

# Pengantar Data Mining

menggali pengetahuan  
dari bongkahan data

Sani Susanto, Ph.D.  
Dedy Suryadi, S.T., M.S.

17.2.11.

# PENGANTAR DATA MINING

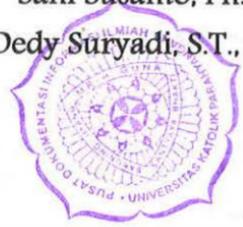
Menggali Pengetahuan dari Bongkahan Data

No. Kelas	006.3 SUS p.
No Induk	129977
Tgl	17.2.11.
Hadiah/Beli	
Dari	Penerbit Andi

# PENGANTAR DATA MINING

## Menggali Pengetahuan dari Bongkahan Data

Sani Susanto, Ph.D.  
Dedy Suryadi, S.T., M.S.



006.3  
SUS  
P

R  
129877 / FTI  
17.2.11.

Penerbit ANDI Yogyakarta

**PENGANTAR DATA MINING Menggali Pengetahuan dari Bongkahan Data**  
**Oleh: Sani Susanto & Dedy Suryadi**

Hak Cipta © 2010 pada Penulis

Editor : Nikodemus WK  
Setting : Alek  
Desain Cover : Bowo  
Korektor : Marsi / Aktor Sadewa

Hak Cipta dilindungi undang-undang.

Dilarang memperbanyak atau memindahkan sebagian atau seluruh isi buku ini dalam bentuk apapun, baik secara elektronis maupun mekanis, termasuk memfotocopy, merekam atau dengan sistem penyimpanan lainnya, tanpa izin tertulis dari Penulis.

Penerbit: C.V ANDI OFFSET (Penerbit ANDI)

Jl. Beo 38-40, Telp. (0274) 561881 (Hunting), Fax. (0274) 588282  
Yogyakarta 55281

Percetakan: ANDI OFFSET

Jl. Beo 38-40, Telp. (0274) 561881 (Hunting), Fax. (0274) 588282  
Yogyakarta 55281

**Perpustakaan Nasional: Katalog dalam Terbitan (KDT)**

Susanto, Sani

PENGANTAR DATA MINING Menggali Pengetahuan dari  
Bongkahan Data/ Sani Susanto & Dedy Suryadi;

– Ed. I. – Yogyakarta: ANDI,

**19 18 17 16 15 14 13 12 11 10**

xx + 124 hlm.; 14 x 21 Cm.

**10 9 8 7 6 5 4 3 2 1**

**ISBN: 978 – 979 – 29 – 1649 – 2**

1. Judul

1. Basis Data/Computer

2. Suryadi, Dedy

**DDC'21 : 005.74**

Sebuah persembahan untuk  
Universitas Katolik Parahyangan

## KATA PENGANTAR

Teknologi Internet telah mengubah dunia. Tentu masih lekat dalam ingatan kita masa-masa sebelum teknologi ini memasyarakat hingga ke pelosok-pelosok planet bumi ini. Pada masa itu, kegiatan mencari, memperoleh, atau memiliki informasi merupakan hal yang tidak mudah diusahakan. Mengapa demikian? Karena bahan baku informasi, yaitu data, pun sulit dicari, sulit diperoleh, dan, terlebih lagi, sulit untuk dimiliki. Kondisi tersebut sangat berbeda dengan kondisi yang kita hadapi dewasa ini. Pada masa kini, kita benar-benar diterpa "tsunami data". Data tersedia secara luar biasa melimpah. Sedemikian melimpahnya data, bahkan disertai dengan sedemikian mudahnya cara memperolehnya (sekalipun mungkin tidak untuk kita miliki), membuat kita semakin tertantang untuk bertanya, "*Pengetahuan* apakah yang dapat dihasilkan dari data atau informasi tersebut?" Buku ini kami beri judul *Pengantar Data Mining: Menggali Pengetahuan dari Bongkahan Data* karena bahasan utama buku ini berpusat pada upaya penggalian pengetahuan yang masih tersimpan dalam bongkahan data.

Dengan terbitnya buku ini, kami mengucapkan rasa syukur pertama-tama kepada Tuhan yang Mahapengasih, tanpa penyertaan dan petunjuk-Nya, proses penulisan buku tidak akan pernah selesai. Kedua, penulis menyampaikan terima kasih yang sebesar-besarnya kepada:

- Dr. Cecilia Lauw, Rektor Universitas Katolik Parahyangan, untuk dorongan tanpa henti kepada kami

untuk senantiasa berkarya bagi Universitas Katolik Parahyangan,

- Dr. Paulus Sukpto, Dekan Fakultas Teknologi Industri, Universitas Katolik Parahyangan, untuk fasilitas fakultas yang diberikan selama proses penulisan, berikut peringatan agar buku ini segera selesai,
- Dr. Kinley Aritonang, Ketua Jurusan Teknik Industri, Universitas Katolik Parahyangan, untuk fasilitas jurusan yang telah diberikan, dan
- Pihak Penerbit ANDI yang membantu menjadikan buku yang kami impikan dapat berada di tangan pembaca sekalian.

Embrio buku ini berasal dari diktat kuliah yang dibuat pada tahun 2007. Diktat kuliah itu dibuat dengan dana yang berasal dari perolehan Program Hibah Kompetisi K-1 yang telah diupayakan oleh Dr. Budi Husodo Bisowarno, Dekan Fakultas Teknologi Industri, Universitas Katolik Parahyangan periode April 2004 – Oktober 2007. Terima kasih untuk keberhasilan upayanya.

Ucapan terima kasih juga penulis ucapkan kepada segenap kolega di Jurusan Teknik Industri Universitas Katolik Parahyangan atas segenap diskusi maupun bantuannya, baik secara langsung maupun tidak langsung.

Kedua penulis juga tidak melupakan hal-hal yang bersifat pribadi. Penulis pertama mengucapkan terima kasih atas kesabaran istri (Julianti Kasih) dan kedua anak penulis (Griselda Raisa Susanto dan Hans Adrian Susanto) yang waktu kebersamaan dalam keluarga telah tersita untuk penyelesaian buku ini. Penulis kedua mengucapkan terima kasih kepada Bapak Sani

Susanto yang telah memunculkan gagasan awal hingga akhirnya buku ini berhasil diselesaikan. Besar harapan agar buku ini bermanfaat dan nyaman untuk dibaca.

Bandung, 20 Mei 2010 (Hari Kebangkitan Nasional)

Sani Susanto, Ph.D.

Dedy Suryadi, S.T., M.S.

# DAFTAR ISI



KATA PENGANTAR.....	v
DAFTAR ISI.....	ix
DAFTAR TABEL .....	xiii
DAFTAR GAMBAR.....	xvii
BAB I PENGANTAR.....	1
1.1 Pengertian Data Mining.....	2
1.2 Fungsi-fungsi dalam Data Mining .....	2
BAB II FUNGSI MINOR PERTAMA DATA MINING: DESKRIPSI.....	5
2.1 Pengantar .....	5
2.2 Berbagai Cara Deskripsi dan Pengetahuan yang Dihasilkannya.....	7
2.2.1 Deskripsi Grafis.....	7
2.2.2 Deskripsi Lokasi .....	10
2.2.3 Deskripsi Keberagaman .....	15
2.3 Deskripsi: Epilog.....	19
2.4 Soal Latihan.....	19
BAB III FUNGSI MINOR KEDUA DATA MINING: ESTIMASI.....	21
3.1 Pengantar .....	21
3.2 Pengenalan Beberapa Istilah Dasar.....	22

3.3 Estimasi Titik dan Pengetahuan yang Dihasilkannya.....	23
3.4 Selang Kepercayaan dan Pengetahuan yang Dihasilkannya.....	25
3.5 Epilog: Estimasi .....	29
3.6 Soal Latihan.....	29
<b>BAB IV FUNGSI MINOR KETIGA DATA MINING: PREDIKSI.....</b>	<b>31</b>
4.1 Pengantar .....	31
4.2 Regresi Linear dan Pengetahuan yang Dihasilkannya.....	33
4.2.1 Regresi Linear Sederhana .....	34
4.2.2 Regresi Linear Berganda ( <i>Multiple Linear Regression</i> ).....	38
4.3 Analisis Garis Regresi Menggunakan Koefisien Determinasi .....	43
4.4 Epilog: Prediksi .....	47
4.5 Soal Latihan.....	48
<b>BAB V FUNGSI MAYOR PERTAMA DATA MINING: KLASIFIKASI .....</b>	<b>51</b>
5.1 Pengantar .....	51
5.2 Pengenalan Beberapa Istilah Dasar.....	53
5.3 Prototipe Masalah Klasifikasi dan Pengetahuan yang Dihasilkannya .....	54
5.4 Algoritme Klasifikasi: CART ( <i>Classification and Regression Trees</i> ) .....	57
5.5 Klasifikasi: Pengembangan Selanjutnya .....	77

---

5.6 Klasifikasi: Epilog .....	78
5.7 Soal Latihan.....	78
<b>BAB VI FUNGSI MAYOR KEDUA DATA MINING:</b>	
<b>PENGELOMPOKAN.....</b>	<b>81</b>
6.1 Pengantar .....	81
6.2 Pengenalan Beberapa Istilah Dasar.....	83
6.3 Prototipe Masalah Pengelompokan dan Pengetahuan yang Dihasilkannya .....	84
6.4 Algoritme Pengelompokan <i>k-means</i> .....	85
6.5 Pengelompokan: Pengembangan Selanjutnya..	92
6.6 Pengelompokan: Epilog.....	93
6.7 Soal Latihan.....	93
<b>BAB VII FUNGSI MAYOR KETIGA DATA MINING:</b>	
<b>ATURAN ASOSIASI.....</b>	<b>95</b>
7.1 Pengantar .....	95
7.2 Pengenalan Beberapa Istilah Dasar.....	97
7.3 Prototipe Masalah Aturan Asosiasi dan Pengetahuan yang Dihasilkannya .....	101
7.4 Algoritme Aturan Asosiasi: MBA ( <i>Market Basket Analysis</i> ) .....	102
7.5 Aturan Asosiasi: Pengembangan Selanjutnya..	108
7.6 Aturan Asosiasi: Epilog.....	108
7.7 Soal Latihan.....	109
<b>DAFTAR PUSTAKA.....</b>	<b>113</b>
<b>BIOGRAFI SINGKAT.....</b>	<b>115</b>

## DAFTAR TABEL

TABEL 2.1	Data tinggi badan 10 siswa dari Kelas Biasa (dalam sentimeter).....	5
TABEL 2.2	Data tinggi badan 10 siswa dari Kelas Plus (dalam sentimeter).....	6
TABEL 2.3	Data tinggi badan 30 siswa dari Kelas Biasa (dalam sentimeter).....	8
TABEL 2.4	Interval tinggi badan siswa dari Kelas Biasa (dalam sentimeter).....	9
TABEL 2.5	Interval dan frekuensi tinggi badan siswa dari Kelas Biasa (dalam sentimeter).....	9
TABEL 2.6	Perincian perhitungan untuk Kelompok I dan II pada Contoh 2.2.....	18
TABEL 3.1	Data volume air minum di dalam botol.....	21
TABEL 3.2	Perincian perhitungan rata-rata dan varians untuk data TABEL 3.1 .....	24
TABEL 4.1	Data waktu pengantaran pesanan .....	31
TABEL 4.2	Perincian perhitungan untuk regresi linear sederhana.....	36
TABEL 4.3	Data waktu pengantaran pesanan yang mengikutsertakan faktor lampu merah .....	40
TABEL 4.4	Perincian perhitungan untuk regresi linear berganda .....	41

TABEL 4.5	Perincian perhitungan koefisien determinasi .	45
TABEL 5.1	Data tabungan, aset, pendapatan, dan risiko kredit nasabah Bank Bhatara Putra.....	51
TABEL 5.2	Daftar calon cabang mutakhir masalah nasabah Bank Bhatara Putra (iterasi-1) .....	60
TABEL 5.3	Perhitungan nilai kesesuaian untuk calon cabang 1, 2, 3, 4, 5, 6, 7, 8, 9.....	62
TABEL 5.4	Daftar calon cabang mutakhir masalah nasabah Bank Bhatara Putra (iterasi-2) .....	67
TABEL 5.5	Perhitungan nilai kesesuaian untuk calon cabang 1, 2, 3, 5, 6, 7, 8, 9.....	68
TABEL 5.6	Daftar calon cabang mutakhir masalah nasabah Bank Bhatara Putra (iterasi-3) .....	73
TABEL 5.7	Perhitungan nilai kesesuaian untuk calon cabang 1, 2, 5, 6, 7, 8, 9.....	74
TABEL 6.1	Data jumlah rumah dan mobil yang dimiliki oleh 8 nasabah Bank Bhatara Putra .....	81
TABEL 6.2	Hasil pengelompokan nasabah ke dalam 3 kelompok.....	84
TABEL 6.3	Perhitungan jarak catatan ke pusat kelompok dan penetapan keanggotaan catatan (iterasi-1) .....	87
TABEL 6.4	Perhitungan jarak catatan ke pusat kelompok dan penetapan keanggotaan catatan (iterasi-2).....	89

---

TABEL 6.5	Perhitungan jarak catatan ke pusat kelompok dan penetapan keanggotaan catatan (iterasi-3).....	91
TABEL 7.1	Daftar item belanja dari 14 pengunjung Toserba Favorit .....	95
TABEL 7.2	Daftar calon aturan asosiasi.....	107
TABEL 7.3	Daftar aturan asosiasi yang memenuhi syarat min ( <i>support</i> ) = 30% dan min ( <i>confidence</i> ) = 70% .....	108

## DAFTAR GAMBAR

Gambar 2.1	Diagram titik untuk tinggi badan siswa dari Kelas Biasa dan Kelas Plus .....	8
Gambar 2.2	Histogram untuk tinggi badan 30 siswa dari Kelas Biasa .....	10
Gambar 3.1	Cara membaca tabel distribusi normal .....	26
Gambar 4.1	Data TABEL 4.1 dengan Jarak sebagai sumbu X dan Waktu sebagai sumbu Y.....	34
Gambar 4.2	Contoh tiga garis lurus untuk mendekati titik-titik data TABEL 4.1.....	35
Gambar 5.1	Pohon keputusan bagi masalah klasifikasi data nasabah Bank Bhatara Putra.....	55
Gambar 5.2	Pohon keputusan bagi masalah klasifikasi data nasabah Bank Bhatara Putra (iterasi-1) ....	66
Gambar 5.3	Pohon keputusan bagi masalah klasifikasi data nasabah Bank Bhatara Putra (iterasi-2) ....	72
Gambar 5.4	Pohon keputusan bagi masalah klasifikasi data nasabah Bank Bhatara Putra (iterasi-3) ....	76

# BAB I

## PENGANTAR



Istilah *gelombang ketiga* atau *gelombang informasi* pertama kali terdengar sekitar seperempat abad yang lalu. Futuris Alvin Toffler-lah yang memperkenalkannya melalui bukunya *The Third Wave*, yang sudah diterjemahkan ke dalam berbagai bahasa. Sekalipun akan, sedang, atau bahkan telah muncul gelombang-gelombang berikutnya, satu hal yang pasti adalah bahwa gelombang ketiga ini belum berlalu, melainkan tetap ada, hanya mungkin disertai dengan hadirnya gelombang yang lain, yaitu *gelombang keempat* atau *gelombang pengetahuan*.

Diakui atau tidak, teknologi Internet telah mengubah dunia. Tentu masih lekat dalam ingatan kita masa-masa sebelum teknologi ini memasyarakat hingga ke pelosok-pelosok planet bumi ini. Pada masa itu, kegiatan mencari, memperoleh, atau memiliki informasi merupakan hal yang tidak mudah diusahakan. Mengapa demikian? Karena bahan baku informasi, yaitu data, pun sulit dicari, sulit diperoleh, dan, terlebih lagi, sulit untuk dimiliki. Kondisi tersebut sangat berbeda dengan kondisi yang kita hadapi dewasa ini. Pada masa kini, kita benar-benar diterpa "tsunami data". Data begitu melimpah. Melimpahnya data, bahkan disertai dengan mudahnya cara untuk memperolehnya (sekalipun mungkin tidak untuk kita miliki), membuat kita semakin tertantang untuk bertanya, "*Pengetahuan* apakah yang dapat dihasilkan dari data atau informasi tersebut?"

## 1.1 PENGERTIAN DATA MINING

Istilah *data mining* memiliki beberapa padanan, seperti *knowledge discovery* ataupun *pattern recognition*. Kedua istilah tersebut sebenarnya memiliki ketepatannya masing-masing. Istilah *knowledge discovery* atau *penemuan pengetahuan* tepat digunakan karena tujuan utama dari data mining memang untuk mendapatkan pengetahuan yang masih tersembunyi di dalam bongkahan data. Istilah *pattern recognition* atau *pengenalan pola* pun tepat untuk digunakan karena pengetahuan yang hendak digali memang berbentuk pola-pola yang mungkin juga masih perlu digali dari dalam bongkahan data yang tengah dihadapi. Bila dalam tulisan ini digunakan istilah *data mining*, hal ini lebih didasarkan pada lebih populernya istilah tersebut dalam kegiatan penggalian pengetahuan data.

Jadi, apakah sebenarnya data mining itu? Banyak definisi bagi istilah ini dan belum ada yang dibakukan atau disepakati semua pihak. Namun demikian, istilah ini memiliki hakikat (*notion*) sebagai disiplin ilmu yang tujuan utamanya adalah untuk menemukan, menggali, atau menambang pengetahuan dari data atau informasi yang kita miliki. Kegiatan inilah yang menjadi garapan atau perhatian utama dari disiplin ilmu data mining.

## 1.2 FUNGSI-FUNGSI DALAM DATA MINING

Lalu, fungsi atau subkegiatan apa sajakah yang ada dalam data mining dalam rangka menemukan, menggali, atau menambang pengetahuan tersebut? Mengacu kepada Larose (2005), terdapat enam fungsi dalam data mining, yaitu

1. fungsi deskripsi (*description*),
2. fungsi estimasi (*estimation*),

3. fungsi prediksi (*prediction*),
4. fungsi klasifikasi (*classification*),
5. fungsi pengelompokan (*classification*), dan
6. fungsi asosiasi (*association*).

Mengacu kepada Berry dan Browne (2006), keenam fungsi data mining tersebut dapat dipilah menjadi:

1. fungsi minor atau fungsi tambahan, yang meliputi ketiga fungsi yang pertama, yaitu deskripsi, estimasi, dan prediksi;
2. fungsi mayor atau fungsi utama, yang meliputi ketiga fungsi berikutnya, yaitu klasifikasi, pengelompokan, dan asosiasi.

Fungsi mayor maupun fungsi minor dari data mining akan dibahas dalam buku ini.

## BAB II

### FUNGSI MINOR PERTAMA

### DATA MINING: DESKRIPSI

#### 2.1 PENGANTAR

Sebagai gambaran kasus, Anda akan diberi data tinggi badan para siswa dari kelas yang tidak diberi makanan suplemen dan dari kelas yang diberi makanan suplemen peninggi badan. Sebut saja keduanya Kelas Biasa dan Kelas Plus. Data tinggi badan para siswa selengkapnya (dalam sentimeter) ditampilkan pada TABEL 2.1 dan TABEL 2.2

**TABEL 2.1** Data tinggi badan 10 siswa dari Kelas Biasa (dalam sentimeter)

Nomor Siswa (Kelas Biasa)	Tinggi Badan
1	168
2	164
3	167
4	164
5	171
6	166
7	169
8	172
9	166
10	166

TABEL 2.2 Data tinggi badan 10 siswa dari Kelas Plus (dalam sentimeter)

Nomor Siswa (Kelas Plus)	Tinggi Badan
1	175
2	176
3	183
4	180
5	177
6	177
7	182
8	179
9	179
10	171

Ketika diberi sekumpulan data, terkadang agak sukar bagi kita untuk menangkap arti kumpulan data tersebut. Sekumpulan angka-angka tersebut perlu dirangkum sedemikian rupa agar dapat “berbicara” sehingga kita memiliki gambaran mengenai kumpulan data tersebut.

Lebih jauh, bayangkan seandainya peneliti memutuskan untuk melibatkan 10.000 siswa. Peneliti tersebut melibatkan begitu banyak orang agar hasil penelitiannya lebih dapat dipercaya. Akibatnya, kita justru akan dihadapkan pada 10.000 angka! Menelusuri deretan angka sebanyak itu dengan mata manusia tentu tidak akan membawa manfaat apa pun. Oleh karena itu, kita memerlukan suatu cara untuk menggambarkan sekumpulan data secara ringkas. Cara tersebut dinamakan **deskripsi**, yang merupakan fungsi minor pertama dari data mining dan akan dibahas pada bab ini.

Subbab selanjutnya akan membahas berbagai cara untuk mendeskripsikan data. Di dalam pembahasan tersebut juga akan