

# Klasifikasi Data Mining: Konsep, Metode, dan Implementasi

Panduan Komprehensif Teori dan Praktik Klasifikasi dalam Data Mining dengan Pendekatan Konvensional hingga Deep Learning Modern



- **Decision Tree (CART, C4.5, ID3)**
- **Random Forest Naïve Bayes**
- **K-Nearest Neighbors (KNN)**
- **Support Vector Machine (SVM)**

- **Logistic Regression**
- **Artificial Neural Networks (ANN)**
- **Gradient Boosting**
- **Deep Learning**

# **Klasifikasi Data Mining: Konsep, Metode, dan Implementasi**

Penulis

BAYU ADITYA PRATAMA

SONDANG NURLINA NATALIA BR MANIK

SULTAN SIDQI

DEWI MARTINI

ROBBY SANDI

SAYUTI RAHMAN

RAHMAD SYAH

DITERBITKAN OLEH:

**UNIVERSITAS MEDAN AREA PRESS**

# **Klasifikasi Data Mining: Konsep, Metode, dan Implementasi**

Penulis : Bayu Aditya Pratama  
Sondang Nurlina  
Natalia Br Manik  
Sultan Sidqi  
Dewi Martini  
Robby Sandi  
Sayuti Rahman  
Rahmad Syah

Desain Cover : Ramadhan Ady Pratama  
Edit Layout : Sayuti Rahman  
Editor : Yuan Anisa

ISBN :  
Uk: 21 x 29,7 cm hlm.121  
Copyright 2026

**Universitas Medan Area Press**

**Address: Jalan Kolam Nomor 1, Kenangan Baru, Kec.  
Percut Sei Tuan, Deliserdang, Sumatera Utara  
Telephone:061-7366878,  
e-mail: [pghc@uma.ac.id](mailto:pghc@uma.ac.id)**

## KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas segala rahmat dan karunia-Nya sehingga buku *Klasifikasi Data Mining: Konsep, Metode, dan Implementasi* ini dapat diselesaikan dengan baik. Buku ini disusun sebagai upaya untuk menghadirkan referensi yang komprehensif, sistematis, dan mudah dipahami mengenai teknik klasifikasi dalam data mining, yang saat ini menjadi salah satu kompetensi penting di era transformasi digital.

Perkembangan teknologi informasi dan ketersediaan data dalam jumlah besar menuntut adanya metode analisis yang efektif dan akurat. Melalui buku ini, penulis berupaya menjembatani kebutuhan pembaca dengan menyajikan pembahasan mulai dari konsep dasar, algoritma klasifikasi konvensional seperti Decision Tree, Naïve Bayes, K-Nearest Neighbors, dan Support Vector Machine, hingga pendekatan modern berbasis Artificial Neural Networks dan Deep Learning. Materi disusun secara bertahap disertai contoh implementasi dan studi kasus agar pembaca tidak hanya memahami teori, tetapi juga mampu menerapkannya secara praktis.

Buku ini ditujukan bagi mahasiswa, dosen, peneliti, serta praktisi yang ingin memperdalam pemahaman di bidang data

mining dan machine learning. Penulis menyadari bahwa buku ini masih memiliki keterbatasan. Oleh karena itu, kritik dan saran yang membangun sangat diharapkan demi penyempurnaan pada edisi berikutnya.

Akhir kata, penulis berharap buku ini dapat memberikan manfaat dan kontribusi nyata dalam pengembangan ilmu pengetahuan, khususnya di bidang data mining dan kecerdasan buatan.

Medan, Maret 2026

Penulis

# DAFTAR ISI

KATA PENGANTAR | III

DAFTAR ISI | V

DAFTAR GAMBAR | XIV

DAFTAR TABEL | XV

BAB 1 | 1

DECISION TREE | 1

- 1.1. Latar Belakang | 1
- 1.2. Konsep Dasar Klasifikasi | 5
- 1.3. Pengertian Decision Tree | 8
- 1.4. Komponen Utama Decision Tree | 12
  - 1.4.1. Root Node | 12
  - 1.4.2. Internal Node | 13
  - 1.4.3. Leaf Node | 13
  - 1.4.4. Cabang (Branch) | 13
  - 1.4.5. Ilustrasi Struktur Komponen Decision Tree | 14
  - 1.4.6. Kedalaman Pohon dan Kompleksitas Model | 15
  - 1.4.7. Pruning | 16
- 1.5. Algoritma ID3 (Iterative Dichotomiser 3) | 16
  - 1.5.1. Konsep Dasar Algoritma ID3 | 17
  - 1.5.2. Entropy sebagai Ukuran Ketidakpastian | 17
  - 1.5.3. Information Gain | 18

- 1.5.4. Contoh Perhitungan Manual ID3 |19
- 1.5.5. Proses Pembentukan Pohon ID3 |20
- 1.5.6. Kelebihan Algoritma ID3 |20
- 1.5.7. Kelemahan Algoritma ID3 |21
- 1.5.8. Ringkasan Subbab |21
- 1.6. Algoritma C4.5 |22
  - 1.6.1. Pengembangan C4.5 dari ID3 |22
  - 1.6.2. Gain Ratio sebagai Kriteria Pemilihan Atribut |23
  - 1.6.3. Penanganan Atribut Kontinu |24
  - 1.6.4. Penanganan Missing Value |24
  - 1.6.5. Pruning pada Algoritma C4.5 |24
  - 1.6.6. Proses Pembentukan Pohon C4.5 |25
  - 1.6.7. Kelebihan Algoritma C4.5 |26
  - 1.6.8. Kelemahan Algoritma C4.5 |26
- 1.7. Algoritma CART (Classification and Regression Tree) |27
  - 1.7.1. Konsep Dasar Algoritma CART |27
  - 1.7.2. Binary Split |28
  - 1.7.3. Gini Index sebagai Ukuran Impurity |28
  - 1.7.4. Contoh Perhitungan Gini Index |29
  - 1.7.5. Pruning pada Algoritma CART |30
  - 1.7.6. Proses Pembentukan Pohon CART |30
  - 1.7.7. Kelebihan Algoritma CART |31
  - 1.7.8. Kelemahan Algoritma CART |31

- 1.8. Perbandingan Algoritma ID3, C4.5, dan CART |32
- 1.9. Aplikasi Decision Tree dalam Penelitian |33
- 1.10. Ringkasan Bab |36

## BAB 2 |39

### RANDOM FOREST DAN NAIVE BAYES |39

- 2.1. Pendahuluan Bab |39
- 2.2. Konsep Dasar Klasifikasi Data |41
- 2.3. Algoritma Naive Bayes |44
  - 2.3.1. Definisi dan Prinsip Dasar Naive Bayes |44
  - 2.3.2. Teorema Bayes |45
  - 2.3.3. Rumus Matematis Naive Bayes |47
  - 2.3.4. Contoh Perhitungan Manual Naive Bayes |51
  - 2.3.5. Varian Naive Bayes |54
- 2.4. Algoritma Random Forest |55
  - 2.4.1. Definisi dan Konsep Dasar Random Forest |56
  - 2.4.2. Struktur dan Cara Kerja Random Forest |57
  - 2.4.3. Dasar Matematis Random Forest |58
  - 2.4.4. Contoh Sederhana Cara Kerja Random Forest |59
  - 2.4.5. Parameter Penting dalam Random Forest |60
  - 2.4.6. Penerapan Random Forest dalam Penelitian |61
- 2.5. Perbandingan Random Forest dan Naive Bayes |62
  - 2.5.1. Perbandingan Konseptual |62
  - 2.5.2. Perbandingan Kinerja |63

2.5.3.	Kelebihan dan Kekurangan	64
2.5.4.	Ringkasan Perbandingan	65
2.6.	Ringkasan Bab	66
BAB 3  69		
K-NEAREST NEIGHBOR (KNN)  69		
3.1.	Pendahuluan	69
3.2.	Konsep Dasar K-Nearest Neighbor	71
3.2.1.	Definisi K-Nearest Neighbor	71
3.2.2.	Karakteristik Utama KNN	72
3.2.3.	Jenis Permasalahan yang Dapat Diselesaikan	73
3.3.	Tahapan Kerja Algoritma KNN	74
3.4.	Metode Perhitungan Jarak pada K-Nearest Neighbor (KNN)	78
3.4.1.	Euclidean Distance	79
3.4.2.	Manhattan Distance	80
3.4.3.	Minkowski Distance	81
3.4.4.	Cosine Similarity	83
3.5.	Contoh Perhitungan Manual K-Nearest Neighbor (KNN)	84
3.5.1.	Data Latih dan Data Uji	84
3.5.2.	Menentukan Nilai K	85
3.5.3.	Tabel Hasil Perhitungan Jarak	86
3.5.4.	Mengurutkan Jarak dari yang Terkecil	87
3.5.5.	Memilih K Tetangga Terdekat ( $K = 3$ )	87
3.5.6.	Menentukan Kelas Mayoritas	88

3.5.7.	Hasil Klasifikasi	88
3.6.	Penentuan Nilai K pada Algoritma K-Nearest Neighbor	88
3.6.1.	Pengaruh NilaiK Kecil dan Nilai K Besar	88
3.6.2.	Risiko Overfitting dan Underfitting	89
3.6.3.	Metode Penentuan Nilai K	90
3.7.	Kelebihan dan Kekurangan Algoritma K-Nearest Neighbor	91
3.7.1.	Kelebihan Algoritma KNN	92
3.7.2.	Kekurangan Algoritma KNN	93
3.8.	Ringkasan Bab	95
BAB 4  98		
Support Vector Machine  98		
4.1.	Pendahuluan	98
4.1.1.	Sejarah Support Vector Machine	99
4.1.2.	Perkembangan Support Vector Machine	100
4.2.	Algoritma SVM	101
4.3.	Kelebihan & Kekurangan Algoritma SVM	104
4.3.1.	Kelebihan Algoritma SVM	105
4.3.2.	Kekurangan Algoritma SVM	106
4.4.	Implementasi Algoritma SVM	108
4.4.1.	Implementasi Algoritma SVM Untuk Kesehatan	108
4.4.2.	Implementasi Algoritma SVM Untuk Sosial & Politik	108

- 4.4.3. Implementasi Algoritma SVM Untuk ioinformatika & Kimia |109
- 4.4.4. Implementasi Algoritma SVM Untuk Teknologi |109
- 4.5. Kesimpulan |109

## BAB 05 |113

### LOGISTIC REGRESSION |113

- 5.1. Konsep Dasar Dan Formulasi Matematis |114
- 5.2. Logistic Regression Dalam Data Mining |122
- 5.3. Asumsi Dalam Logistic Regression |123
- 5.4. Jenis-Jenis Logistic Regression |123
- 5.5. Implementasi Dan Studi Kasus |128
- 5.6. Implementasi Dan Studi Kasus |131
- 5.7. Strategi Peningkatan Model Dan Regulasi |134
- 5.8. Regularisasi L1 (LOSS) |136
- 5.9. Regularisasi L2 (RIDGE) |140

## BAB 6 |143

### ARTIFICIAL NEURAL NETWORKS |143

- 6.1. Latar Belakang Artificial Neural Networks |143
- 6.2. Konsep Dasar Artificial Neural Networks |145
- 6.3. Perkembangan Artificial Neural Networks |146
- 6.4. Pengertian Dan Komponen Artificial Neural Networks |147
- 6.5. Arsitektur Artificial Neural Networks |150
- 6.6. Cara Kerja Artificial Neural Networks |152
- 6.7. Algoritma Perceptron |154
  - 6.7.1. Struktur Dasar Algoritma Perceptron |156

- 6.7.2. Penggunaan Perceptron Dalam Kehidupan |159
- 6.7.3. Contoh Perhitungan Algoritma Perceptron |161

## BAB 7 |167

### GRADIENT BOOSTING |167

- 7.1. Pendahuluan |167
  - 7.1.1. Sejarah & Perkembangan Gradient Boosting |168
  - 7.1.2. Pengertian & Prinsip Kerja Gradient Boosting |170
- 7.2. Algoritma *XGBoost* |172
- 7.3. Algoritma LightGBM |174
- 7.4. Algoritma CatBoost |176
- 7.5. Kelebihan & Kekurangan Gradient Boosting |178
  - 7.5.1. Kelebihan Gradient Boosting |179
  - 7.5.2. Kekurangan Gradient Boosting |180
- 7.6. Implementasi Gradient Boosting |183
  - 7.6.1. Implementasi Gradient Boosting Untuk Keuangan |184
  - 7.6.2. Implementasi Gradient Boosting Untuk Keamanan |184
  - 7.6.3. Implementasi Gradient Boosting Untuk Pemasaran |184
  - 7.6.4. Implementasi Gradient Boosting Untuk Industri |185
- 7.7. Kesimpulan |185

## BAB 8 |187

### Deep Learning |187

- 8.1. Pengenalan Deep Learning |187
- 8.2. Definisi dan Konsep Dasar Deep Learning |188
  - 8.2.1. Perbedaan Machine Learning dan Deep Learning |190
- 8.3. Multilayer Perceptron (MLP) |192
  - 8.3.1. Definisi dan Arsitektur Multilayer Perceptron |192
  - 8.3.2. Rumus dan Perhitungan Manual MLP (Sederhana) |193
  - 8.3.3. Penelitian Terkait Multilayer Perceptron dalam Klasifikasi |195
- 8.4. Convolutional Neural Network (CNN) |197
  - 8.4.1. Konsep dan Karakteristik Convolutional Neural Network |197
  - 8.4.2. Komponen Utama Convolutional Neural Network |198
  - 8.4.3. Contoh Perhitungan Manual Convolution |200
  - 8.4.4. Penelitian CNN dalam Klasifikasi |202
- 8.5. Recurrent Neural Network (RNN) |203
  - 8.5.1. Definisi dan Mekanisme Recurrent Neural Network |203
  - 8.5.2. Struktur dan Alur Kerja RNN |204
  - 8.5.3. Contoh Perhitungan Manual RNN Sederhana |205
  - 8.5.4. Permasalahan Vanishing Gradient pada RNN |206

8.5.5.	Penelitian RNN dalam Klasifikasi	206
8.6.	Long Short-Term Memory (LSTM)	207
8.6.1.	Konsep Dasar Long Short-Term Memory	207
8.6.2.	Struktur Gate pada LSTM	208
8.6.3.	Contoh Alur Kerja LSTM (Konseptual)	210
8.6.4.	Ilustrasi Arsitektur Long Short-Term Memory (LSTM)	211
8.6.5.	Penelitian LSTM dalam Klasifikasi	212
8.6.6.	Posisi LSTM dalam Deep Learning Modern	213
8.7.	Transformer	213
8.7.1.	Latar Belakang Transformer	213
8.7.2.	Self-Attention Mechanism	214
8.7.3.	Peran Transformer dalam Klasifikasi Modern	215
8.7.4.	Penelitian Transformer	216
8.8.	Perbandingan Metode Deep Learning untuk Klasifikasi	217
8.9.	Tantangan dan Arah Penelitian Deep Learning	220
8.10.	Ringkasan Bab	222
DAFTAR PUSTAKA		223

## DAFTAR GAMBAR

- Gambar 1.1.** *Struktur dan Komponen Utama Decision Tree*  
| **7***Error! Bookmark not defined.*
- Gambar 3.1.** Diagram alur KNN |42
- Gambar 3.2.** Ilustrasi titik data dua dimensi (2D) |43
- Gambar 4.1.** Diagram Model Support Vector Machine |56
- Gambar 4.2.** Diagram Alur Algoritma SVM |58
- Gambar 6.1.** Arsitektur Artificial Neural Network (ANN)  
|85
- Gambar 6.2.** Alur Kerja Feedforward pada ANN |86
- Gambar 6.3.** Skema Proses Pelatihan ANN |87
- Gambar 6.4.** Struktur Dasar Perceptron |88
- Gambar 7.1.** Diagram Model Gradient Boosting |97
- Gambar 7.2.** Diagram Alur Algoritma XGBoost |98
- Gambar 7.3.** Diagram Alur Algoritma LightGBM |99
- Gambar 7.4.** Diagram Alur Algoritma CatBoost |81
- Gambar 8.1.** *Deep learning* |105
- Gambar 8.2.** *Convolutional Neural Network* |111
- Gambar 8.3.** *Arsitektur Long Short-Term Memory (LSTM)*  
|118

## DAFTAR TABEL

**Tabel 1.1** Perbandingan ID3, C4.5, dan CART |17

**Tabel 2.1.** Data Latih |28

**Tabel 2.2.** Tabel Perbandingan Varian Navie Bayes |30

**Tabel 2.3.** Tabel Perbandingan Navie Bayes Dan Random  
Forest |36

**Tabel 3.1** Jarak Data Uji terhadap Data Latih |48

**Tabel 3.2** Urutan Jarak Terdekat |48

**Tabel 8.1** Perbandingan Metode Deep Learning |121