasil peramalan penjualan di tahun 2016 adalah 193,4.

2.4 Menghitung Forecast Error

Menghitung kesalahan *forecasting* sering pula disebut dengan menghitung ketepatan pengukuran (*accuracy measures*). Dalam praktek ada beberapa alat ukur yang digunakan untuk menghitung kesalahan prediksi. Berikut ini ada 2 cara untuk menghitung kesalahan prediksi :

- **Mean Absolut Error (MAD)**

Mean Absolute Deviation (MAD) adalah rata-rata nilai absolute dari kesalahan meramal (tidak dihiraukan tanda positif atau negatifnya). Berikut ini adalah persamaannya :

$$MAD = \frac{1}{n} \sum_{t=1}^n |Y_t - \hat{Y}_t| \dots \dots \dots (2.4)$$

Persamaan 2.4 digunakan untuk menghitung kesalahan error berdasarkan rata-rata nilai absolut.

- **Mean Absolute Percentage Error**

Persamaan berikut sangat berguna untuk menghitung kesalahan-kesalahan peramalan dalam bentuk persentase daripada jumlah. Mean Absolute Percentage Error (MAPE) dihitung dengan menggunakan kesalahan absolut pada tiap periode dibagi dengan nilai observasi yang nyata untuk periode itu. Kemudian, merata-rata kesalahan persentase absolut tersebut. Pendekatan ini berguna ketika ukuran atau besar variabel ramalan itu penting dalam mengevaluasi ketepatan ramalan. MAPE dapat dihitung dengan rumus sebagai berikut:

$$MAPE = \frac{100}{n} \sum_{t=1}^n \frac{|Y_t - \hat{Y}_t|}{Y_t} \dots\dots\dots (2.5)$$

Persamaan 2.5 digunakan untuk menghitung kesalahan error dengan cara dipersenkan.

Keterangan :

Y_t : nilai aktual pada periode waktu t .

\hat{Y}_t : nilai ramalan untuk periode waktu t .

n : banyak data hasil ramalan

2.5 Penelitian Sebelumnya

Penulis mengkaji hasil-hasil penelitian yang memiliki kesamaan topik dengan yang sedang diteliti oleh penulis. Adapun beberapa kajian yang berhubungan dengan topik yang sedang diteliti :

1. Mantauli Simanjuntak, 1011448, “*Sistem Informasi Prediksi Jumlah Buah Sawit Kebutuhan Produksi Pada PT. TASIK RAJA(POM) Menggunakan Metode Least Square*”. Tahun 2014, STMIK Budi Darma Medan.

Kesimpulan dari penulisan ilmiah diatas adalah sebagai berikut :

Metode *Least Square* dapat diterapkan pada prediksi kebutuhan jumlah buah sawit dengan menggunakan data jumlah buah tahun sebelumnya.

2. Muhammad Ihsan Fauzi Rambe, 1111456, “*Perancangan Aplikasi Peramalan Persediaan Obat-Obatan Menggunakan Metode Least Square*”

(*Studi Kasus : Apotik Mutiara Hati*)". Tahun 2014, STMIK Budi Darma Medan. Kesimpulan dari penulisan ilmiah diatas adalah sebagai berikut :

Analisis peramalan menggunakan metode *Least Square* dapat dipergunakan untuk meramalkan penjualan obat di periode yang akan datang, dan dapat menghasilkan hasil ramalan dengan kesalahan yang minimum (*Forecast Error*) tingkat penjualan obat-obatan pada Apotek.

3. Joko Widodo, 10204526, "*Ramalan Penjualan Sepeda Motor Honda Pada CV. RODA MITRA LESTARI*". Tahun 2008, Fakultas Ekonomi, Universitas Gunadarma Jakarta. Ramalan penjualan sepeda motor ini menggunakan metode *Least Square* dan menghasilkan ramalan dengan tingkat kesalahan MAD (*Mean Absolut Deviation*) 0,1.