

PENENTUAN LOYALITAS PELANGGAN PADA DISTRIBUTOR PULSA ELEKTRONIK MENGUNAKAN ALGORITMA C4.5 DAN NAIVE BAYES

Eni Irfiani

Abstract— Customers are important assets to support the passage of the Company. Company engaged in telecommunications have a high level of competition. To reduce the number of loyal customers who are not required analysis to predict the causes and out of its customers to be high. One way that can be used to analyze the database in large numbers to produce meaningful output is to use data mining. Prediction customer loyalty needed to predict the behavior of the customers most likely to cause customers to move to other competitors. To determine the predictive classification techniques used in data mining. In addition to predicting customer loyalty, in this paper will compare two data mining models are decision tree and Naive Bayes. Through testing dataset customers will get the best model to predict customer loyalty and gain a loyal customer forecasting. The measurement results show that the C4.5 algorithm is an algorithm that best predicted the customer loyalty that is equal to 88.14% by 0954 AUC values.

Intisari- Pelanggan merupakan aset penting untuk mendukung berjalannya Perusahaan. Perusahaan yang bergerak dibidang telekomunikasi mempunyai tingkat persaingan yang tinggi. Untuk mengurangi jumlah pelanggan yang tidak loyal dibutuhkan analisa untuk memprediksi faktor penyebab keluar nya pelanggan menjadi tinggi. Salah satu cara yang dapat digunakan untuk menganalisa database dalam jumlah besar untuk menghasilkan output yang bermakna adalah menggunakan data mining. Prediksi loyalitas pelanggan diperlukan untuk memprediksi perilaku pelanggan yang paling mungkin menyebabkan pelanggan berpindah ke kompetitor lain. Untuk menentukan prediksi tersebut digunakan teknik klasifikasi pada data mining. Selain memprediksi loyalitas pelanggan, dalam penulisan ini akan membandingkan dua model data mining yaitu decision tree dan naive bayes. Melalui pengujian dataset pelanggan akan didapatkan model yang paling baik dalam memprediksi loyalitas pelanggan dan memperoleh peramalan pelanggan yang loyal. Hasil pengukuran menunjukkan bahwa algoritma C4.5 merupakan algoritma yang paling baik untuk memprediksi loyalitas pelanggan yaitu sebesar 88.14% dengan nilai AUC 0.954.

Kata kunci : Loyalitas, Pelanggan, Algoritma C4.5, Naive Bayes

Program Studi Manajemen Informatika AMIK BSI Jakarta; Jl. Kramat Raya No. 18 Jakarta Pusat DKI Jakarta. Telp: 021-3100413 Fax: 021-3144869 ; e-mail: eni.enf@bsi.ac.id

I. PENDAHULUAN

Berkembangnya perusahaan-perusahaan telekomunikasi seluler saat ini mendorong pertumbuhan bisnis distributor pulsa elektronik. Persaingan bisnis pada industri telekomunikasi seluler semakin pesat didukung oleh berkembangnya perusahaan-perusahaan telekomunikasi seluler dan kebutuhan akan sarana telekomunikasi untuk urusan bisnis, keluarga ataupun keperluan lainnya.

Bisnis pulsa elektronik merupakan peluang bisnis yang menjanjikan dikarenakan rendahnya resiko, tidak mudah mengalami kerusakan, tidak ada masa kadaluarsa dan merupakan kebutuhan utama dalam komunikasi. Dengan kemudahan tersebut menjadikan bisnis pulsa elektronik diminati oleh investor-investor.

Data mining digunakan untuk menganalisis database dalam jumlah besar untuk menghasilkan output yang bermakna. Salah satu aplikasi yang paling umum dari data mining disebut sebagai *churn analysis* yang digunakan untuk memprediksi perilaku pelanggan yang paling mungkin untuk berpindah ke kompetitor lain dan untuk membantu manajemen dalam pengambilan keputusan. Tujuan dari penulisan ini adalah untuk menentukan pelanggan yang tidak loyal pada distributor server elektronik di daerah Depok.

Permasalahan yang ada pada bisnis pulsa elektronik adalah tingginya jumlah pelanggan yang berhenti menggunakan layanan perusahaan dan berpindah ke perusahaan kompetitor lain [2]. Berkurangnya jumlah pelanggan berpengaruh terhadap pendapatan yang diterima oleh perusahaan.

Dalam jurnal sebelumnya analisis loyalitas pelanggan pada industri telekomunikasi dengan metode klasifikasi data mining menggunakan teknik klasifikasi antara lain: decision tree, SVM, neural network. Prediksi loyalitas pelanggan pada Service Provider di Iran menggunakan model decision tree. Untuk memprediksi loyalitas pelanggan pada perusahaan telekomunikasi di Indonesia menggunakan model decision tree [2]. Perbandingan teknik klasifikasi menggunakan decision tree, regression dan neural network pada perusahaan telekomunikasi [7]. Komparasi beberapa model klasifikasi yaitu decision tree, neural network, SVM pada perusahaan telekomunikasi [6]. Dari berbagai macam teknik klasifikasi data mining yang digunakan dapat diketahui model mana yang paling baik untuk menentukan loyalitas pelanggan.

II. KAJIAN LITERATUR

Pelanggan (*customer*) adalah seseorang yang menjadi terbiasa untuk membeli produk. Kebiasaan itu terbentuk melalui pembelian dan interaksi yang sering selama periode waktu tertentu [3]. Orang yang melakukan pembelian tanpa ada hubungan yang kuat dan pembelian berulang tidak dapat dikategorikan sebagai pelanggan melainkan sebagai pembeli.

Loyalitas pelanggan didefinisikan berdasarkan perilaku membeli. Pelanggan yang loyal adalah orang yang melakukan pembelian berulang secara teratur, membeli antar lini produk atau jasa, mereferensikan kepada orang lain, menunjukkan kekebalan terhadap tarikan dari pesaing [3]. Bagi Perusahaan menjaga pelayanan dengan sangat baik akan meningkatkan rasa kepuasan pelanggan sehingga pada akhirnya pelanggan akan memberikan kesetiaan. Kesetiaan pelanggan tidak dapat dipaksakan oleh Perusahaan, kesetiaan diperoleh dari kepuasan terhadap pelayanan yang diberikan Perusahaan seiring berjalannya waktu dan usaha.

Data mining atau *knowledge discovery in database* (KDD) adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola dan hubungan dalam set data berukuran besar. Keluaran dari data mining dapat dipakai untuk memperbaiki pengambilan keputusan di masa depan [13].

Data Mining didefinisikan sebagai proses penemuan pola dalam data. Menurut Daryl Pregibons dalam [1] "Data mining adalah perpaduan dari ilmu statistik, kecerdasan buatan, dan penelitian bidang database". Nama data mining berasal dari kemiripan antara pencarian informasi yang bernilai dari database yang besar dengan menambang sebuah gunung untuk sesuatu yang bernilai.

Berikut ini penjelasan singkat langkah-langkah dalam KDD:

1. Pembentukan pemahaman domain aplikasi

Pada tahap ini menentukan tujuan dari *end-user* dan bagian terkait dimana KDD dilakukan.

2. Memilih dan membuat *data set* dimana proses penemuan *knowledge* akan dilakukan.

Penentuan data yang akan digunakan untuk proses KDD dilakukan pada tahap ini. Mencari data yang tersedia, memperoleh data tambahan yang dibutuhkan, mengintegrasikan semua data untuk KDD ke dalam sebuah data set, termasuk atribut yang diperlukan dalam proses KDD.

3. *Preprocessing* dan *cleansing*

Dalam tahap ini kehandalan data ditingkatkan. Termasuk *data clearing*, seperti menangani data yang tidak lengkap, menghilangkan gangguan atau *outlier*.

4. Transformasi data

Pada tahap ini, membuat data menjadi lebih baik menggunakan metode reduksi dimensi dan transformasi atribut.

5. Memilih tugas *data mining* yang cocok

Pada tahap ini ditentukan tipe *data mining* yang akan digunakan, apakah klasifikasi, regresi, atau *clustering*, tergantung pada tujuan KDD dan tahap sebelumnya.

6. Memilih algoritma *data mining*

Pada tahap ini dilakukan pemilihan algoritma yang paling tepat untuk menemukan pola.

7. Penggunaan algoritma *data mining*

Pada tahap ini dilakukan implementasi dari algoritma *data mining* yang telah ditentukan pada tahap sebelumnya.

8. Evaluasi

Pada tahap ini dilakukan evaluasi dan penerjemahan dari pola yang didapat.

9. Penggunaan pengetahuan yang didapat

Pada tahap ini, pengetahuan dimasukkan ke sistem lain dan mengaktifkan sistem tersebut serta mengukur hasilnya.

Berdasarkan tugasnya, data mining dikelompokkan menjadi [8]:

1. Deskripsi

Mencari cara untuk menggambarkan pola dan trend yang terdapat dalam data. Sebagai contoh, seorang pengumpul suara mengungkap bukti bahwa mereka yang diberhentikan dari jabatannya saat ini, akan kurang mendukung dalam pemilihan presiden. Untuk deskripsi ini bisa dilakukan dengan *exploratory data analysis*, yaitu metode grafik untuk menelusuri data dalam mencari pola dan tren.

2. Estimasi

Estimasi mirip seperti klasifikasi tapi sasaran adalah numerik. Model dibuat menggunakan record yang lengkap, juga ada targetnya. Kemudian untuk data baru, estimasi nilai target dibuat berdasarkan nilai prediktor. Contoh, untuk estimasi tekanan darah pada pasien, prediktornya umur, jenis kelamin, berat badan, dan tingkat sodium darah. Hubungan antara tekanan darah, dan variabel prediktor pada data training akan menghasilkan model kemudian diaplikasikan pada data baru. Untuk melakukan estimasi bisa digunakan neural network atau metode statistik seperti point estimation dan confidence interval estimations, simple linear regression dan correlation, dan multiple regression.

3. Prediksi

Prediksi mirip seperti klasifikasi dan estimasi, tapi hasilnya untuk memprediksi masa depan. Contoh, memprediksi harga barang tiga bulan mendatang, memprediksi presentasi kenaikan angka kematian karena kecelakaan tahun mendatang jika kecepatan berkendara dinaikkan. Metode dan teknik untuk klasifikasi dan estimasi, jika cocok, bisa juga digunakan untuk prediksi, termasuk metode statistik. Algoritma untuk prediksi antara lain regression tree dan model tree [4].

4. Klasifikasi

Dalam klasifikasi, sarannya adalah kategori, misalkan atribut penghasilan, yang bisa dikategorikan menjadi tiga kelas atau kategori yaitu, tinggi, sedang, dan rendah.

Model data mining membaca sejumlah besar record tiap record berisi informasi pada variabel target. Contoh, dari sebuah data set misalkan mau mengklasifikasikan penghasilan seseorang yang datanya tidak terdapat pada dataset, berdasarkan karakteristik yang berhubungan dengan orang itu seperti, umur, jenis kelamin, dan pekerjaan. Tugas klasifikasi ini cocok untuk metode dan teknik data mining. Algoritma akan mengolah dengan cara membaca data set yang berisi variabel predictor dan variabel target yang telah diklasifikasikan, yaitu penghasilan. Di sini algoritma (software) “mempelajari” kombinasi variabel mana yang berhubungan dengan penghasilan yang mana. Data ini disebut training set. Kemudian algoritma akan melihat ke data baru yang belum termasuk klasifikasi manapun. Berdasarkan klasifikasi pada data set kemudian algoritma akan memasukkan data baru tersebut ke dalam klasifikasi yang mana. Misalkan seorang professor wanita berusia 63 tahun bisa jadi diklasifikasikan ke dalam kelas penghasilan tinggi. Algoritma klasifikasi yang banyak digunakan secara luas untuk klasifikasi antara lain, decision tree, bayesian classifier, dan neural network [1].

Clustering mengacu pada pengelompokan record-record, observasi, atau kasus-kasus ke dalam kelas-kelas dari objek yang mirip. Pada clustering tidak ada variabel sasaran. Sebuah cluster adalah koleksi record yang mirip satu sama lain, dan tidak mirip dengan record pada cluster. Tidak seperti klasifikasi, pada clustering tidak ada variabel target. Clustering tidak mengklasifikasi atau mengestimasi atau memprediksi tetapi mencari untuk membagi seluruh data set ke subgroup yang relative sejenis atau cluster, dimana kemiripan record di dalam cluster dimaksimalkan dan kemiripan dengan record di luar cluster diminimalkan. Contoh clustering, untuk akunting dengan tujuan audit untuk mensegmentasi financial behaviour ke dalam kategori ramah dan curiga, sebagai alat reduksi dimensi ketika data set memiliki ratusan atribut, untuk clustering ekspresi gen, dimana kuantitas gen bisa terlihat mempunyai behavior yang mirip. Algoritma untuk clustering antara lain, hierarchical agglomerative clustering, Bayesian clustering, self-organizing feature maps, growing hierarchical self-organizing [14].

6. Asosiasi

Tugas asosiasi untuk data mining adalah kegiatan untuk mencari atribut yang “go together.” Dalam dunia bisnis, asosiasi dikenal sebagai affinity analysis atau market basket analysis, tugas asosiasi adalah membuka rules untuk pengukuran hubungan antara dua atribut atau lebih. Contoh asosiasi, prediksi degradasi dalam jaringan komunikasi, menemukan barang apa di supermarket yang dibeli bersama dengan barang lain yang tidak pernah dibeli bersama, menemukan proporsi kasus dimana obat baru akan memperlihatkan efek samping yang berbahaya.

Untuk menemukan association rules, bisa dilakukan dengan algoritma a priori dan algoritma GRI (Generalized Rule Induction) [8].

Model dari algoritma klasifikasi dalam data mining antara lain:

1. Decision/classification trees

Algoritma *decision tree* yang dikenal luas antara lain *Hunt*, *CART (C&RT)*, *ID3*, *C4.5&C5.0*, *SLIQ*, *SPRINT*, *QUEST*, *DTREG*, *THAID*, *CHAID*

2. Bayesian classifiers/ Naïve Bayes classifiers

Klasifikasi Bayes juga dikenal dengan *Naïve Bayes*, *idiot's Bayes*, *simple Bayes*, dan *independence Baye*.

3. Neural networks

Dalam neural network terdapat model *Support Vector Machine (SVM)*

4. Analisa Statistik

Beberapa metode estimasi antara lain *linear regression* dan *multiple regression*.

Metode klasifikasi yang akan digunakan dalam penulisan ini yaitu decision tree, naïve bayes dan neural network.

2.1. Algoritma C4.5

Algoritma C4.5 merupakan struktur pohon dimana terdapat simpul yang mendeskripsikan atribut-atribut, setiap cabang menggambarkan hasil dari atribut yang diuji, dan setiap daun menggambarkan kelas. Algoritma C4.5 secara rekursif mengunjungi setiap simpul keputusan, memilih pembagian yang optimal, sampai tidak bisa dibagi lagi. Algoritma C4.5 menggunakan konsep *information gain* atau *entropy reduction* untuk memilih pembagian yang optimal [4]. Tahapan dalam pembuatan sebuah pohon keputusan dengan algoritma C4.5 [13] yaitu:

1. Menentukan atribut / data training. Atribut biasanya diambil dari data histori yang pernah terjadi sebelumnya yang bisa memisahkan obyek menurut kelasnya. Atribut yang dipilih adalah atribut yang menghasilkan simpul yang paling bersih. Semakin bersih suatu cabang semakin baik.
2. Menentukan atribut mana yang akan dipilih terlebih dahulu, dengan cara memilih atribut yang menghasilkan *information gain* paling besar. Sebelum menghitung *information gain* terlebih dahulu menghitung entropi yaitu :

$$Entropi(y) = \sum_{i=1}^n -p_i * \log_2 p_i$$

Keterangan:

y : sekumpulan obyek dengan label/output y

n : jumlah partisi S

p_i : proporsi dari S_i terhadap S

3. Hitung nilai Gain dengan metode *information gain* :
4. Untuk menghitung *information gain* dari output data atau dependent y yang dikelompokkan berdasarkan atribut A, dinotasikan dengan gain (y,A).

$$\sum_{i=0}^n \frac{|yc|}{|y|}$$

5. Gain(y,A)= Entropi(y)- entropi (y_c)

Keterangan:

y : sekumpulan obyek dengan label/output y

A : atribut

n : jumlah nilai atribut

|y_c| : jumlah kasus pada masing-masing pecahan atribut

|y| : jumlah label/output

6. Ulangi langkah ke-2 hingga semua atribut terpisah atau terpecah.

7. Kriteria pemisahan obyek di setiap simpul berdasarkan information gain yang paling besar.

2.2. Naive Bayes

Naive Bayes, yang juga disebut *idiot's Bayes*, *simple Bayes*, dan *independence Bayes*, adalah metode yang baik karena mudah dibuat, tidak membutuhkan skema estimasi parameter perulangan yang rumit, ini berarti bisa diaplikasikan untuk *data set* berukuran besar [14].

Penggunaan teorema Bayes pada algoritma Naive Bayes yaitu dengan mengkombinasikan *prior probability* dan probabilitas bersyarat dalam sebuah rumus yang bisa digunakan untuk menghitung probabilitas tiap klasifikasi yang mungkin. Klasifikasi Bayes didasarkan pada teorema Bayes, yaitu

$$P(x|y) = \frac{P(y|x)P(x)}{P(y)}$$

keterangan :

y = data dengan kelas yang belum diketahui

x = hipotesis data y merupakan suatu kelas spesifik

$P(x|y)$ = probabilitas hipotesis x berdasar kondisi y (*posteriori probability*)

$P(x)$ = probabilitas hipotesis x (*prior probability*)

$P(y|x)$ = probabilitas y berdasarkan kondisi pada hipotesis x

$P(y)$ = probabilitas dari y

Setelah dilakukan klasifikasi model data, maka tahap selanjutnya melakukan pengujian data untuk memprediksi akurasi data uji. Metode yang digunakan untuk menganalisa model klasifikasi Confusion Matrix dan Kurva ROC.

2.3. Confusion Matrix

Confusion matrix merupakan metode yang digunakan mengevaluasi prediksi benar atau tidak data uji berdasarkan model klasifikasi [1]. Evaluasi model ini menggunakan tabel berikut ini:

Tabel 1. Tabel *Performance* Estimasi

Parameter	Maksud
<i>True Positive</i> (TP)	Pelanggan Aktif
<i>False Positive</i> (FP)	Pelanggan Aktif tetapi tidak Loyal
<i>True Negative</i> (TN)	Tidak Aktif dan Loyal
<i>False Negative</i> (FN)	Tidak Aktif dan Tidak Loyal

Sumber: Hasil Penelitian (2014)

Data uji dimasukkan ke dalam *confusion matrix*, hitung nilai-nilai yang telah dimasukkan tersebut untuk menghitung jumlah *sensitivity*, *specificity*, *precision* dan *accuracy*. *Sensitivity* digunakan untuk membandingkan jumlah *true positives* terhadap jumlah tupel yang *positives* sedangkan *specificity* adalah perbandingan jumlah *true negatives* terhadap jumlah tupel yang *negatives*. Untuk menghitung digunakan persamaan di bawah ini:

$$sensitivity = \frac{t_pos}{pos}$$

$$specificity = \frac{t_neg}{neg}$$

$$precision = \frac{t_pos}{t_pos + f_pos}$$

$$specificity = \frac{t_neg}{neg}$$

$$accuracy = sensitivity \frac{pos}{(pos+neg)} + specificity \frac{neg}{(pos+neg)}$$

dimana:

t_pos = jumlah *true positives*

t_neg = jumlah *true negatives*

pos = jumlah tupel positif

neg = jumlah tupel negatif

f_pos = jumlah *false positives*

Untuk menganalisa hasil prediksi klasifikasi dapat dilihat pada AUC (Area Under the ROC Curve). Penentuan hasil prediksi klasifikasi dilihat dari batasan nilai AUC sebagai berikut (Gorunescu, 2011):

a. Nilai AUC 0.90-1.00 = *excellent classification*

b. Nilai AUC 0.80-0.90 = *good classification*

c. Nilai AUC 0.70-0.80 = *fair classification*

d. Nilai AUC 0.60-0.70 = *poor classification*

III. METODE PENELITIAN

Metode pengumpulan data yang digunakan oleh penulis dalam memprediksi loyalitas pelanggan adalah:

a. Data Primer

Adalah sumber data yang diperoleh secara langsung dari sumber, misalnya data-data pelanggan, aktivitas transaksi dan hasil interview.

b. Data Sekunder

Merupakan data yang diperoleh secara tidak langsung, misalnya dari dokumentasi, literatur, buku, jurnal, dan informasi lainnya yang ada hubungannya dengan masalah yang diteliti.

Data dikumpulkan mulai dari bulan april sampai juli 2011 dengan jumlah data sebanyak 643 data. Data historis tersebut digunakan sebagai data training.

Data training tersebut akan melalui *preprocessing* dimana:

- Data training di *cleansing* digunakan untuk membersihkan nilai yang kosong, tidak konsisten atau ada data yang kosong (*missing value* dan *noisy*).
- Kemudian data akan di *integration* digunakan untuk menyatukan tempat penyimpanan yang berbeda-beda ke dalam satu tempat penyimpanan yang sama. Dalam hal ini data diambil dari data *warehouse* yaitu data pelanggan dan data transaksi penjualan.
- Data *reduction*, dimana jumlah data tupel yang digunakan untuk data training terlalu besar dan hanya

beberapa atribut yang diperlukan, sehingga atribut yang tidak perlu akan dihapus

IV. HASIL DAN PEMBAHASAN

Penerapan data mining pada penelitian sebelumnya tentang loyalitas pelanggan kebanyakan digunakan untuk menyelesaikan masalah penentuan pola. Untuk memprediksi model pelanggan digunakan algoritma klasifikasi data mining. Klasifikasi adalah metode pembelajaran, dengan menggunakan data set *sampel* data sehingga memperoleh *classifier* kecenderungan untuk memprediksi loyalitas pelanggan. Tujuan dari penelitian ini melakukan prediksi loyalitas pelanggan dengan menggunakan algoritma C4.5 dan Naive Bayes dan menentukan algoritma yang paling baik digunakan.

Dari data training yang ada terbentuk tabel berdasarkan kategori atribut yaitu:

Tabel 2. Kategori atribut data training

Atribut	Nilai Angka	Kategori
Area	Kota	Kota
	Pinggir kota	pinggirKota
	Desa	Desa
Lama berlangganan	< 1 tahun	Baru
	>= 1 tahun	Lama
Frekuensi transaksi	1-10	sering
	11-150	sedang
	>150	jarang
Saldo Deposit	<= Rp.10.000	minim
	>Rp.10.000,	cukup
	< Rp.100.000	
	>=Rp.100.000	banyak

Sumber: Hasil Penelitian (2014)

Dari data eksperimen akan diujikan dengan menggunakan metode 10-fold *cross-validation*, dimana data secara acak (*random*) akan dibagi menjadi 10 bagian. Pembagian menjadi 10 bagian merupakan metode yang paling tepat untuk mendapatkan estimasi terbaik menentukan kesalahan. Pengujian data diperlukan untuk memprediksi akurasi data uji.

Hasil penelitian dengan menggunakan algoritma C4.5 adalah:

1. Menghitung jumlah kasus class aktif dan jumlah class tidak aktif. Entropi dari semua kasus dan kasus yang dibagi berdasarkan atribut area, deposit, lama berlangganan,

frekuensi transaksi. Kemudian lakukan penyeleksian atribut dengan menghitung Gain tertinggi.

Baris Total kolom Entropi dihitung sebagai berikut:

$$Entropi (Total) = \left(- \frac{39}{54} * \log_2 \left(\frac{39}{54} \right) \right) + \left(- \frac{15}{54} * \log_2 \left(\frac{15}{54} \right) \right)$$

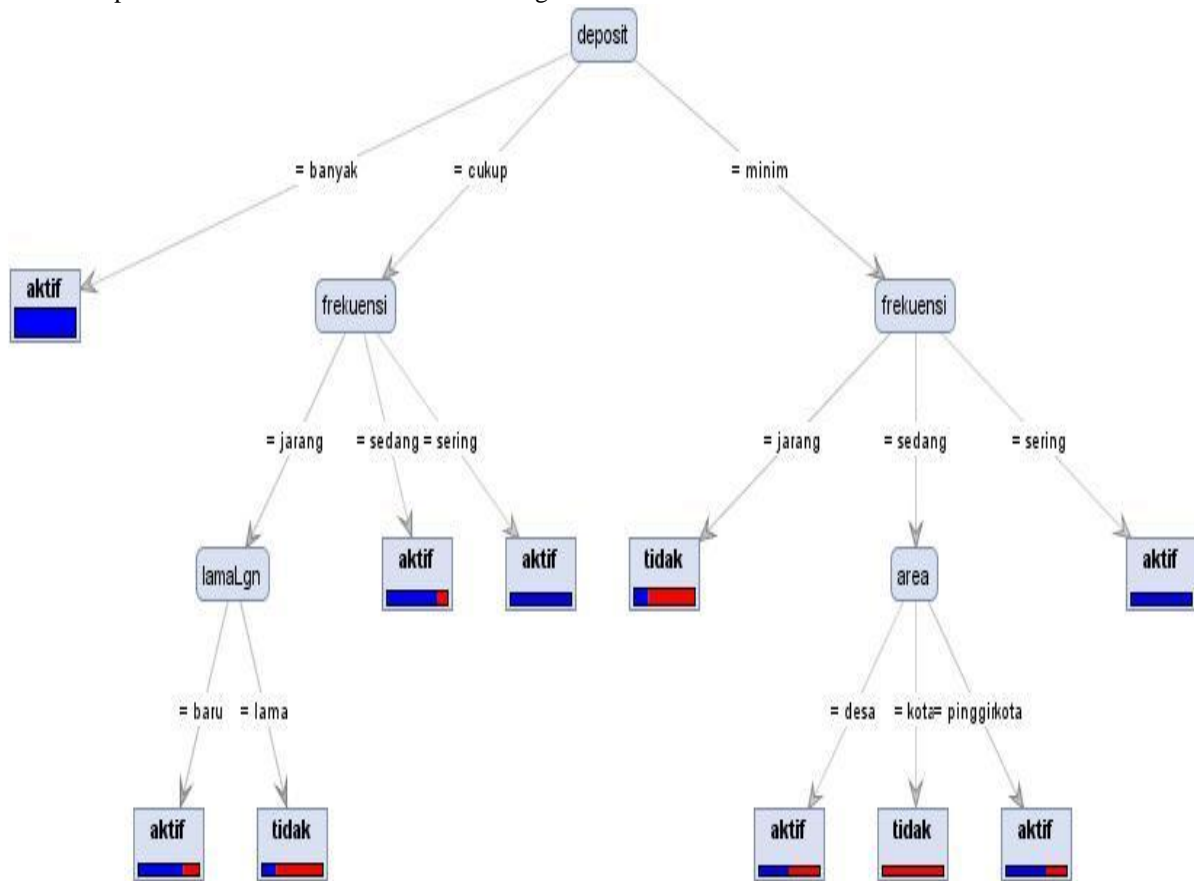
$$Entropi (Total) = 0.852$$

2. Dengan menggunakan persamaan Gain, dapat menghitung jumlah kasus dari masing-masing atribut dan hitung gain. Hasil perhitungan information gain yaitu:

Atribut area memiliki entropi = 0.8389722, information gain= 0.01343, atribut deposit memiliki entropi = 0.635283, information gain= 0.217121, atribut lama berlangganan

memiliki entropi = 0.831, information gain= 0.02, atribut frekuensi memiliki entropi = 0.68, information gain= 0.1715. Dengan demikian atribut yang menempati node tertinggi adalah atribut deposit karena memiliki information gain

tertinggi. Untuk perhitungan node selanjutnya menggunakan algoritma yang sama, sehingga dihasilkan pohon keputusan sebagai berikut:



Sumber: Hasil Penelitian (2014)

Gambar 1. Decision tree prediksi loyalitas pelanggan

Dari gambar 1 dapat diprediksi bahwa kriteria loyalitas pelanggan berdasarkan saldo deposit, frekuensi transaksi, lama berlangganan dan area tempat tinggal pelanggan.

Sedangkan dengan menggunakan naive bayes dari data training yang digunakan didapat

Tabel 3. Perhitungan data naive bayes

Class	jumlah (%)
Aktif	0.712
Tidak Aktif	0.288

Sumber: Hasil Penelitian (2014)

Berdasarkan tabel 3 diketahui bahwa prediksi loyalitas pelanggan menggunakan naive bayes dikategorikan baik dikarenakan jumlah pelanggan aktif diatas 70%.

Dari dua model tersebut didapat hasil sebagai berikut:

Tabel 4. Perbandingan akurasi C4.5 dan Naive Bayes

Model	TP	FP	TN	FN	Akurasi	AUC
Algoritma C 4.5	39	3	4	13	88.14%	0.954
Naive Bayes	38	2	5	14	88.02%	0.947

Sumber: Hasil Penelitian (2014)

Berdasarkan tabel 4 ke dua model tersebut mempunyai nilai AUC antara 0.8 sampai 0.9 yang menunjukkan bahwa hasil prediksi dari data yang ada menunjukkan *Excellent classification*.

V. KESIMPULAN

Dari penulisan ini dapat ditarik kesimpulan atas dasar melakukan prediksi pelanggan menggunakan dua teknik klasifikasi algoritma c4.5 dan naive bayes. Dari hasil prediksi

didapat kriteria-kriteria untuk mendapatkan pelanggan yang potensial sebagai acuan pengambilan keputusan manajemen. Komparasi dari dua teknik tersebut dilakukan untuk menemukan metode yang paling baik digunakan untuk memprediksi loyalitas pelanggan pada distributor pulsa elektronik. Metode klasifikasi yang berbeda memiliki cara dan pengetahuan masing-masing dalam menyelesaikan masalah. Kesimpulan dari penelitian ini adalah:

1. Untuk memprediksi loyalitas pelanggan menggunakan metode klasifikasi Algoritma C4.5 mempunyai nilai akurasi lebih baik dibandingkan dengan Naive Bayes yaitu 88.14%
2. Representasi dari algoritma C4.5 mudah dimengerti dibandingkan dengan Naive Bayes akan tetapi output dari algoritma C4.5 berupa kategori dari atribut.
3. Kedua model tersebut dapat digunakan untuk klasifikasi pada database dengan jumlah yang besar

Berdasarkan kesimpulan tersebut, penulis memberikan saran tentang metode algoritma C4.5 yang merupakan metode yang baik digunakan namun ada beberapa hal yang harus ditambahkan antara lain:

1. Sebaiknya perlu dilakukan komparasi dengan metode klasifikasi data mining yang lain seperti Neural Network, SVM, dan lain-lain. Sehingga akan didapatkan suatu metode yang paling baik untuk memprediksi loyalitas pelanggan pada distributor pelanggan.
2. Perlu adanya penelitian lanjutan untuk mengembangkan penelitian dengan menambah atribut dari model yang digunakan.
3. Untuk mendukung berjalannya sistem informasi diperlukan hardware dan infrastruktur yang mendukung terhadap proses pengambilan keputusan.

REFERENSI

- [1] Gorunescu. Data Mining Concepts, Models and Techniques. Romania: Springer-Verlag Berlin Heidelberg. 2011.
- [2] Govindaraju, R & Simatupang, T & Samadhi, Ari. Perancangan Sistem Prediksi Churn Pelanggan PT. Telekomunikasi Seluler Dengan Memanfaatkan Proses Data Mining. Jurnal Informatika Vol 9 No 1, 33-42. 2008.
- [3] Griffin, Jill. Customer Loyalty. Jakarta: Erlangga. 2005
- [4] Han, J & Kamber, M. Data Mining Concept and Techniques. India: New Age International Limited. 2006
- [5] Jahromi, A. Tamaddoni & Moesini, Mehrad & Akbari, Issrar. A Dual-Step Multi-Algorithm Approach For Churn Prediction In Pre-Paid Telecommunications Service Providers. Journal of Innovation and Sustainability. ISSN:2179-3565. 2010
- [6] Jin-hui, Lei & Jian-jun, He. The Comparative analysis and study of Mobile-based Customer Data Churn Prediction Model. World Congress on Software Engineering. IEEE:978-0-7695-3570-8/09. 2009
- [7] Kraljevic, Goran & Gotovac, Sven. Modelling Data Mining Applications for Prediction of Prepaid Churn in Telecommunication Services. ATKAFF 51(3), 275-283. ISSN 0005-1144. 2010
- [8] Larose, D. T. Discovering Knowledge in Data. New Jersey: John Wiley & Sons, Inc. 2005.

- [9] Liao T, Warren. Recent Advances in Data Mining of Enterprise Data Algorithms and Applications. Louisiana: World Scientific. 2007
- [10] Moore, Patricia. Menguasai CRM. Jakarta: Prestasi Pustaka Publisher. 2005
- [11] Oseman, Khalida. Data Mining in Churn Analysis Model for Telecommunication Industry. Journal of Statistical and Analytics Vol 1 No 19-27. 2010
- [12] Rahmayanty, Nina. Manajemen Pelayanan Prima. Yogyakarta: Graha Ilmu. 2010
- [13] Santoso, Budi. Data Mining Teknik Pemanfaatan Data Untuk Keperluan Bisnis. Yogyakarta: Graha Ilmu. 2007
- [14] Wu, Xindong & Kumar, Vipin. The Top Ten Algorithms in Data Mining. Boca Raton: CRC Press. 2009



Eni Irfiani, M.Kom. Tahun 1999 Lulus dari Program Strata Satu (S1) Program Studi Sistem Informasi. Universitas Gunadarma. Tahun 2011 lulus dari Program Strata Dua (S2) Magister Ilmu Komputer STMIK Nusa Mandiri. Aktif menulis paper di beberapa jurnal antara lain Jurnal Paradigma, Jurnal Sistem Informasi. Aktif sebagai anggota Asosiasi Perguruan Tinggi Informatika dan Komputer (APTIKOM).