

SISTEM PENDUKUNG KEPUTUSAN BERBASIS *DECISION TREE* DALAM PEMBERIAN BEASISWA STUDI KASUS: AMIK “BSI YOGYAKARTA”

Anik Andriani

*Program Studi Manajemen Informatika, AMIK BSI Jakarta
Jl. RS Fatmawati 24, Pondok Labu, Jakarta Selatan
Telp. (021) 7500282
E-mail: anik.andriani2011@gmail.com*

ABSTRACT

Scholarships reserved for students who are economically lessable to pay for college tuition. Many criteria used for determining the scholarship made it difficult to make a decision on the granting of scholarships and the time it takes too long. The purpose of this studi is the classification of scholarships recipients by using decision tree with the algorithm C4.5. Classification results are evaluated and validated by Confusion Matrix and ROC curves to determine the level of accuracy decision tree in making the classification scholarship. The resulting classification is used to create a decision support system in the awarding of scholarships. The system used was made with Microsoft Visual Basic 6.0. The decision support system is to simplify and speed up decision making for the scholarships.

Keywords: Scholarship, Decision Tree, Decision Support System

ABSTRAK

Pemberian beasiswa diperuntukkan untuk mahasiswa yang secara ekonomi kurang mampu untuk membayar biaya studi di perguruan tinggi. Banyaknya kriteria yang digunakan untuk penentuan pemberian beasiswa mengakibatkan sulitnya membuat keputusan pemberian beasiswa. Selain itu waktu yang dibutuhkan juga lama. Tujuan penelitian ini adalah membuat klasifikasi mahasiswa penerima beasiswa dengan Decision Tree yang menggunakan Algoritma C4.5. Hasil klasifikasi dievaluasi dan divalidasi dengan Confusion Matrix dan Kurva ROC untuk mengetahui tingkat akurasi Decision Tree dalam membuat klasifikasi beasiswa. Hasil klasifikasi digunakan untuk membuat sistem pendukung keputusan dalam pemberian beasiswa. Sistem yang digunakan dibuat dengan Microsoft Visual Basic 6.0. Dengan adanya sistem pendukung keputusan ini dapat mempermudah dan mempercepat pengambilan keputusan untuk pemberian beasiswa.

Kata Kunci: Beasiswa, Decision Tree, Sistem Pendukung Keputusan

1. PENDAHULUAN

1.1 Latar Belakang Masalah

Tingginya biaya kuliah bagi sebagian mahasiswa kurang mampu menjadi kendala dalam melanjutkan studi di perguruan tinggi. Hal ini dapat mengakibatkan mahasiswa sering mengajukan cuti akademik maupun *dropout*. Pemberian beasiswa pada perguruan tinggi mempunyai beberapa kriteria yang digunakan sebagai bahan pertimbangan dalam mengambil sebuah keputusan apakah akan diberikan beasiswa atau tidak.

Akademi Manajemen Informatika dan Komputer (AMIK) “BSI Yogyakarta” memberikan beasiswa kepada para mahasiswanya yang secara ekonomi kurang mampu dan mahasiswa yang mempunyai prestasi akademik yang bagus, sehingga para mahasiswa tersebut dapat terus melanjutkan studi.

Banyaknya kriteria yang diseleksi dalam menentukan keputusan pemberian beasiswa mengakibatkan pihak manajemen kesulitan dalam mengambil sebuah keputusan dan waktu yang dibutuhkan menjadi lebih lama.

Klasifikasi merupakan sebuah proses untuk

menemukan suatu model atau fungsi yang menggambarkan dan membedakan sebuah kelas data atau konsep yang mempunyai tujuan dapat menggunakan model hasil klasifikasi tersebut untuk membuat prediksi kelas objek dimana kelas labelnya tidak diketahui (Han & Kamber, 2007). Salah satu teknik klasifikasi yang populer digunakan adalah *Decision Tree* (Gorunescu, 2011).

1.2 Rumusan Masalah

Berdasarkan uraian dari latar belakang masalah sebelumnya, maka dapat dirumuskan beberapa permasalahan sebagai berikut:

- Apakah klasifikasi beasiswa yang terbentuk dengan *Decision Tree* memiliki tingkat akurasi yang tinggi?
- Apakah sistem pendukung keputusan untuk penentuan pemberian beasiswa lebih efisien?

1.3 Tujuan Penelitian

Penelitian ini bertujuan membuat klasifikasi mahasiswa penerima beasiswa dengan *Decision Tree* yang selanjutnya hasil klasifikasi digunakan untuk

membuat sistem pendukung keputusan dalam penentuan pemberian beasiswa.

2. TINJAUAN PUSTAKA

2.1 Sistem Pendukung Keputusan

Decision Support System atau Sistem Pendukung Keputusan merupakan sebuah sistem informasi yang berbasis komputer yang digunakan untuk proses pengambilan keputusan yang menyediakan dukungan informasi yang interaktif bagi manajer dan praktisi bisnis. Dalam sistem ini model yang digunakan adalah model analitis, *database*, penilaian dan pandangan pembuat keputusan dan proses pemodelan berbasis komputer yang interaktif guna mendukung pengambilan keputusan yang semi terstruktur (Turban & Aronson, 2001).

2.2 Decision Tree

Algoritma *decision tree* didasarkan pada pendekatan *divide-and-conquer* untuk klasifikasi suatu masalah. Algoritma tersebut bekerja dari atas ke bawah, mencari pada setiap tahap atribut untuk membaginya ke dalam bagian terbaik *class* tersebut, dan memproses secara rekursif submasalah yang dihasilkan dari pembagian tersebut. Strategi ini menghasilkan sebuah *decision tree* yang dapat diubah menjadi satu set *classification rules* (Witten et al, 2011).

Decision tree merupakan salah satu metode klasifikasi yang menggunakan representasi struktur pohon (*tree*) di mana setiap *node* merepresentasikan atribut, cabangnya merepresentasikan nilai dari atribut, dan daun merepresentasikan kelas. *Node* yang paling atas dari *decision tree* disebut sebagai *root* (Gorunescu, 2011).

Pada *decision tree* terdapat 3 jenis *node*, yaitu:

a. *Root Node*, merupakan *node* paling atas, pada *node* ini tidak ada *input* dan bisa tidak mempunyai *output* atau mempunyai *output* lebih dari satu.

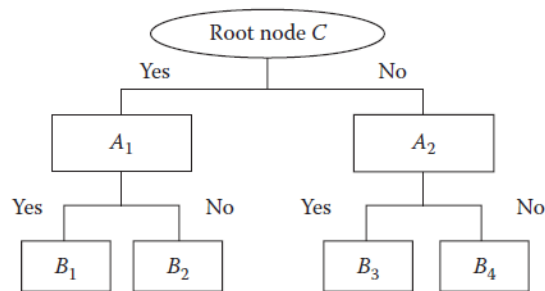
b. *Internal Node*, merupakan *node* percabangan, pada *node* ini hanya terdapat satu *input* dan mempunyai *output* minimal dua.

c. *Leaf node* atau *terminal node*, merupakan *node* akhir, pada *node* ini hanya terdapat satu *input* dan tidak mempunyai *output*.

Seperti ditunjukkan dalam Gambar 1, *decision tree* tergantung pada aturan *if-then*, tetapi tidak membutuhkan parameter dan metrik. Struktur sederhana dan dapat ditafsirkan memungkinkan *decision tree* untuk memecahkan masalah atribut *multi-type*. *Decision tree* juga dapat mengelola nilai-nilai yang hilang atau data *noise* (Dua & Xian, 2011).

Banyak algoritma yang dapat dipakai dalam pembentukan *Decision Tree*, antara lain ID3, CART, dan C4.5 (Larose, 2005). Algoritma C4.5 dan pohon keputusan merupakan dua model yang tak terpisahkan, karena untuk membangun sebuah pohon keputusan, dibutuhkan algoritma C4.5 (Han &

Kamber, 2006).



Gambar 1 Contoh Struktur *Decision Tree*

Sumber: Dua & Xian, 2011

Membangun klasifikasi dengan *Decision Tree* yang menggunakan Algoritma C4.5, melalui beberapa tahapan sebagai berikut (Larose, 2005):

- Pertama siapkan data training yang biasanya diambil dari data histori atau data masa lampau yang kemudian dibuat ke dalam kelas-kelas tertentu.
- Menghitung nilai *entropy* yang akan digunakan untuk menghitung nilai *gain* dari masing-masing atribut sehingga diperoleh atribut dengan nilai *gain* yang tertinggi yang selanjutnya akan digunakan menjadi akar pohon. Rumus menghitung *entropy* dan *gain* seperti yang ditunjukkan dalam persamaan (1) dan (2).

$$\text{Entropy}(S) = -\sum_{i=1}^n p_i \log p_i \quad (1)$$

Keterangan:

S= Himpunan kasus

n = jumlah partisi S

P_i= proporsi S_i terhadap S

$$\text{Gain}(S, A) = \text{entropy}(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} \cdot \text{Entropy}(S_i) \quad (2)$$

Keterangan:

S = Himpunan Kasus

A = Fitur

n = jumlah partisi atribut A

|S_i| = Proporsi S_i terhadap S

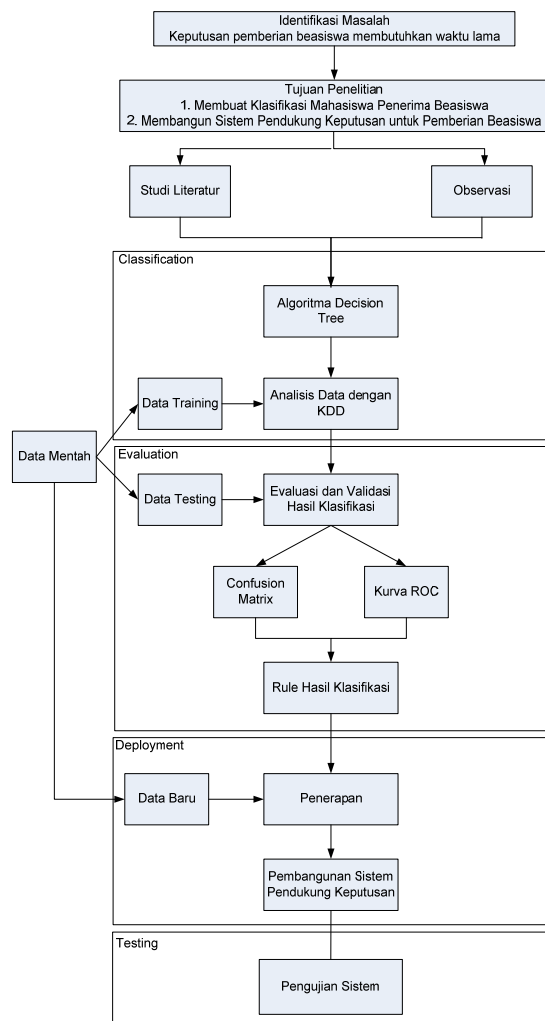
|S| = jumlah kasus dalam S

- Ulangi terus langkah sebelumnya yaitu menghitung nilai tiap atribut berdasarkan nilai *gain* yang tertinggi hingga semua record terpartisi.
- Proses dari *Decision Tree* ini akan berhenti jika semua record dalam simpul N mendapat kelas yang sama, tidak ada atribut di dalam *record* yang dipartisi lagi, dan tidak ada *record* di dalam cabang yang kosong.

3. METODE PENELITIAN

Penelitian eksperimen menggunakan data dalam penelitiannya dan menghasilkan kesimpulan yang mampu dibuktikan oleh pengamatan atau percobaan

(Kothari, 2004). Penelitian ini termasuk penelitian eksperimen dengan menggunakan data mahasiswa di AMIK “BSI Yogyakarta” sejumlah 350 data yang dibagi menjadi data *training* dan data *testing* dengan perbandingan 80% dan 20%, sehingga diperoleh 280 data *training* dan 70 data *testing*. Data *training* digunakan untuk memperoleh hasil klasifikasi mahasiswa penerima beasiswa dalam bentuk *decision tree*, sedangkan data *testing* digunakan untuk mengukur tingkat akurasi dari hasil klasifikasi tersebut.



Gambar 2. Kerangka Pemikiran

4. HASIL DAN PEMBAHASAN

4.1 Analisis Data

Analisis data yang digunakan untuk proses klasifikasi dapat menggunakan *Knowledge Discovery in Databases* (KDD) yang terdiri dari sembilan langkah yang dimulai dari tahap pemahaman data yang akan digunakan hingga tahap terciptanya sebuah pengetahuan tentang klasifikasi beasiswa dan penggunaannya. KDD sendiri diartikan sebagai proses terorganisir untuk mengidentifikasi pola dalam data yang besar dan kompleks dimana pola data tersebut ditemukan yang bersifat sah, baru, dan dapat bermanfaat serta dapat

dimengerti (Maimon & Rokach, 2010)

Sembilan langkah dalam KDD yang digunakan dalam analisis data untuk klasifikasi yaitu:

- Developing an understanding of the application domain* yang merupakan tahap untuk memahami apa yang akan dilakukan dalam penelitian ini.
- Selecting and creating a data set on which discovery will be performed*, merupakan tahap untuk pemilihan set data dan mempersiapkannya untuk digunakan dalam penelitian.
- Preprocessing and cleansing*, merupakan tahap untuk meningkatkan kehandalan data dengan cara membersihkan data yang tidak lengkap (*missing value*) dan data yang tidak benar (*noise*).
- Data transformation*, merupakan tahap untuk menyusun dan mengembangkan set data menjadi lebih baik sehingga tahap ini membutuhkan proses kreatif dan sangat bergantung pada jenis atau pola informasi yang akan dicari dalam basis data, seperti mengkategorikan data kedalam beberapa kategori dan membagi data menjadi dua bagian yaitu data *training* dan *testing*
- Choosing the appropriate Data Mining task*, pada tahap ini memilih teknik data mining yang digunakan yaitu klasifikasi.
- Choosing the Data Mining Algorithm*, tahap ini memilih jenis algoritma yang akan digunakan dalam klasifikasi yaitu *Decision Tree* dengan Algoritma C4.5.
- Employing the Data Mining Algorithm*, merupakan tahap membuat klasifikasi beasiswa dengan menggunakan algoritma yang telah dipilih dari proses sebelumnya.
- Evaluation*, tahap evaluasi dilakukan dengan menggunakan data *testing* untuk mengukur tingkat akurasi pola data yang diperoleh hasil klasifikasi dengan data *training* dengan menggunakan *confusion matrix* dan kurva ROC.
- Using the discovered knowledge*, merupakan tahap menggunakan *knowledge* yang diperoleh dari hasil klasifikasi dengan *Decision Tree* dan menerapkannya dalam sistem pendukung keputusan pemberian beasiswa.

4.2 Evaluasi Hasil

Evaluasi dari hasil klasifikasi mahasiswa penerima beasiswa dengan *decision tree* dapat menggunakan *confusion matrix* dan kurva ROC/AUC (*Area Under Curve*).

1. Confusion Matrix

Confusion Matrix merupakan sebuah metode untuk evaluasi yang menggunakan tabel matrix seperti pada tabel 1.

Pada tabel 1 dapat kita lihat bahwa jika dataset terdiri dari dua kelas, kelas yang satu dianggap sebagai positif dan yang lainnya negatif (Bramer, 2007). Evaluasi dengan *confusion matrix* menghasilkan nilai *accuracy*, *precision*, dan *recall*.

Nilai *accuracy* merupakan persentase jumlah record data yang diklasifikasikan secara benar oleh sebuah algoritma dapat membuat klasifikasi setelah dilakukan pengujian pada hasil klasifikasi tersebut (Han & Kamber, 2006). Nilai *precision* atau dikenal juga dengan nama *confidence* merupakan proporsi jumlah kasus yang diprediksi positif yang juga positif benar pada data yang sebenarnya. Sedangkan nilai dari *recall* atau *sensitivity* merupakan proporsi jumlah kasus positif yang sebenarnya yang diprediksi positif secara benar (Powers, 2011).

Tabel 1. Model *Confusion Matrix*

Correct Classification	Classified as	
	+	-
+	True positives	False negatives
-	False positives	True negatives

Sumber Han & Kamber (2006)

True Positive adalah jumlah *record* positif yang diklasifikasikan sebagai positif, *false positive* adalah jumlah *record negative* yang diklasifikasikan sebagai positif, *false negative* adalah jumlah *record* positif yang diklasifikasikan sebagai negatif, *true negative* adalah jumlah *record negative* yang diklasifikasikan sebagai negatif, kemudian masukkan data uji. Setelah data uji dimasukkan ke dalam *confusion matrix*, hitung nilai-nilai yang telah dimasukkan tersebut untuk dihitung jumlah *sensitivity (recall)*, *Specifity*, *precision*, dan *accuracy*. *Sensitivity* digunakan untuk membandingkan jumlah *t_pos* terhadap jumlah *record* yang positif sedangkan *Specifity*, *precision* adalah perbandingan jumlah *t_neg* terhadap jumlah *record* yang negatif. Untuk menghitung digunakan persamaan dibawah ini (Han dan Kamber).

$$\text{Sensitivity} = \frac{t_{pos}}{p} \quad (3)$$

$$\text{Specifity} = \frac{t_{neg}}{n} \quad (4)$$

$$\text{Precision} = \frac{p}{p+f_{pos}} \quad (5)$$

$$\text{accuracy} = \frac{p}{p+n} + \text{Specifity} \frac{n}{p+n} \quad (6)$$

Keterangan :

t_pos : jumlah *true* positif
t_neg : jumlah *true* negatif
p : jumlah *record* positif
n : jumlah tupel negatif
f_pos : jumlah *false* positif

Hasil evaluasi *confusion matrix* dalam klasifikasi menunjukkan tingkat akurasi hasil klasifikasi seperti yang ditunjukkan dalam tabel 2. Dari tabel tersebut dapat kita ketahui tingkat akurasi hasil klasifikasi

penerima beasiswa sebesar 71,43%.

Tabel 2. Hasil Evaluasi nilai *accuracy* dengan *Confusion Matrix*

accuracy: 71.43%			
	true tidak	true beasiswa	class precision
pred. tidak	27	13	67.50%
pred. beasiswa	7	23	76.67%
class recall	79.41%	63.89%	

Berdasarkan tabel 3 hasil evaluasi dengan *confusion matrix* menunjukkan nilai *precision* sebesar 76,67%.

Tabel 3. Hasil Evaluasi nilai *precision* dengan *Confusion Matrix*

precision: 76.67% (positive class: beasiswa)			
	true tidak	true beasiswa	class precision
pred. tidak	27	13	67.50%
pred. beasiswa	7	23	76.67%
class recall	79.41%	63.89%	

Berdasarkan tabel 4 hasil evaluasi dengan *confusion matrix* menunjukkan nilai *recall* sebesar 63,89%.

Tabel 4. Hasil Evaluasi nilai *recall* dengan *Confusion Matrix*

recall: 63.89% (positive class: beasiswa)			
	true tidak	true beasiswa	class precision
pred. tidak	27	13	67.50%
pred. beasiswa	7	23	76.67%
class recall	79.41%	63.89%	

2. Kurva ROC

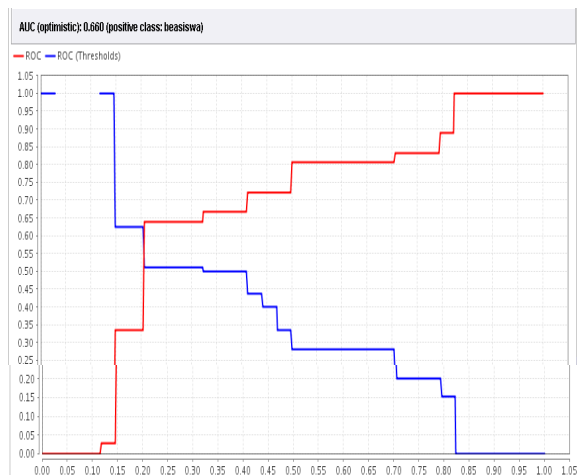
Kurva ROC menunjukkan akurasi dan membandingkan klasifikasi secara visual. ROC mengekspresikan *confusion matrix*. ROC adalah grafik dua dimensi dengan *false positives* sebagai garis horizontal dan *true positive* sebagai garis vertical (Vercellis, 2009).

Hasil perhitungan divisualisasikan dengan kurva ROC (*Receiver Operating Characteristic*) atau AUC (*Area Under Curve*). ROC memiliki tingkat nilai diagnosa yaitu, Gorunescu[5]:

- Akurasi bernilai 0.90 – 1.00 = *excellent classification*
- Akurasi bernilai 0.80 – 0.90 = *good classification*

- c. Akurasi bernilai 0.70 – 0.80 = *fair classification*
- d. Akurasi bernilai 0.60 – 0.70 = *poor classification*
- e. Akurasi bernilai 0.50 – 0.60 = *failure*

Hasil yang didapat dari pengolahan ROC untuk algoritma C4.5 dengan menggunakan data *training* sebesar 0.660 dapat dilihat pada gambar 3 dengan tingkat diagnosa *poor classification*.



Gambar 3. Hasil Evaluasi dengan Kurva ROC/AUC

4.3 Pembangunan Sistem

Pembangunan sistem pendukung keputusan pemberian beasiswa menggunakan Microsoft Visual Basic 6.0 dengan menerapkan *rule* hasil klasifikasi dengan *decision tree*, hasilnya sebagai berikut:

Gambar 4. Form Pembuat Keputusan Pemberian Beasiswa

5. KESIMPULAN DAN SARAN

Dari hasil dan pembahasan dapat diambil beberapa kesimpulan sebagai berikut:

1. Model *Decision Tree* dapat digunakan dalam membuat klasifikasi sebagai dasar dalam pembangunan sistem pendukung keputusan pemberian beasiswa.

2. Evaluasi hasil klasifikasi penerima beasiswa dengan model *decision tree* menggunakan algoritma C4.5 yang dievaluasi dengan *confusion matrix* menghasilkan tingkat akurasi hasil klasifikasi penerima beasiswa dengan *decision tree* sebesar 71,43%, sedangkan evaluasi dengan kurva ROC sebesar 0,660.
3. Penerapan sistem pendukung keputusan dapat mempercepat waktu pengambilan keputusan dalam menentukan pemberian beasiswa kepada mahasiswa yang mengajukan beasiswa di kampus AMIK “BSI Yogyakarta”

Beberapa saran yang dapat dijadikan pertimbangan untuk penelitian selanjutnya yaitu:

1. Menambahkan beberapa atribut dalam menentukan pemberian beasiswa selain prestasi akademik yang dalam hal ini menggunakan IPK dengan prestasi non- akademik seperti prestasi dalam bidang olahraga, seni, music, dan lain-lain.
2. Menerapkan teknik penyeleksian atribut yang paling berpengaruh dengan *chi-square* sehingga tingkat akurasi bisa lebih tinggi.
3. Melakukan komparasi dari beberapa algoritma dalam klasifikasi untuk memperoleh algoritma dengan tingkat akurasi yang paling tinggi dalam membuat klasifikasi pemberian beasiswa.

PUSTAKA

- Bramer, Max. 2007. *Principles of Data Mining*. London: Springer. ISBN-10: 1-84628-765-0, ISBN-13: 978-1-84628-765-7.
- C.R.Kothari. 2004. *Research Methodology Methods and Techniques*. India: New Age International Limited. ISBN (13) : 978-81-224-2488-1
- Dua, S. & Xian Du. 2011. *Data Mining and Machine Learning in Cybersecurity*. USA: Taylor & Francis Group. ISBN-13: 978-1-4398-3943-0
- Gorunescu, F. 2011. *Data Mining Concept Model and Techniques*. Berlin: Springer. ISBN 978-3-642-19720-8
- Han, J., & Kamber, M. 2006. *Data Mining Concept and Tehniques*. San Fransisco: Morgan Kauffman. ISBN 13: 978-1-55860-901-3
- Larose, D. T. 2005. *Discovering Knowledge in Data*. New Jersey: John Wiley & Sons, Inc. ISBN 0-471-66657-2.
- Maimon, Oded., & Rokach, Lior. 2010. *Data Mining and Knowledge Discovery Handbook, 2nd Edition*. New York: Springer. ISBN 978-0-387-09822-7
- Powers, D.M.W. 2011. Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation. *Journal of Machine Learning Technologies*, ISSN: 2229-3981 & ISSN: 2229-399X, Volume 2, Issue 1, 2011, pp-37-63.

- Turban, E., and Aronson, J.E., 2001, *Decision Support System and Intelligent System*, 6th Edition. New York. Prentice Hall, Inc.
- Vercellis, Carlo. 2009. *Business Intelligence: Data Mining and Optimization for Decision Making*. United Kingdom: John Willey & Son.
- Witten, I. H., Frank, E., Hall, M. A. 2011. *Data Mining Practical Machine Learning Tools and Techniques* (3rd ed). USA: Elsevier