

## Comparison of C4.5 and C5.0 Algorithm Classification Tree Models for Analysis of Factors Affecting Auction

### Perbandingan Model Pohon Klasifikasi Algoritma C4.5 dan C5.0 untuk Analisis Faktor yang Mempengaruhi Keberhasilan Lelang\*

Mohammad Fajri<sup>1</sup>, Iut Tri Utami<sup>2</sup>, Muh. Maruf<sup>3</sup>

<sup>1,2</sup>Statistics Studies Program, Tadulako University, Indonesia  
‡corresponding author: [m.fajri@untad.ac.id](mailto:m.fajri@untad.ac.id)

Copyright © 2022 Mohammad Fajri, Iut tri Utami, and Muh. Maruf. This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

#### Abstract

Auction in Indonesia is carried out by the Office of State Assets and Auction Services (KPKNL). Goods auctioned at KPKNL are quite diverse including land, wood, inventory, vehicles, and other goods. However, not all of the items auctioned were sold. Because not a few items have been auctioned but no one has made an offer. The Purpose of this study is to compare two classification methods, C4.5 and C5.0 algorithm and to determine which items were successfully auctioned with those that did not and its factors. The methods that used were comparing the classification tree C4.5 algorithm and C5.0 algorithm with cross validation. From the results of the comparison of the two methods, it was found that the C5.0 Algorithm method was rated better than the C4.5 algorithm in classifying the auction results with an accuracy of 96.43% and 92.86% respectively. In this case, C5.0 has a higher precision than C4.5.

**Keywords:** auction, C4.5 Algorithm, C5.0 Algorithm, decision tree.

## 1. Pendahuluan

Lelang di Indonesia sudah dikenal cukup lama, muncul saat masa pemerintahan Hindia Belanda dengan mengeluarkan aturan lelang (*Vendu Reglement Staatsblad* tahun 1908 nomor 189) dan peraturan tersebut masih berlaku hingga sekarang. Lelang secara histori berasal dari kata *auktion* (latin), merupakan penaikan harga secara berangsur-angsur. Lelang telah diketahui ada sejak 450 tahun sebelum Masehi oleh para ahli. Lelang merupakan suatu pemasaran barang kepada publik dengan harga barang yang ditawarkan kepada penawar naik setiap saat (Lusk & Shogren, 2007; Salim, 2014).

Pelaksanaan lelang di Indonesia dilakukan oleh Kantor Pelayanan Kekayaan Negara dan Lelang berdasarkan Permenkeu Nomor 170/PMK/01/2012 Tentang Organisasi dan Tata Kerja Instansi Vertikal Direktorat Jenderal Kekayaan Negara Pasal 30. Barang yang dilelang di KPKNL ada berbagai macam jenis diantaranya tanah, kayu, inventaris, kendaraan dan barang lainnya. Meskipun barang-barang tersebut dilelang, namun tidak semua dari barang lelang laku terjual untuk umum. Salah satu cara untuk mengklasifikasi keberhasilan lelang barang tersebut adalah dengan menggunakan pohon klasifikasi (Chen & Chung, 2015).

Pohon klasifikasi merupakan suatu metodologi inovatif untuk menganalisis data dengan ukuran besar melalui proses pemilahan biner yang digunakan untuk menggambarkan hubungan antar variabel yaitu variabel dependen dan independen (Brieman et al., 1984). Ada beberapa algoritma pohon klasifikasi yang bisa digunakan diantaranya algoritma C4.5 dan algoritma C5.0.

(Ente et al., 2020) menggunakan algoritma C4.5 dalam mengklasifikasi penyakit diabetes mellitus dengan menggunakan teknik *Cross validation* untuk menentukan model terbaik yang dihasilkan. Teknik *Cross validation* akan mengkombinasikan beberapa kelompok data pengamatan yang telah dipartisi ke dalam dua bagian, yakni data untuk membangun model dan data untuk verifikasi model (Blockeel & Struyf, 2002).

Pada penelitian ini akan menggunakan dua algoritma, yaitu algoritma C4.5 dan C5.0, serta juga akan menggunakan teknik *Cross validation* sebagai penentu model terbaik sehingga kombinasi yang dihasilkan juga lebih banyak. Hasil yang diharapkan dari penelitian ini ialah untuk memodelkan kasus keberhasilan lelang di KPKNL Palu menggunakan algoritma C4.5 dan algoritma C5.0 serta mengetahui model terbaik yang dihasilkan.

## 2. Metode Penelitian

Klasifikasi adalah suatu metode untuk menentukan model yang menegaskan atau memperjelas perbedaan konsep maupun kelas data, dimana bertujuan untuk dapat memprediksikan kelas dari suatu amatan yang kelasnya tak diketahui (Tan et al., 2006). Lazimnya, klasifikasi dilakukan dalam dua jenjang, yakni proses *training* dan *testing*. Pada proses *training*, prosedur klasifikasi menganalisis data *train* untuk mendapatkan suatu model. Kemudian model akan diuji pada proses *testing*, dengan menggunakannya untuk mengklasifikasi kelas dari amatan baru dalam proses pengambilan keputusan (Quinlan, 2014).

Data yang dieproleh Dalam penelitian adalah sekunder Kantor Pelayanan Kekayaan Negara Dan Lelang (KPKNL) Kota Palu, dengan variabel yang digunakan yaitu :

- a. Keterangan Penjualan Lelang (Y)  
Barang dikatakan laku bila ada yang melakukan penawaran. Variabel ini dibagi menjadi 2 kategori yaitu:  
0 : Laku  
1 : Tidak Laku
- b. Waktu Pelelangan  
Waktu pelelangan adalah saat pertama kali barang tersebut dimuat untuk dilakukan pelelangan. Variabel ini dibagi menjadi 3 kategori yaitu:  
0 : Awal Bulan  
1: Tengah Bulan  
2 : Akhir Bulan
- c. Jenis Lelang  
Jenis lelang didasarkan pada pemohon lelang. Variabel ini dibagi menjadi 4 kategori :  
0 : eksekusi  
1 : non-eksekusi wajib BMD  
2 : non-eksekusi wajib BMN  
3 : sukarela
- d. Teknis Lelang  
Teknis lelang adalah cara pelaksanaan lelang. Variable ini dibagi menjadi 3 kategori yaitu:  
0 : conventional  
1 : e-conventional  
2 : e-auction
- e. Jenis Barang  
Variabel ini dibagi menjadi 5 kategori yaitu:  
0: Tanah  
1: Kayu  
2: Inventaris  
3: Kendaraan  
4: Barang lain
- f. Nilai Limit  
Harga awal dari barang yang dijual. Variabel ini berbentuk numerik.

## 2.1 Algoritma C4.5

Sebelum membentuk model, terlebih dahulu data dibentuk menjadi dua bagian yaitu data *training* dan data *testing*. Data *training* digunakan dalam rangka membangun model, adapun data *testing* digunakan untuk menentukan akurasi. Adapun tahapan algoritma C4.5 yaitu:

1. Membentuk data *training*.
2. Menghitung nilai entropy dengan rumus:

$$Entropy(S) = -\sum_{i=1}^n p_i \log_2 p_i \quad (1)$$

Keterangan:

$S$  = himpunan Kasus

$n$  = jumlah partisi  $S$

$p_i$  = proporsi  $S_i$  terhadap  $S$

3. Menghitung nilai Gain.
4. Rumus yang digunakan untuk menghitung Gain sebagai berikut:

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

Keterangan:

$S$  = himpunan kasus

$A$  = atribut

$n$  = jumlah partisi atribut  $A$

$S_i$  = proporsi  $S_i$  terhadap  $S$

$S$  = jumlah kasus dalam  $S$

5. Kemudian menghitung nilai Split Info:

$$Split\ Information(S, A) = - \sum_{i=1}^n \frac{|S_i|}{|S|} \log_2 \frac{|S_i|}{|S|} \quad (3)$$

Keterangan:

$S$  = himpunan kasus

$A$  = atribut

$S_i$  = proporsi  $S_i$  terhadap  $S$

$S$  = jumlah kasus dalam  $S$

6. Setelah mendapatkan nilai Gain dan Split Info, lalu mencari nilai Gain Ratio dengan rumus sebagai berikut:

$$Gain\ Ratio(S, A) = \frac{Gain(S, A)}{SplitInformation(S, A)} \quad (4)$$

Keterangan:

$Gain$  = information gain pada atribut  $A$

$Split\ Information$  = Split Information pada atribut  $A$

7. Nilai Gain Ratio tertinggi akan digunakan sebagai atribut akar. Dengan itu akan terbentuk pohon keputusan sebagai node 1.
8. Ulangi proses ke-2 sampai semua cabang memiliki kelas yang sama.
9. Maka akan terbentuk pohon keputusan.
10. Dari pohon keputusan yang terbentuk maka dapat ditentukan *Rule-Rule*. (Lakshmi et al., 2013)

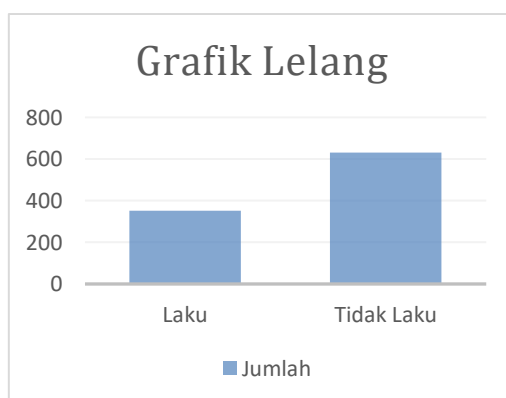
## 2.2 Algoritma C5.0

Algoritma C5.0 adalah penyempurnaan dari algoritma ID3 dan C4.5. Dalam proses pembentukan pohon keputusan nilai *information gain* yang paling tinggi akan dipilih sebagai root bagi node berikutnya. Algoritma ini diawali dengan semua data dijadikan akar dari pohon keputusan dan atribut yang dipilih akan menjadi pembagi bagi sampel tersebut (Han et al., 2011). Formula ukuran atribut dapat dihitung sesuai persamaan entrophy pada algoritma C4.5 (1), sedangkan untuk mendapatkan nilai *information gain* dapat menggunakan persamaan *information gain* pada algoritma C4.5 (2). Atribut A dengan *information gain* tertinggi, akan dijadikan root node. Algoritma C5.0 memiliki akurasi yang baik, dapat digunakan pada berbagai macam tipe data serta mempunyai kelebihan dalam memproses data dengan jumlah yang besar (Bujlow et al., 2012)

## 3. Hasil dan Pembahasan

### 3.1 Karakteristik Penjualan Lelang Tanah

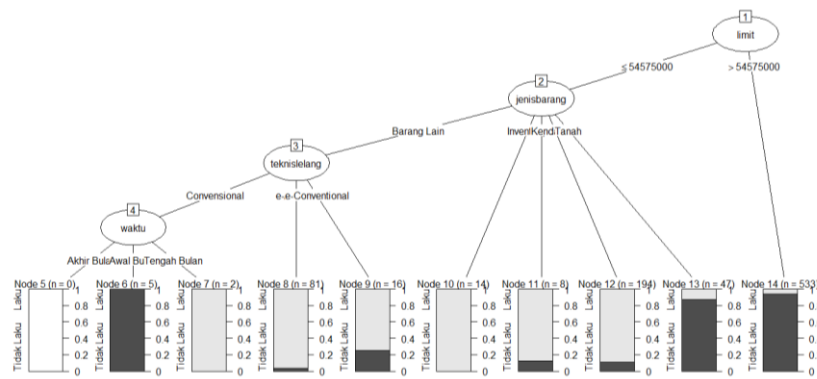
Pada Gambar 1, diketahui bahwa total barang yang laku saat dilakukan pelelangan di Sulawesi Tengah tahun 2018 berjumlah 353 sementara total barang yang tidak laku terjual sebanyak 631. Dengan jumlah total keseluruhan lelang yang dilakukan sebanyak 984.



Gambar 1: Penjualan lelang di sulawesi tengah tahun 2018

### 3.2 Analisis Pohon Klasifikasi Dengan Algoritma C4.5

Model pohon klasifikasi dengan algoritma C4.5 menghasilkan ketepatan klasifikasi sebesar 92,86%, dengan hasil sebagai berikut:



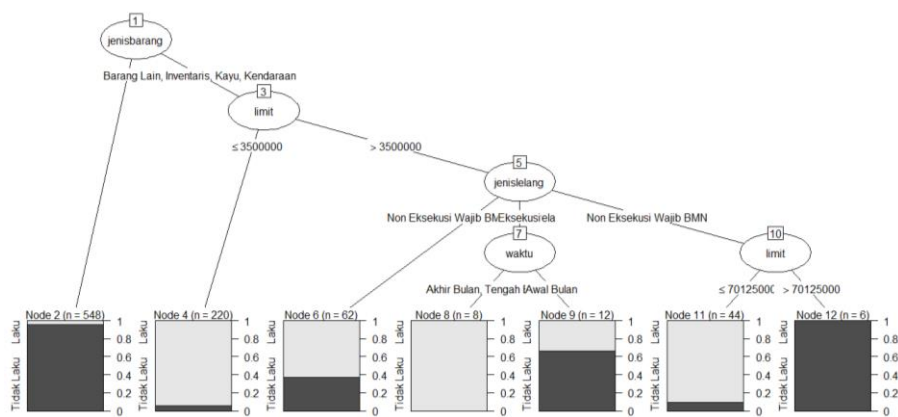
Gambar 2: Model pohon klasifikasi dengan algoritma C4.5

Interpretasi :

- Jika nilai limit  $> 54.575.000$  maka kemungkinan barang tersebut tidak laku sebesar 94 %
- Jika nilai limit  $\leq 54.575.000$ , dan jenis barangnya adalah tanah maka kemungkinan barang tersebut tidak laku sebesar 88%
- Jika nilai limit  $\leq 54.575.000$ , dan jenis barangnya adalah kendaraan maka kemungkinan barang tersebut laku sebesar 90%
- Jika nilai limit  $\leq 54.575.000$ , dan jenis barangnya adalah kayu maka kemungkinan barang tersebut laku sebesar 88%
- Jika nilai limit  $\leq 54.575.000$ , dan jenis barangnya adalah inventaris maka kemungkinan barang tersebut laku sebesar 100%
- Jika nilai limit  $\leq 54.575.000$ , jenis barangnya adalah barang lain, dan teknis lelangnya adalah e-convensional, maka kemungkinan barang tersebut laku sebesar 80 %
- Jika nilai limit  $\leq 54.575.000$ , jenis barangnya adalah barang lain, dan teknis lelangnya adalah e-auction, maka kemungkinan barang tersebut laku sebesar 96 %
- Jika nilai limit  $\leq 54.575.000$ , jenis barangnya adalah barang lain, teknis lelangnya adalah konvensional, dan waktu pelelangannya tengah bulan, maka kemungkinan barang tersebut laku sebesar 100 %
- Jika nilai limit  $\leq 54.575.000$ , jenis barangnya adalah barang lain, teknis lelangnya adalah konvensional, dan waktu pelelangannya awal bulan, maka kemungkinan barang tersebut tidak laku sebesar 100 %
- Jika nilai limit  $\leq 54.575.000$ , jenis barangnya adalah barang lain, teknis lelangnya adalah konvensional, dan waktu pelelangannya akhir bulan, maka kemungkinan barang tersebut laku sebesar 0 %.

### 3.3 Analisis Pohon Klasifikasi Dengan Algoritma C5.0

Model pohon klasifikasi dengan algoritma C5.0 menghasilkan ketepatan klasifikasi sebesar 96,43% dengan hasil sebagai berikut:



Gambar 3: Model pohon klasifikasi dengan algoritma C5.0

Interpretasi :

- Jika jenis barangnya adalah barang lain, inventaris, kayu dan kendaraan, nilai limit  $> 3.500.000$ , jenis lelangnya adalah non eksekusi wajib BMN dan limit  $> 70.125.000$  maka kemungkinan barang tersebut tidak laku sebesar 100%.
- Jika jenis barangnya adalah barang lain, inventaris, kayu dan kendaraan, nilai limit  $> 3.500.000$ , jenis lelangnya adalah non eksekusi wajib BMN dan limit  $\leq 70.125.000$  maka kemungkinan barang tersebut laku sebesar 91% .
- Jika jenis barangnya adalah barang lain, inventaris, kayu dan kendaraan, nilai limit  $> 3.500.000$ , jenis lelangnya adalah eksekusi dan waktu lelangnya adalah awal bulan maka kemungkinan barang tersebut tidak laku sebesar 67% .
- Jika jenis barangnya adalah barang lain, inventaris, kayu dan kendaraan, nilai limit  $> 3.500.000$ , jenis lelangnya adalah eksekusi dan waktu lelangnya adalah tengah bulan dan akhir bulan maka kemungkinan barang tersebut laku sebesar 100% .
- Jika jenis barangnya adalah barang lain, inventaris, kayu dan kendaraan, nilai limit  $> 3.500.000$  dan jenis lelangnya adalah non eksekusi BMD dan sukarela, maka kemungkinan barang tersebut laku sebesar 63%.
- Jika jenis barangnya adalah barang lain, inventaris, kayu dan kendaraan dan nilai limit  $\leq 3.500.000$ , maka kemungkinan barang tersebut laku sebesar 94%.
- Jika jenis barangnya adalah tanah maka kemungkinan barang tersebut tidak laku sebesar 95%.

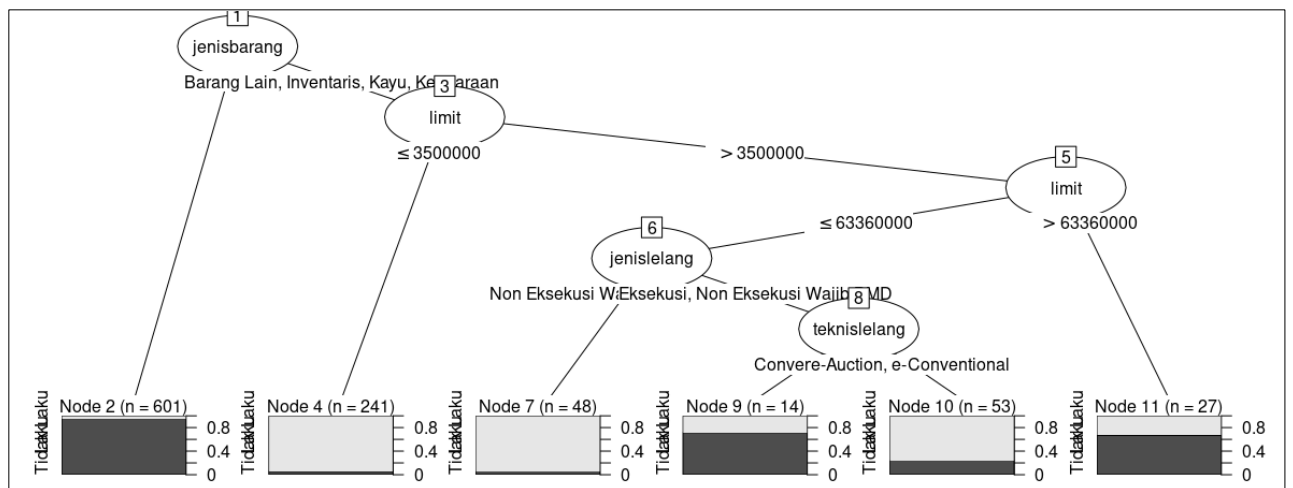
### 3.4 Perbandingan Ketepatan Klasifikasi *Cross Validation* Algoritma C4.5 dan Algoritma C5.0

Adapun ketepatan klasifikasi hasil dari *cross validation* algoritma C4.5 dan algoritma C5.0 secara lengkap dapat dilihat pada tabel berikut.

Tabel 1 :Ketepatan Klasifikasi Cross Validation

Algoritma	Ketepatan Klasifikasi (%)										
	Fold1	Fold2	Fold3	Fold4	Fold5	Fold6	Fold7	Fold8	Fold9	Fold10	
C4.5	92,86	94	89,79	93,87	94,89	92,92	94,89	90,81	91,83	91,83	92,92
C5.0	96,43	92	92,85	93,87	94,89	85,85	100	89,79	94,89	89,79	95,95

Dapat dilihat pada Tabel 1 di atas bahwa nilai akurasi pada tiap fold untuk masing-masing algoritma C4.5 maupun algoritma C5.0 bervariasi. Secara umum cross validation algoritma C5.0 lebih baik dibandingkan cross validation algoritma C4.5 dengan nilai akurasi pada fold ke enam sebesar 100%. Adapun hasil pohon klasifikasi yang terbentuk pada fold ke enam tersebut disajikan pada Gambar 4.



Gambar 4 : Model pohon klasifikasi dengan algoritma C5.0 Cross validation

Interpretasi :

- Jika jenis barang adalah tanah maka kemungkinan barang tersebut tidak laku sebesar 95 %
- Jika jenis barang adalah barang lain atau inventaris atau kayu atau kendaraan dengan limit  $\leq 3.500.000$  maka kemungkinan barang tersebut laku sebesar 94%
- Jika jenis barang adalah barang lain atau inventaris atau kayu atau kendaraan dengan limit  $> 3.500.000, > 63.360.000$  maka kemungkinan barang tersebut tidak laku sebesar 75%
- Jika jenis barang adalah barang lain atau inventaris atau kayu atau kendaraan dengan limit  $\leq 63.360.000$ , jenis lelang Non Eksekusi Wajib BMN, Sukarela maka kemungkinan barang tersebut laku sebesar 94 %
- Jika jenis barang adalah barang lain atau inventaris atau kayu atau kendaraan dengan  $\leq 63.360.000$ , jenis lelang eksekusi, Non Eksekusi Wajib BMD dan teknis lelang konvensional maka kemungkinan barang tersebut tidak laku sebesar 77 %
- Jika jenis barang adalah barang lain atau inventaris atau kayu atau kendaraan dengan  $\leq 63.360.000$ , jenis lelang eksekusi, Non Eksekusi Wajib BMD dan teknis lelang e-Auction, e-Convensional maka kemungkinan barang tersebut laku sebesar 80 %



#### 4. Simpulan

Dari hasil perbandingan kedua metode, didapatkan bahwa setiap metode memiliki model pohon klasifikasi yang relatif berbeda serta metode Algoritma C5.0 *Cross validation* dinilai lebih baik daripada algoritma C4.5, algoritma C5.0 dan algoritma C4.5 *Cross validation* dalam mengklasifikasikan hasil pelelangan dengan mempunyai akurasi sebesar 100%, lebih besar dibandingkan akurasi dari beberapa metode lainnya.

**Ucapan terima Kasih.** Ucapan terima kasih disampaikan kepada Pimpinan Fakultas MIPA Universitas Tadulako atas bantuan pendanaan sehingga penelitian ini dapat terlaksana dengan baik.

#### Daftar Pustaka

- Blockeel, H., & Struyf, J. (2002). *Efficient Algorithms for Decision Tree Cross-validation*. 30.
- Brieman, L., Friedman, J., & Olshen, R. A. (1984). *Classification and Regression Trees*. CRC press.
- Bujlow, T., Riaz, T., & Pedersen, J. M. (2012). A method for classification of network traffic based on C5. 0 Machine Learning Algorithm. *International Conference on Computing, Networking and Communications*, 237–241.
- Chen, C. C., & Chung, M.-C. (2015). Predicting the success of group buying auctions via classification. *Knowledge-Based Systems*, 89, 627–640. <https://doi.org/10.1016/j.knosys.2015.09.009>
- Ente, D. R., Thamrin, S. A., Arifin, S., Kuswanto, H., & Andreza, A. (2020). Klasifikasi Faktor-Faktor Penyebab Penyakit Diabetes Melitus di Rumah Sakit Unhas Menggunakan Algoritma c4.5. *Indonesian Journal of Statistics and Its Applications*, 4(1), 80–88.
- Han, J., Pei, J., & Kamber, M. (2011). *Data Mining: Concepts and Techniques*. Elsevier.
- Lakshmi, T. M., Martin, A., Begum, R. M., & Venkatesan, V. P. (2013). An Analysis on Performance of Decision Tree Algorithms Using Student's Qualitative Data. *International Journal of Modern Education and Computer Science*, 5(5), 18.
- Lusk, J., & Shogren, J. F. (2007). *Experimental auctions: Methods and applications in economic and marketing research*. Cambridge University Press.

Quinlan, J. R. (2014). *C4. 5: Programs for Machine Learning*. Elsevier.

Salim, H. S. (2014). *Perkembangan Hukum Jaminan di Indonesia. Divisi Buku Perguruan*. PT RajaGrafindo Persada.

Tan, P. N., Steinbach, M., & Kumar, V. (2006). *Introduction to Data Mining*. Pearson.