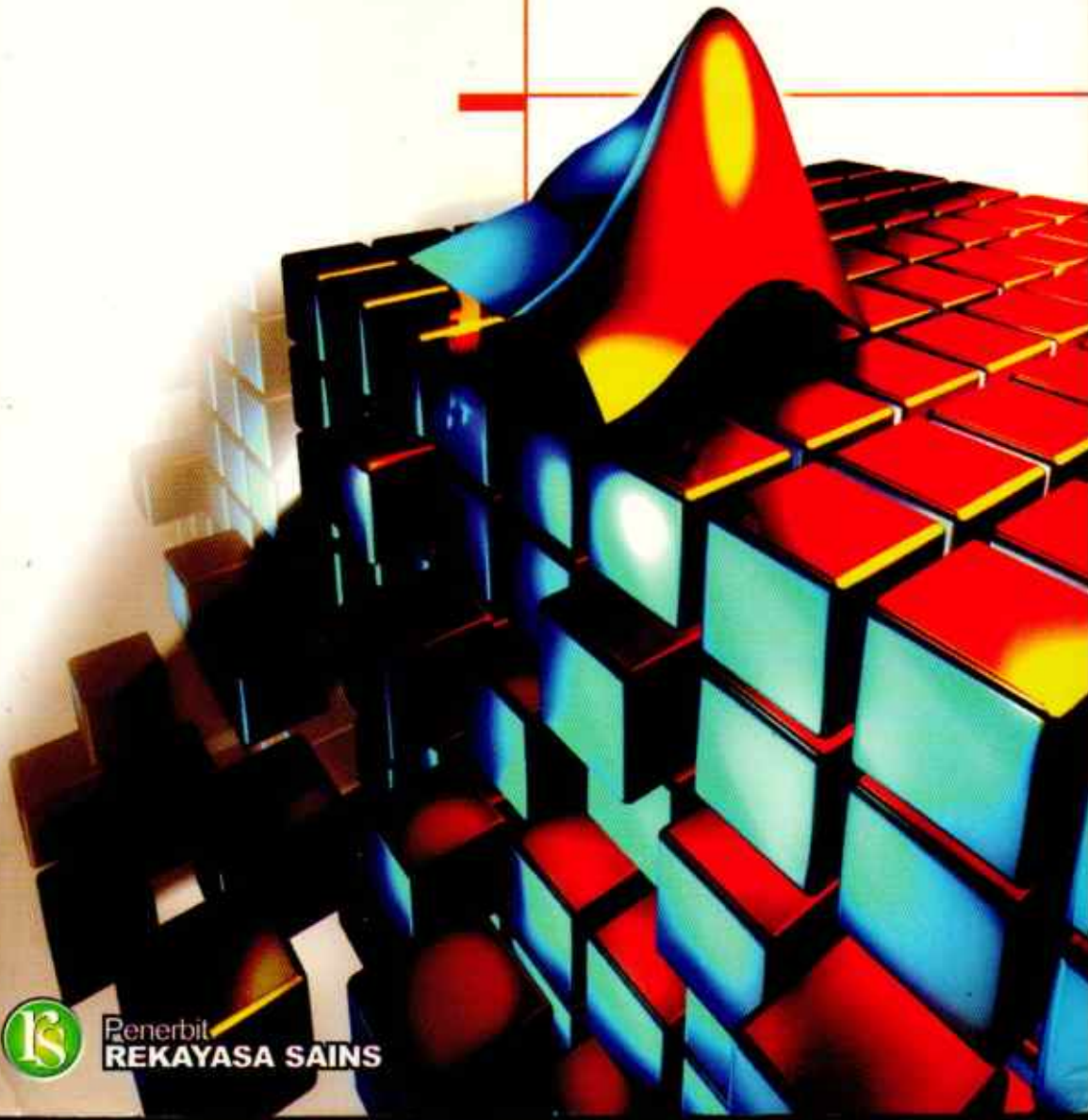


PRABOWO PUDJO WIDODO
RAHMADYA TRIAS HANDAYANTO
HERLAWATI

PENERAPAN DATA MINING DENGAN MATLAB

BAHASAN MELIPUTI:

- Konsep Data Mining
- Pohon Keputusan
- Adaptive Neuro-Fuzzy Inference System (ANFIS)
- Jaringan Syaraf Tiruan (JST)
- Fuzzy C-Means Clustering (FCM)
- Support Vector Machine (SVM)
- Pengujian Sistem
- Instalasi MATLAB 7
- Dasar-Dasar MATLAB
- Mengompilasi Program dengan MATLAB



Penerbit
REKAYASA SAINS

PENERAPAN DATA MINING DENGAN MATLAB

Prabowo Pudjo Widodo
Rahmadya Trias Handayanto
Herlawati



Rekayasa Sains, Bandung

PENERAPAN DATA MINING DENGAN MATLAB

Prabowo Pujjo Widodo

Rahmadya Trias Handayanto

Herlawati

@ 2013, Penerbit Rekayasa Sains, Bandung

Pemasaran:BI-Obses Pasar Buku Palasari No. 82 Bandung 40264

Telp.(022) 7317812 - Fax. (022) 7317896

Prabowo Pujjo Widodo
Rahmadya Trias Handayanto
Herlawati

Cetakan Pertama: April 2013

ISBN: 978-979-3784-60-1

KATA PENGANTAR

Puji syukur kami panjatkan ke hadirat Allah atas selesainya penulisan buku Data Mining ini. Buku ini merupakan kelanjutan dari buku sebelumnya tentang *Soft Computing*. Banyaknya pembaca yang berminat membuat sistem berbasis mesin pembelajaran (*machine learning*) membuat kami membuat buku yang kebanyakan diambil dari kasus-kasus yang telah dibuat oleh mahasiswa-mahasiswa tingkat sarjana dan pascasarjana.

Data Mining mengharuskan tersedianya data yang akan digali dan dicari informasi-informasi tersembunyi yang bermanfaat bagi pengambil keputusan. Namun untuk mempermudah proses belajar, dalam buku ini kami hanya menyediakan data-data sederhana yang sengaja dirancang mirip dengan data-data yang jumlahnya banyak yang diperoleh lewat riset di institusi tertentu. Akan tetapi, walaupun sederhana, prinsipnya dapat diterapkan langsung dengan data riil.

Buku ini disusun dalam bentuk teori dasar singkat yang dilanjutkan dengan terapannya untuk kasus tertentu dengan bahasa pemrograman Matlab. Diharapkan pembaca melihat

DAFTAR ISI

| | |
|---|----------|
| Kata Pengantar | iii |
| Daftar Isi | v |
| Bab 1. Konsep Data Mining | 1 |
| 1. Pendahuluan | 1 |
| 1.1 Manfaat <i>Data Mining</i> | 3 |
| 1.2 Proses <i>Data Mining</i> | 4 |
| 1.3 Jenis Permasalahan <i>Data Mining</i> | 5 |
| 1.3.1 Klasifikasi | 5 |
| 1.3.2 Analisa Kluster (<i>Cluster Analysis</i>) | 8 |
| 1.3.3 Pencarian Hukum Asosiasi (<i>Association Rule Discovery</i>)..... | 9 |
| 1.3.4 Pencarian Pola Berurutan (<i>Sequential Pattern Discovery</i>)..... | 10 |
| 1.3.5 Regresi..... | 11 |
| 1.3.6 Deteksi Penyimpangan (<i>Deviation/ Anomaly Detection</i>)..... | 11 |
| 1.4 Pemodelan <i>Data Mining</i> | 12 |
| 1.5 Aplikasi-Aplikasi <i>Data Mining</i> | 16 |
| 1.6 Terminologi <i>Data Mining</i> | 19 |

| | |
|---|--|
| Bab 2. Pohon Keputusan | |
| 2.1 Teori Dasar | |
| 2.2 Studi Kasus Pohon Keputusan | |
| 2.2.1 Pembentukan Pohon Keputusan..... | |
| 2.2.2 Membuat Aplikasi Berbasis Pohon Keputusan..... | |
| | |
| Bab 3. Adaptive Neuro-Fuzzy Inference System (ANFIS) | |
| 3.1 Teori Dasar | |
| 3.2 Studi Kasus <i>Data Mining</i> dengan ANFIS | |
| 3.3 Membuat FIS Melalui Pelatihan (<i>Learning</i>).... | |
| 3.4 Membuat Aplikasi Berbasis ANFIS | |
| | |
| Bab 4. Jaringan Syaraf Tiruan (JST) | |
| 4.1 Teori Dasar | |
| 4.1.1 Fungsi Aktivasi | |
| 4.1.2 Metode Pembelajaran | |
| 4.1.3 Tipe Jaringan Syaraf Tiruan | |
| 4.1.4 Perambatan Balik (<i>Backpropagation</i>).. | |
| 4.1.5 <i>Fuzzy Neural Network</i> (FNN) | |
| 4.1.6 <i>Radial Basis Function</i> (RBF) <i>Neural Network</i> | |
| 4.2 Studi Kasus <i>Data Mining</i> dengan JST | |
| 4.2.1 Membuat Aplikasi untuk <i>Backpropagation Learning</i> | |
| 4.2.2 Membuat Aplikasi Pengambilan Keputusan | |
| 4.2.3 Memadukan Sistem Pembelajaran dengan <i>Decision Support System</i> (DSS) | |

| | |
|---|------------|
| Bab 5. Fuzzy C-Means Clustering (FCM) | 79 |
| 5.1 Teori Dasar | 79 |
| 5.1.1 Pengklusteran Kasar dengan <i>K-Nearest Neighbourhood</i> (KNN)..... | 80 |
| 5.1.2 Pengklustern Halus dengan <i>Fuzzy C-Means Clustering</i> (FCM)..... | 83 |
| 5.2 Studi Kasus | 85 |
| 5.2.1 Penyelesaian dengan K-NN | 86 |
| 5.2.2 Penyelesaian dengan FCM..... | 94 |
| a. Menentukan Titik Pusat Kluster | 94 |
| b. Membuat Aplikasi Berbasis FCM.. | 97 |
| b.1 Aplikasi Berbasis Text..... | 97 |
| b.2 Aplikasi Berbasis GUI | 101 |
| Bab 6. Support Vector Machine (SVM) | 107 |
| 6.1 Teori Dasar | 107 |
| 6.1.1 Kasus Data yang Terpisah Secara Linear | 108 |
| 6.1.2 Kasus Data yang Tidak Terpisah Secara Linear | 112 |
| 6.2 Studi Kasus <i>Data Mining</i> dengan SVM..... | 116 |
| 6.2.1 Kasus Kelas Biner (Dua Kelas)..... | 116 |
| 6.2.2 Kasus Kelas Jamak (<i>Multiclass</i>)..... | 121 |
| a. Membuat persamaan svmStruct1 antara kelas IPA dan IPS..... | 122 |
| b. Membuat Persamaan svmStruct2 antara IPA dan Bahasa..... | 125 |
| c. Membuat Persamaan svmStruct3 antara IPS dan Bahasa..... | 127 |
| 6.2.3 Membuat Kode Program Untuk Klasifikasi dengan Basis Text (<i>Text-based</i>) | 131 |
| 6.2.4 Membuat Kode Program Untuk Klasifikasi dengan GUI (<i>Graphical User Interface</i>) | 133 |

| | |
|---|--|
| BAB 7. Pengujian Sistem | |
| 7.1 Pendahuluan | |
| 7.2 Data Training dan Data Testing | |
| 7.3 Validasi Silang (Cross Validation)..... | |
| 7.4 Kurva Receiver Operating Characteristic (ROC) | |
| 7.5 Studi Kasus Cross Validation | |

| | |
|---|--|
| Lampiran A. Instalasi MATLAB 7 | |
|---|--|

| | |
|---|--|
| Lampiran B. Dasar-Dasar MATLAB | |
|---|--|

| | |
|-------------------------|--|
| A. Command Window | |
| B. Kalkulus..... | |
| C. Aljabar Linear | |
| D. Statistik..... | |
| E. Grafik..... | |
| F. Excel Link..... | |

| | |
|--|--|
| Lampiran C. Mengompilasi Program dengan MATLAB. | |
|--|--|

| | |
|-------------------------------|--|
| Daftar Referensi | |
|-------------------------------|--|

BAB

I

Konsep *Data Mining*

1. Pendahuluan

Beberapa literatur yang membahas *data mining* banyak kita jumpai saat ini dan layak untuk dibaca (lihat daftar referensi). Tidak ada buku yang sanggup memberikan informasi lengkap karena *Data Mining* merupakan bidang yang sangat luas penerapannya. Untuk perkuliahan ini beberapa materi diambil dari buku referensi tersebut (kompilasi) sesuai dengan perkembangan terkini yang cenderung mengarah ke sistem *Soft Computing* yang lebih baik dalam mengatasi permasalahan yang ada saat ini.

BAB

2

Pohon Keputusan

2.1 Teori Dasar

Pohon keputusan (*decision tree*) merupakan salah satu teknik terkenal dalam *data mining* dan merupakan salah satu metode yang populer dalam menentukan keputusan suatu kasus. Hal ini karena metode ini tidak memerlukan proses pengelolaan pengetahuan terlebih dahulu dan dapat menyelesaikan dengan sederhana kasus-kasus yang memiliki dimensi yang besar. Akurasi sangat baik asalkan data yang akan dijadikan patokan merupakan data yang akurat. Metode ini banyak diterapkan dalam berbagai bidang antara lain: kesehatan dan pengobatan, finansial, produksi, astronomi, hingga biologi

BAB

4

Jaringan Syaraf Tiruan (JST)

1 Teori Dasar

beda dengan *Fuzzy Inference System* (FIS), atau yang lebih dikenal dengan *fuzzy* saja, ANFIS membuat *rule* berdasarkan *rule* yang di-*training* lewat mekanisme mirip Jaringan Syaraf Tiruan (JST). Jenis *rule* yang bisa dilayani hanyalah yang tipe Takagi-Sugeno-Kang (TSK), atau dikenal dengan istilah Sugeno saja. Jenis Mamdani tidak dapat diterapkan pada ANFIS.

Jaringan syaraf tiruan adalah sistem pemroses informasi yang karakteristik dan performa yang mendekati syaraf

BAB

7

Pengujian Sistem

Pendahuluan

asi adalah kunci ketika membuat aplikasi berbasis *data*
g. Ada berbagai macam cara dalam melakukan evaluasi.
asi ternyata tidak sesederhana yang kita bayangkan. Jika
memiliki data yang kita gunakan dalam proses pelatihan,
tidak serta merta kita menjadikan data tersebut sebagai
tor keberhasilan aplikasi yang kita buat. Oleh karena itu,
membutuhkan metode tertentu guna memprediksi performa
sarkan eksperimen untuk berbagai macam data selain
raining tersebut.