

# DETEKSI TRANSAKSI PENCUCIAN UANG DENGAN ALGORITMA KLASIFIKASI C4.5

Kahfi Heryandi Suradiradja

<sup>1</sup>Dosen Teknik Informatika STMIK Eresha  
Jl. Raya Puspitek Serpong No. 10 Tangerang Selatan Banten

Email : [kheryandi@gmail.com](mailto:kheryandi@gmail.com)

## ABSTRACT

*Money laundering is an activity to disguise the origins of money or assets obtained as if it came from a legitimate activity is not derived from the proceeds of crime. For the prevention of these activities, the government through the Pusat Pelaporan dan Analisis Transaksi Keuangan (PPATK) require the banking institutions and other financial services in Indonesia to report any suspicious transactions. To assist the banking and financial services, needed a model aids in detecting money laundering transactions by financial daily transactions that deviate from the profile, characteristics, or customary pattern of customer transactions.*

*The model algorithm used in this study is the classification Decision Tree by analyzing the optimal parameters in order to get a good accurate in detecting transactions of suspected money laundering. Parameter or criterion used are C4.5 (Gain Ratio), ID3 (Information Gain) and the Gini Index and the parameter Pre-Pruning and Pruning the decision tree formation. The research data is categorized imbalanced dataset of 93,365 records of data are used only 37 suspects, so the measurement model used is the AUC.*

*The results of this study showed the best accuracy results in detecting suspected money laundering transactions using C4.5 classification algorithm with Pre-Pruning and Pruning has 0.936 accuracy.*

*Keywords Money Laundering, Decision Tree, C4.5, imbalanced dataset, AUC*

## ABSTRAK

Kegiatan pencucian uang merupakan kegiatan menyamarkan asal-usul uang atau harta kekayaan yang diperoleh seolah-olah berasal dari kegiatan yang sah bukan berasal dari hasil tindak pidana. Sebagai usaha pencegahan kegiatan tersebut, pemerintah melalui Pusat Pelaporan dan Analisis Transaksi Keuangan (PPATK) mewajibkan kepada institusi perbankan dan jasa keuangan lainnya di Indonesia agar melaporkan setiap transaksi yang mencurigakan. Guna membantu perbankan dan jasa keuangan, diperlukan model alat bantu dalam mendeteksi transaksi pencucian uang berdasarkan transaksi harian keuangan yang menyimpang dari profil, karakteristik, atau kebiasaan pola transaksi dari nasabah.

Model algoritma yang digunakan dalam penelitian ini adalah klasifikasi Decision Tree dengan menganalisa parameter yang optimal guna mendapatkan akurat yang baik dalam mendeteksi transaksi-transaksi yang dicurigai sebagai pencucian uang. Parameter atau criterion yang digunakan adalah C4.5 (Ratio Gain), ID3 (Information Gain) dan Gini Index serta parameter Pre-Pruning dan Pruning dalam pembentukan decision tree. Data penelitian termasuk kategori imbalanced dataset dari 93.365 record yang digunakan hanya 37 data dicurigai, sehingga pengukuran model yang digunakan adalah AUC.

Hasil penelitian ini menunjukkan hasil akurasi terbaik dalam mendeteksi transaksi yang dicurigai pencucian uang menggunakan algoritma klasifikasi C4.5 dengan Pre- Pruning dan Pruning memiliki akurasi 0.936.

Kata kunci :

Pencucian Uang, Decision Tree, C4.5, imbalanced dataset, AUC

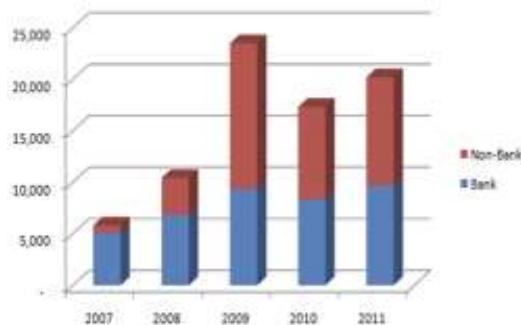
## 1. PENDAHULUAN

- Pencucian uang adalah suatu proses atau perbuatan yang bertujuan untuk menyembunyikan atau menyamarkan asal-usul uang atau harta kekayaan yang diperoleh dari hasil tindak pidana yang kemudian diubah menjadi harta kekayaan yang seolah-olah berasal dari kegiatan yang sah.
- Deteksi transaksi yang dicurigai sebagai pencucian uang sangat diperlukan

sederhana dan cepat. Secara umum tingkat akurasi [SEP] tinggi.

- Kelemahan model C4.5 dalam akurasi pengukuran jika menggunakan [SEP] imbalance dataset.
- Penelitian ini menggunakan metode algoritma C4.5 dengan kriteria atribut [SEP] data penelitian berbeda dan imbalance dataset.

• LTKM PJK 2007-2011



• Hasil Analisa PPATK 2011

DUGAAN TINDAK PIDANA	JUMLAH
Korupsi	178
Narkotika	2
Pembalakan Liar	3
Penggelapan	9
Penggelapan Pajak	3
Di Bidang Perbankan	1
Penipuan	4
Penyuapan	25
Perjudian	2
Teroris	3
Tidak Teridentifikasi / dll	67

sebagai langkah pencegahan, diatur dalam UU No 25 Tahun 2003.

- Model klasifikasi C4.5 banyak digunakan dalam penelitian pencucian uang di bidang finansial seperti:
  - 1) Penelitian (Wang & Yang, 2007: p283-286) untuk anti pencucian uang dengan menggunakan metode Decision Tree.
  - 2) Research anti pencucian uang yang dilakukan oleh (Ju & Zheng, 2009 : p525-528) menggunakan algoritma klasifikasi privacy-preserve. [SEP]
  - 3) Penelitian yang dilakukan oleh (da Rocha & Timóteo, 2010 : p162- 169) guna melihat apakah Decision Tree dapat digunakan untuk membuat keputusan, terutama dalam penanganan data yang besar.
- Kelebihan menggunakan model C4.5 ini:
  - 1) Hasil pohon keputusannya sederhana dan mudah dimengerti
  - 2) Menurut (Han & Kamber, 2006 : p292) proses learning dan klasifikasi algoritma C4.5

## 2. METODE PENELITIAN

Tahapan Penelitian:

- Business Understanding
- Data Understanding
- Data Preparation
- Modeling
- Evaluation
- Deployment

### 1. Business Understanding

- Pendeteksian transaksi nasabah yang dicurigai melihat karakteristik transaksi diluar kebiasaan dari history transaksi nasabah tersebut diatas Rp. 500.000.000,-
- Keakuratan deteksi terhadap transaksi tersebut sangat diperlukan [SEP] sebagai bahan laporan pihak bank ke PPATK.

### 2. Data Understanding

- Data Januari s.d Maret 2012.
- Kolom data history yang dibutuhkan penelitian:
  - a) ID Rekening
  - b) Golongan Pemilik
  - c) Kode Transaksi
  - d) Tanggal Transaksi
  - e) Frekuensi Debet
  - f) Frekuensi Kredit

- g) Jumlah Total Debet
- h) Jumlah Total Kredit
- i) Rata-Rata Transaksi
- j) Standar Deviasi
- k) Suspect

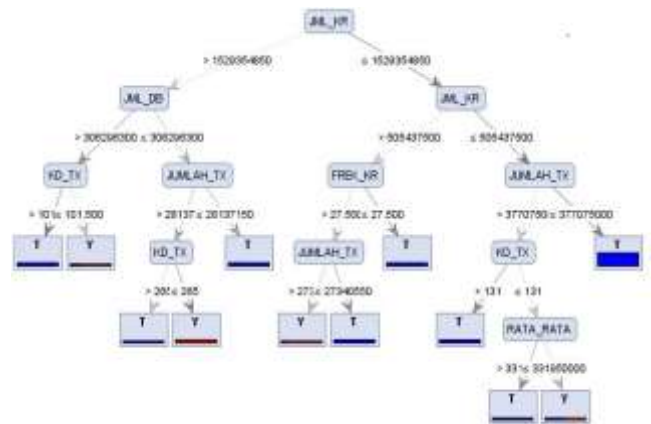
### 3. Data Preparation

- Data Cleaning membersihkan dari reversal transaksi
- Data Integration menggabungkan transaksi dari data histori transaksi tabungan dan giro.
- Data reduction jumlah data yang digunakan hanya difokuskan kepada data – data transaksi setoran tunai, tarikan tunai, pemindahan bukuan dimana data – data transaksi yang tidak diperlukan seperti biaya administrasi, posting bunga, posting jasa giro, pajak jasa giro dan lain- lain akan dihapus.
- Total data 93.365 dan yang dicurigai 37 data

## 3. HASIL PENELITIAN DAN PEMBAHASAN

### 1. Modeling Decision Tree

- Gain Ratio Pruning & Pre-Pruning
- Gain Ratio No Pruning
- Gain Ratio No Pre-Pruning
- Gain Ratio No Pre-Pruning dan No Pruning
- Information Gain Pruning & Pre-Pruning
- Information Gain No Pruning
- Information Gain No Pre-Pruning
- Information Gain No Pruning dan No Pre-Pruning
- Gini Index Pruning & Pre-Pruning
- Gini Index No Pre Pruning
- Gini Index No Pruning
- Gini Index No Pre-Pruning dan No Pruning



### Classification Rule:

```

IF JML_KR > 1529354850 THEN
  IF JML_DB > 306296300 THEN
    IF KD_TX > 101,500 THEN TIDAK
    IF KD_TX ≤ 101,500 THEN YA
  IF JML_DB ≤ 306296300
    IF JUMLAH_TX > 28137150
      IF KD_TX > 265 THEN TIDAK
      IF KD_TX ≤ 265 THEN YA
    IF JUMLAH_TX ≤ 28137150 THEN TIDAK
  IF JML_KR ≤ 1529354850 THEN
    IF JML_KR > 505437500 THEN
      IF FREK_KR > 27,500 THEN
        IF JUMLAH_TX > 27340550 THEN YA
        IF JUMLAH_TX ≤ 27340550 THEN TIDAK
      IF FREK_KR ≤ 27,500 THEN TIDAK
    IF JML_KR ≤ 505437500 THEN
      IF JUMLAH_TX > 377075000 THEN
        IF KD_TX > 131 THEN TIDAK
        IF KD_TX ≤ 131 THEN
          IF RATA_RATA > 331950000 THEN TIDAK
          IF RATA_RATA ≤ 331950000 THEN YA
      IF JUMLAH_TX ≤ 377075000 THEN TIDAK
  
```

### 2. Evaluation

Pada tahap evaluasi ini adalah tahap menentukan akurasi pengujian, dimana dalam tahap pengujiannya melihat hasil akurasi dari proses klasifikasi algoritma C4.5 menggunakan metode evaluasi dengan *ROC Curve*.

Pada Tabel *Confusion Matrix* dibawah ini dapat dilihat bahwa jumlah *True Positive* (TP) adalah 27, *False Negative* (FN) adalah 10, untuk *False Positive* (FP) adalah 10, dan untuk *True Negative* adalah 93318.

Tabel *Confusion Matrix Gain Ratio*

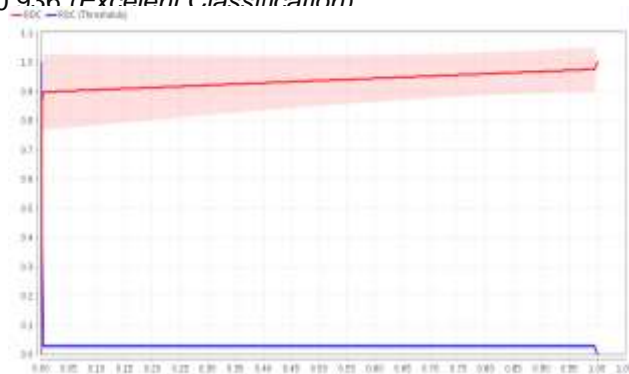
		Predictive Class	
		TIDAK	YA
Actual Class	TIDAK	93318	10
	YA	10	27

Maka untuk nilai akurasi adalah sebagai berikut:

$$\text{Accuracy (Overall success rate)} = \frac{27 + 93318}{27 + 93318 + 10 + 10}$$

Karena dipengaruhi oleh *imbalance* data maka hasil perhitungan berdasarkan *Confusion Matrix* tersebut memiliki nilai akurasi 99.98%.

Pada Gambar AUC Grafik ROC Decision Tree Gain Ratio berikut menunjukkan grafik ROC dengan nilai AUC (*Area Under Curve*) sebesar 0.936 (*Excelent Classification*)



Gambar AUC Grafik ROC *Decision Tree Gain Ratio*

Hasil dari pengujian akan menentukan nilai akurasi AUC (*Area Under Curve*) dari keempat perbandingan eksperimen setting parameter algoritma yang meliputi *Criterion*, *Pre-Pruning* dan *Pruning* yang paling baik agar dapat diterapkan untuk mendeteksi transaksi mencurigakan pencucian uang.

Tabel Summary Hasil Perbandingan Penelitian

Criterion	Pre Pruning	Pruning	Measurement	
			Accuracy	AUC
Gain Ratio	Ya	Tidak	99.98%	0.907
Gain Ratio	Tidak	Ya	99.98%	0.500
Gain Ratio	Tidak	Tidak	99.98%	0.500
<b>Gain Ratio</b>	<b>Ya</b>	<b>Ya</b>	<b>99.98%</b>	<b>0.936</b>
Information Gain	Ya	Tidak	99.99%	0.704
Information Gain	Tidak	Ya	99.99%	0.500

Information Gain	Tidak	Tidak	99.99%	0.500
Information Gain	Ya	Ya	99.99%	0.865
Gini Index	Ya	Tidak	99.98%	0.500
Gini Index	Tidak	Ya	99.97%	0.907
Gini Index	Tidak	Tidak	99.98%	0.500
Gini Index	Ya	Ya	99.97%	0.907

Pada tabel diatas adalah summary hasil perbandingan penelitian algoritma C4.5 dengan *Criterion Gain Ratio* dan *Information Gain* baik menggunakan *Pruning* ataupun *Pre-Pruning*. Dari hasil penelitian tersebut, hasil pengukuran accuracy selalu memiliki nilai diatas 99% dikarenakan data penelitian adalah *imbalance*, sehingga untuk mendapatkan nilai accuracy dengan persamaan

$$\text{Accuracy (Overall success rate)} = \frac{TP + TN}{TP + TN + FP + FN}$$

Selisih TP (Actual & Predictive Ya) + TN (Actual & Predictive Tidak) dengan seluruh nilai pengujian adalah sedikit sehingga menghasilkan nilai diatas 99%. Karena itulah untuk mendapatkan nilai accuracy tersebut menggunakan AUC. Seperti disebutkan dalam (Weng & Poon, 2006 : p2) bahwa ROC merupakan salah satu pengukuran datamining yang paling relevan untuk *imbalance* dataset.

Untuk mendapatkan nilai AUC apabila data transaksi perbankan tersebut menggunakan model klasifikasi C4.5 dengan parameter tanpa *Pre-Pruning* dan tanpa *Pre-Pruning* selalu menghasilkan 0.500 atau *failure*. Nilai yang terbaik dari aspek akurasi nilai AUC adalah menggunakan *Criterion Gain Ratio* menggunakan *Pre-Pruning* dan menggunakan *Pruning*, yaitu memiliki nilai 0.936

#### 4. KESIMPULAN DAN SARAN

##### 1. Kesimpulan

- Akurasi terbaik untuk C4.5 menggunakan *Pre-Pruning* dan *Pruning* yakni dengan nilai AUC 0.936 (*Excelent Classification*). [SEP]
- Pengaruh penerapan *Criterion Gain Ratio* pada C4.5 dengan menggunakan *Pruning* dan *Pre-Pruning* sebagai

parameter meningkatkan akurasi algoritma C4.5

## 2. Saran

- Membandingkan tingkat akurasi dengan model algoritma lain seperti Naive Bayes atau Support Vector Machine.
- Membandingkan hasil akurasi dengan data yang bukan imbalance
- Membandingkan pengukuran akurasi dengan model evaluasi lain seperti Precision and Recall atau Cost-sensitive Measure

## DAFTAR PUSTAKA

Alpaydm, E. , 2010, "*Introduction to Machine Learning*", London, England: The MIT Press.

Bramer, M., 2007, "*Principles of Data Mining*", London: Springer.

Chawla, N. , 2003, "*C4.5 and Imbalanced Data sets: Investigating the effect of sampling method, probabilistic estimate, and decision tree structure*", Washington DC: ICML.

da Rocha, B. C., & Timóteo, R. d. , 2010, "Identifying Bank Frauds Using CRISP-DM And Decision Tress. *International journal of computer science & information Technology (IJCSIT)* , 2 Issue 5, 162-169"

Gorunescu, F., 2011 , "*Data Mining Concepts, Model and Techniques*", Berlin: Spinger,

Han, J., & Kamber, M., 2006, "*Data Mining: Concepts and Techniques*", Morgan Kaufmann.

Ju, C., & Zheng, L., 2009, "Research on Suspicious Financial Transactions Recognition Based on Privacy-Preserving of Classification Algorithm. *Education Technology and Computer Science*, 2, pp. 525 - 528.", Hangzhou.

Larose, D. T., 2005, "*Discovering Knowledge In Data An Introduction to Data Mining*", New Jersey: John Wiley & Sons, Inc.

Maimon, O., & Rokach, L., 2010, "*Data Mining and Knowledge Discovery*", New York: Springer.

Pemerintah, 2003, "*Undang Undang No.25*"

PPATK, 2011, "*Laporan Tahunan PPATK*", Jakarta.

PPATK, 2003, "*Pedoman Umum Pencegahan dan Pemberantasan Tindak Pidana Pencucian Uang bagi Penyedia Jasa Keuangan*", Jakarta.

Wang, S.-N., & Yang, J.-G., 2007, " A Money Laundering Risk Evaluation Method Based on Decision Tree. *Machine Learning and Cybernetics*. 1, pp. 283 – 286", Hong Kong: IEEE Xplore.

Weng, C., & Poon, J., 2006, "*A New Evaluation Measure for Imbalanced Datasets*", Sidney: University of Sidney.

Witten, I. H., Frank, E., & Hall, M. A., 2011 , "*Data mining practical machine learning tools and technique*", Morgan Kaufmann Publisher.

## BIODATA PENULIS

Kahfi Heryandi Suradiradja memperoleh gelar Ahli Madya (A.Md) pada Jurusan Teknik Komputer Politeknik Negeri Bandung Tahun 1997, memperoleh gelar Sarjana Sain (S.Si) pada Jurusan Matematika Bidang Ilmu Komputer Universitas Padjadjaran Bandung Tahun 1998. Dan memperoleh gelar Magister Komputer (M.Kom) pada Program Pasca Sarjana Magister Komputer STMIK Eresha, 2012. Saat ini menjadi Dosen Tetap di STMIK Eresha.