

## TEKNIK KLASIFIKASI POHON KEPUTUSAN UNTUK MEMPREDIKSI KEBANGKRUTAN BANK BERDASARKAN RASIO KEUANGAN BANK

<sup>1</sup>Nurma Jayanti  
<sup>2</sup>Sulistyo Puspitodjati  
<sup>3</sup>Tety Elida

<sup>1</sup>jurusan sistem informasi, universitas gunadarma ([umabult@yahoo.com](mailto:umabult@yahoo.com))  
<sup>2</sup>jurusan sistem informasi, universitas gunadarma ([sulistyo@staff.gunadarma.ac.id](mailto:sulistyo@staff.gunadarma.ac.id))  
<sup>3</sup>jurusan sistem informasi, universitas gunadarma ([tety@staff.gunadarma.ac.id](mailto:tety@staff.gunadarma.ac.id))

### ABSTRAK

*Penelitian ini bertujuan membentuk model klasifikasi pohon keputusan untuk memprediksi kebangkrutan bank dan melihat variabel yang paling berpengaruh dari rasio keuangan bank antara bank yang dikategorikan bangkrut dan tidak bangkrut. Obyek dalam penelitian ini adalah semua Bank Umum Swasta Nasional Devisa di Indonesia yang tercantum dalam Buku Direktori Perbankan Indonesia periode tahun 2001/2002-2002/2003. Data yang digunakan adalah laporan keuangan berdasarkan perhitungan rasio keuangan bank tahun 2001-2003. Pembentukan model prediksi kebangkrutan bank ini menggunakan metode pohon keputusan dengan menggunakan perangkat lunak open source WEKA (Waikato Environment for Knowledge Analysis) yang dikembangkan di Universitas Waikato. Algoritma yang digunakan ialah C4.5 (J48 implementasi di WEKA), kemudian model tersebut direpresentasikan dalam bentuk aturan (rule). Kondisi bank yang diprediksi yaitu status bank bangkrut dan tidak bangkrut merupakan output (variabel tidak bebas). Sedangkan sebagai inputnya (variabel bebas) adalah rasio-rasio keuangan bank yaitu rasio solvabilitas (Capital adequacy Ratio (CAR), Aktiva tetap terhadap modal), rasio rentabilitas (Return on Assets (ROA), Return on Equity (ROE), Net Interest Margin (NIM), Rasio Biaya Operasional (BOPO)) dan rasio likuiditas (Loan to Deposit Ratio (LDR). Dari pohon keputusan yang terbentuk, diketahui CAR merupakan variabel yang menentukan kebangkrutan suatu bank.*

**Kata Kunci:** *pohon keputusan, algoritma C4.5, kebangkrutan bank, model prediksi kebangkrutan*

### 1. PENDAHULUAN

Klasifikasi sebagai fungsi prediksi dapat diaplikasikan untuk meramalkan kebangkrutan bank. Analisis kebangkrutan diperlukan untuk memperoleh peringatan awal dari kebangkrutan. Semakin awal tanda-tanda kebangkrutan bank diketahui semakin baik karena pihak manajemen dapat melakukan perbaikan-perbaikan.

Selain itu, pihak kreditur dan pihak pemegang saham dapat melakukan persiapan untuk mengatasi berbagai kemungkinan yang buruk.

Secara garis besar ada dua teknik yang umum digunakan untuk meramalkan kebangkrutan yaitu teknik statistika seperti analisa regresi, korelasi, analisa diskriminasi, model logit dan model probit. Teknik yang lain adalah menggunakan

komputer intelegensia seperti pohon keputusan (*decision tree*), *Artificial Neural Network* (ANN) atau *Support Vector Machine* (SVM). Menurut Luciana (2006) tidak mudah mengambil kesimpulan bahwa teknik yang satu lebih baik dari teknik yang lain.

Pohon keputusan ialah salah satu metode klasifikasi yang paling populer karena mudah untuk diinterpretasi oleh manusia. Algoritma C4.5 adalah algoritma klasifikasi data dengan metode pohon keputusan yang memiliki kelebihan misalnya dapat mengolah data numerik (kontinyu) dan diskret, dapat menangani nilai atribut yang hilang, menghasilkan aturan-aturan yang mudah diinterpretasikan dan tercepat diantara algoritma-algoritma yang menggunakan memori utama di komputer.

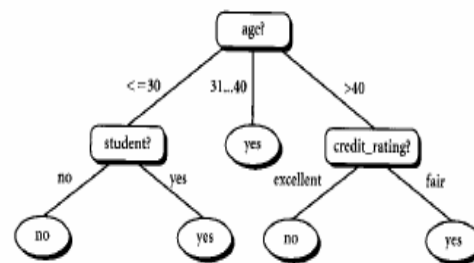
## 2. TINJAUAN PUSTAKA

Pohon keputusan adalah model prediksi menggunakan struktur pohon atau struktur berhirarki. Contoh dari pohon keputusan dapat dilihat di Gambar 1. Disini setiap percabangan menyatakan kondisi yang harus dipenuhi dan tiap ujung pohon menyatakan kelas data. Contoh di Gambar 1 adalah identifikasi pembeli komputer, dari pohon keputusan tersebut diketahui bahwa salah satu kelompok yang potensial membeli komputer adalah orang yang berusia di bawah 30 tahun dan juga pelajar.

Setelah sebuah pohon keputusan dibangun maka dapat digunakan untuk mengklasifikasikan *record* yang belum ada kelasnya. Dimulai dari *node root*, menggunakan tes terhadap atribut dari *record* yang belum ada kelasnya tersebut lalu mengikuti cabang yang sesuai dengan

hasil dari tes tersebut, yang akan membawa kepada *internal node* (*node* yang memiliki satu cabang masuk dan dua atau lebih cabang yang keluar), dengan cara harus melakukan tes lagi terhadap atribut atau *node* daun. *Record* yang kelasnya tidak diketahui kemudian diberikan kelas yang sesuai dengan kelas yang ada pada *node* daun. Pada pohon keputusan setiap simpul daun menandai label kelas

Gambar 1. Contoh Pohon Keputusan



Sumber: Pramudiono(2008)

Proses dalam pohon keputusan yaitu mengubah bentuk data (tabel) menjadi model pohon (*tree*) kemudian mengubah model pohon tersebut menjadi aturan (*rule*).

### ALGORITMA C4.5

Salah satu algoritma induksi pohon keputusan yaitu ID3 (Iterative Dichotomiser 3). ID3 dikembangkan oleh J. Ross Quinlan. Dalam prosedur algoritma ID3, input berupa sampel training, label training dan atribut. Algoritma C4.5 merupakan pengembangan dari ID3. Sedangkan pada perangkat lunak *open source* WEKA mempunyai versi sendiri C4.5 yang dikenal sebagai J48.

Tabel 1  
Algoritma C4.5

|   |
|---|
| <p>Algoritma C4.5<br/>Input : sampel training, label training, atribut</p> <ul style="list-style-type: none"><li>• Membuat simpul akar untuk pohon yang dibuat</li><li>• Jika semua sampel positif, berhenti dengan suatu pohon dengan satu simpul akar, beri label (+)</li><li>• Jika semua sampel negatif, berhenti dengan suatu pohon dengan satu simpul akar beri label (-)</li><li>• Jika atribut kosong, berhenti dengan suatu pohon dengan satu simpul akar, dengan label sesuai nilai yang terbanyak yang ada pada label training</li><li>• Untuk yang lain, Mulai<ul style="list-style-type: none"><li>▪ <math>A \leftarrow</math> atribut yang mengklasifikasikan sampel dengan hasil terbaik (berdasarkan <i>gain ratio</i>)</li><li>▪ Atribut keputusan untuk simpul akar <math>\leftarrow A</math></li><li>▪ Untuk setiap nilai, <math>v_i</math>, yang mungkin untuk <math>A</math>,<ul style="list-style-type: none"><li>– Tambahkan cabang dibawah akar yang berhubungan dengan <math>A = v_i</math></li><li>– Tentukan sampel <math>S_{v_i}</math> sebagai subset dari sampel yang mempunyai nilai <math>v_i</math> untuk atribut <math>A</math></li><li>– Jika sampel <math>S_{v_i}</math> kosong,<ul style="list-style-type: none"><li>* Dibawah cabang tambahkan simpul daun dengan label = nilai yang terbanyak yang ada pada label training</li><li>* Yang lain, tambah cabang baru dibawah cabang yang sekarang</li></ul></li></ul></li></ul></li><li>• Berhenti</li></ul> |
|---|

Sumber : Mitchell (1997)

Pohon dibangun dengan cara membagi data secara rekursif hingga tiap bagian terdiri dari data yang berasal dari kelas yang sama. Bentuk pemecahan (*split*) yang digunakan untuk membagi data tergantung dari jenis atribut yang digunakan dalam *split*. Algoritma C4.5 dapat menangani data numerik (kontinyu) dan diskret. *Split* untuk atribut numerik yaitu mengurutkan contoh berdasarkan atribut kontinyu  $A$ , kemudian membentuk minimum permulaan (*threshold*)  $M$  dari contoh-contoh yang ada dari kelas mayoritas pada setiap partisi yang bersebelahan, lalu menggabungkan partisi-partisi yang bersebelahan tersebut dengan kelas mayoritas yang sama. *Split* untuk atribut diskret  $A$  mempunyai bentuk *value* ( $A$ )  $\in X$  dimana  $X \subset domain(A)$ .

Jika suatu set data mempunyai beberapa pengamatan dengan *missing value* yaitu *record* dengan beberapa nilai variabel tidak ada, Jika jumlah pengamatan terbatas maka atribut dengan

*missing value* dapat diganti dengan nilai rata-rata dari variabel yang bersangkutan.[Santosa,2007]

Untuk melakukan pemisahan obyek (*split*) dilakukan tes terhadap atribut dengan mengukur tingkat ketidakmurnian pada sebuah simpul (*node*). Pada algoritma C.45 menggunakan rasio perolehan (*gain ratio*). Sebelum menghitung rasio perolehan, perlu menghitung dulu nilai informasi dalam satuan bits dari suatu kumpulan objek. Cara menghitungnya dilakukan dengan menggunakan konsep entropi.

$$E(S) = -p_+ \log_2 p_+ - p_- \log_2 p_-$$

$S$  adalah ruang (data) sampel yang digunakan untuk pelatihan,  $p_+$  adalah jumlah yang bersolusi positif atau mendukung pada data sampel untuk kriteria tertentu dan  $p_-$  adalah jumlah yang bersolusi negatif atau tidak mendukung pada data sampel untuk kriteria tertentu. Entropi( $S$ ) sama dengan 0, jika semua

contoh pada  $S$  berada dalam kelas yang sama. Entropi( $S$ ) sama dengan 1, jika jumlah contoh positif dan negative dalam  $S$  adalah sama. Entropi( $S$ ) lebih dari 0 tetapi kurang dari 1, jika jumlah contoh positif dan negative dalam  $S$  tidak sama [Mitchell,1997]. Entropi split yang membagi  $S$  dengan  $n$  record menjadi himpunan-himpunan  $S_1$  dengan  $n_1$  baris dan  $S_2$  dengan  $n_2$  baris adalah :

$$E(S_1, S_2) = \frac{n_1}{n} E(S_1) + \frac{n_2}{n} E(S_2)$$

Kemudian menghitung perolehan informasi dari output data atau variabel dependent  $y$  yang dikelompokkan berdasarkan atribut  $A$ , dinotasikan dengan  $gain(y, A)$ . Perolehan informasi,  $gain(y, A)$ , dari atribut  $A$  relative terhadap output data  $y$  adalah :

$$Gain(y, A) = entropi(y) - \sum_{c \in \text{nilai}(A)} \frac{y_c}{y} entropi(y_c)$$

nilai ( $A$ ) adalah semua nilai yang mungkin dari atribut  $A$ , dan  $y_c$  adalah subset dari  $y$  dimana  $A$  mempunyai nilai  $c$ . Term pertama dalam persamaan diatas adalah  $entropi$  total  $y$  dan term kedua adalah  $entropi$  sesudah dilakukan pemisahan data berdasarkan atribut  $A$ .

Untuk menghitung rasio perolehan perlu diketahui suatu term baru yang disebut pemisahan informasi ( $SplitInfo$ ). Pemisahan informasi dihitung dengan cara :

$$SplitInfo(S, A) = - \sum_{i=1}^c \frac{S_i}{S} \log_2 \frac{S_i}{S}$$

bahwa  $S_1$  sampai  $S_c$  adalah  $c$  subset yang dihasilkan dari pemecahan  $S$  dengan menggunakan atribut  $A$  yang mempunyai sebanyak  $c$  nilai. Selanjutnya rasio perolehan (gain ratio) dihitung dengan cara :

$$GainRatio(S, A) = \frac{Gain(S, A)}{SplitInformation(S, A)}$$

### Indikator Kepailitan

Rasio Solvabilitas digunakan untuk mengukur kemampuan bank dalam memenuhi kewajiban jangka panjangnya atau mengetahui perbandingan antara jumlah dana yang diperoleh dari berbagai hutang (jangka pendek dan jangka panjang). Analisa ini dapat dilihat berdasarkan rasio :

- Capital Adequacy Ratio (CAR): rasio kinerja bank untuk mengukur kecukupan modal yang dimiliki bank untuk menunjang aktiva yang mengandung resiko.
- Aktiva Tetap terhadap modal (ATTM): rasio kinerja bank untuk mengukur kecukupan aktiva tetap terhadap modal yang dimiliki bank.

Rasio Rentabilitas adalah alat untuk menganalisis tingkat efisiensi usaha dan profitabilitas yang dicapai bank yang bersangkutan. Rentabilitas bank dapat dilihat dari beberapa rasio berikut :

- Return on Assets (ROA) mengukur kemampuan manajemen bank dalam memperoleh keuntungan secara keseluruhan.
- Return on Equity (ROE) adalah perbandingan antara laba bersih bank dengan modal sendiri.
- Rasio Biaya Operasional (BOPO) mengukur tingkat efisiensi dan kemampuan bank dalam melakukan kegiatan operasinya.
- Net Interest Margin (NIM) menggambarkan tingkat keuntungan yang diperoleh bank dibandingkan dengan pendapatan yang diterima dari kegiatan operasionalnya.

Rasio Likuiditas adalah analisis yang dilakukan terhadap kemampuan bank dalam memenuhi kewajiban jangka pendek atau kewajiban yang sudah jatuh tempo. Rasio yang dapat digunakan untuk menganalisa likuiditas bank yaitu :

- Loan to Deposit Ratio (LDR) yaitu menyatakan seberapa jauh kemampuan bank dalam membayar kembali penarikan dana yang dilakukan deposan dengan mengandalkan kredit yang diberikan sebagai sumber likuiditasnya.

### WEKA

WEKA (*Waikato Environment for Knowledge Analysis*) adalah suatu perangkat lunak pembelajaran mesin yang populer ditulis dengan *Java*, yang dikembangkan di Universitas Waikato. WEKA adalah perangkat lunak gratis yang tersedia di bawah GNU General Public License. Pada WEKA menyediakan penggunaan teknik klasifikasi menggunakan pohon keputusan dengan algoritma J48. Teknik klasifikasi dan algoritma yang digunakan di WEKA disebut *classifier*.

### 3. METODE PENELITIAN

Obyek penelitian ini adalah semua Bank Umum Swasta Nasional Devisa di Indonesia yang tercantum dalam buku Direktori Perbankan Indonesia periode tahun 2001/2002-2002/2003 yaitu sebanyak 29 bank yang tidak gagal dan 7 bank yang gagal. Data yang digunakan adalah laporan keuangan bank berdasarkan Perhitungan Rasio Keuangan selama tiga tahun, mulai dari tahun 2001 sampai tahun 2003.

Variabel bebas atau variabel input (prediktor) pada penelitian ini adalah rasio solvabilitas, rasio rentabilitas, dan rasio likuiditas.

Rasio solvabilitas yang digunakan adalah :

1. CAR

2. ATTM

Rasio rentabilitas yang digunakan adalah :

3. ROA

4. ROE

5. NIM

6. BOPO

Rasio Likuiditas yang digunakan adalah :

7. LDR

Sedangkan variabel tidak bebasnya (output) adalah status kepailitan bank. Variabel ini menyatakan status bank yang dikategorikan sebagai bank yang pailit dan tidak pailit. Data set tersebut disajikan dalam bentuk tabel yang terdiri dari 22 kolom (atribut) dan 36 baris (record).

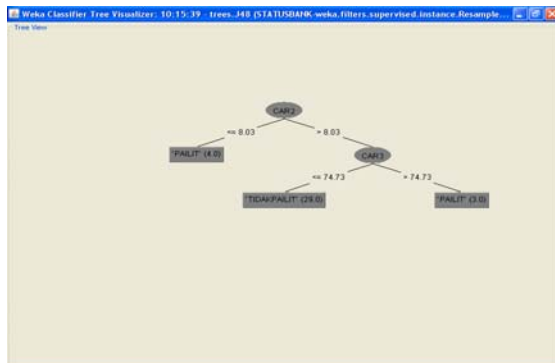
### Pemrosesan data pada WEKA

Data\_BUSND.arff menggunakan WEKA *filter supervised instance* untuk memperoleh interval yang baik pada setiap nilai kelas (atribut) dan memilih *Resample* untuk menghasilkan suatu subcontoh acak dari suatu dataset, percontohnya dengan penggantian untuk memelihara distribusi kelas. Faktor terbatasnya jumlah data untuk data pelatihan dan data uji maka pada penelitian ini memilih *cross-validation* untuk mengevaluasi *classifier*. Data akan digunakan sebagai data pelatihan dan data uji. Pemisahkan data pelatihan ke dalam  $n$  bagian (*folds*) menggunakan 10 bagian (*folds*), setiap bagian akan digunakan sebagai data tes dan sisanya  $n-1$  bagian (*folds*) dari data akan digunakan sebagai data training untuk membentuk model dan memunculkan hasil evaluasi. Hasil evaluasi terakhir akan dikumpulkan dari hasil masing-masing bagian.

### 4. HASIL DAN PEMBAHASAN

Dari pohon keputusan yang terbentuk, hasil kriteria perhitungan *gain ratio* paling besar dari atribut

Data\_BUSND ialah CAR<sub>2</sub> yang dijadikan sebagai simpul akar pada pohon keputusan tersebut. Pada model pohon keputusan tersebut dapat diperoleh pengetahuan (*knowledge*) bahwa variabel yang berpengaruh untuk memodelkan kebangkrutan ialah Capital adequacy Ratio (CAR) pada rasio solvabilitas.



Gambar 2. Pohon keputusan yang terbentuk

Sumber : hasil keluaran di WEKA (2008)

Aturan-aturan yang diperoleh dari pohon keputusan tersebut untuk membentuk model prediksi kebangkrutan berdasarkan perhitungan rasio keuangan tahun 2001 sampai 2003 pada penelitian ini, yaitu :

- IF CAR<sub>2</sub> <= 8.03 THEN status bank = PAILIT
- IF CAR<sub>2</sub> > 8.03 AND CAR<sub>3</sub> <= 74.73 THEN status bank = TIDAKPAILIT
- IF CAR<sub>2</sub> > 8.03 AND CAR<sub>3</sub> > 74.73 THEN status bank = PAILIT

```
Test mode:      10-fold cross-validation
=== Classifier model (full training set) ===
J48 pruned tree
-----
CAR2 <= 8.03: 'PAILIT' (4.0)
CAR2 > 8.03
| CAR3 <= 74.73: 'TIDAKPAILIT' (29.0)
| CAR3 > 74.73: 'PAILIT' (3.0)

Number of Leaves :      3
Size of the tree :      5
```

Gambar 3. Model pohon keputusan direpresentasikan dalam bentuk aturan (rule)

Sumber : hasil keluaran di WEKA (2008)

```
--- Summary ---
Correctly Classified Instances      34      94.4444 %
Incorrectly Classified Instances     2      5.5556 %
Kappa statistic                     0.8227
Mean absolute error                 0.0556
Root mean squared error             0.2357
Relative absolute error              17.0536 %
Root relative squared error         59.044 %
Total Number of Instances          36

--- Detailed Accuracy By Class ---
TP Rate  FP Rate  Precision  Recall  F-Measure  ROC Area  Class
0.857    0.034    0.857     0.857   0.857     0.911    'PAILIT'
0.966    0.143    0.966     0.966   0.966     0.911    'TIDAKPAILIT'

--- Confusion Matrix ---
a b  <-- classified as
6 1 | a = 'PAILIT'
1 28 | b = 'TIDAKPAILIT'
```

Gambar 4. Hasil evaluasi metode pohon keputusan dengan algoritma J48

Sumber : hasil keluaran di WEKA (2008)

Dari 36 record Data\_BUSND yang terdiri dari 29 bank tidak gagal dan 7 bank gagal pada model pohon keputusan yang terbentuk ternyata menghasilkan keakurasian yang baik terhadap hasil prediksi pada data secara keseluruhan. Hal ini dapat diketahui dari evaluasi indeks yang ditunjukkan pada area *Classifier output*, meliputi : banyaknya jumlah yang diprediksikan benar (*Correctly Classified Instances*) sebanyak 34 record yaitu 94.4444 % sedangkan banyaknya jumlah yang diprediksikan salah (*Incorrectly Classified Instances*) sebanyak 2 record yaitu 5.5556 % dari keseluruhan data. *Kappa statistic* sebesar 0.8227, *Mean absolute error* sebesar 0,0556, *Root mean squared error* sebesar 0.2357, *Relative absolute error* sebesar 17.0536 %, dan *Root relative squared error* sebesar 59.044 %.

Pada *Confusion matrix* tersebut diketahui status bank pailit yang diprediksikan benar pailit (*true positive*) sebanyak 6 bank, status bank tidak pailit yang diprediksikan benar tidak pailit (*true negative*) sebanyak 28 bank, nilai yang diprediksikan salah sebagai tidak pailit padahal nilai sebenarnya pailit (*False Positive*) sebanyak 1 bank dan nilai yang diprediksikan salah sebagai pailit padahal nilai sebenarnya tidak pailit (*False Negative*) sebanyak 1 bank. Dari

keterangan-keterangan tersebut, jumlah data yang diprediksikan benar adalah penjumlahan antara status bank pailit yang diprediksikan benar pailit dengan status bank tidak pailit yang diprediksikan benar tidak pailit sebanyak 34 bank dari keseluruhan data. Hal ini juga menjelaskan bahwa pohon keputusan yang terbentuk memiliki kesalahan keakurasian yang minimal terhadap keseluruhan data. Sedangkan perhitungan untuk representasi keakurasian yang baik juga ditunjukkan dengan evaluasi indeks pada masing-masing kelas meliputi *True Positive (TP) rate*, *False Positive (FP) rate*, *Precision*, *Recall*, dan *F-Measure*.

## 5. PENUTUP

Berdasarkan pohon keputusan yang terbentuk dapat diketahui bahwa model prediksi kebangkrutan bank dengan menggunakan metode pohon keputusan menunjukkan bahwa rasio keuangan yang paling berpengaruh ialah Capital Adequacy Ratio (CAR) dua tahun sebelum terjadi kebangkrutan.

Penggunaan pohon keputusan dalam memprediksi kebangkrutan kemungkinan bukan merupakan metode yang tepat, jika dilihat dari hasil penelitian ini, yaitu hanya CAR sebagai variabel penentu kebangkrutan.

Penelitian lanjutan penggunaan metode ini untuk memprediksi kebangkrutan dapat dilakukan dengan penggunaan jumlah data yang lebih banyak untuk meningkatkan keakurasian hasil pencarian pengetahuan (*knowledge*) dan perlu adanya uji validasi terhadap hasil pengetahuan akhir tentang status bank pada data baru.

## 6. DAFTAR PUSTAKA

- Almillia, Luciana Spica. 2006. Prediksi Kondisi Financial Distress Perusahaan GoPublic Dengan Menggunakan Analisis Multinomial Logit. *Jurnal Ekonomi dan Bisnis*. Vol.XII No 1, ISSN: 0854-9087.
- Lan,lin dan Lichun Zhu. 2006. *The Exploration and Application of WEKA Data Miner*. Project report.
- Mitchell, Tom M. 1997. *Machine Learning*. McGraw-Hill. United States of America.
- Moertini,Veronica Sri. 2007. *Pengembangan skalabilitas algoritma klasifikasi C4.5 dengan pendekatan konsep operator relasi (Studi kasus: Prapengolah-an dan klasifikasi citra batik*, Institut Teknologi Bandung.
- Pramudiono, Iko. *Pengantar Data Mining: Menambang Permata Pengetahuan di Gunung Data*. <http://www.ilmukomputer.com>. Diakses tanggal 2 Mei 2008.
- Riyanto, Bambang. 1997. *Dasar-dasar Pembelanjaan Perusahaan*. BPFE. Yogyakarta.
- Santosa, Budi. 2007. *Data Mining : Teknik Pemanfaatan Data untuk keperluan Bisnis*. Graha Ilmu. Yogyakarta.
- Tan, Pang-Ning, Michael Steinbach, and Vipin Kumar. 2004. *Introduction to Data Mining*.
- Website WEKA. <http://www.cs.waikato.ac.nz/ml/weka/>. Diakses tanggal 2 Mei 2008.
- Witten, Ian H. dan Eibe Frank. 2005. *Data Mining: Practical machine learning tools and techniques, 2<sup>nd</sup> Edition*. Morgan Kaufmann. San Francisco.