



## Implementasi Algoritma C5.0 pada Penilaian Kinerja Mitra BPS Kota Salatiga

Ulya Fatmawati<sup>1\*</sup>, Bowo Winarno<sup>2</sup>, Tri Atmojo Kusmayadi<sup>3</sup>  
<sup>1,2,3.</sup> Universitas Sebelas Maret

---

### INFO ARTICLES

#### Article History:

Received: 22-06-2023  
Revised: 22-06-2023  
Approved: 27-06-2023  
Publish Online: 30-06-2023

#### Key Words:

BPS; Decision Tree;  
C5.0 Algorithm; RStudio.



This article is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

**Abstract:** Partner of a government agency named the Statistics Indonesia (BPS) Kota Salatiga, must provide maximum service to the community, because the partners of the BPS Kota Salatiga are directly involved in providing services to the community, especially in terms of data collection in the field. With this important role, good partner management is needed. Partner performance must be measured to determine whether the partner's professionalism has been maximized. Performance appraisal is needed for purposes that require selection to enable partners who have participated in activities at BPS to re-register for further BPS activities. C5.0 algorithm is one of the decision tree algorithms that can help determine partners who are worth recommending. Therefore the C5.0 algorithm is implemented using the RStudio programming application in this research. The partner assessment dataset uses the attributes of Quality of Work, Quantity of Work, Leadership, Responsibility, and Teamwork. The results of the research show that the C5.0 algorithm is able to determine partner data that is worthy of being recommended by 94.33%.

**Abstrak:** Mitra dari lembaga pemerintah yaitu Badan Pusat Statistik (BPS) Kota Salatiga harus mengedepankan pelayanan yang maksimal kepada masyarakat, karena mitra BPS Kota Salatiga ikut terjun langsung dalam pelayanan kepada masyarakat terutama dalam hal pengambilan data di lapangan. Dengan adanya peran penting tersebut diperlukan pengelolaan mitra yang baik. Kinerja mitra perlu diukur untuk menentukan apakah profesionalisme mitra telah tercapai dengan maksimal. Penilaian kinerja diperlukan untuk keperluan yang membutuhkan seleksi yang memungkinkan mitra yang telah mengikuti kegiatan di BPS mendaftar kembali di kegiatan BPS selanjutnya. Algoritma C5.0 adalah salah satu algoritma pohon keputusan yang dapat membantu untuk menentukan mitra yang layak direkomendasikan. Oleh karena itu diimplementasikan Algoritma C5.0 dengan menggunakan aplikasi perogramaran RStudio pada penelitian ini. Dataset penilaian mitra menggunakan atribut Kualitas Kerja, Kuantitas Kerja, Kepemimpinan, Tanggung Jawab, dan Kerjasama Tim. Hasil penelitian menunjukkan bahwa algoritma C5.0 mampu menentukan data mitra layak direkomendasikan sebesar 94,33%.

---

**Correspondence Address:** Jln. Ir. Sutami No. 36A, Kota Surakarta, 57126, Jawa Tengah, Indonesia; e-mail: [ulyafatmawati@student.uns.ac.id](mailto:ulyafatmawati@student.uns.ac.id)

**How to Cite:** Fatmawati, U., Winarno, B., & Kusmayadi, T. A. (2023). Implementasi algoritma C5.0 pada penilaian kinerja mitra BPS kota Salatiga. *Himpunan: Jurnal Ilmiah Mahasiswa Pendidikan Matematika*, 3(1), 95-102.

**Copyright:** Ulya Fatmawati, Bowo Winarno, Tri Atmojo Kusmayadi. (2023).

---

## PENDAHULUAN

Badan Pusat Statistik atau dapat disebut BPS merupakan lembaga pemerintah di luar Kementerian, bertanggungjawab kepada Presiden. Salah satu tugas BPS yang berdasarkan UU No. 16 Tahun 1997 adalah memenuhi kebutuhan informasi pemerintah dan masyarakat. Data tersebut berasal dari sensus atau wawancara mandiri dan juga diperoleh dari kementerian atau lembaga pemerintah lainnya sebagai data sekunder. Peran BPS tidak terlepas dari keberadaan mitra BPS sebagai komponen yang membantu pelaksanaan sensus atau survei.

Sebagai mitra dari lembaga pemerintah yaitu Badan Pusat Statistik (BPS) Kota Salatiga, mitra BPS Kota Salatiga tentunya harus mengedepankan pelayanan yang maksimal kepada masyarakat, karena mitra BPS Kota Salatiga ikut terjun langsung dalam pelayanan kepada masyarakat terutama dalam hal pengambilan data di lapangan. Dengan adanya peran penting tersebut diperlukan pengelolaan mitra yang baik. Profesionalisme mitra ditentukan dengan mengukur kinerja mitra.

Penilaian kinerja diperlukan untuk keperluan yang membutuhkan seleksi. Pada saat ini penilaian masih dilakukan secara manual dan belum ada parameter tertentu yang ditetapkan. Hal tersebut tentunya akan membuat proses pengelompokan data menjadi lebih lama terlebih lagi jika mitra yang terlibat semakin banyak. *Data mining* digunakan untuk mengatasi kesulitan dalam pengelompokan data. Teknologi tersebut digunakan pada penelitian ini untuk mengolah data mitra BPS Kota Salatiga.

Definisi data mining menurut Witten dan Frank (2011) adalah proses penggalian data/pola dalam informasi atau informasi dari sejumlah besar data. Penambangan data diproses (disaring) sehingga data membentuk pola/informasi baru yang bermanfaat. *Data mining* juga dapat diartikan sebagai proses pencarian pola data yang tidak diketahui atau tidak diharapkan. Peran data mining diperlukan dalam mengolah data menjadi informasi/model/informasi yang berguna. Peran-peran tersebut secara umum ada 5, yaitu estimasi, prediksi, klasifikasi, klustering, dan asosiasi. Teknik klasifikasi yang digunakan salah satunya merupakan pohon keputusan (*decision tree*).

Menurut Suntoro, J, pohon keputusan adalah pohon rekursif top-down algoritma induktif yang menerapkan metrik pemilihan atribut untuk memilih atribut yang akan diuji. *Decision tree* memudahkan pengguna untuk melihat hubungan antar faktor yang mempengaruhi suatu masalah. *Decision tree* dapat menganalisis nilai suatu informasi yang ada pada suatu alternatif pemecahan masalah. *Decision tree* dapat membuat prediksi dengan struktur pohon, yaitu dengan cara mengubah data menjadi *decision tree* dan beberapa *rule* atau aturan keputusan.

Algoritma C5.0 diterapkan pada penelitian ini seperti yang sudah pernah dilakukan pada penelitian sebelumnya namun terdapat perbedaan pada variabel. Penelitian sebelumnya yang telah dilakukan dengan menggunakan algoritma C5.0 memperoleh hasil yang menunjukkan bahwa dalam proses klasifikasi, algoritma C5.0 lebih efektif saat digunakan seperti pada penelitian mengenai perbandingan penggunaan algoritma *decision tree* C5.0, *Classification and Regression Tree* (CART), dan *Chi-square Automatic Interaction Detection* (CHAID), dimana pada hasil tersebut menunjukkan bahwa algoritma C5.0 lebih akurat (Putu Wirya Kastawan dkk.).

Menurut Moss dan Said (2000), algoritma C5.0 adalah versi lebih sempurna dari algoritma ID3 dan C4.5. Pada proses pembangunan *decision tree* tersebut *information gain* terbesar dipilih menjadi akar(*root*) pada *node* selanjutnya. Semua data dijadikan akar dari *decision tree* terlebih dahulu kemudian atribut terpilih menjadi partisi sampel.

## METODE

### 1. Data Penelitian

Data yang digunakan yaitu data mitra BPS Kota Salatiga yang telah dipilah sehingga dapat dipastikan tidak ada data null atau data kosong. Data yang digunakan adalah data penilaian 300 mitra BPS Kota Salatiga dengan menggunakan 5 atribut penilaian yaitu Kualitas

Kerja, Kuantitas Kerja, Kepemimpinan, Tanggung Jawab, dan Kerjasama Tim. Setiap atribut memiliki 3 kategori penilaian, yaitu: Rendah, Sedang, dan Tinggi.

## 2. Langkah Penelitian

Pembentukan pohon keputusan dan aturan dengan algoritma C5.0 melalui langkah-langkah berikut.

a. Menyiapkan dataset

Dataset penilaian 300 mitra BPS Kota Salatiga dalam bentuk *file excel*.

b. Membagi data

Data yang ada diacak terlebih dahulu kemudian dibagi menjadi 2, yaitu data latih (*data training*) dan data uji (*data testing*) dengan perbandingan 80:20 atau sebesar 80% untuk *data training* dan 20% untuk *data testing* dari keseluruhan data.

c. Membuat model

Pembuatan model *decision tree* dan prediksi pada *tree*.

d. Pembentukan *tree* C5.0 dan aturan keputusan

Data diolah dengan aplikasi RStudio untuk mendapatkan *tree* dan aturan keputusan.

e. Penghitungan akurasi dengan *confusion matrix*

*Confusion matrix* adalah sebuah cara evaluasi yang menggunakan tabel *matrix* untuk menghitung akurasi.

Algoritma C5.0 dibentuk dengan nilai *entropy* dan *information gain*(informasi *gain*) dihitung terlebih dahulu. *Entropy* total sampel dihitung dengan persamaan berikut.

$$I(s_1, s_2, \dots, s_m) = - \sum_{i=1}^n p_i * \log_2(p_i) \quad (1)$$

Dengan S adalah himpunan kasus, n adalah jumlah sampel dan  $p_i$  adalah proporsi kelas. Kemudian dihitung *entropy* tiap atribut dengan persamaan

$$E(A) = \sum_{j=1}^y \frac{s_{1j} + \dots + s_{mj}}{S} I(s_{1j}, s_{2j}, \dots, s_{mj}) \quad (2)$$

Dengan  $s_{1j} + \dots + s_{mj}$  adalah banyaknya subset  $j$  yang akan dibagi dengan banyaknya sampel S. Selanjutnya menghitung nilai *information gain* dengan persamaan

$$Gain(A) = I(s_1, s_2, \dots, s_m) - E(A) \quad (3)$$

Dengan A adalah atribut kasus dan S adalah himpunan kasus. Kemudian yang terakhir yaitu menghitung *gain ratio* dengan persamaan

$$Gain Ratio = \frac{Gain(A)}{\sum_{i=1}^n I(s_{1j}, s_{2j}, \dots, s_{mj})} \quad (4)$$

Dengan  $Gain(A)$  adalah nilai *gain* dari atribut dan  $\sum_{i=1}^n I(s_{1j}, s_{2j}, \dots, s_{mj})$  adalah banyaknya nilai *entropy* per atribut.

Setelah pohon keputusan dan aturan keputusan terbentuk, dilakukan evaluasi tingkat akurasi dengan *confusion matrix*. Nilai akurasi dihasilkan dari evaluasi dengan *confusion matrix* tersebut. Setelah diuji, banyaknya data yang diklasifikasikan benar dinyatakan sebagai akurasi. Akurasi dihitung dengan persamaan

$$Akurasi = \frac{TP + TN}{TP + FP + TN + FN} \times 100\% \quad (5)$$

Dimana TP atau *True Positive* merupakan banyaknya data positif yang teridentifikasi benar pada sistem, TN atau *True Negative* merupakan banyaknya data negatif yang teridentifikasi benar pada sistem, FN atau *False Negative* merupakan banyaknya data negatif namun teridentifikasi salah pada sistem dan FP atau *False Positive* merupakan banyaknya data positif namun teridentifikasi salah pada sistem.

## HASIL PENELITIAN

Dari data yang terdapat pada dataset, *root node* ditentukan dengan menghitung nilai *entropy*, *information gain*, dan *gain ratio*. Nilai *entropy* total dihitung dengan rumus (1), didapatkan *entropy* total sebesar 0,609840305. Perhitungan *entropy* tiap atribut dihitung dengan rumus (2). Perhitungan nilai *information gain* menggunakan rumus (3), kemudian *gain ratio* dihitung dengan menggunakan rumus (4). Hasil perhitungan untuk menentukan *root node* pada tabel 1.

**Tabel 1. Hasil Perhitungan Data Kinerja Mitra**

Atribut	Kategori	Jumlah data	Rekomendasi		Entropy	Gain	Gain Ratio
			Ya	Tidak			
Total		300	255	45	0,609840305		
Kualitas Kerja	Rendah	63	39	24	0,958711883	0,081393405	0,051210054
	Sedang	182	162	20	0,499584407		
	Tinggi	55	54	1	0,13110653		
				Total	1,58940282		
Kuantitas Kerja	Rendah	48	33	15	0,896038233	0,042561119	0,024972804
	Sedang	189	161	28	0,605186577		
	Tinggi	63	61	2	0,20307393		
				Total	1,704298739		
Kepemimpinan	Rendah	63	42	21	0,918295834	0,066697414	0,042125044
	Sedang	189	167	22	0,51892966		
	Tinggi	48	47	1	0,14609425		
				Total	1,583319744		
Tanggung Jawab	Rendah	56	31	25	0,991703308	0,113590807	0,077509526
	Sedang	197	177	20	0,473804387		
	Tinggi	47	47	0	0		
				Total	1,465507695		
Kerjasama Tim	Rendah	64	36	28	0,988699408	0,148159765	0,106348921
	Sedang	186	171	15	0,404448386		
	Tinggi	50	50	0	0		
				Total	1,393147795		

Sumber: Diolah dari data penelitian, 2023.

Pada tabel tersebut ditunjukkan bahwa nilai *information gain* tertinggi ada pada atribut Kerjasama Tim, maka atribut tersebut terpilih menjadi *root node* atau akar dari pohon keputusan. Selanjutnya dilakukan proses yang sama pada atribut selanjutnya sehingga menghasilkan pohon keputusan dan beberapa aturan keputusan.

Proses pembentukan *decision tree* dimulai dengan menginput dataset pada RStudio. Data tersebut selanjutnya dibagi menjadi data latih (*data training*) sebanyak 80% dari data keseluruhan dan *data testing*, kemudian data yang berbentuk karakter pada setiap atribut diubah menjadi bentuk faktor agar memudahkan program untuk membentuk prediksi.

Dari *data training* yang telah diolah dengan algoritma C5.0, *decision tree* terbentuk dengan output 3 atribut utama yang paling berpengaruh dalam menentukan aturan keputusan pada Gambar 1.

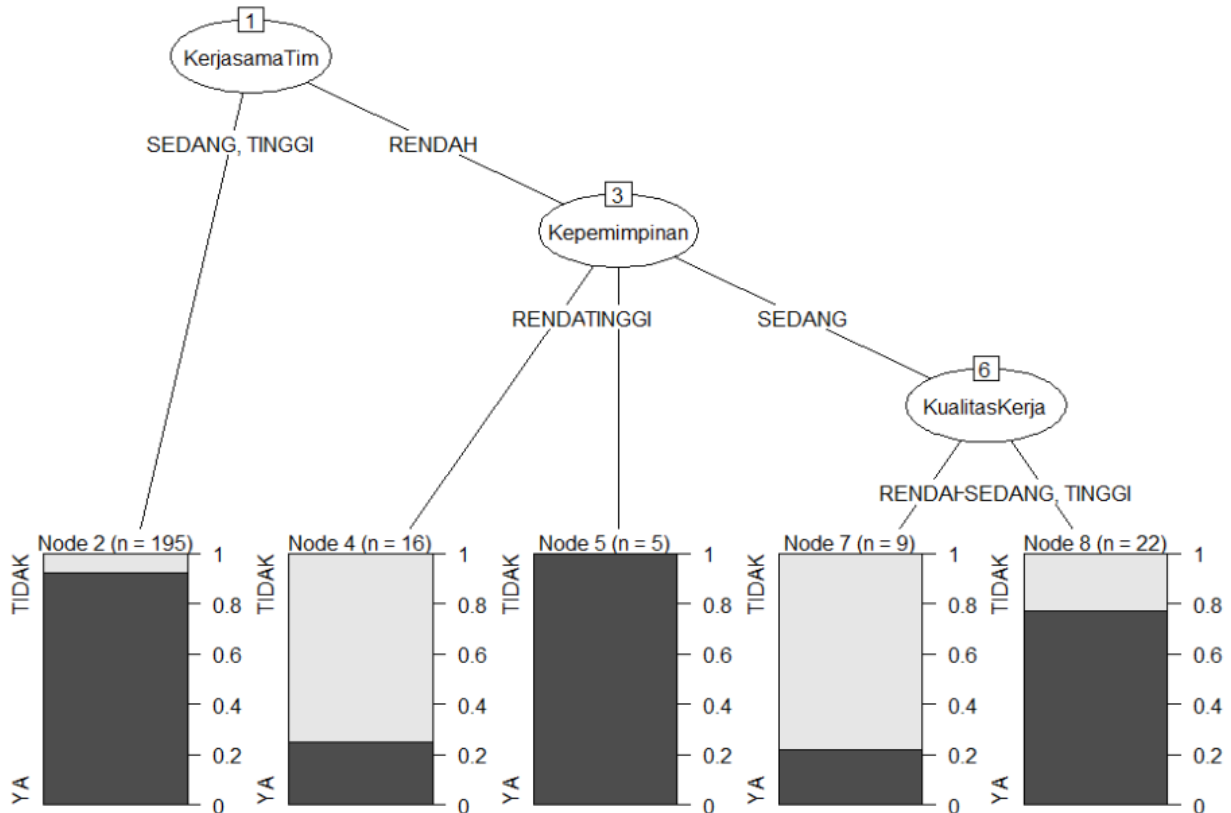
### Decision tree:

```

KerjasamaTim in {SEDANG,TINGGI}: YA (195/15)
KerjasamaTim = RENDAH:
...Kepemimpinan = RENDAH: TIDAK (16/4)
  Kepemimpinan = TINGGI: YA (5)
  Kepemimpinan = SEDANG:
...KualitasKerja = RENDAH: TIDAK (9/2)
  kualitaskerja in {SEDANG,TINGGI}: YA (22/5)
  
```

**Gambar 1.** *Decision Tree* Algoritma C5.0

Tampilan plot *decision tree* algoritma C5.0 tersebut terlihat pada Gambar 2.



**Gambar 2.** Plot *Decision Tree* Algoritma C5.0

### PEMBAHASAN HASIL PENELITIAN

*Decision tree* dapat terbentuk dengan 3 atribut yang paling mempengaruhi hasil pohon keputusan, yaitu Kerjasama Tim, Kepemimpinan, dan Kualitas Kerja. Dari pohon keputusan tersebut berlaku aturan keputusan yang telah ditentukan dengan pemrograman RStudio. Adapun aturan keputusan tersebut adalah

#### Rules:

```
Rule 1: (13/2, lift 5.1)
        KualitasKerja = RENDAH
        KerjasamaTim = RENDAH
        -> class TIDAK [0.800]

Rule 2: (16/4, lift 4.6)
        Kepemimpinan = RENDAH
        KerjasamaTim = RENDAH
        -> class TIDAK [0.722]

Rule 3: (40/1, lift 1.1)
        Kepemimpinan = TINGGI
        -> class YA [0.952]

Rule 4: (123/7, lift 1.1)
        KualitasKerja in {SEDANG, TINGGI}
        Kepemimpinan = SEDANG
        -> class YA [0.936]

Rule 5: (195/15, lift 1.1)
        KerjasamaTim in {SEDANG, TINGGI}
        -> class YA [0.919]

Default class: YA
```

**Gambar 3. Aturan Keputusan Algoritma C5.0**

Berdasarkan gambar tersebut didapatkan aturan keputusan berikut.

1. Jika Kualitas Kerja rendah dan Kerjasama Tim rendah, maka mitra tidak direkomendasikan.
2. Jika Kepemimpinan rendah dan Kerjasama Tim rendah, maka mitra tidak direkomendasikan.
3. Jika Kepemimpinan tinggi maka mitra direkomendasikan.
4. Jika Kualitas Kerja sedang dan tinggi serta Kepemimpinan sedang maka mitra direkomendasikan.
5. Jika Kerjasama sedang dan tinggi maka mitra direkomendasikan.

Selanjutnya menghitung prediksi pada *data testing* untuk ditentukan nilai akurasi.

treec5predict		TIDAK	YA
TIDAK	4	1	
YA	2	46	

**Gambar 4. Prediksi Data Uji (*Data Testing*) pada Pohon Keputusan Algoritma C5.0**

Perhitungan akurasi dilakukan dengan menggunakan rumus 5, sehingga didapatkan nilai akurasi sebesar 94,33%.

## SIMPULAN

Berdasarkan hasil dan pembahasan tersebut didapatkan kesimpulan bahwa Algoritma C5.0 dapat membentuk pohon keputusan secara cepat dan dapat membantu BPS Kota Salatiga dalam menentukan mitra yang layak untuk direkomendasikan sehingga dapat dijadikan pertimbangan apabila mitra yang ada pada daftar tersebut mendaftar kembali di kegiatan BPS Kota Salatiga selanjutnya. Selanjutnya hasil akurasi yang diperoleh dari data penilaian mitra BPS Kota Salatiga sebesar 94,33%.

## DAFTAR RUJUKAN

- Amrieh, E. A., Hamtini, T., & Aljarah, I. (2016) Mining Educational Data to Predict Student's Academic Performance using Ensemble Methods. *International Journal of Database Theory and Application*, pp. 119-136. <http://dx.doi.org/10.14257/ijdt.2016.9.8.13>
- Badan Pusat Statistik. *Tentang Profil BPS*. Retrieved from <https://www.bps.go.id/menu/1/tentangprofil-bps.html>
- Dyah A. W. D., Cholissodin, I., & Sutrisno. (2019). Klasifikasi Penyimpangan Tumbuh Kembang Anak Menggunakan Algoritme C5.0. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer Universitas Brawijaya*, 10258-10265. <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/6631>
- Hadi, F. (2017). Penerapan Data Mining Dalam Menganalisa Pemberian Pinjaman Dengan Menggunakan Metode Algoritma C5.0 (Studi Kasus: Koperasi Jasa Keuangan Syariah Kelurahan Lambung Bukik. *Jurnal KomTekInfo*, pp. 214–223. <http://journal.uad.ac.id/index.php/JITEKI/article/view/11416>
- Jiawei, H., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques (3rd ed)*. Morgan Kaufmann, Burlington.
- Kastawan, P. W., Wiharta, D. M., & Sudarma, I. M. (2018). Implementasi Algoritma C5.0 pada Penilaian Kinerja Pegawai Negeri Sipil. *Majalah Ilmiah Teknologi Elektro*, 17(1), 371-376. <https://ojs.unud.ac.id/index.php/JTE/article/view/41634>
- Larose, & Daniel T. (2005) *Discovering Knowledge in Data: An Introduction to Data Mining*, New Jersey : John Willey Sons, Inc.
- Moss L. T., & Adelman, S. (2000) *Data Warehouse Project Mangement*. Canada: Addison Wesley.
- Patil, P. N., Lathi, R. , & Chitre, V. (2012). "Comparison of C5 . 0 CART Classification algorithms using pruning technique" *Int. J. Eng. Res. Technol.*, pp. 1–5.
- Salsabila C. E., Budi N., & Fetty T. A.(2021). Implementasi Fase Boosting pada Algoritma C5.0 dalam Menentukan Prestasi Akademik Siswa. *Seminar Nasional Informatika Bela Negara (SANTIKA)*, Vol. 2.

- Sammut G. I., Claude, & Webb. (2011). *“Encyclopedia of Machine Learning and Data Mining”*. Springer.
- Sokolova, M., & Lapalme, G.(2009) *A systematic analysis of performance measures for classification tasks*, Inf. Process. Manag.,pg. 427–437.
- Suntoro, J. (2018). *Data Mining Algoritme dan Implementasi Menggunakan Pemrograman PHP*.
- Suntoro, J., & Indah, C. N. (2017). Average Weight Information Gain Untuk Menangani Data Berdimensi. *Jurnal Buana Informatika.*, pp. 131–140. <https://ojs.uajy.ac.id/index.php/jbi/article/view/1315>
- Sutoyo, I. (2018). F. T. Informasi, U. Bina, S. Informatika, *Implementasi Algoritma Decision Tree Untuk Klasifikasi*, pp. 217–224, 2018.
- Tajrin, Iron, P. H., Ria, A. S., & Husien, C. (2021). Implementasi Algoritma C5.0 dalam Keputusan Pemberian Beasiswa di SMA Negeri 1 Adiankoting Kabupaten Tapanuli Utara. *Jurnal Sistem Informasi Komputama (JSIK)*.
- Witten, Ian H., & Frank, E. H. (2011). *Data Mining Practical Machine Learning Tools and Techniques* (Vol. 1). Elsevier and Morgan Kaufmann.