



# DATA MINING DAN PENERAPAN ALGORITMA

Ni Luh Wiwik Sri Rahayu Ginantra • Fatimah Nur Arifah • Anggi Hadi Wijaya  
Ri Sabti Septarini • Nazaruddin Ahmad • Dewa Putu Yudhi Ardiana  
Fariel Effendy • Akbar Iskandar • Hazriani • Ika Yusnita Sari  
Zelvi Gustiana • Cahyo Prianto • Dudih Gustian • Edi Surya Negara



# **DATA MINING DAN PENERAPAN ALGORITMA**

## UU 28 tahun 2014 tentang Hak Cipta

### Fungsi dan sifat hak cipta Pasal 4

Hak Cipta sebagaimana dimaksud dalam Pasal 3 huruf a merupakan hak eksklusif yang terdiri atas hak moral dan hak ekonomi.

### Pembatasan Perlindungan Pasal 26

Ketentuan sebagaimana dimaksud dalam Pasal 23, Pasal 24, dan Pasal 25 tidak berlaku terhadap:

- a. penggunaan kutipan singkat Ciptaan dan/atau produk Hak Terkait untuk pelaporan peristiwa aktual yang ditujukan hanya untuk keperluan penyediaan informasi aktual;
- b. Penggandaan Ciptaan dan/atau produk Hak Terkait hanya untuk kepentingan penelitian ilmu pengetahuan;
- c. Penggandaan Ciptaan dan/atau produk Hak Terkait hanya untuk keperluan pengajaran, kecuali pertunjukan dan Fonogram yang telah dilakukan Pengumuman sebagai bahan ajar; dan
- d. penggunaan untuk kepentingan pendidikan dan pengembangan ilmu pengetahuan yang memungkinkan suatu Ciptaan dan/atau produk Hak Terkait dapat digunakan tanpa izin Pelaku Pertunjukan, Produser Fonogram, atau Lembaga Penyiaran.

### Sanksi Pelanggaran Pasal 113

1. Setiap Orang yang dengan tanpa hak dan/atau tanpa izin Pencipta atau pemegang Hak Cipta melakukan pelanggaran hak ekonomi Pencipta sebagaimana dimaksud dalam Pasal 9 ayat (1) huruf c, huruf d, huruf f, dan/atau huruf h untuk Penggunaan Secara Komersial dipidana dengan pidana penjara paling lama 3 (tiga) tahun dan/atau pidana denda paling banyak Rp500.000.000,00 (lima ratus juta rupiah).
2. Setiap Orang yang dengan tanpa hak dan/atau tanpa izin Pencipta atau pemegang Hak Cipta melakukan pelanggaran hak ekonomi Pencipta sebagaimana dimaksud dalam Pasal 9 ayat (1) huruf a, huruf b, huruf e, dan/atau huruf g untuk Penggunaan Secara Komersial dipidana dengan pidana penjara paling lama 4 (empat) tahun dan/atau pidana denda paling banyak Rp1.000.000.000,00 (satu miliar rupiah).

# **Data Mining dan Penerapan Algoritma**

Ni Luh Wiwik Sri Rahayu Ginantra, Fatimah Nur Arifah  
Anggi Hadi Wijaya, Ri Sabti Septarini, Nazaruddin Ahmad  
Dewa Putu Yudhi Ardiana, Faried Effendy, Akbar Iskandar  
Hazriani, Ika Yusnita Sari, Zelvi Gustiana  
Cahyo Prianto, Dudih Gustian, Edi Surya Negara



Penerbit Yayasan Kita Menulis

# Data Mining dan Penerapan Algoritma

Copyright © Yayasan Kita Menulis, 2021

Penulis:

Ni Luh Wiwik Sri Rahayu Ginantra, Fatimah Nur Arifah  
Anggi Hadi Wijaya, Ri Sabti Septarini, Nazaruddin Ahmad  
Dewa Putu Yudhi Ardiana, Faried Effendy, Akbar Iskandar  
Hazriani, Ika Yusnita Sari, Zelvi Gustiana  
Cahyo Prianto, Dudih Gustian, Edi Surya Negara

Editor: Ronal Watrianthos & Janner Simarmata

Desain Sampul: Devy Dian Pratama, S.Kom.

Penerbit

Yayasan Kita Menulis

Web: [kitamenulis.id](http://kitamenulis.id)

e-mail: [press@kitamenulis.id](mailto:press@kitamenulis.id)

WA: 0821-6453-7176

Anggota IKAPI: 044/SUT/2021

Ni Luh Wiwik Sri Rahayu Ginantra, dkk.

Data Mining dan Penerapan Algoritma

Yayasan Kita Menulis, 2021

xiv; 202 hlm; 16 x 23 cm

ISBN: 978-623-342-073-0

Cetakan 1, Mei 2021

- I. Data Mining dan Penerapan Algoritma
- II. Yayasan Kita Menulis

## Katalog Dalam Terbitan

Hak cipta dilindungi undang-undang

Dilarang memperbanyak maupun mengedarkan buku tanpa  
izin tertulis dari penerbit maupun penulis

# Kata Pengantar

Puji syukur kehadiran Tuhan Yang Maha Esa yang telah memberikan Rahmat dan HidayahNya, sehingga kami dapat menyelesaikan buku Data Mining dan Penerapan Algoritmanya

Perkembangan data yang pesat membutuhkan pengolahan data untuk tidak hanya sekedar menghasilkan informasi namun juga memberikan pengetahuan baru. Proses penggalian/ menambang data dengan teknik yang ada dalam data mining dapat menemukan pola tersembunyi pada data tersebut seperti penggunaan klasifikasi dan clustering.

Buku Data Mining dan Penerapan Algoritma ini terdiri dari 14 bab yaitu:

Bab 1 Pengenalan Data Mining

Bab 2 Proses Knowledge Discovery in Databases (KDD)

Bab 3 Arsitektur dan Model Data Mining

Bab 4 Jenis Model Data Mining dan Fungsinya

Bab 5 Fungsi-Fungsi Data Mining

Bab 6 Teknik-Teknik Data Mining

Bab 7 Teknik Clustering Dalam Data Mining

Bab 8 Eksplorasi Data

Bab 9 Analisis Asosiasi Dalam Data Mining

Bab 10 Algoritma Data Mining Klasifikasi Menggunakan Naïve Bayes Classifier

Bab 11 Algoritma Data Mining Klasifikasi Menggunakan Decision Tree C4.5

Bab 12 Prediksi Menggunakan Regresi Linear Sederhana dan Berganda

Bab 13 Association Dengan A Priori

Bab 14 Aplikasi dan Tren Data Mining

Akhir kata penulis menyampaikan terima kasih kepada teman-teman sejawat atas motivasi dan masukan yang bersifat membangun selama penulisan buku ini.

Denpasar April 2021

# Daftar Isi

Kata Pengantar .....	v
Daftar Isi .....	vii
Daftar Gambar .....	xi
Daftar Tabel .....	xiii

## **Bab 1 Pengenalan Data Mining**

1.1 Pendahuluan .....	1
1.2 Pengertian Data Mining .....	2
1.3 Kegunaan Data Mining .....	4
1.4 Arsitektur Data Mining .....	5
1.5 Siklus Hidup Data Mining .....	7
1.6 Penerapan Data Mining .....	8
1.6.1 Penerapan Data Mining dalam Bisnis .....	8
1.6.2 Penerapan Data Mining dalam Sains dan Teknik .....	9

## **Bab 2 Proses Knowledge Discovery in Databases (KDD)**

2.1 Pendahuluan .....	11
2.2 Knowledge Discovery in Databases .....	12
2.3 Cross – Industry Standard Process for Data Mining (CRISP-DM) .....	15

## **Bab 3 Arsitektur dan Model Data Mining**

3.1 Pendahuluan .....	19
3.2 Model Data Mining .....	21
3.2.1 Description Methods .....	28

## **Bab 4 Jenis Model Data Mining dan Fungsinya**

4.1 Decision Tree (Pohon Keputusan) .....	31
4.1.1 Algoritma Iterative Dichotomiser 3 (ID3) .....	34
4.1.2 Algoritma Classification 4.5 (C4.5) .....	37
4.2 Algoritma K-Means .....	42
4.2.1 Teknik Pengelompokan Clustering .....	43
4.3 Algoritma Naive Bayes .....	47
4.4 Algoritma Regresi Linear .....	52
4.4.1 Algoritma Regresi Linear Sederhana .....	53



---

4.5 Algoritma Apriori.....	58
<b>Bab 5 Fungsi-Fungsi Data Mining</b>	
5.1 Pestisida.....	63
5.2 Fungsi-Fungsi Data Mining.....	64
<b>Bab 6 Teknik-Teknik Data Mining</b>	
6.1 Pendahuluan.....	69
6.2 Proses Data Mining.....	73
6.3 Teknik Data Mining.....	75
6.4 Metode Deskriptif.....	75
6.4.1 Teknik Clustering.....	76
6.4.2 Teknik Association Rule.....	78
6.5 Metode Prediktif.....	79
6.5.1 Teknik Classification.....	80
6.5.2 Teknik Regression.....	81
<b>Bab 7 Teknik Clustering Dalam Data Mining</b>	
7.1 Pendahuluan.....	83
7.2 Clustering pada Data Mining.....	85
7.2.1 Clustering.....	85
7.2.2 Metode Clustering Data Mining.....	86
<b>Bab 8 Eksplorasi Data</b>	
8.1 Pendahuluan.....	95
8.2 Kegunaan Analisis Eksplorasi Data.....	96
8.3 Jenis Data dan Skala Pengukuran.....	98
8.4 Jenis Analisis Eksplorasi Data.....	99
8.5 Eksplorasi Data Dalam Machine Learning.....	100
8.6 Eksplorasi Data Interaktif.....	103
8.7 Eksplorasi Data di GIS.....	104
<b>Bab 9 Analisis Asosiasi Dalam Data Mining</b>	
9.1 Pendahuluan.....	105
9.2 Aturan Asosiasi (Association Rules).....	106
9.3 Algoritma Asosiasi (Association Algorithm).....	108
9.3.1 Algoritma Apriori.....	108
9.3.2 Algoritma Fp-Growth.....	110
9.4 Penerapan Association Rule.....	111

**Bab 10 Algoritma Data Mining Klasifikasi Menggunakan Naïve Bayes Classifier**

10.1 Pendahuluan.....	113
10.2 Naive Bayes Classifier (NBC) .....	115
10.2.1 Karakteristik Naïve Bayes Classifier.....	116
10.2.2 Formula Perhitungan Naïve Bayes Classifier.....	118
10.2.3 Kekuatan dan Kelemahan Naïve Bayes Classifier.....	119
10.2.4 Macam-Macam Naïve Bayes Classifier.....	119
10.3 Penerapan Metode Naïve Bayes Classifier.....	120

**Bab 11 Algoritma Data Mining Klasifikasi Menggunakan Decision Tree C4.5**

11.1 Pendahuluan.....	127
11.2 Algoritma C 4.5 .....	128
11.3 Entropi, Information Gain, dan Gain Ratio .....	131
11.4 Contoh Kasus.....	133

**Bab 12 Prediksi Menggunakan Regresi Linear Sederhana dan Berganda**

12.1 Pendahuluan.....	141
12.2 Regresi Linear Sederhana .....	143
12.3 Regresi Linear Berganda .....	147

**Bab 13 Association Dengan A Priori**

13.1 Pendahuluan.....	153
13.2 Algoritma A Priori.....	154
13.3 Contoh Kasus.....	158
13.4 Implementasi Dengan Rapid Miner .....	162

**Bab 14 Aplikasi dan Tren Data Mining**

14.1 Pendahuluan.....	171
14.2 Aplikasi Data Mining Pada Kehidupan Nyata.....	172
14.3 Aplikasi Data Mining Pada Bidang Keuangan .....	175
14.4 Aplikasi Data Mining Pada Bidang Pemasaran .....	177
14.5 Aplikasi Data Mining Pada Bidang Kesehatan.....	178
14.6 Tren Data Mining .....	179

Daftar Pustaka .....	181
Biodata Penulis .....	195



# Daftar Gambar

Gambar 1.1: Pertemuan disiplin ilmu dalam data mining .....	2
Gambar 1.2: Arsitektur sistem data mining .....	6
Gambar 1.3: Siklus Hidup Data Mining .....	8
Gambar 2.1: Proses Tahapan dalam KDD.....	13
Gambar 2.2: Proses Data Mining CRISP-DM .....	15
Gambar 3.1: Arsitektur Data Mining .....	20
Gambar 3.2: Model Data Mining .....	22
Gambar 3.3: Clasification Tree.....	23
Gambar 3.4: Regression Tree .....	23
Gambar 3.5: Jaringan Syaraf Tiruan .....	27
Gambar 3.6: Klastering .....	28
Gambar 3.7: Association Rule.....	29
Gambar 3.8: Rangkaian klik pada sebuah website berita .....	30
Gambar 4.1: Pohon Keputusan.....	33
Gambar 4.2: Metode Clustering K-Means .....	47
Gambar 4.3: Naïve Bayes .....	48
Gambar 5.1: Tanggungan.....	65
Gambar 6.1: Data menjadi informasi .....	70
Gambar 6.2: Piramida Data-Information-Knowledge-Wisdom (DIKW) ....	71
Gambar 6.3: Cross-industry standard process for data mining (CRISP-DM) ..	73
Gambar 6.4: Teknik data mining .....	75
Gambar 6.5: Penggambaran Teknik Clustering .....	76
Gambar 6.6: Perbedaan intra-cluster dengan inter-cluster.....	76
Gambar 6.7: Contoh rekomendasi produk pada situs amazon .....	79
Gambar 6.8: Ilustrasi Teknik Classification .....	80
Gambar 7.1: Ilustrasi metode partitioning clustering .....	86
Gambar 7.2: Ilustrasi metode hierarchical clustering.....	89
Gambar 7.3: Ilustrasi metode density-based clustering .....	90
Gambar 7.4: Ilustrasi metode distribution-based clustering .....	93
Gambar 7.5: Ilustrasi metode constraint-based clustering.....	94
Gambar 8.1: Scatter plot.....	101

---

Gambar 8.2: Categorical & Categorical.....	102
Gambar 8.3: Categorical and Continuous.....	102
Gambar 8.4: Eksplorasi data interaktif.....	104
Gambar 8.5: Eksplorasi data GIS .....	104
Gambar 9.1: Pseudocode Algoritma Apriori.....	109
Gambar 9.2: Pembentukan Fp-Tree .....	111
Gambar 10.1: Jurnal SAINTIKOM, Diagram Penyelesaian Masalah Naive Bayes .....	116
Gambar 10.2: Skema Klasifikasi Algoritma NBC.....	117
Gambar 11.1: Contoh Pohon Keputusan .....	128
Gambar 11.2: Contoh Atribut Kontinu .....	129
Gambar 11.3: Pohon Keputusan Node 1 (Root Node).....	136
Gambar 11.4: Pohon Keputusan Analisis Node 1.1.....	138
Gambar 11.5: Pohon Keputusan Akhir.....	139
Gambar 12.1: Hubungan dua variabel yang bersifat linear .....	142
Gambar 13.1: Grafik asosiasi penjualan .....	161
Gambar 13.2: Menu tambah data dalam reporsitory.....	162
Gambar 13.3: Pemilihan lokasi data yang akan diinput.....	163
Gambar 13.4: Proses pemilihan data.....	163
Gambar 13.5: Proses penyimpanan data.....	164
Gambar 13.6: Histori hasil .....	164
Gambar 13.7: Proses pemasukan data ke dalam proses.....	165
Gambar 13.8: Penyetingan parameters .....	165
Gambar 13.9: Proses pembentukan model .....	166
Gambar 13.10: Menghubungkan antar operator .....	166
Gambar 13.11: Menghubungkan operator ke output .....	167
Gambar 13.12: Penyetingan parameters assosiasi.....	167
Gambar 13.13: Hasil rule assosiasi.....	168
Gambar 13.14: Hasil rule dalam graph .....	168
Gambar 13.15: Rule akhir assosiasi .....	169
Gambar 14.1: Metode Continuously Improve Dengan Data Mining Pada Penyedia Layanan Seluler .....	173
Gambar 14.2: Proses Data Warehouse dan Data Mining Pada Data Bidang Keuangan.....	176

# Daftar Tabel

Tabel 5.1: Tanggungan.....	65
Tabel 9.1: Contoh Data Transaksi Pembelian Barang.....	107
Tabel 9.2: Data Tabular Transaksi pembelian obat.....	109
Tabel 9.3: Hasil Perhitungan Support dan Confidence.....	109
Tabel 9.4: Konstruksi Fp-tree dari Transaksi Database.....	110
Tabel 10.1: Tabel Keputusan.....	121
Tabel 11.1: Learning Dataset.....	134
Tabel 11.2: Hasil Perhitungan pada Dataset.....	134
Tabel 11.3: Analisis Kejadian Nilai, Entropi, dan Gain Node 1.....	135
Tabel 11.4: Data yang memiliki Kelembaban = Tinggi.....	136
Tabel 11.5: Hasil Analisis pada Node 1.1.....	137
Tabel 11.6: Data yang Memiliki Kelembaban = Tinggi dan Cuaca = Hujan.....	138
Tabel 11.7: Hasil Analisis pada Node 1.1.2.....	138
Tabel 12.1: Penamaan Variabel x dan y.....	142
Tabel 12.2: Data Biaya Promosi dan Laba Bersih (Data Dalam Jutaan).....	144
Tabel 12.3: Perhitungan Koefisien Regresi.....	145
Tabel 12.4: Data perbandingan biaya promosi, biaya perawatan, dan laba bersih (semua biaya dalam jutaan).....	149
Tabel 12.5: Tabel bantu perhitungan persamaan normal.....	150
Tabel 13.1: Contoh kasus pasangan makanan dan minuman.....	158
Tabel 13.2: Transaksi pembelian untuk masing-masing item kombinasi.....	158
Tabel 13.3: Jumlah frekuensi transaksi kombinasi item.....	159
Tabel 13.4: Kombinasi item.....	159
Tabel 13.5: Jumlah kombinasi item.....	160
Tabel 13.6: Nilai support dari kombinasi.....	160
Tabel 13.7: Nilai confident dan final asosiasi.....	161
Tabel 13.8: Konfigurasi kemungkinan asosiasi.....	161
Tabel 13.9: Contoh kasus dengan A Priori.....	162

# DATA MINING DAN PENERAPAN ALGORITMA

Perkembangan data yang pesat membutuhkan pengolahan data untuk tidak hanya sekedar menghasilkan informasi namun juga memberikan pengetahuan baru. Proses penggalian/ menambang data dengan teknik yang ada dalam data mining dapat menemukan pola tersembunyi pada data tersebut seperti penggunaan klasifikasi dan *clustering*.

Buku Data Mining dan Penerapan Algoritma ini terdiri dari 14 bab yaitu:

Bab 1 Pengenalan Data Mining

Bab 2 Proses Knowledge Discovery in Databases (KDD)

Bab 3 Arsitektur dan Model Data Mining

Bab 4 Jenis Model Data Mining dan Fungsinya

Bab 5 Fungsi-Fungsi Data Mining

Bab 6 Teknik-Teknik Data Mining

Bab 7 Teknik Clustering Dalam Data Mining

Bab 8 Eksplorasi Data

Bab 9 Analisis Asosiasi Dalam Data Mining

Bab 10 Algoritma Data Mining Klasifikasi Menggunakan Naïve Bayes Classifier

Bab 11 Algoritma Data Mining Klasifikasi Menggunakan Decision Tree C4.5

Bab 12 Prediksi Menggunakan Regresi Linear Sederhana dan Berganda

Bab 13 Association Dengan A Priori

Bab 14 Aplikasi dan Tren Data Mining



YAYASAN KITA MENULIS  
press@kitamenulis.id  
www.kitamenulis.id

