



# Pengantar DATA MINING & STATISTIK

Sari Ningsih, S.Si, MM., Miftahul Jannah, S.Kom., M.Kom.,  
Djarot Hindarto, S.Kom., M.Kom., Dr. Nur Putri Erdianti, SKM., MKM.,  
Erni Rihyanti, S.Si., M.M., Ratih Titi Komala Sari, S.Si, M.MSI.,  
Ira Diana Sholihati, S. Si. , MMSI, Ir. Endah Tri Esthi Handayani, M.MSI.,  
Dr. Fauziah, S.Kom., MMSI, Panca Dewi Pamungkasari, S.T., M.T., Ph.D.,  
Raden Muhamad Firzatullah, M.Kom., Loso Judijanto, S.Si, M.M., M.Stats.,  
Arie Gunawan, S.Kom. MMSI, Dr. Ir. Arief Wibowo, M.Kom.

# Pengantar DATA MINING & STATISTIK

Sari Ningsih, S.Si, MM., Miftahul Jannah, S.Kom., M.Kom.,  
Djarot Hindarto, S.Kom., M.Kom., Dr. Nur Putri Erdianti, SKM., MKM.,  
Erni Rihyanti, S.Si, M.M., Ratih Titi Komala Sari, S.Si, M.MSI,  
Ira Diana Sholihati, S. Si. , MMSI, Ir. Endah Tri Esthi Handayani, M.MSI,  
Dr. Fauziah, S.Kom., MMSI, Panca Dewi Pamungkasari, S.T., M.T., Ph.D.,  
Raden Muhamad Firzatullah, M.Kom., Loso Judijanto, S.Si, M.M., M.Stats.,  
Arie Gunawan, S.Kom. MMSI, Dr. Ir. Arief Wibowo, M.Kom.



## **Pengantar Data Mining & Statistik**

Copyright© PT Penerbit Penamuda Media, 2024

### **Penulis:**

Sari Ningsih., S.Si., MM., Miftahul Jannah, S.Kom., M.Kom.,  
Djarot Hindarto, S.Kom., M.Kom., Dr. Nur Putri Erdianti, SKM., MKM.,  
Erni Rihyanti, S.Si., M.M., Ratih Titi Komala Sari., S.Si, M.MSI.,  
Ira Diana Sholihati, S. Si. , MMSI., Ir. Endah Tri Esthi Handayani., M.MSI.,  
Dr. Fauziah, S.Kom., MMSI., Panca Dewi Pamungkasari, S.T., M.T., Ph.D., Raden  
Muhamad Firzatullah, M.Kom., Loso Judijanto, S.Si., M.M., M.Stats.,  
Arie Gunawan, S.Kom.. MMSI., Dr. Ir. Arief Wibowo, M.Kom.

### **ISBN:**

978-623-8586-66-0

### **Desain Sampul:**

Tim PT Penerbit Penamuda Media

### **Tata Letak:**

Enbookdesign

### **Diterbitkan Oleh**

#### **PT Penerbit Penamuda Media**

Casa Sidoarium RT 03 Ngentak, Sidoarium Dodeam Sleman Yogyakarta

HP/Whatsapp : +6285700592256

Email : penamudamedia@gmail.com

Web : www.penamuda.com

Instagram : @penamudamedia

Cetakan Pertama, Juni 2024

x + 267, 15x23 cm

*Hak cipta dilindungi oleh undang-undang*

*Dilarang memperbanyak sebagian atau seluruh isi buku  
tanpa izin Penerbit*



# Kata Pengantar

**S**eiring dengan pesatnya perkembangan teknologi informasi, penting bagi para pembaca untuk memahami konsep dari data mining dan statistik. Buku ini hadir sebagai panduan yang komprehensif untuk memahami kedua bidang ini, serta bagaimana penggabungannya dapat memberikan wawasan bernilai dalam pengambilan keputusan.

# Daftar Isi

Kata Pengantar.....	v
Daftar Isi.....	vi
<b>Bab 1 - Konsep Dasar Data Mining.....</b>	<b>1</b>
A. Data Mining dan KDD ( Knowledge Discovery in Database ).....	3
B. Asal Usul Data Mining.....	3
C. Tugas Data Mining ( <i>Data Mining Task</i> ).....	5
D. Data.....	10
<b>Bab 2 - Pendekatan Statistik dalam Data Mining.....</b>	<b>14</b>
A. Statistik dan Data Mining .....	14
B. Algoritma pada Data Mining.....	20
<b>Bab 3 - Preprocessing Data untuk Data Mining .....</b>	<b>23</b>
A. Pengenalan preprocessing data.....	23
B. Pentingnya preprocessing data dan tujuan preprocessing data.....	25

C. Pemahaman Data.....	26
D. Deskripsi data.....	31
E. Eksplorasi data .....	32
F. Evaluasi kualitas data .....	36
G. Pembersihan Data (Data Cleaning).....	37
<b>Bab 4 - Analisis Deskriptif dan Inferensial .....</b>	<b>72</b>
A. Analisis Deskriptif .....	72
B. Analisis Inferensial.....	78
<b>Bab 5 - Regresi dan Korelasi dalam Data Mining .....</b>	<b>85</b>
A. Regresi .....	85
B. Korelasi.....	93
C. Konsep Dasar Tau Kendall.....	98
<b>Bab 6 - Analisis Kluster dalam Data Mining .....</b>	<b>101</b>
A. Konsep Dasar Analisis Kluster.....	102
B. Metodologi Analisis Kluster .....	108
C. Aplikasi dan Evaluasi Analisis Kluster .....	117
<b>Bab 7 - Analisis Asosiasi dan Frekuensi .....</b>	<b>125</b>
A. Analisis Asosiasi.....	126
B. Analisis Frekuensi .....	126

C. Aturan Asosiasi .....	127
D. Algoritma Apriori.....	130

**Bab 8 - Penggunaan Statistik dalam Pengelompokan Data ..... 139**

A. Pengenalan terhadap pengelompokan data dan signifikansinya dalam statistik dan analisis data .....	140
B. Definisi pengelompokan data dan perbedaannya dengan klasifikasi.....	142
C. Konsep dasar statistik variabel, distribusi, ukuran tendensi sentral, dan dispersi .....	144
D. Metodologi Pengelompokan Data .....	151

**Bab 9 - Penggunaan Statistik dalam Klasifikasi Data..... 160**

A. Data dalam Statistik.....	160
B. Dasar Klasifikasi Data Statistik .....	164
C. Metode Klasifikasi .....	167

**Bab 10 - Evaluasi Model dalam Data Mining ..... 173**

A. Kualitas Data .....	174
B. Model Evaluasi pada Klasifikas .....	175
C. Model Evaluasi Cluster.....	185

**Bab 11 - Teknik Sampling dalam Data Mining..... 187**

**Bab 12 - Interpretasi Hasil *Data Mining* dengan Pendekatan Statistika..... 199**

A. Tahap Interpretasi Hasil *Data Mining*..... 199

B. Teknik Statistika yang Lazim Digunakan..... 202

C. Interpretasi dengan Teknik Statistika..... 204

**Bab 13 - Visualisasi Data untuk Data Mining ..... 218**

A. Pengenalan Visualisasi Data..... 218

B. Peran Visualisasi dalam Proses Data Mining ..... 220

C. Jenis-Jenis Visualisasi Data..... 221

D. Metode Visualisasi Data Univariate ..... 223

E. Metode Visualisasi Data Multivariate ..... 223

F. Penggunaan Visualisasi untuk Eksplorasi Data ..... 224

G. Teknik Visualisasi untuk Pemodelan Data..... 225

H. Pemahaman Pola dengan Visualisasi..... 225

I. Interaktif dan Dinamika dalam Visualisasi Data..... 226

J. Penerapan Visualisasi pada Algoritma Data Mining.. 227

**Bab 14 - Tantangan dan Peluang Integrasi Data Statistik dalam Data Mining..... 229**

A. Tantangan Integrasi Data Statistik dan Data Mining 230

B. Peluang Integrasi Data Statistik dan Data Mining ..... 236

C. Kesimpulan dan Rekomendasi .....	241
<b>Daftar Pustaka .....</b>	<b>243</b>
<b>Tentang Penulis .....</b>	<b>258</b>



# Konsep Dasar Data Mining

Sari Ningsih.,S.Si., MM.

**K**emajuan pesat dalam pengumpulan dan teknologi penyimpanan data telah memungkinkan organisasi mengumpulkan data dalam jumlah besar. Namun, mengekstraksi informasi yang berguna terbukti sangat menantang. Seringkali, alat dan teknik analisis data tradisional tidak dapat digunakan karena ukuran kumpulan data yang sangat besar. Terkadang, sifat data yang non-tradisional menyebabkan pendekatan tradisional tidak dapat diterapkan meskipun kumpulan datanya relatif kecil. Dalam situasi lain, pertanyaan yang perlu dijawab tidak dapat dijawab dengan menggunakan teknik analisis data yang ada, sehingga perlu dikembangkan metode baru.

Data mining adalah penggunaan teknologi yang menggabungkan metode analisis data konvensional dengan algoritma yang canggih untuk mengolah volume data yang besar. Teknologi ini juga membuka kesempatan untuk menggali dan menganalisis jenis data baru serta menerapkan pendekatan baru dalam menganalisis data yang sudah ada.

Dalam bagian pendahuluan ini, kami menyajikan gambaran tentang data mining dan menjelaskan topik-topik utama yang akan dibahas dalam buku ini. Kami memulai dengan menggambarkan beberapa aplikasi terkemuka yang memerlukan teknik baru dalam analisis data.

### **Apa itu Data Mining?**

Data Mining merupakan langkah menemukan informasi dari data yang bermanfaat secara otomatis dalam kumpulan /repositori data besar. Teknik data mining dimanfaatkan untuk mengeksplorasi basis data yang luas guna menemukan pola baru dan berarti yang mungkin belum teridentifikasi sebelumnya. Selain itu, teknik ini juga memberikan ke-mampuan untuk meramalkan hasil pengamatan di masa yang akan datang, seperti meramalkan apakah pelanggan baru akan melakukan pembelian lebih dari \$100 di pusat perbelanjaan.

Tidak semua tugas penemuan informasi dianggap sebagai data mining. Misalnya, mencari catatan individual menggunakan sistem manajemen basis data atau menemukan halaman web tertentu melalui query ke mesin pencari internet adalah tugas yang berkaitan dengan bidang pengambilan informasi. Walaupun tugas-tugas tersebut memiliki signifikansi yang besar dan mungkin memerlukan penerapan algoritma dan struktur data yang maju, mereka masih mengandalkan pada prinsip-prinsip tradisional ilmu komputer serta karakteristik data yang jelas untuk membentuk struktur indeks guna mengatur dan mengambil informasi dengan efisien. Meskipun begitu, teknik data mining telah diterapkan untuk meningkatkan sistem pencarian informasi.

# Pengantar DATA MINING & STATISTIK

Buku ini menjelaskan secara mendalam konsep-konsep kunci dalam data mining dan statistik, mulai dari dasar-dasar hingga aplikasi canggihnya dalam berbagai bidang. Dengan penekanan pada analisis data yang relevan dan teknik modeling yang efektif, pembaca akan dibimbing melalui studi kasus praktis yang mengilustrasikan kekuatan gabungan antara data mining dan statistik dalam mengungkap pola penting dan membuat prediksi yang akurat. Dengan membaca buku ini, diharapkan para pembaca dapat meningkatkan pemahaman mereka tentang cara menggali informasi berharga dari data dan mengoptimalkan penggunaannya dalam berbagai konteks.

ISBN 978-623-8586-66-0



9 786238 586660



PT Penerbit Penamuda Media  
Godean, Yogyakarta  
085700592256  
@penamuda\_media  
penamuda.com