



# DATA MINING

## PRAKTIS UNTUK PEMULA



Desiyanna Lasut

# Data Mining Praktis untuk Pemula

Desiyanna Lasut, M. Kom.



Universitas Buddhi Dharma

# DATA MINING PRAKTIS UNTUK PEMULA

ISBN:

Hak Cipta 2025 pada Penulis

Hak penerbitan pada UNIVERSITAS BUDDHI DHARMA. Bagi mereka yang ingin memperbanyak sebagian isi buku ini dalam bentuk atau cara apapun harus mendapatkan izin tertulis dari penulis dan penerbit UNIVERSITAS BUDDHI DHARMA.

**Penulis:**

Desiyanna Lasut, M.Kom.

**Editor:**

Yusuf Kurnia, M.Kom., M.M.

**Layout**

Lidya Lunardi, S.Kom.

**Desain sampul:**

Aditiya Hermawan, M.Kom., M.M



**Penerbit:**

UNIVERSITAS BUDDHI DHARMA

Gedung Vipassi Lt. 1 Universitas Buddhi Dharma

Jl. Imam Bonjol No 41 Karawaci Ilir, Tangerang 15115

Telp. (021) 5517853

E-Mail: [lp3m@buddhidharma.ac.id](mailto:lp3m@buddhidharma.ac.id)

Hak Cipta dilindungi Undang-undang

All Right Reserved

Cetakan I, \_\_\_\_\_ 2025

## KATA PENGANTAR

Buku ini lahir dari kebutuhan untuk memfasilitasi orang yang ingin belajar data mining secara konsep dan praktis, dapat menyentuh data, mencoba model dan tahapan data mining tanpa perlu repot instalasi dan dapat dilakukan cukup dari browser. Materi disusun agar dapat dijalankan dari Google colab atau bahkan online notebook provider apapun.

Isi dari buku ini membahas konsep, algoritma dan tahapan yang dilakukan untuk dapat melakukan data mining secara konsep dan praktek. Mulai dari membangun model klasifikasi dan regresi, memahami clustering, association rule mining, hingga dimensionality reduction untuk visualisasi data berdimensi tinggi. Pada tahap terakhir, disertakan carap raktis menyimpan model dan membangun API untuk dapat digunakan atau di-integrasikan dengan aplikasi lain.

Buku ini ditujukan untuk mahasiswa, pengajar, peneliti, dan praktisi yang ingin meningkatkan keterampilan analitik. Dengan syarat yang disarankan sebelum memulai, berupa dasar Python (list, dict, fungsi), statistik tingkat dasar (mean, varians, korelasi), dan rasa ingin eksperimen. Namun dengan teknologi saat ini, buku ini dapat langsung digunakan bahkan tanpa pengetahuan dasar yang disebutkan

Akhir kata, buku ini disusun sebagai tahapan praktikal, yang bisa menjadi dasar pengetahuan untuk melakukan datamining, sehingga bisa menjembatani pengetahuan teori dan cara mempraktekkan teknik data minig yang relevan dengan data yang dimiliki

# DAFTAR ISI

KATA PENGANTAR.....	iii
DAFTAR ISI.....	iv
BAB 1 PENDAHULUAN.....	1
A. Apa itu Data Mining.....	1
B. Data Mining, Statistik, dan Machine Learning.....	2
C. Mengapa Belajar Data Mining ?.....	4
D. Lingkup Buku ini.....	7
E. Evaluasi.....	7
BAB 2 DATA MINING DENGAN GOOGLE COLAB.....	8
A. Apa itu Google Colab.....	8
B. Mengapa menggunakan Google Colab.....	8
C. Menulis code di Google Colab.....	9
D. Mengelola Notebook.....	10
E. Contoh Singkat dengan Colab.....	11
F. Evaluasi.....	13
BAB 3 MATEMATIKA DASAR DAN STATISTIK.....	15
A. Probabilitas.....	15
B. Distribusi.....	17
C. Statistik Deskriptif dan inferensial.....	17
D. Hands-on : Probability Distribution (scipy.stats).....	19
E. Evaluasi.....	22
BAB 4 DATA WRANGLING & PREPROCESSING.....	23
A. Apa itu Data Wrangling ?.....	23
B. Data Wrangling dengan Python.....	23
C. Contoh Kasus : Pengolahan Data.....	26
BAB 5 EXPLORATORY DATA ANALYSIS.....	34
A. Apa itu EDA ?.....	34
B. Contoh Kasus : Kelulusan Tepat Waktu.....	35
C. Contoh Kasus: Melakukan EDA.....	37

D. Evaluasi.....	46
BAB 6 KLASIFIKASI.....	47
A. Klasifikasi.....	47
B. Logistic Regression .....	48
C. Decision Tree.....	52
D. Metrik Evaluasi .....	55
E. Dasar perhitungan: Confusion Matrix.....	55
F. ROC (Receiver Operating Characteristic).....	56
G. Ilustrasi Perhitungan Akurasi.....	57
BAB 7 Regresi .....	58
A. Regresi.....	58
B. Metrik Evaluasi (Regresi).....	59
C. Studi kasus : Prediksi harga rumah.....	60
D. Evaluasi.....	63
BAB 8 Clustering.....	64
A. Clustering.....	64
B. Algoritma k-means.....	65
C. Contoh kode .....	66
D. Evaluasi.....	70
BAB 9 Assosiasi.....	71
A. Association Rule Mining.....	71
B. Algoritma Assosiasi.....	71
C. Algoritma FP-Growth.....	72
D. Contoh kode .....	74
E. Evaluasi.....	75
BAB 10 Reduksi Dimensi.....	77
A. Reduksi Dimensi.....	77
B. PCA, t-SNE, UMAP .....	78
C. Contoh Kode.....	79
BAB 11 FEATURE ENGINEERING.....	84
A. Feature Engineering.....	84

B. Pipeline & Column Transformer .....	85
C. Pipeline untuk Klasifikasi .....	86
D. Evaluasi .....	97
BAB 12 DEPLOYMENT .....	99
A. Menyimpan Model .....	99
B. Menyajikan Model dengan API .....	102
C. Evaluasi .....	106
DAFTAR PUSTAKA .....	108
GLOSARIUM .....	110
INDEKS .....	114
HASIL SCANNING SIMILARITY .....	116
BIOGRAFI PENULIS .....	117

# DATA MINING

## PRAKTIS UNTUK PEMULA



Buku ajar ini membawa pembaca menelusuri perjalanan lengkap dunia data mining, mulai dari konsep dasar hingga penerapan nyata menggunakan Python dan Google Colab. Setiap bab dirancang berurutan agar pembaca dapat memahami proses pengelolaan data, eksplorasi, hingga pembangunan model seperti klasifikasi, regresi, clustering, dan asosiasi. Fondasi matematika dan statistik juga disertakan untuk memastikan setiap teknik dapat dipahami secara logis, sementara berbagai studi kasus membuat pembelajaran terasa relevan dan aplikatif.

Melangkah lebih jauh, buku ini tidak hanya mengajarkan cara membangun model, tetapi juga bagaimana menyiapkannya untuk digunakan di dunia nyata. Pembaca diajak berlatih feature engineering, membuat pipeline yang rapi, mereduksi dimensi data, hingga menyimpan dan menyajikan model melalui API. Dengan kombinasi teori, praktik langsung, dan contoh kode yang dapat diulang, buku ini menjadi pendamping ideal bagi siapa saja yang ingin menguasai data mining secara terstruktur dan berorientasi proyek.



UNIVERSITAS BUDDHI DHARMA  
Gedung Vipassi Lt. 1  
Jl. Imam Bonjol No. 41 Karawaci Ilir  
Tangerang 15115  
Telp. (021) 5517853  
E-mail: lp3m@buddhidharma.ac.id





**DATA MINING** PRAKTIS UNTUK PEMULA

Desiyanna Lasut

