

Metode C45 Untuk Mengklarifikasi Pelanggan Perusahaan Telekomunikasi Seluler

Akhmad Zulkifli

Program Studi Sistem Informasi, STMIK Hang Tuah
Jl. Mustafa Sari No. 05 Tangkerang Selatan Pekanbaru
Hp. 085265444014
Zulkifli.akhmad@gmail.com

Abstrak, Penelitian ini dilakukan untuk mengklasifikasikan pelanggan perusahaan telekomunikasi seluler guna menemukan decision tree dengan keputusan masih berlangganan atau berhenti berlangganan. Klasifikasi menggunakan algoritma C45 yang akan memproses 54.979 record menggunakan Rapidminer. Data dari perusahaan telekomunikasi yang berupa MS Excel dilakukan pembersihan dari record yang tak lengkap, data ganda, dan data yang salah. Variabel input yang akan diproses untuk klasifikasi adalah Umur Pelanggan, Lama Berlangganan, Penggunaan Data, Pemakaian telepon ke Internasional, Asal kota (besar/kecil). Sedangkan variabel target adalah Berhenti Berlangganan atau Masih Berlangganan.

Kata Kunci : pelanggan telekomunikasi seluler, klasifikasi, algoritma C45.

Abstract, The research was carried out to classify customers in order to find a mobile telecommunications company decision tree with decisions still subscribe or unsubscribe. The Classification use C45 algorithm to process 54,979 records using Rapidminer. The data from mobile telecommunications company was served in MS Excel format, the process start from cleaning the incomplete records, duplicate data, and incorrect data. Input variables to be processed for classification is Customer Age, Length of Service, Usage of International Call, Usage of Data, and City of Origin (big city / town). While the target variable is Still Subscribe or Unsubscribe.

Keywords : mobile telecommunication customer, classification, algorithm C45

PENDAHULUAN

Jumlah pelanggan telekomunikasi seluler di Sumatera bagian tengah telah mencapai 10 juta pelanggan. Dengan jumlah pelanggan sebanyak itu tentunya data yang tercipta juga sangat besar. Namun, hingga saat ini belum dilakukan eksplorasi terhadap kumpulan data (*database*) tersebut.

Dengan ukuran database pelanggan yang besar (*huge database*) yang dimiliki oleh Perusahaan Telekomunikasi Seluler, sangat dimungkinkan dilakukan penggalian data atau *Data Mining*. Menurut Thian Shyug Lee *et al.*, (2004) *Data mining* (DM), bisa juga disebut sebagai *Knowledge Discovery in Database* (KDD). KDD adalah pendekatan sistematis untuk menemukan pola dasar, *trend*, dan hubungan yang tersembunyi di dalam data. Selama ini tim *Marketing* dan *Sales* belum menggunakan pola (*pattern*) dalam melakukan kegiatannya.

Salah satu teknik yang bisa dilakukan dalam KDD adalah Klasifikasi menggunakan algoritma C45. Menurut Veronica S. Moertini, (2003), C45 adalah algoritma yang sudah banyak dikenal dan digunakan untuk klasifikasi data yang memiliki atribut-atribut numerik dan kategorial. Hasil dari proses klasifikasi yang berupa aturan-aturan dapat digunakan untuk memprediksi nilai atribut bertipe diskret dari *record* yang baru.

Pada penelitian ini, data pelanggan di wilayah sumatera bagian tengah (data tagihan, data pelanggan yang berhenti berlangganan) akan dibersihkan (*cleaning*) dan disaring (*filtering*), lalu dikelompokkan berdasarkan pelanggan kota besar dan kota kecil. Jumlah Tagihan juga akan dikategorikan ke dalam tagihan besar, sedang, dan kecil. Sedangkan pengambilan paket berlangganan akan dikelompokkan berdasarkan jumlah paket yang banyak, sedang, dan *standard*. Pola yang dihasilkan dari algoritma ini adalah sebuah pohon keputusan (*decision tree*) dengan output : berhenti berlangganan atau masih berlangganan. Penelitian ini akan

menggunakan data dari pelanggan 3 kota besar, dan 3 kota kecil. Hasil dari penelitian ini akan mendeskripsikan pola pelanggan kota besar dan kota kecil untuk membantu dalam pengambilan keputusan di Perusahaan Telekomunikasi Seluler untuk mencari solusi pengelolaan data pelanggan perlu mencari langkah-langkah menyusun strategi untuk melakukan klasifikasi pelanggan perusahaan telekomunikasi seluler di 3 kota besar dan 3 kota kecil di wilayah Sumatera bagian tengah kemudian bagaimana hasil pola klasifikasi pelanggan perusahaan telekomunikasi seluler di kota besar dan kota kecil

C45 adalah algoritma yang sudah banyak dikenal dan digunakan untuk klasifikasi data yang memiliki atribut-atribut numerik dan kategorial. Hasil dari proses klasifikasi yang berupa aturan-aturan dapat digunakan untuk memprediksi nilai atribut bertipe diskret dari record yang baru. Beberapa Jurnal dibawah ini membahas teknik-teknik klasifikasi data secara umum, seperti analisis hasil eksperimen yang menggunakan C4.5 untuk mengklasifikasi data perbankan.

Menurut Veronica S. Moertini (2003), Database perbankan kaya dengan data. Bank dapat melakukan penelitian untuk mengklasifikasikan pelanggan-pelanggan mereka. Hasil penelitian tersebut dapat digunakan dalam bisnis misalnya untuk menentukan perilaku pelanggan yang baik atau yang berpotensi “tidak baik”. Penelitian yang bertujuan untuk menghasilkan *knowledge* ini telah banyak dilakukan dan metode yang digunakan menggunakan algoritma c4.5

Algoritma C4.5 dapat digunakan untuk klasifikasi data pelanggan perbankan, mahasiswa, di bidang sumber daya manusia, bahkan di bidang Jaringan Komputer. Visualisasi *decision tree* mengubah fakta yang sangat besar menjadi pohon keputusan yang merepresentasikan aturan. Aturan-Aturan atau Pola tersebut dapat dengan mudah dipahami dengan bahasa alami. Selain itu, pohon keputusan juga berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah calon variabel input dengan sebuah variabel target.

a. Pelanggan Perusahaan Telekomunikasi Seluler

Bisnis usaha telekomunikasi seluler membagi pelanggannya dalam 2 segmen, *retail* dan korporat, tiap segmen memiliki karakteristik perilaku masing-masing. Pelanggan *retail* didominasi oleh pelanggan pra-bayar dengan variasi penggunaan produk yang sedikit, sedangkan pelanggan korporat sepenuhnya adalah pelanggan paska-bayar dengan berbagai pilihan produk jasa yang luas.

Strategi teritorial membangun jaringan produk dan layanan telekomunikasi seluler membagi wilayah Indonesia dalam 3 regional; Sumatera, Jawa, dan Indonesia Timur. Regional Sumatera dibagi menjadi 3 wilayah, yakni Sumatera Bagian Utara, Sumatera Bagian Tengah, dan Sumatera Bagian Selatan. Objek tesis ini adalah data pelanggan korporat yang berada di Wilayah Sumatera Bagian Tengah yang membawahi 3 propinsi; Sumatera Barat, Riau, Riau Kepulauan, dengan jumlah pelanggan korporat sekitar 63 ribu.

b. Layanan Jasa Telekomunikasi Seluler

Layanan jasa telekomunikasi seluler terus mengalami pembaharuan dan peningkatan, berawal dari layanan tradisional berupa *Voice* dan SMS (*short message service*) hingga layanan Data, *Network*, dan Aplikasi. Berikut ini adalah beberapa layanan dasar yang dipakai oleh pelanggan korporat :

a. SMS

Short Message Service (SMS) adalah layanan pesan singkat berupa *text* dengan maksimal 160 huruf/karakter. Layanan ini adalah layanan dasar yang menawarkan alternatif lain bagi pelanggan untuk berkomunikasi melalui pesan singkat tertulis.

b. MMS

Multimedia Messaging Service (MMS) adalah sebuah standar layanan yang memungkinkan untuk mengirim pesan yang mengandung objek multimedia, seperti gambar, audio, video, dan *rich text*.

c. GPRS

GPRS (*General Packet Radio Service*) adalah layanan yang memungkinkan pengiriman dan penerimaan data lebih cepat dibandingkan dengan penggunaan teknologi *Circuit Switch*

Data atau CSD. Penggabungan layanan telepon seluler dengan GPRS menghasilkan generasi baru yang disebut 2.5G/3G. Sistem GPRS dapat digunakan untuk transfer data (dalam bentuk paket data) yang berkaitan dengan *e-mail*, data gambar (MMS), *Wireless Application Protocol* (WAP), dan *World Wide Web* (WWW).

- d. **Panggilan Telepon / Call**
Panggilan telpon keluar bisa dilakukan ke seluruh nomor operator telekomunikasi lain, terdiri dari panggilan Lokal, Interlokal/SLJJ, dan SLI (Sambungan Langsung Internasional). Jika pelanggan sedang berada di luar negeri, maka jenis layanannya disebut *International Roaming*.
- e. **BlackBerry**
Layanan *BlackBerry* dimulai sejak tahun 2005 untuk pelanggan korporat. Kemudian layanan *BlackBerry* ini berkembang pesat dan mulai dipasarkan ke semua pelanggan, baik itu pelanggan korporat maupun pelanggan retail.
- f. **Internet**
Layanan koneksi internet di perusahaan telekomunikasi seluler dikenal dengan nama *Flash* yang merupakan layanan internet tanpa kabel (wireless) dan didukung dengan teknologi HSDPA/3G/EDGE/GPRS yang dapat menghasilkan kecepatan *download* sampai dengan 3.2 Mbps.

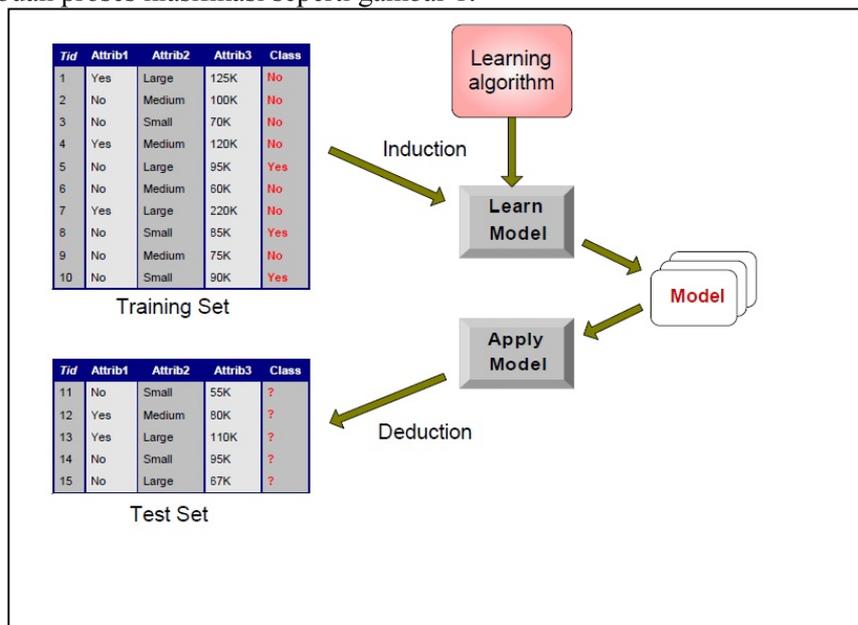
a. **Klasifikasi**

Menurut Tan, *et al* (2009), klasifikasi adalah proses untuk menyatakan suatu objek ke salah satu kategori yang sudah didefinisikan sebelumnya. Klasifikasi juga bisa didefinisikan sebagai proses pembelajaran fungsi target (model klasifikasi) yg memetakan setiap sekumpulan atribut x (input) ke salah satu klas y yang didefinisikan sebelumnya.

Lebih lanjut, *input* didefinisikan sebagai sekumpulan *record* (*training set*), dan setiap *record* terdiri atas sekumpulan atribut, salah satu atribut adalah klas. Adapun model klasifikasi digunakan untuk antara lain :

1. Pemodelan Deskriptif sebagai perangkat penggambaran untuk membedakan objek-objek dari klas berbeda
2. Pemodelan Prediktif digunakan untuk memprediksi label klas untuk *record* yang tidak diketahui atau tidak dikenal

Ilustrasi sebuah proses klasifikasi seperti gambar 1.



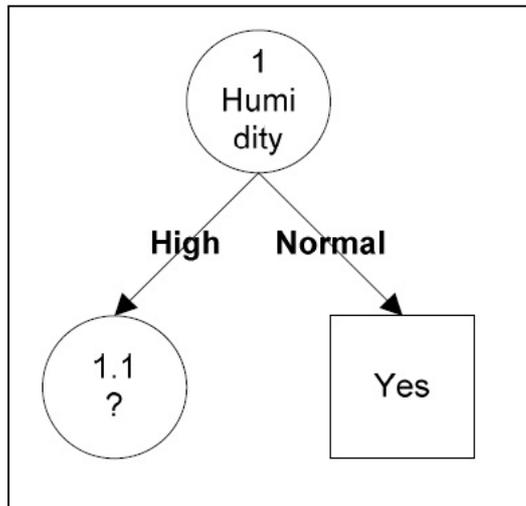
Menurut Bertalya (2009), beberapa Teknik Klasifikasi yang sudah dikenal luas antara lain :

Dari hasil perhitungan pertama, maka didapat hasil seperti tabel 2.2 :

Tabel 2.2 Perhitungan Tahap Pertama

Node		Jml Kasus (S)	Tidak (S ₁)	Ya (S ₂)	Entropy	Gain
1	TOTAL	14	4	10	0.8631	
	OUTLOOK					0.2585
	CLOUDY	4	0	4		
	RAINY	5	1	4	0.7219	
	SUNNY	5	3	2	0.9709	
	TEMPERATURE					0.1839
	COOL	4	0	4	0	
	HOT	4	2	2	1	
	MILD	6	2	4	0.9183	
	HUMIDITY					0.3705
	HIGH	7	4	3	0.9852	
	NORMAL	7	0	7	0	
	WINDY					0.0059
	FALSE	8	2	6	0.8113	
	TRUE	6	4	2	0.9183	

Dari hasil pada Tabel 2.2 dapat diketahui bahwa atribut dengan Gain tertinggi adalah HUMIDITY yaitu sebesar 0.3705. Dengan demikian HUMIDITY dapat menjadi node akar. Ada 2 nilai atribut dari HUMIDITY yaitu HIGH dan NORMAL. Dari kedua nilai atribut tersebut, nilai atribut NORMAL sudah mengklasifikasikan kasus menjadi 1 yaitu keputusannya **YES**, sehingga tidak perlu dilakukan perhitungan lebih lanjut, tetapi untuk nilai atribut HIGH masih perlu dilakukan perhitungan lagi. Dari hasil tersebut dapat digambarkan pohon keputusan sementara yang tampak seperti gambar 2.2 :



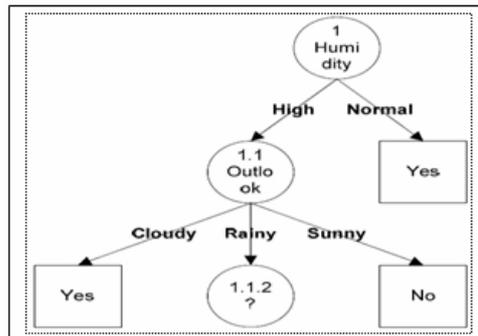
Gambar 2.2 : Pohon dari Hasil Perhitungan pertama

Karena pada HUMIDITY = HIGH masih terdapat keputusan yang belum seragam, maka akan dilakukan penghitungan jumlah kasus untuk keputusan **YES**, jumlah kasus untuk keputusan **NO**, dan Entropy dari semua kasus yang dibagi berdasarkan atribut OUTLOOK, TEMPERATURE dan WINDY yang dapat menjadi node akar dari nilai atribut HIGH tersebut. Setelah itu lakukan penghitungan Gain untuk masing-masing atribut. Hasil perhitungan ditunjukkan oleh Tabel 2.3.

Tabel 2.3. Perhitungan Untuk Node 1.1

Node			Jml Kasus (S)	Tidak (S1)	Ya (S2)	Entropy	Gain
1.1	HUMIDITY-HIGH		7	4	3	0.9852	
	OUTLOOK						0.6995
		CLOUDY	2	0	2	0	
		RAINY	2	1	1	1	
		SUNNY	3	3	0	0	
	TEMPERATURE						0.0202
		COOL	0	0	0	0	
		HOT	3	2	1	0.9183	
		MILD	4	2	2	1	
	WINDY						0.0202
		FALSE	4	2	2	1	
		TRUE	3	2	1	0.9183	

Dari hasil pada Tabel 2.3 dapat diketahui bahwa atribut dengan Gain tertinggi adalah OUTLOOK yaitu 0.6995. Dengan demikian OUTLOOK dapat menjadi node cabang dari nilai atribut HIGH. Ada 3 nilai atribut dari OUTLOOK yaitu CLOUDY, RAINY dan SUNNY. Dari ketiga nilai atribut tersebut, nilai atribut CLOUDY sudah mengklasifikasikan kasus menjadi 1 yaitu keputusannya **YES**, SUNNY = **NO**, sehingga tidak perlu dilakukan perhitungan lebih lanjut, tetapi untuk RAINY masih perlu dilakukan perhitungan lagi. Pohon keputusan yang terbentuk sebagai berikut :



Gambar 2.3 Pohon Dari Hasil Perhitungan 1.1

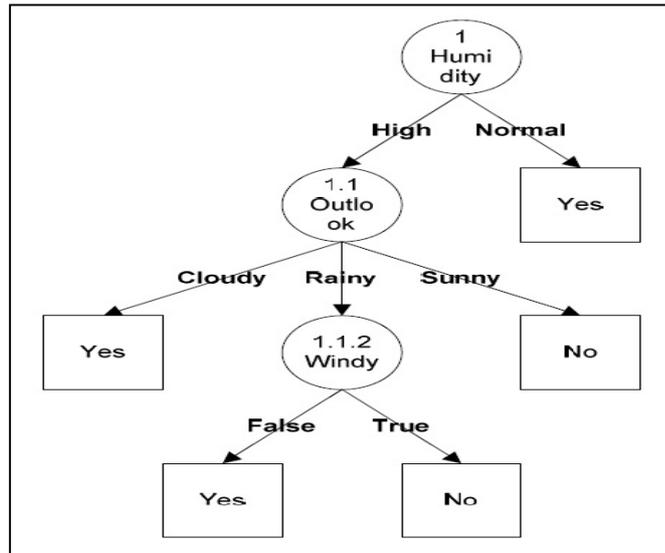
Langkah berikutnya yaitu menghitung jumlah kasus (YES atau NO), dan Entropy berdasarkan atribut TEMPERATURE dan WINDY yang dapat menjadi node cabang dari nilai atribut RAINY, lalu menghitung Gain untuk masing-masing atribut. Hasil perhitungan ditunjukkan oleh tabel 2.4 :

Tabel 2.4. Perhitungan Node 1.1.2

Node			Jml Kasus (S)	Tidak (S1)	Ya (S2)	Entropy	Gain
1.1.2	HUMIDITY-HIGH dan OUTLOOK - RAINY		2				
	TEMPERATURE						0
		COOL	0	0	0	0	
		HOT	0	0	0	0	
		MILD	2	1	1	1	
	WINDY						1
		FALSE	1	0	1	0	
		TRUE	1	1	0	0	

Dari hasil pada tabel 2.4 dapat diketahui bahwa atribut dengan Gain tertinggi adalah WINDY yaitu sebesar 1. Dengan demikian WINDY dapat menjadi node cabang dari nilai atribut RAINY. Ada 2 nilai atribut dari WINDY yaitu FALSE dan TRUE. Dari kedua nilai atribut tersebut, nilai atribut FALSE sudah mengklasifikasikan kasus menjadi 1 yaitu keputusannya **YES** dan nilai atribut TRUE sudah mengklasifikasikan kasus menjadi satu dengan keputusan **NO**, sehingga tidak perlu dilakukan perhitungan lebih lanjut untuk nilai atribut ini.

Akhirnya sebuah pohon keputusan sudah bisa dibuat lengkap karena semua kasus sudah masuk ke dalam klas, untuk itu, pohon keputusan yang final adalah seperti gambar 2.4.



Gambar 2.4 Pohon Keputusan Yang Final

Kurva ROC (Receiver Operating Characteristic)

Untuk melakukan validasi terhadap *tree* yang dihasilkan digunakan nilai dari area dibawah kurva ROC (AUC/ *Area Under Curve*). Menurut Gorunescu (2011), hasil perhitungan yang divisualisasikan dengan kurva ROC (*Receiver Operating Characteristic*) atau AUC (*Area Under Curve*).

ROC memiliki tingkat nilai diagnosa yaitu :

- Akurasi bernilai 0.90 – 1.00 = *excellent classification*
- Akurasi bernilai 0.80 – 0.90 = *good classification*
- Akurasi bernilai 0.70 – 0.80 = *fair classification*
- Akurasi bernilai 0.60 – 0.70 = *poor classification*
- Akurasi bernilai 0.50 – 0.60 = *failure*

METODOLOGI PENELITIAN

Hasil dalam kegiatan penelitian ini adalah penemuan model dari pola perilaku pelanggan perusahaan telekomunikasi seluler, sehingga pola tersebut dapat bermanfaat bagi banyak pihak. Penelitian ini akan menggunakan aplikasi RapidMiner dalam mengolah data pelanggan menggunakan metode C4.5.

Kegiatan pendahuluan dalam penelitian ini dilakukan dengan mengajukan permohonan mendapatkan data resmi pada perusahaan telekomunikasi seluler wilayah sumatera bagian tengah di Batam. Setelah dilakukan wawancara dengan staf dan manajemen didapatkan data awal dari masalah yang dihadapi.

Masalah yang sudah diidentifikasi ini kemudian ditelaah untuk dilakukan penentuan metode yang akan digunakan. Dengan pertimbangan jumlah data yang besar maka diputuskan untuk menggunakan teknik *data mining* C4.5. Setelah melalui studi literatur lalu dipilih *tools* yang digunakan dalam pemrosesan data, yakni RapidMiner.

1. Mengidentifikasi Masalah

Ruang lingkup masalah yang akan diteliti akan diidentifikasi terlebih dahulu, serta menentukan dan mendefinisikan batasan masalah penelitian, hal ini dilakukan agar jelas dan fokus pada langkah pemecahan masalah, dan pembahasan tidak melebar kemana-mana.

2. Menganalisa Masalah

Dengan menggunakan data pelanggan dalam format *MS Excel* akan ditransformasikan ke dalam *Repository* di dalam aplikasi *RapidMiner*. Pada tahap ini, pola klasifikasi pelanggan telekomunikasi seluler di kota besar dan kota kecil, penentuan jumlah tagihan besar dan kecil, maupun jumlah paket yang diambil oleh pelanggan telekomunikasi seluler akan dijadikan atribut dalam melakukan analisa.

3. Mengumpulkan Data Tagihan Pelanggan dan Data Pelanggan yang Berhenti Pengumpulan

data merupakan proses kompilasi dari beberapa data yang terpisah, sebuah *file* tidak memiliki semua atribut yang diperlukan, untuk itu harus didapatkan dari *file* yang berbeda lalu digabungkan.

4. Pengolahan Data untuk *Filtering*

Tidak semua atribut dalam *file* akan digunakan, untuk itu akan dilakukan proses penyaringan seperti : membuang atribut yg tidak perlu, membuang data ganda, membuang data yang tidak lengkap.

5. Pengolahan Data untuk Model Karakteristik

Data harus diolah sebagai bagian dari pra pengolahan. Beberapa atribut yang digunakan harus dilakukan klasifikasi menggunakan proses statistik.

6. Menyiapkan *Tool RapidMiner*

RapidMiner adalah aplikasi data mining berbasis open-source yang terkemuka dan ternama. Didalamnya terdapat aplikasi yang berdiri sendiri untuk analisis data dan sebagai mesin data mining seperti untuk *loading* data, transformasi data, pemodelan data, dan metode visualisasi data. *RapidMiner* akan dipasang pada Sistem Operasi *MS Windows 7*.

7. Validasi Hasil Klasifikasi

Kebenaran hasil klasifikasi harus divalidasi, dan salah satu metoda yang bisa digunakan untuk memeriksa validitas hasil adalah *AUC (Area Under Cover)*, nilai akurasi yang bagus adalah 0.9 hingga 1.0.

8. Mengevaluasi Hasil

Berdasarkan hasil klasifikasi dari beberapa kombinasi data yang berbeda akan dilakukan evaluasi, pada tahap ini sudah didapat *Knowledge*.

9. Kesimpulan

Tahap ini merupakan kesimpulan dari rangkaian proses penelitian, memaparkan tingkat ketercapaian tujuan, dan juga memaparkan kelemahan dan kekurangan dari proses penelitian, untuk perbaikan jika ada penelitian lanjutan.

HASIL DAN PEMBAHASAN

a. Klasifikasi dengan *Rapidminer*

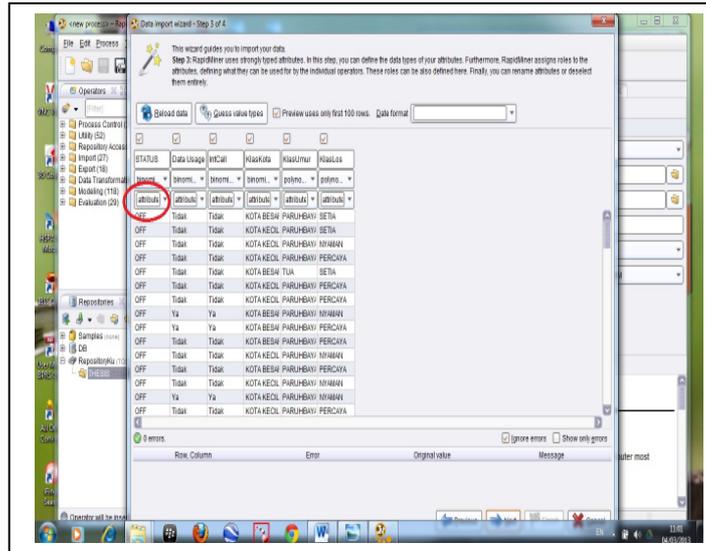
Klasifikasi dilakukan pada data dengan kriteria sebagai berikut :

- a. Jumlah kasus = 54.979 *record*
- b. Jumlah Status OFF = 10.309 *record*
- c. Jumlah Status ON = 44.670 *record*

Tahapan yang dilakukan adalah sebagai berikut :

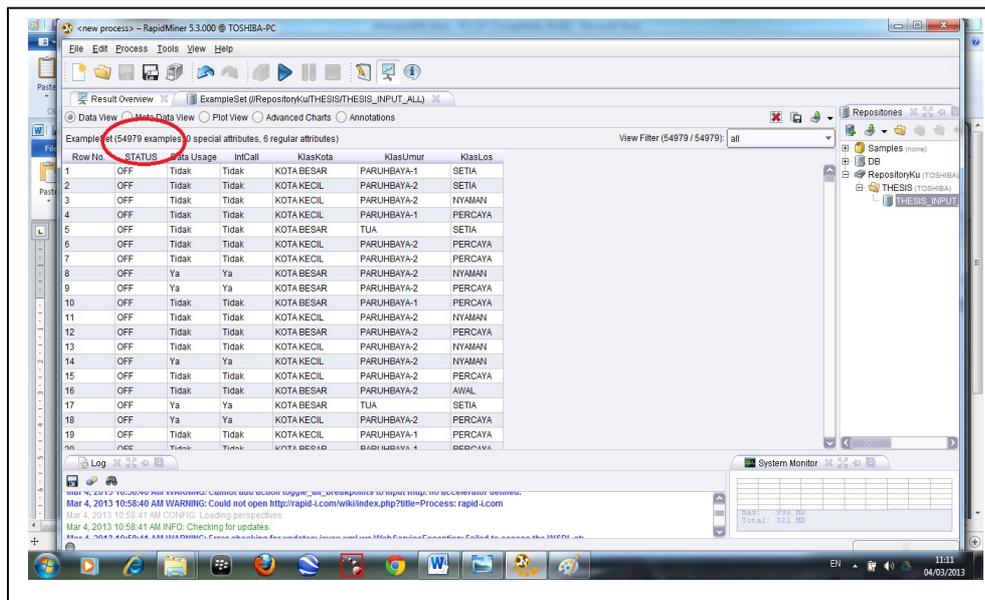
1. *Loading* data ke *RapidMiner*

Data dalam format *MS Excel* dimasukkan ke dalam *Rapidminer* dan ditentukan *field STATUS* sebagai keputusan. Atribut *feld* keputusan diganti menjadi label, seperti terlihat pada gambar 5.1.



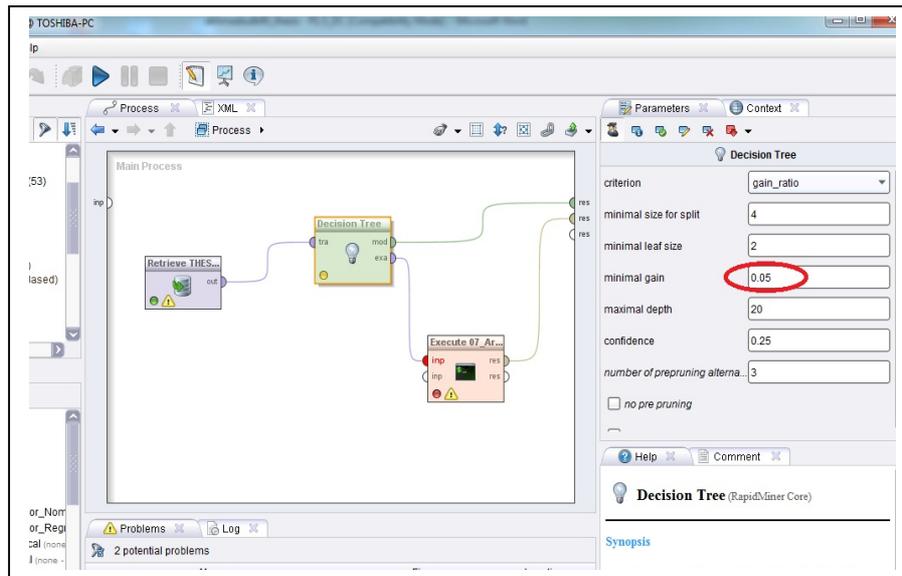
Gambar 5.1 Data Loading

2. Verifikasi Data yang sudah sempurna masuk ke dalam Rapidminer untuk diverifikasi jumlah *record* nya, dalam tesis ini terdapat 54.979 *record* seperti terlihat pada gambar 5.2.



Gambar 5.2 Verifikasi Hasil Data Loading

3. Klasifikasi C4.5
 Data yang sudah diverifikasi kemudian diproses dengan algoritma c45 yang didalam *tool* Rapidminer berada pada menu : Operator -> Data Modelling -> Tree Induction -> Decision Tree Untuk melakukan validasi terhadap *tree* yang dihasilkan menggunakan AUC dengan menggunakan *tool* yang ada di menu sebagai berikut :
 Repositories -> Samples -> Processes -> Validation -> 07 Area Under Curve
 Persiapan proses klasifikasi kemudian terlihat di kotak "Main Process" seperti pada gambar 5.3.

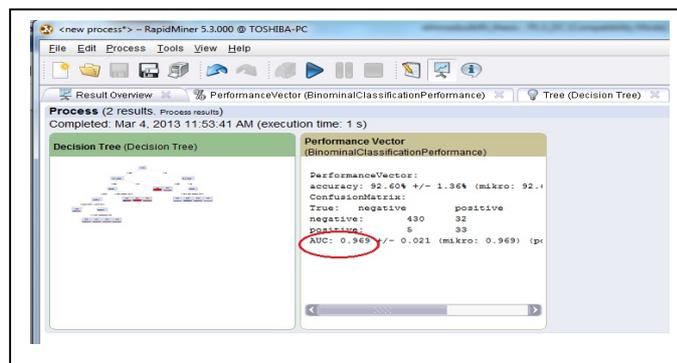


Gambar 5.3 Persiapan Klasifikasi C4.5

Nilai *default* pada proses klasifikasi dengan *Decision Tree* menggunakan nilai "Minimal Gain" = 0.1, namun pada proses ini ditentukan "Minimal Gain" adalah 0.05 untuk lebih melihat kompleksitas dari *tree* yang dihasilkan.

Setelah sampai pada tahap ini, maka langkah berikutnya adalah dengan menjalankan proses dengan menekan tombol "Run" yang ada di *menu bar* Rapidminer.

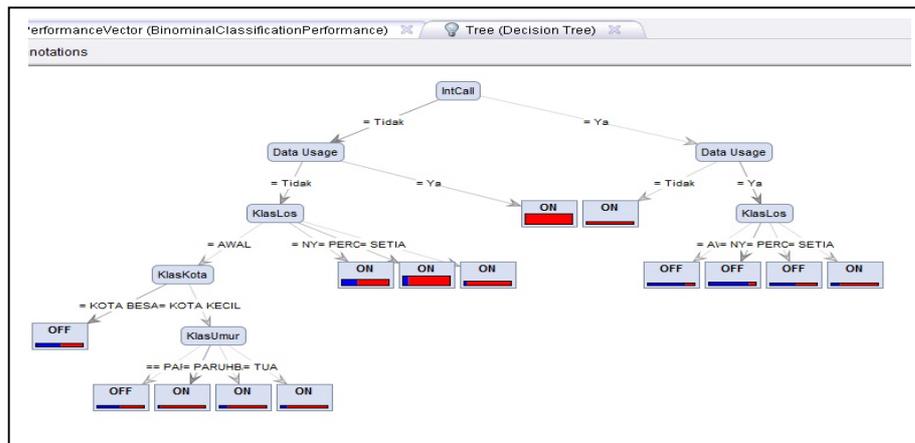
Hasil klasifikasi ini dapat dilihat pada menu "Result Overview" dengan tampilan awal berupa *history* yang berisi tanggal dan waktu pelaksanaan klasifikasi. Hal ini dapat dilihat pada gambar 5.4.



Gambar 5.4 Result Overview

Pada *Result Overview* ini dihadirkan 2 kotak hasil, yakni gambaran *tree* dalam ukuran kecil, dan hasil AUC dalam kotak *Performance Vector*. Pada kotak *Performance Vector* level AUC pada nilai 0.969 yang menurut Gorunescu (2011) tergolong sebagai *Excellent Classification*.

Tampilan hasil *tree* dengan ukuran yang sesungguhnya ada di *window Tree (Decision Tree)*. Label "ON" digambarkan dengan warna kotak merah, dan Label "OFF" digambarkan dengan kotak berwarna biru. Hal ini dapat dilihat pada gambar 5.5.



Gambar 5.5 Tree Yang Dihasilkan

Text View dari Tree pada gambar 5.5. dapat dilihat di gambar 5.6.

```

Result Overview PerformanceVector (BinominalClassificationPerformance) Tree (Deci
Graph View Text View Annotations
Tree
IntCall = Tidak
| Data Usage = Tidak
| | KlasLos = AWAL
| | | KlasKota = KOTA BESAR: OFF {OFF=1473, ON=1241}
| | | KlasKota = KOTA KECIL
| | | | KlasUmur = MUDA: OFF {OFF=23, ON=23}
| | | | KlasUmur = PARUH BAYA-1: ON {OFF=36, ON=532}
| | | | KlasUmur = PARUH BAYA-2: ON {OFF=18, ON=79}
| | | | KlasUmur = TUA: ON {OFF=6, ON=33}
| | | KlasLos = NYAMAN: ON {OFF=2895, ON=5506}
| | | KlasLos = PERCAYA: ON {OFF=1981, ON=12671}
| | | KlasLos = SETIA: ON {OFF=322, ON=4007}
| | Data Usage = Ya: ON {OFF=0, ON=17927}
IntCall = Ya
| Data Usage = Tidak: ON {OFF=0, ON=1388}
| Data Usage = Ya
| | KlasLos = AWAL: OFF {OFF=279, ON=65}
| | KlasLos = NYAMAN: OFF {OFF=2565, ON=398}
| | KlasLos = PERCAYA: OFF {OFF=628, ON=482}
| | KlasLos = SETIA: ON {OFF=83, ON=318}
    
```

Gambar 5.6 Text View

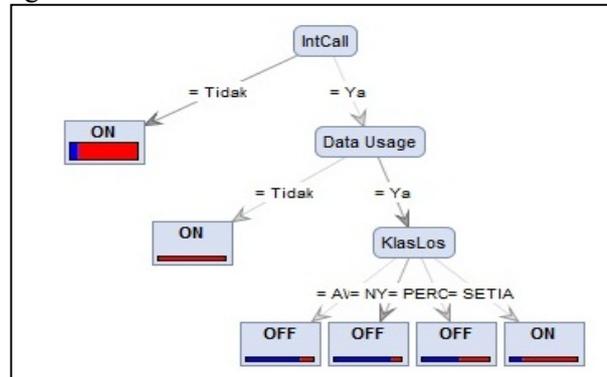
Makna dari Hasil Klasifikasi

Dari pohon keputusan yang dihasilkan dapat dimaknai sebagai berikut :

- a. Hal pertama yang membedakan apakah pelanggan itu ON atau OFF adalah atribut panggilan ke internasional (IntCall)
- b. Pelanggan yang 100% masih berlangganan (ON) adalah :
 1. Pelanggan yang tidak menggunakan telepon selulernya untuk menelepon ke internasional (IntCall = Tidak) namun menggunakan telepon tersebut untuk mengakses data (Data Usage = Ya). Perbandingan jumlah pelanggan yang ON dan OFF pada kategori ini adalah 17.927 berbanding 0.
 2. Pelanggan yang menggunakan telepon selulernya untuk menelepon ke internasional (IntCall = Ya) namun tidak menggunakannya untuk mengakses data (Data Usage = Tidak). Perbandingan jumlah pelanggan yang ON dan OFF pada kategori ini adalah 1.388 berbanding 0.
- c. Jumlah pelanggan tertinggi yang berhenti berlangganan (OFF) terjadi pada Lama Berlangganan = NYAMAN (2 sampai 4 tahun), yakni 2.895 pelanggan. Hal ini terjadi pada IntCall = Tidak dan Data Usage = Tidak.
- d. Pelanggan yang berhenti(OFF) paling sedikit adalah pada kategori LoS = SETIA (berlangganan diatas 10 tahun)
 - IntCall = Tidak dan Data Usage = Tidak
AWAL(1556), NYAMAN(2895), PERCAYA(1981), SETIA(322)
 - IntCall = Ya dan Data Usage = Ya
AWAL(279), NYAMAN(2565), PERCAYA(628) SETIA(83)

Validasi Hasil

Dengan mengubah nilai MINIMAL GAIN menjadi 0,1 (default), maka *Tree* yang dihasilkan seperti pada gambar 5.7.



Gambar 5.7. *Tree* dengan Minimal Gain = 0.1 (default)

Dari gambar 5.7. didapat bahwa dengan mengganti nilai *Minimal Gain* didapat :

- root* pada pohon keputusan tetap IntCall. Dengan demikian, maka *root* = IntCall adalah akar dari pohon keputusan yang benar, dan dengan memperbesar atau memperkecil nilai minimal Gain tidak mengubah *root*. Memperkecil nilai Minimal Gain akan membuat pohon keputusan menjadi lebih lengkap. Dalam penelitian ini, penulis menggunakan nilai minimal Gain = 0,05 dan hasil pohon keputusan tetap menempatkan IntCall sebagai *root*.
- Pada Validasi ini, nilai AUC tetap yakni 0,969.

KESIMPULAN DAN SARAN

Dari *Tree* yang didapat dari klasifikasi 54.979 data, maka dapat disimpulkan bahwa :

- Jumlah kasus dan karakteristiknya pada sebuah klasifikasi bisa mempengaruhi bentuk pohon keputusan
- Variabel Input IntCall (Pemakaian Sambungan Langsung Internasional) merupakan *root* dari pohon keputusan
- Kategori Kota Besar dan Kota Kecil tidak akan berpengaruh pada keputusan ON dan OFF jika :
 - Pelanggan tersebut menggunakan telepon selulernya untuk melakukan panggilan ke internasional (IntCall = Ya)
 - Pelanggan tersebut menggunakan telepon selulernya untuk mengakses data (*Data Usage* = Ya)
- Pelanggan yang mutlak paling banyak berhenti berlangganan adalah pada kategori Lama Berlangganan = NYAMAN (2 tahun sampai 4 tahun).

DAFTAR PUSTAKA

Florence Gorunescu, 2011, "Data Mining – Concepts, Models, and Tehiniques", Springer Verlag Berlin Heidelberg

bertalya.staff.gunadarma.ac.id/Downloads/files/16884/Klasifikasi-Pohon+Keputusan.pdf

<http://home.unpar.ac.id/~integral/Volume8/Integral 8 No.2/ C45 Algorithm.PDF>

<http://www.enggjournals.com/ijcse/doc/IJCSE10-02-08-029.pdf>

<http://home.unpar.ac.id/~integral/Volume8/Integral 8 No.2/C45 Algorithm.PDF>

bertalya.staff.gunadarma.ac.id/Downloads/files/16884/Klasifikasi-Pohon+Keputusan.pdf

<http://home.unpar.ac.id/~integral/Volume8/Integral 8 No.2/ C45 Algorithm.PDF>