

**DETEKSI TRANSAKSI PENCUCIAN UANG DENGAN  
ALGORITMA KLASIFIKASI C4.5**

**T E S I S**

**Diajukan Sebagai Salah Satu Syarat Untuk Menyelesaikan  
Program Strata Dua (S2) Magister Komputer**



**OLEH :**

**KAHFI HERYANDI SURADIRADJA**

**3612091078**

**PROGRAM STUDI TEKNIK INFORMATIKA  
PROGRAM PASCA SARJANA (S2) MAGISTER KOMPUTER  
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER ERESHA  
JAKARTA  
2012**

## **PERSETUJUAN TESIS**

Nama : Kahfi Heryandi Suradiradja  
NPM : 3612091078  
Konsentrasi : Software Engineering  
Judul tesis : Deteksi transaksi Pencucian Uang dengan Algoritma  
Klasifikasi C4.5

Telah disetujui untuk disidangkan pada Sidang Tesis pada Program Pasca Sarjana  
(S2) Magister Komputer, Program Studi Teknik Informatika Sekolah Tinggi  
Manajemen Informatika dan Komputer Eresha.

Jakarta, 17 September 2012

Pembimbing Utama

Pembimbing Pendamping

(Romi Satria Wahono, B.Eng., M.Eng) (Dr. Rufman Iman Akbar E., SE, MM, M.Kom)

Mengetahui :

Direktur

Ketua

Program Pasca Sarjana

(Ir. Damsiruddin Siregar, MMT) (Dr. Rufman Iman Akbar E., SE, MM, M.Kom)

## **PENGESAHAN TESIS**

Nama : Kahfi Heryandi Suradiradja  
NPM : 3612091078  
Konsentrasi : Software Engineering  
Judul tesis : Deteksi transaksi Pencucian Uang dengan Algoritma  
Klasifikasi C4.5

Telah disidangkan dan dinyatakan Lulus Sidang Tesis pada Program Pasca Sarjana (S2) Magister Komputer, Program Studi Teknik Informatika Sekolah Tinggi Manajemen Informatika dan Komputer Eresha pada tanggal 27 September 2012

Nama Penguji	Tanda Tangan
Romi Satria Wahono, B.Eng., M.Eng (Ketua)	.....
Ir. Damsiruddin Siregar, MMT (Sekretaris)	.....
Didik Setiyadi, M.Kom (Anggota)	.....

Mengetahui:  
Direktur  
Program Pasca Sarjana

(Dr. Rufman Iman Akbar E., SE, MM, M.Kom)

## PERNYATAAN KEASLIAN TESIS

Nama : Kahfi Heryandi Suradiradja  
NPM : 3612091078  
Konsentrasi : Software Engineering  
Judul tesis : Deteksi Transaksi Pencucian Uang Dengan Algoritma  
Klasifikasi C4.5

Dengan ini saya menyatakan bahwa dalam Tesis ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Jakarta, 27 September 2012

***Materai 6000***

( Kahfi Heryandi Suradiradja )

Kahfi Heryandi Suradiradja, 3612091078

Deteksi Transaksi Pencucian Uang Dengan Algoritma Klasifikasi C4.5; dibawah bimbingan Romi Satrio Wahono, B.Eng, M.Eng dan Dr.Rufman Iman Akbar E., MM,M.Kom.

67 + x hal / 25 tabel / 30 gambar / 15 pustaka ( 2003 – 2011 )

## 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

Kahfi Heryandi Suradiradja, 3612091078

Detection of Money Laundering Transactions With C4.5 classification algorithm;  
under the guidance of Romi Satrio Wahono, B.Eng, M.Eng dan Dr.Rufman Iman  
Akbar E., MM,M.Kom.

67 + x pages / 25 tables / 30 pictures / 15 literatures ( 2003 – 2011 )

## **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

## KATA PENGANTAR

Dengan memanjatkan puji syukur kehadiran Allah SWT yang telah melimpahkan segala rahmat dan hidayahnya kepada penulis, sehingga tersusunlah Tesis yang berjudul **“Deteksi Transaksi Pencucian Uang Dengan Algoritma Klasifikasi C4.5”**.

Tesis tersebut melengkapi salah satu persyaratan yang diajukan dalam rangka menempuh ujian akhir untuk memperoleh gelar Magister Komputer (M.Kom.) pada Program Pasca Sarjana (S2), Program Studi Teknik Informatika di Sekolah Tinggi Manajemen Informatika dan Komputer Eresha

Penulis sungguh sangat menyadari, bahwa penulisan Tesis ini tidak akan terwujud tanpa adanya dukungan dan bantuan dari berbagai pihak. Maka, dalam kesempatan ini penulis menghaturkan penghargaan dan ucapan terima kasih yang sebesar - besarnya kepada yang terhormat:

1. Bapak Romi Satrio Wahono, B.Eng, M.Eng selaku pembimbing utama tesis yang telah menyediakan waktu, pikiran dan tenaga dalam membimbing penulis dalam menyelesaikan tesis ini.
2. Keluarga yang telah memberikan dukungan kepada penulis.
3. Bantuan semua pihak yang tak bisa saya sebutkan satu persatu.

Akhir kata penulis mohon maaf atas kekeliruan dan kesalahan yang terdapat dalam Tesis ini dan berharap semoga Tesis ini dapat memberikan manfaat bagi khasanah pengetahuan Teknologi Informasi di Indonesia.

Kahfi Heryandi Suradiradja

Penulis

# DAFTAR ISI

	Hal
Persetujuan Tesis.....	i
Pengesahan Tesis .....	ii
Pernyataan Keaslian Tesis.....	iii
Abstrak .....	iv
Abstract .....	v
Kata Pengantar .....	vi
Daftar Isi.....	vii
Daftar Gambar.....	ix
Daftar Tabel .....	x
<b>BAB 1 PENDAHULUAN .....</b>	<b>1</b>
1.1 Latar Belakang Masalah .....	1
1.2 Permasalahan Penelitian .....	4
1.2.1 Identifikasi Masalah .....	4
1.2.2 Ruang Lingkup Masalah .....	4
1.2.3 Rumusan Masalah .....	4
1.3 Tujuan dan Manfaat Penelitian .....	4
1.4 Sistematika Penulisan .....	5
<b>BAB 2 LANDASAN TEORI DAN KERANGKA PEMIKIRAN .....</b>	<b>6</b>
2.1 Tinjauan Studi .....	6
2.2 Landasan Teori .....	8
2.2.1 Algoritma C4.5 Dalam Klasifikasi Data Mining.....	8
2.2.2 Metode Evaluasi .....	18
2.2.3 Pencucian Uang .....	20
2.3 Kerangka Pemikiran .....	23
<b>BAB 3 METODE PENELITIAN.....</b>	<b>24</b>
3.1 Analisa Kebutuhan.....	24
3.2 Perancangan Penelitian .....	24
3.3 Teknik Analisis .....	25
3.3.1 Business Understanding .....	25
3.3.2 Data Understanding .....	25



3.3.3 Data Preparation .....	26
3.3.4 Modeling .....	29
3.3.5 Evaluation.....	48
BAB 4 HASIL DAN PEMBAHASAN.....	49
4.1 Hasil .....	49
4.1.1 Evaluasi Confusion Matrix dan AUC Pada Algoritma C4.5..	49
4.1.2 Evaluasi Confusion Matrix dan AUC C4. 5 Gain Ratio No Pruning .....	50
4.1.3 Evaluasi Decision Tree Gain Ratio No Pre-Pruning .....	51
4.1.4 Evaluasi Decision Tree Gain Ratio No Pre-Pruning dan No Pruning .....	52
4.1.5 Evaluasi Decision Tree Information Gain.....	53
4.1.6 Evaluasi Decision Tree Information Gain No Pruning .....	54
4.1.7 Evaluasi Decision Tree Information Gain No Pre-Pruning....	55
4.1.8 Evaluasi Decision Tree Information Gain No Pruning No Pre-Pruning .....	56
4.1.9 Evaluasi Decision Tree Gini Index .....	57
4.1.10 Evaluasi Decision Tree Gini Index No Pre-Pruning .....	58
4.1.11 Evaluasi Decision Tree Gini Index No Pruning.....	59
4.1.12 Evaluasi Decision Tree Gini Index No Pruning No Pre-Pruning .....	60
4.2 Pembahasan .....	61
4.3 Implikasi Penelitian .....	62
BAB 5 KESIMPULAN DAN SARAN.....	63
5.1 Kesimpulan .....	63
5.2 Saran .....	63
DAFTAR PUSTAKA .....	64
DAFTAR RIWAYAT HIDUP.....	66

## DAFTAR GAMBAR

	Hal
Gambar 1.1 Penerimaan Transaksi Keuangan Mencurigakan PPATK .....	1
Gambar 2.1 Bentuk penyajian (a) classification rules, (b) neural network dan (c) <i>Decision Tree</i> , (Gambar diolah dari (Han & Kamber, 2006, p.25) ).....	11
Gambar 2.2 <i>Decision Tree</i> Contoh Kasus Node 1 .....	15
Gambar 2.3 <i>Decision Tree</i> Contoh Kasus Node 1.1 .....	16
Gambar 2.4 Final <i>Decision Tree</i> Contoh Kasus.....	17
Gambar 2.5 Grafik Receiver Operating Curve .....	20
Gambar 2.6 Kerangka Penelitian Deteksi Transaksi Pencucian Uang .....	23
Gambar 3.1 Decision Tree Gain Ratio .....	30
Gambar 3.2 Decision Tree Gain Ratio No Pruning .....	31
Gambar 3.3 Decision Tree Gain Ratio No Pre-Pruning.....	33
Gambar 3.4 Decision Tree Gain Ratio No Pre-Pruning dan No Pruning .....	35
Gambar 3.5 Decision Tree Information Gain .....	37
Gambar 3.6 Decision Tree Information Gain No Pruning.....	38
Gambar 3.7 Decision Tree Information Gain No Pre-Pruning .....	39
Gambar 3.8 Decision Tree Information Gain No Pruning dan No Pre-Pruning.....	40
Gambar 3.9 Decision Tree Gini Index .....	41
Gambar 3.10 Decision Tree Gini Index No Pre Pruning .....	43
Gambar 3.11 Decision Tree Gini Index No Pruning .....	45
Gambar 3.12 Decision Tree Gini Index Criterion No Pre-Pruning No Pruning.....	47
Gambar 4.1 AUC Grafik ROC <i>Decision Tree Gain Ratio</i> .....	49
Gambar 4.2 AUC Grafik ROC <i>Decision Tree Gain Ratio No Pruning</i> .....	50
Gambar 4.3 AUC Grafik ROC <i>Decision Tree Gain Ratio Ratio No Pre-Pruning</i> .....	51
Gambar 4.4 AUC Grafik ROC <i>Decision Tree Gain Ratio Ratio No Pre-Pruning</i> <i>dan No Pruning</i> .....	52
Gambar 4.5 AUC Grafik ROC <i>Decision Tree Information Gain Criterion</i> .....	53
Gambar 4.6 AUC Grafik ROC <i>Decision Tree Information Gain No Pruning</i> .....	54
Gambar 4.7 AUC Grafik ROC <i>Decision Tree Information Gain No Pre-Pruning</i> .....	55
Gambar 4.8 AUC Grafik ROC <i>Decision Tree Information Gain No Pruning &amp; No</i> <i>Pre Pruning</i> .....	56
Gambar 4.9 AUC Grafik ROC <i>Decision Tree Gini Index Criterion</i> .....	57
Gambar 4.10 AUC Grafik ROC <i>Decision Tree Gini Index No Pruning</i> .....	58
Gambar 4.11 AUC Grafik ROC <i>Decision Tree Gini Index No Pruning</i> .....	59
Gambar 4.12 AUC Grafik ROC <i>Decision Tree Gini Index No Pruning &amp; No Pre</i> <i>Pruning</i> .....	60

## DAFTAR TABEL

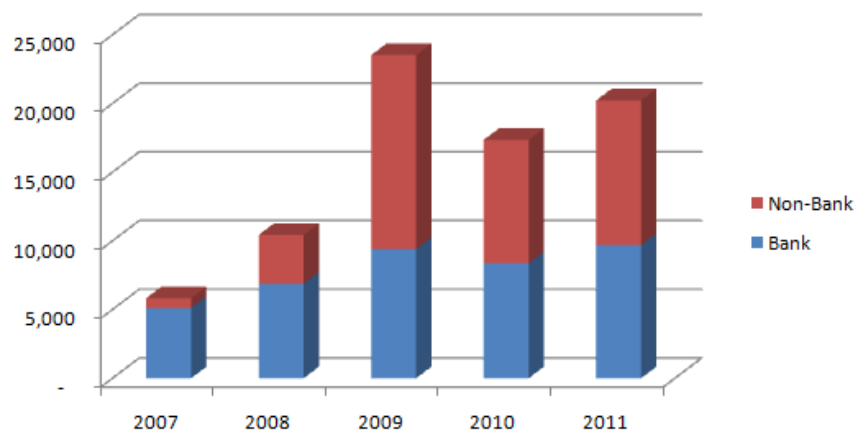
	Hal
Tabel 1.1 Hasil Analisis PPATK, 2011 .....	2
Tabel 2.1 Tabel Penelitian Anti Money Laundering Dengan <i>Decision Tree</i> .....	7
Tabel 2.2 Data Testing Kasus Keputusan Membeli Komputer di Pameran .....	13
Tabel 2.3 Perhitungan Node 1 Contoh Kasus .....	14
Tabel 2.4 Perhitungan Node 1.1 Contoh Kasus .....	15
Tabel 2.5 Perhitungan Node 1.1.1 Contoh Kasus .....	17
Tabel 2.6 Perbedaan Hasil Dua Class Prediction ( Tabel diolah dari: (Witten, Frank, & Hall, 2011 : p.164)).....	18
Tabel 2.7 Atribut dan Sifat Nilai Dataset Penelitian.....	22
Tabel 3.1 Spesifikasi Hardware dan Software .....	24
Tabel 3.2 Data Transaksi Bank Januari - Maret 2012.....	27
Tabel 3.3 Kandidat Split Algoritma C4.5 .....	28
Tabel 4.1 <i>Confusion Matrix Criterion Gain Ratio</i> .....	49
Tabel 4.2 <i>Confusion Matrix C4.5 Gain Ratio No Pruning</i> .....	50
Tabel 4.3 <i>Confusion Matrix C4.5 Gain Ratio No Pre-Pruning</i> .....	51
Tabel 4.4 <i>Confusion Matrix C4.5 Gain Ratio No Pre-Pruning dan No Pruning</i> .....	52
Tabel 4.5 <i>Confusion Matrix C4.5 Information Gain Criterion</i> .....	53
Tabel 4.6 <i>Confusion Matrix C4.5 Information Gain Criterion No Pruning</i> .....	54
Tabel 4.7 <i>Confusion Matrix C4.5 Information Gain Criterion No Pre-Pruning</i> .....	55
Tabel 4.8 <i>Confusion Matrix C4.5 Information Gain No Pruning &amp; No Pre Pruning</i> .....	56
Tabel 4.9 <i>Confusion Matrix C4.5 Gini Index Criterion</i> .....	57
Tabel 4.10 <i>Confusion Matrix C4.5 Gini Index Criterion No Pre-Pruning</i> .....	58
Tabel 4.11 <i>Confusion Matrix C4.5 Gini Index Criterion No Pruning</i> .....	59
Tabel 4.12 <i>Confusion Matrix C4.5 Gini Index Criterion No Pruning &amp; No Pre Pruning</i> .....	60
Tabel 4.13 Summary Hasil Perbandingan Penelitian.....	61

# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang Masalah

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 sesuai (Undang Undang No 25, 2003 : p3). Tindak pidana yang menjadi pemicu terjadinya pencucian uang meliputi dari hasil korupsi, penyuapan, penyelundupan barang/ tenaga kerja/ imigran, perbankan, narkoba, psikotropika, perdagangan budak/ wanita/ anak/ senjata gelap, penculikan, terorisme, pencurian, penggelapan, dan penipuan (PPATK, Pedoman Umum Pencegahan dan Pemberantasan Tindak Pidana Pencucian Uang bagi Penyedia Jasa Keuangan, 2003 : p4).



**Gambar 1.1**  
**Perkembangan Penerimaan Lap. Transaksi Keuangan Mencurigakan PPATK**

Diambil sumber dari (PPATK, Laporan Tahunan, 2011 : p12), dalam Gambar 1.1 menunjukkan grafik peningkatan dari tahun ke tahun untuk laporan transaksi keuangan mencurigakan (LTKM) tahun 2011 dari perusahaan keuangan perbankan berdasarkan kumulatif LTKM hingga 31 Desember 2011 sebesar 54.7% . Dimana LTKM jumlah terlapor berdasarkan hasil analisis proaktif menurut dugaan tindak pidana selama tahun 2011 sebanyak 297 kasus (PPATK, Laporan Tahunan, 2011 : p17). Detil hasil analisa tersebut dapat dilihat pada Tabel 1.1.

**Tabel 1.1**  
**Hasil Analisis PPATK, 2011**

DUGAAN TINDAK PIDANA	JUMLAH	%
Korupsi	178	60%
Narkotika	2	1%
Pembalakan Liar	3	1%
Penggelapan	9	3%
Penggelapan Pajak	3	1%
Di Bidang Perbankan	1	0%
Penipuan	4	1%
Penyuapan	25	8%
Perjudian	2	1%
Teroris	3	1%
Tidak Teridentifikasi / dll	67	23%

Pendeteksian transaksi yang dicurigai pencucian uang sangat diperlukan karena sebagai pencegahan terhadap pencucian uang sebagaimana diatur dalam undang undang tersebut, pemerintah melalui Pusat Pelaporan dan Analisis Transaksi Keuangan (PPATK) mewajibkan kepada institusi perbankan dan jasa keuangan lainnya agar melaporkan setiap transaksi yang mencurigakan. Laporan Transaksi Keuangan Mencurigakan (LTKM) dari institusi perbankan kepada PPATK tidak hanya melihat dari besarnya transaksi diatas 500 juta rupiah namun juga berdasarkan transaksi-transaksi yang terjadi diluar kebiasaan perilaku nasabah (Undang Undang No 25, 2003 : p8). Karena itulah dibutuhkan sebuah model datamining dan algoritma untuk mendeteksi transaksi mencurigakan yang akan dilaporkan bank kepada PPATK.

Model algoritma yang banyak digunakan dalam penelitian ini adalah model klasifikasi *decision tree* C4.5. Kelebihan menggunakan model klasifikasi *decision tree* ini adalah hasil pohon keputusannya sederhana dan mudah dimengerti. Menurut (Han & Kamber, 2006 : p292) proses learning dan klasifikasi pada algoritma *Decision Tree* sederhana dan cepat. Secara umum model algoritma klasifikasi *Decision Tree* mempunyai tingkat akurasi yang tinggi. Namun masalahnya menurut (Chawla, 2003) klasifikasi *Decision tree* biasanya dievaluasi dengan akurasi prediksi yang mempertimbangkan semua kesalahan yang sama akan tetapi akurasi prediksi mungkin tidak sesuai jika jika menggunakan *imbalance* dataset.

Beberapa penggunaan metode klasifikasi *Decision Tree* C4.5 telah dilakukan beberapa orang untuk bidang finansial seperti:

- a. Penelitian (Wang & Yang, 2007 : p283-286) untuk anti pencucian uang dengan menggunakan metode *Decision Tree* C4.5.
- b. Research anti pencucian uang yang dilakukan oleh (Ju & Zheng, 2009 : p525-528) menggunakan algoritma klasifikasi *privacy-preserve*.
- c. Penelitian yang dilakukan oleh (da Rocha & Timóteo, 2010 : p162-169) guna melihat apakah *Decision Tree* C4.5 dapat digunakan untuk membuat keputusan, terutama dalam penanganan data yang besar.

Karena penelitian pencucian uang pernah dilakukan beberapa peneliti sebelumnya banyak menggunakan klasifikasi *decision tree* C4.5, maka pada penelitian ini akan diukur akurasi untuk mendeteksi transaksi mencurigakan pencucian uang menggunakan algoritma klasifikasi *decision tree* yang tepat dengan kriteria atribut data yang digunakan berbeda dan termasuk ke dalam *imbalanced* dataset. Penerapan pemilihan parameter *criterion*, *Pre-Pruning* dan *Pruning* yang sesuai serta optimal, sehingga akan menghasilkan prediksi yang lebih akurat.

## **1.2 Permasalahan Penelitian**

### *1.2.1 Identifikasi Masalah*

Berdasarkan uraian pada latar belakang diatas, yang menjadi permasalahan adalah:

- a. Diperlukan model guna membantu pendeteksian transaksi keuangan mencurigakan pencucian uang berdasarkan transaksi harian keuangan yang menyimpang dari profil, karakteristik, atau kebiasaan pola transaksi dari nasabah.
- b. Proses algoritma klasifikasi sederhana dan cepat namun memiliki kelemahan akurasi pada *imbalance* data set yang besar sehingga diperlukan pemilihan parameter yang optimal guna mendapatkan prediksi yang akurat.

### *1.2.2 Ruang Lingkup Masalah*

Terkait dengan Laporan Transaksi Keuangan Mencurigakan (LTKM) yang dibuat oleh institusi perbankan harus dilaporkan kepada PPATK, dalam Tesis ini ruang lingkup yang diteliti hanya pada deteksi transaksi yang dicurigai pencucian uang saja.

### *1.2.3 Rumusan Masalah*

Yang menjadi rumusan masalah dalam penelitian ini adalah:

- a. Bagaimana akurasi algoritma klasifikasi C4.5 apabila digunakan mendeteksi transaksi mencurigakan pencucian uang?
- b. Bagaimana pengaruh penerapan *Pruning* dan *Pre-Pruning* sebagai parameter guna meningkatkan akurasi algoritma?

## **1.3 Tujuan dan Manfaat Penelitian**

Tujuan dari penelitian ini adalah:

- a. Penerapan metode algoritma klasifikasi C4.5 ini untuk mendeteksi transaksi mencurigakan pencucian uang.
- b. Penerapan parameter *Criterion*, *Pruning* dan *Pre-Pruning* berpengaruh pada nilai akurasi deteksi transaksi mencurigakan pencucian uang.

Beberapa manfaat dari penelitian ini adalah:

- a. Manfaat secara teoritis memberikan sumbangan dengan metode Algoritma C4.5 apat membantu perbankan dalam mendeteksi transaksi mencurigakan pencucian uang.
- b. Manfaat praktis yang diharapkan untuk memberikan kemudahan bagi dunia perbankan dalam membuat laporan pencucian uang.

Manfaat kebijakannya yakni metode klasifikasi C4.5 menjadi sebuah pilihan untuk mendeteksi transaksi mencurigakan pencucian uang.

#### **1.4 Sistematika Penulisan**

Dalam sistematika penulisan disajikan terdiri dari lima bab, serta masing-masing bab memiliki beberapa sub-sub bab yang terdiri dari:

- a. Bab I Pendahuluan  
Bab ini berisi tentang latar belakang permasalahan, identifikasi masalah, tujuan dan manfaat penelitian dan sistematika penulisan.
- b. Bab II Landasan Teori dan Kerangka Pemikiran  
Bab ini berisi tentang tinjauan studi, landasan teori penelitian dan kerangka pemikiran.
- c. Bab III Metode Penelitian  
Bab ini berisi tentang metode penelitian yang membahas tentang analisa kebutuhan, perancangan penelitian, teknik analisis, business understanding, data understanding, data preparation, modeling dan evaluation.
- d. Bab IV Hasil dan Pembahasan  
Bab ini berisi tentang hasil, pembahasan serta implikasi penelitian.
- e. Bab V Kesimpulan dan Saran  
Bab ini berisi kesimpulan dari hasil penelitian serta saran untuk pengembangan sistem lebih lanjut.



## BAB 2

### LANDASAN TEORI DAN KERANGKA PEMIKIRAN

#### 2.1 Tinjauan Studi

Metode klasifikasi *Decision Tree* yang digunakan dalam penelitian (Wang & Yang, 2007 : p283-286) untuk membuat aturan penentuan risiko pencucian uang dengan profil pelanggan bank komersial di Cina. Latar belakang penelitian ini ialah system anti pencucian uang untuk bank komersial kecil dan menengah di Cina sangat diperlukan. Penelitian ini menggunakan algoritma *Decision Tree* dengan ID3 atau *information gain criterion* tanpa menggunakan *pre-pruning* dan *pruning*. Data yang digunakan dalam penelitian ini sebanyak 160 ribu nasabah menggunakan empat atribut klasifikasi yakni *Industry*, *Location*, *Business Size* dan *Product Transaction Risk*. Hasilnya dari 28 sample (21 data training dan 7 data testing) terdapat hanya 1 record data yang tidak seduai dengan actual, sehingga akurasi mencapai 96.4%. Secara total dari data penelitian, ditemukan 12% nasabah memiliki resiko tinggi *money laundering*.

Penelitian oleh (Ju & Zheng, 2009 : p525-528) ini mempelajari bagaimana untuk mengidentifikasi transaksi keuangan yang mencurigakan di departemen perdagangan dan pengawasan. Penelitian ini menggunakan algoritma klasifikasi *Decision Tree* C4.5 dengan *Pruning* *privacy-preserve*. Atribut yang digunakan meliputi Total Volume Transaction, Tax dan Age serta label atribut suspicious. Hasil penelitiannya menunjukkan bahwa tree yang terbentuk berdasarkan metode algoritma klasifikasi digabungkan dengan *privacy-preserve* ini dapat digunakan untuk mengidentifikasi transaksi pencucian uang bagi lembaga keuangan.

Penelitian yang dilakukan oleh (da Rocha & Timóteo, 2010 : p162-169) ini difokuskan untuk melihat bagaimana *Decision Tree* C4.5 ini dapat membantu dalam proses pengambilan keputusan analisis dalam mengidentifikasi penipuan transaksi bank. Metode penelitian dengan CRISP-DM menggunakan real data penipuan dan data transaksi bank sebanyak 17.753 record transaksi. Tidak disebutkan atribut yang digunakan, peneliti hanya menyebutkan data hanya digunakan untuk kepentingan akademik dan telah menyamarkan data nama –

nama nasabah bank tersebut. Algoritma C4.5 (*Gain Ratio Criterion*) yang digunakan tidak menggunakan *pre-pruning* maupun *pruning*. Hasil penelitian menghasilkan akurasi 95.6% ini dapat disimpulkan bahwa *Decision Tree* C4.5 dengan digunakan untuk membuat keputusan identifikasi penipuan transaksi bank.

Dari ketiga penelitian sebelumnya dapat disimpulkan pada Tabel 2.1

**Tabel 2.1**  
**Tabel Penelitian Anti Money Laundering Dengan Decision Tree**

Peneliti	Tahun	Model	Attribute	Hasil
Wang & Yang	2007	Klasifikasi <i>Decision Tree</i> ID3 tanpa Pre-Pruning & Pruning	<ul style="list-style-type: none"> <li>• Industry</li> <li>• Location</li> <li>• Business Size</li> <li>• Product Trx Risk</li> </ul>	Akurasi 96.4%
Ju & Zheng	2009	Klasifikasi <i>Decision Tree</i> C4.5 dengan pruning	<ul style="list-style-type: none"> <li>• Total Volume</li> <li>• Tax</li> <li>• Age</li> <li>• Suspicious</li> </ul>	<i>Decision Tree</i> C4.5 dpt digunakan
Rocha & Timóteo	2010	Klasifikasi <i>Decision Tree</i> C4.5 tanpa Pre-Pruning & Pruning	Tidak disebutkan dengan alasan kerahasiaan data.	Akurasi 95.6%

Pada penelitian ini yang akan dilakukan membandingkan hasil akurasi terbaik dari beberapa variasi model algoritma *Decision Tree* dengan variasi criterion C4.5 (*Gain Ratio*), ID3 (*Information Gain*) dan *Gini Index*. Serta variasi penggunaan *Pruning* dan *Pre-Pruning* dalam pembentukan *Decision Tree*.

## 2.2 Landasan Teori

### 2.2.1 Algoritma C4.5 Dalam Klasifikasi Data Mining

Kegiatan memprediksi dengan menggunakan data history dalam sebuah datawarehouse, adalah salah satu kategori dari beberapa kegiatan data mining atau istilah lainnya sering disebut knowledge discovery atau juga pattern recognition. Disebut knowledge discovery karena mencari pengetahuan yang tersembunyi dari sebuah datawarehouse. Demikian pula istilah pattern recognition karena pengetahuan yang dicari tersebut berbentuk suatu pola-pola yang perlu digali lagi lebih dalam lagi dari sebuah datawarehouse. Pemberian istilah data mining lebih populer untuk menggali pengetahuan dalam datawarehouse, dimana Gartner Group dalam (Larose, 2005 : p2) menyebutkan ”Data mining adalah proses menemukan korelasi baru yang bermakna, pola dan tren dengan memilah-milah data besar yang tersimpan dalam repositori, menggunakan teknologi pengenalan pola serta teknik-teknik statistik dan matematika.”

Berdasarkan (Larose, 2005 : p11), kegiatan - kegiatan dalam data mining dikategorikan sebagai berikut:

- a. Deskripsi (Description)
- b. Estimasi (Estimation)
- c. Prediksi (Prediction)
- d. Klasifikasi (Classification)
- e. Pengelompokan (Clustering)
- f. Asosiasi (Association)

#### a. Description

Model data mining harus setransparan mungkin. Artinya, hasil dari model data mining harus menerangkan pola dan penjelasan. Deskripsi berkualitas tinggi sering dapat dicapai dengan eksplorasi analisis data dengan metode grafis untuk mengeksplorasi data dalam mencari pola dan tren. Contoh untuk mendeskripsi sekumpulan data besar dengan deskripsi grafis seperti diagram titik, histogram, lokasi (mean, median, modus, kuartil dan persentil) dan keberagaman (range, varians dan standar deviasi).

b. Estimasi

Estimasi mirip dengan klasifikasi, namun pada estimasi variabel target tersebut lebih mengarah kepada nilai numerik daripada kategori. Model yang dibangun menggunakan record yang menyediakan nilai estimasi variabel target. Kemudian, untuk pengamatan estimasi nilai dari variabel target yang dibuat didasarkan pada nilai-nilai prediksi. Contoh dalam memperkirakan pembacaan tekanan darah sistolik dari seorang pasien rumah sakit, berdasarkan usia pasien, jenis kelamin, indeks massa tubuh, dan kadar natrium. Hubungan antara tekanan darah sistolik dan variabel prediksi dalam training set akan memberikan sebuah model estimasi.

c. Prediksi

Prediksi ini mirip dengan klasifikasi dan estimasi, namun untuk prediksi, nilai hasil akan ada di masa depan. Metode dan teknik yang digunakan untuk klasifikasi dan estimasi dapat juga digunakan untuk prediksi, termasuk metode statistik tradisional, regresi linear dan regresi berganda, serta data mining dan metode penemuan pengetahuan seperti neural network, *Decision Tree*, K-Nearest Neighbors. Contoh prediksi dalam bisnis dan penelitian yaitu prediksi harga saham tiga bulan ke depan, prediksi persentase kenaikan kematian lalu lintas tahun depan jika batas kecepatan meningkat atau memprediksi pemenang seri baseball musim gugur tahun ini, berdasarkan perbandingan statistik tim.

d. Klasifikasi

Dalam klasifikasi memiliki variabel target kategori. Contohnya, penggolongan pendapatan dibagi dalam kategori pendapatan tinggi, berpenghasilan menengah, dan pendapatan rendah berdasarkan karakteristik yang terkait usia, jenis kelamin, dan pekerjaan. Algoritma akan memeriksa data set yang berisi variabel prediktor dan target variable klasifikasi pendapatan. Learning algoritma pada kombinasi variabel yang berhubungan dengan kategori pendapatan pada kumpulan data / training set yang akan mengklasifikasi ke rekor baru. Misalnya hasil dari rekor baru tersebut, seorang profesor wanita berusia 63 tahun harus diklasifikasikan dalam klasifikasi berpenghasilan tinggi. Contoh klasifikasi lainnya dalam bisnis dan penelitian yakni