



Fajar Astuti Hermawati

DATA MINING

DAFTAR ISI

Kata Pengantar	-----	iii
Daftar Isi	-----	v

BAB 1. Pendahuluan	-----	1
1.1. Definisi <i>Data mining</i>	-----	3
1.2. Operasi <i>Data mining</i>	-----	5
1.3. Gudang Data (<i>Data Warehouse</i>)	-----	6
1.3.1. Karakteristik Data Warehouse	-----	7
1.3.2. Proses Dalam Data Warehouse	-----	7
1.3.3. Data Warehouse dan Sistem OLTP	-----	9
1.3.4. Model Data Warehouse	-----	10
1.4. Permasalahan Dalam <i>Data mining</i>	-----	13
1.5. Teknik <i>Data mining</i>	-----	14
1.5.1. Klasifikasi	-----	14
1.5.2. Regresi	-----	15
1.5.3. Klasterisasi (clustering)	-----	16
1.5.4. Kaidah Asosiasi (association rules)	-----	17
1.5.5. Pencarian pola sekuensial (Sequence Mining)	-----	18
1.6. Tantangan Dalam <i>Data mining</i>	-----	19
Soal Evaluasi Bab 1	-----	20
 BAB 2. Data	-----	23
2.1. Definisi dan Tipe Data	-----	23
2.2. Kualitas Data	-----	29
2.3. Pengolahan Awal (<i>Preprocessing</i>) Data	-----	31
2.3.1. Pengumpulan (Aggregation)	-----	31
2.3.2. Penarikan Contoh (Sampling)	-----	32
2.3.3. Pengurangan Dimensi (Dimensionality Reduction)	-----	33
2.3.4. Pemilihan Fitur (Feature subset selection)	-----	34
2.3.5. Pembuatan Fitur (Feature Creation)	-----	35



2.3.6. Pendiskritan dan Pembineran (Discretization and Binarization)	-----	35
2.3.7. Transformasi Atribut (Attribute Transformation)	-----	39
2.4. Similaritas dan Jarak	-----	40
2.4.1. Pengukuran Jarak (Distance) dari Dua Objek	-----	41
2.4.2. Pengukuran Similaritas dari Dua Objek	-----	45
Soal Evaluasi Bab II	-----	50
 BAB 3. Klarifikasi	 -----	 55
3.1. Klarifikasi	-----	55
3.2. Pendekatan Umum	-----	56
3.3. Pohon Keputusan (<i>Decision Tree</i>)	-----	58
3.3.1. Menentukan Kondisi Tes Atribut	-----	61
3.3.1.1. Percabangan Berdasarkan Atribut Nominal atau Ordinal	-----	61
3.3.1.2. Percabangan Berdasarkan Atribut Kontinyu	-----	63
3.3.2. Menentukan Percabangan Terbaik (Best Splitting)	-----	64
3.3.3. Kondisi Penghentian (Stopping Criteria)	-----	71
3.3.4. Model Overfitting	-----	78
3.5.4.1. Noise Overfitting	-----	79
3.5.4.2. Overfitting Karena Data Yang Tidak Mencukupi	-----	81
3.3.4.3. Estimasi Generalization Error	-----	82
Soal Evaluasi Bab III	-----	87
 BAB 4. Analisa Asosiasi	 -----	 93
4.1. Pengertian	-----	93
4.2. <i>Frequent Itemset</i>	-----	94
4.3. Association Rule	-----	97
4.4. Association Rule Mining	-----	97
4.4.1. FrequentItemset Generation	-----	99
4.4.2. Rule Generation	-----	108
4.5. Representasi <i>FrequentItemset</i> yang Kompak	-----	110
4.5.1. Maximal Frequent Itemset	-----	110
4.5.2. ClosedFrequentItemset	-----	111
Soal Evaluasi Bab IV	-----	112



BAB 5. Analisa Cluster ----- 117

5.1. Pengertian ----- 117

5.2. Tipe *Clustering* ----- 119

5.3. Tipe-Tipe *Cluster* ----- 119

5.4. Algoritma *Clustering* ----- 121

5.4.1. K-Means Clustering ----- 121

5.4.2. Hierarchiecal Clustering ----- 126

5.4.2.1. Anglomerative Clustering ----- 126

5.4.2.2. Divisive Clustering ----- 137

Soal Evaluasi Bab IV ----- 142

BAB 6. Deteksi Anomali ----- 145

6.1. Pengertian ----- 145

6.2. Skema Deteksi Anomali ----- 147

6.2.1. Pendekatan Grafik ----- 147

6.2.2. Pendekatan Statistik ----- 148

6.2.3. Pendekatan Jarak (Distance-based Approaches) ----- 150

Daftar Pustaka ----- 153

LAMPIRAN A

A.1. Persamaan Linier & Matriks ----- 155

A.2. Penyelesaian Persamaan Linear dengan Matriks ----- 156

A.2.1. Bentuk Eselon-Baris ----- 156

A.2.2. Operasi Eliminasi Gauss ----- 157

A.2.3. Operasi Eliminasi Gauss-Jordan ----- 159

A.3. Operasi Dalam Matriks ----- 161

A.3.1. Matriks Balikan (Invers) ----- 162

A.3.2. Matriks Transpose ----- 164

A.3.3. Matriks Diagonal ----- 164

A.3.4. Matriks Segitiga ----- 166

A.3.5. Matriks Simetris ----- 167