

DAFTAR ISI

Kata Pengantar	v
Panduan Penggunaan CD	vii
Daftar Isi	ix
BAB 1. KONSEP DATA MINING	1
1.1. Pendahuluan.....	1
1.2. Manfaat <i>Data Mining</i>	3
1.3. Proses <i>Data Mining</i>	3
1.4. Jenis Permasalahan <i>Data Mining</i>	4
1.4.1 Klasifikasi.....	4
1.4.2 Analisis Klaster (<i>Cluster Analysis</i>).....	6
1.4.3 Pencarian Hukum Asosiasi (<i>Association Rule Discovery</i>).....	8
1.4.4 Pencarian Pola Berurutan (<i>Sequential Pattern Discovery</i>).....	8
1.4.5 Regresi.....	9
1.4.6 Deteksi Penyimpangan (<i>Deviation/Anomaly Detection</i>).....	9
1.5. Pemodelan <i>Data Mining</i>	10
1.6. Aplikasi-Aplikasi <i>Data Mining</i>	14
1.7. Bahasa Pemrograman untuk <i>Data Mining</i>	15
1.8. Terminologi <i>Data Mining</i>	16
BAB 2. MEMPERSIAPKAN MATLAB.....	19
2.1. Dasar-Dasar Penggunaan Matlab	19
2.1.1 <i>Command Window</i>	19
2.1.2 Kalkulus.....	20
2.1.3 Aljabar Linear	25
2.1.4 Statistik	30
2.1.5 Grafik.....	31
2.1.6 <i>Excel Link</i>	34
2.2. Kompilasi Program pada Matlab.....	37
2.3. Instalasi Matlab	44

BAB 3. MEMPERSIAPKAN PYTHON	55
3.1 Mengunduh dan Menginstal Paket Anaconda	56
3.2 Menjalankan Anaconda Navigator	61
3.3 Menggunakan Jupyter Notebook	63
3.3.1 Menjalankan Jupyter Notebook	63
3.3.2 Menjalankan Jupyter Notebook <i>Online</i> (Google Colab)	69
3.4 Memasang TensorFlow	72
3.5 Mengelola Data dengan Python	77
3.5.1 Mengimpor Data Berformat Excel	81
3.5.2 Mengimpor Data berformat CSV	84
3.5.3 Mengimpor Data CSV dengan Google Colab	89
3.5.4 Mengunggah Data CSV pada Google Colab	91
3.6 Menggunakan Prosesor-Prosesor Google	93
3.7 <i>Graphical User Interface</i> (GUI) Berbasis Desktop Python	95
3.8 <i>Graphical User Interface</i> (GUI) Berbasis Web Pada Python	97
3.8.1 Mengunduh Pustaka Flask dan Jinja2	67
3.8.2 Menguji Framework Flask	68
BAB 4. METODE DATA MINING BERBASIS STATISTIKA 105	
4.1 Regresi dan Peramalan	105
4.1.1 Menyiapkan Dataset	107
4.1.2 Proses Pemodelan	110
4.1.3 Pembuatan Grafik/Plot	112
4.1.4 Pembuatan <i>Graphical User Interface</i> (GUI)	114
4.2 Pohon Keputusan	116
4.2.1 Studi Kasus Pohon Keputusan	121
4.2.2 Pembentukan Pohon Keputusan	122
4.2.3 Membuat Aplikasi Berbasis Pohon Keputusan	125
4.3 Penyelesaian Klasifikasi dengan Naïve Bayes	128
menggunakan Python	128
4.3.1 Menyiapkan Dataset	129
4.3.2 Memanggil Naïve Bayes Classifier	130
4.4 Penyelesaian Klasifikasi dengan KNN	133
4.4.1 Klasifikasi dengan KNN Matlab	134
4.4.2 Klasifikasi dengan KNN Python	140
4.4.3 Menggunakan KNN dengan Google Colab	146

BAB 5. ADAPTIVE NEURO-FUZZY INFERENCE SYSTEM (ANFIS)	151
5.1 Teori Dasar	151
5.2 Studi Kasus Data Mining dengan ANFIS	153
5.3 Membuat FIS melalui Pelatihan (<i>Learning</i>)	154
5.4 Membuat Aplikasi Berbasis ANFIS	161
BAB 6. JARINGAN SYARAF TIRUAN (JST).....	167
6.1 Teori Dasar	170
6.1.1 Fungsi Aktivasi	170
6.1.2 Metode Pembelajaran	171
6.1.3 Tipe Jaringan Syaraf Tiruan	172
6.1.4 Perambatan Balik (<i>Backpropagation</i>)	173
6.1.5 <i>Fuzzy Neural Network</i> (FNN)	175
6.1.6 Radial Basis <i>Function</i> (RBF)	176
6.2 Studi Kasus Data Mining dengan JST	178
6.2.1 Membuat Aplikasi untuk Backpropagation Learning	179
6.2.2 Membuat Aplikasi Pengambilan Keputusan	186
6.2.3 Memadukan Sistem Pembelajaran dengan <i>Decision Support System</i> (DSS)	189
6.3 Jaringan Syaraf Tiruan (JST) dengan Python	191
6.3.1 Menyiapkan Pustaka (<i>Library</i>)	191
6.3.2 Menyiapkan Dataset	194
6.3.3 Menyiapkan Model JST	197
6.3.4 Validasi Model JST	199
6.3.5 Menyimpan Model JST	201
6.3.6 Menjalankan Model JST Hasil Pelatihan (<i>Training</i>)	201
6.4 JST dengan Google <i>Interactive Notebook</i> (Google Colab)	203
6.4.1 Pelatihan JST	204
6.4.2 Menggunakan JST Hasil Pelatihan dengan GUI Desktop	208
6.5 Menggunakan JST Hasil Pelatihan dengan Web-based Python	212
6.6 Pemodelan JST Untuk Kasus Kelas Jamak (<i>Multi-class</i>)	219
BAB 7. KLASTERISASI (CLUSTERING)	223
7.1 Teori Dasar	223
7.1.1 Klasterisasi Halus dengan Fuzzy C-Means (FCM)	226

7.1.2	FCM dengan Matlab.....	228
7.1.3	Mencari Pusat Klaster	229
7.2	Klasterisasi K-Means dengan Python.....	231
7.2.1	Menyiapkan Dataset	233
7.2.2	Memanggil K-Means	233
7.3	Klasterisasi Sebagai Alat Bantu Aplikasi Klasifikasi	236
7.3.1	Membuat Aplikasi Berbasis Matlab	236
7.3.2	Aplikasi Berbasis Text	236
7.3.3	Aplikasi Berbasis GUI	240
BAB 8.	SUPPORT VECTOR MACHINE (SVM)	245
8.1	Teori Dasar	245
8.1.1	Kasus Data yang Terpisah secara Linear	246
8.1.2	Kasus Data yang Tidak Terpisah Secara Linear.....	249
8.2	Studi Kasus <i>Data Mining</i> dengan SVM	253
8.2.1	Kasus Kelas Biner (Dua Kelas).....	253
8.2.2	Kasus Kelas Jamak (<i>Multiclass</i>)	257
	a. Membuat persamaan svmStruct1 antara kelas IPA dan IPS.	258
	b. Membuat Persamaan svmStruct2 antara IPA dan Bahasa.	260
	c. Membuat Persamaan svmStruct3 antara IPS dan Bahasa	263
8.2.3	Membuat Kode Program untuk Klasifikasi dengan <i>Basis Text (Text-based)</i>	266
8.2.4	Membuat Kode Program untuk Klasifikasi dengan <i>Graphical User Interface (GUI)</i>	268
8.3.	Support Vector Machine (SVM) dengan Python.....	272
8.3.1	Menyiapkan Dataset	273
8.3.2	Memanggil SVM.....	274
8.3.3	SVM Kelas Jamak (<i>Multi-Class</i>).....	275
8.3.4	Klasifikasi dengan SVM dengan GUI Desktop	277
8.3.5	Klasifikasi dengan SVM dengan <i>Web-based</i> Python.	282
8.3.6	<i>Support Vector Regression (SVR)</i>	286

BAB 9.	VALIDASI DAN PENGUJIAN MODEL	289
9.1	Pendahuluan.....	289
9.2	Data Training dan <i>Data Testing</i>	290
9.3	Validasi Silang (<i>Cross Validation</i>)	291
9.4	Kurva <i>Receiver Operating Characteristic (ROC)</i>	292
9.5	Studi Kasus Validasi Silang dengan Matlab	294
9.6	Studi Kasus Validasi Silang dengan Python.....	298
9.7	Kalkulasi Akurasi pada Python	300
BAB 10.	DATA MINING WORLD WIDE WEB	301
10.1	Pendahuluan.....	301
10.2	Menggunakan Google Colab untuk Menggali Data Twitter	302
10.3	Mengambil Data Twitter	302
10.4	Sentimen Analysis	304
10.5	Mengambil Data Web.....	307
DAFTAR PUSTAKA	313	

1.1 Pendahuluan

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

Data mining menurut David Hand, Heikki Mannila, dan Padhraic Smyth dari MIT adalah analisis terhadap data (biasanya data yang berukuran besar) untuk menemukan hubungan yang jelas serta menyimpulkannya yang belum diketahui sebelumnya dengan cara yang dipahami dan berguna bagi pemilik data tersebut (Larose, 2006). Disebutkan bahwa cara yang digunakan adalah cara terkini (*Novel*) karena perkembangan teknologi yang terus berkembang dan database administrator harus mengikuti perkembangan yang