

PREDIKSI TINGKAT LOYALITAS PELANGGAN MENGGUNAKAN ALGORITMA C4.5 BERBASIS BACKWARD ELIMINATION

Syaifuddin¹, Purwanto², Catur Supriyanto³

¹²³Pascasarjana Teknik Informatika Universitas Dian Nuswantoro

ABSTRACT

Customer loyalty is one of the capital to maintain the company's business strategy in the long run. In the last two decades of Customer Relationship Management (CRM) has grown to become one of the major trends in marketing, both in education and in the world practice. CRM is a comprehensive business strategy of a company that enables the company to effectively manage the company's relationship with the customer. Automatic feature selection algorithm is used with the aim of selecting a subset of the features in the dataset in order to reach the maximum level of accuracy in classification. The use of data mining techniques to predict customer loyalty combines C4.5 algorithm with feature selection Backward Elimination. C4.5 algorithm based backward elimination can improve the accuracy in the prediction of customer loyalty, compared with C4.5 algorithm without feature selection. C4.5 algorithm based backward elimination generate income per month attribute, type of subscription, registration fee, the cost of the bill, and the old subscription.

Keywords: CRM, C4.5 algorithm, Backward Elimination

1. PENDAHULUAN

1.1. Latar Belakang

Loyalitas pelanggan merupakan salah satu modal strategi untuk mempertahankan bisnis perusahaan dalam waktu jangka panjang. Dalam dua dekade terakhir Customer Relationship Management (CRM) telah berkembang menjadi salah satu tren utama dalam pemasaran, baik dalam dunia pendidikan maupun dalam dunia praktek. *Customer Relationship Management* (CRM) adalah sebuah strategi bisnis menyeluruh dalam suatu perusahaan yang memungkinkan perusahaan tersebut secara efektif mengelola hubungan dengan para pelanggan. Pelanggan adalah suatu aset perusahaan untuk mendapatkan keuntungan dari hasil penjualan, perusahaan harus fokus terhadap pelanggan dan membuat setiap divisi untuk mempertahankan mereka, hal ini menjadi penting karena mempertahankan pelanggan menjadi lebih murah dibandingkan mencari pelanggan baru.

Algoritma seleksi fitur otomatis digunakan dengan tujuan memilih subset dari fitur yang ada di dataset sehingga tercapai tingkat akurasi maksimal dalam klasifikasi. Penggunaan teknik data mining dalam memprediksi loyalitas pelanggan menggabungkan algoritma C4.5 dengan seleksi fitur *Backward Elimination*. Algoritma C4.5 berbasis backward elimination mampu meningkatkan akurasi dalam prediksi loyalitas pelanggan, dibandingkan dengan algoritma C4.5 tanpa seleksi fitur. Algoritma C4.5 berbasis *backward elimination* menghasilkan atribut penghasilan per bulan, jenis langganan, biaya registrasi, biaya tagihan, dan lama berlangganan.

1.2. Rumusan Masalah

Berdasarkan latar belakang penelitian yang telah diuraikan maka yang menjadi masalah dalam penelitian ini adalah :

- Pengolahan data pada PT. Cipta visual prima belum sepenuhnya memanfaatkan fasilitas komputer secara maksimal untuk prediksi loyalitas pelanggan.
- Algoritma C4.5 belum maksimal dalam prediksi loyalitas pelanggan tanpa menggunakan fitur seleksi *Backward Elimination*.

1.3. Tujuan Penelitian

Berdasarkan rumusan masalah maka tujuan dari penelitian ini adalah :

- Memanfaatkan fasilitas komputer untuk pengolahan data PT. Cipta Visual Prima lebih mudah prediksi loyalitas pelanggan.
- Menghasilkan model yang lebih akurat dengan menggunakan fitur seleksi *Backward Elimination* dalam prediksi loyalitas pelanggan.

1.4. Manfaat Penelitian

Berdasarkan tujuan penelitian maka tujuan dari penelitian ini adalah :

- Memudahkan PT. Cipta visual prima dalam prediksi loyalitas pelanggan untuk menentukan pelanggan yang bisa dipertahankan dan pelanggan yang tidak bisa dipertahankan.
- Didapatkan hasil akurat dalam prediksi loyalitas pelanggan menggunakan algoritma C4.5 berbasis *Backward Elimination*.

2. LANDASAN TEORI

2.1. Algoritma C4.5

Ada beberapa tahap dalam membuat pohon keputusan yaitu :

- Menyiapkan data *training*. Data training biasanya diambil dari data histori yang pernah terjadi sebelumnya dan sudah dikelompokkan ke dalam kelas – kelas tertentu.
- Menentukan akar dari pohon. Dalam pembentukan pohon keputusan menggunakan metode algoritma *decision tree*, akar diambil dari atribut yang terpilih dengan cara menghitung nilai gain dari masing – masing atribut, sebelum menghitung nilai gain dari atribut, hitung dahulu nilai *entropy* dengan menggunakan rumus :

$$Entropy(S) = - \sum_{i=1}^n p_i * \log_2 p_i \dots\dots\dots (2.1)$$

Keterangan :

- S = Himpunan kasus
 n = Jumlah partisi S
 P_i = Proporsi dari S_i terhadap S

- Kemudian menghitung nilai gain dengan metode *information gain* :

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \dots\dots\dots (2.2)$$

Keterangan :

- S = Himpunan kasus
 A = Atribut
 n = Jumlah partisi atribut A
 |S_i| = Jumlah kasus pada partisi ke-i
 |S| = Jumlah kasus dalam S

$$SplitInfo(S, A) = - \sum_{i=1}^n \frac{S_i}{S} \log_2 \frac{S_i}{S} \dots\dots\dots (2.3)$$

Keterangan :

S = Himpunan kasus

A = Atribut

c = Jumlah partisi atribut A

S_i = Jumlah kasus pada partisi ke- i

S = Jumlah kasus dalam S

Selanjutnya *gain ratio* dihitung dengan cara [18] :

$$Gain\ Ratio\ (S, A) = \frac{Gain(S, A)}{SplitInfo(S, A)} \dots\dots\dots (2.4)$$

Keterangan :

S = Himpunan kasus

A = Atribut

2.2. Confusion Matrix

Tabel 1. *Confusion Matrix*

		Predicted Class	
		Class=Yes/Positive	Class=No/Negative
Observed Class	Class=Yes/Positive	A	b
		(TP=True Positive)	(FN=False Negative)
	Class=No/Negative	C	d
		(FP=False Positive)	(TN=True Negative)

Keterangan :

TP = Prediksi positif yang positif

FN = Prediksi positif yang negatif

FP = Prediksi negatif yang positif

TN = Prediksi negatif yang negatif

Untuk menghitung nilai *accuracy*, *precision*, dan *recall* dengan rumus perhitungan sebagai berikut.

$$Accuracy = \left(\frac{a+d}{a+b+c+d} \right) = \frac{TP+TN}{TP+TN+FP+FN} \dots\dots\dots (2.7)$$

Precision didefinisikan sebagai rasio item relevan yang dipilih terhadap semua item yang terpilih.

$$Precision = \frac{TP}{TP+FP} \dots\dots\dots (2.8)$$

Sedangkan *Recall* didefinisikan sebagai rasio dari item relevan yang dipilih terhadap total jumlah item relevan yang tersedia.

$$Recall = \frac{TP}{TP+FN} \dots\dots\dots (2.9)$$

3. METODE PENELITIAN

3.1. Pengumpulan Data

Data yang digunakan pada penelitian ini dikumpulkan berdasarkan data pelanggan PT. Cipta Visual Prima Kabupaten Bantaeng, perusahaan tersebut bergerak dibidang jasa layanan TV kabel. Data yang didapatkan

adalah berupa data mentah yang didalamnya terdapat 1053 record, dalam dataset yang digunakan dalam penelitian ini berisi tentang informasi pelanggan jasa layanan TV kabel.

Beberapa variabel yang terdapat dalam dataset tersebut adalah :

- a. Nama Pelanggan
- b. Lokasi Pemasangan
- c. Lokasi Penagihan
- d. Start Langganan
- e. Registrasi Pelanggan (Paralel, Non Paralel)
- f. Biaya Registrasi (Rp. 500.000, Rp. 300.000,-)
- g. Jenis Langganan (Paralel, Non Paralel)
- h. Biaya Tagihan (Rp. 35.000, Rp. 25.000)
- i. Lama Berlangganan
- j. Penghasilan Per Bulan:
 - 1) < Rp 1.000.000,-
 - 2) < Rp 2.000.000,-
 - 3) > Rp 2.000.000,-
- k. Status Langganan (Aktif, Tidak Aktif)

3.2. Pengolahan Awal

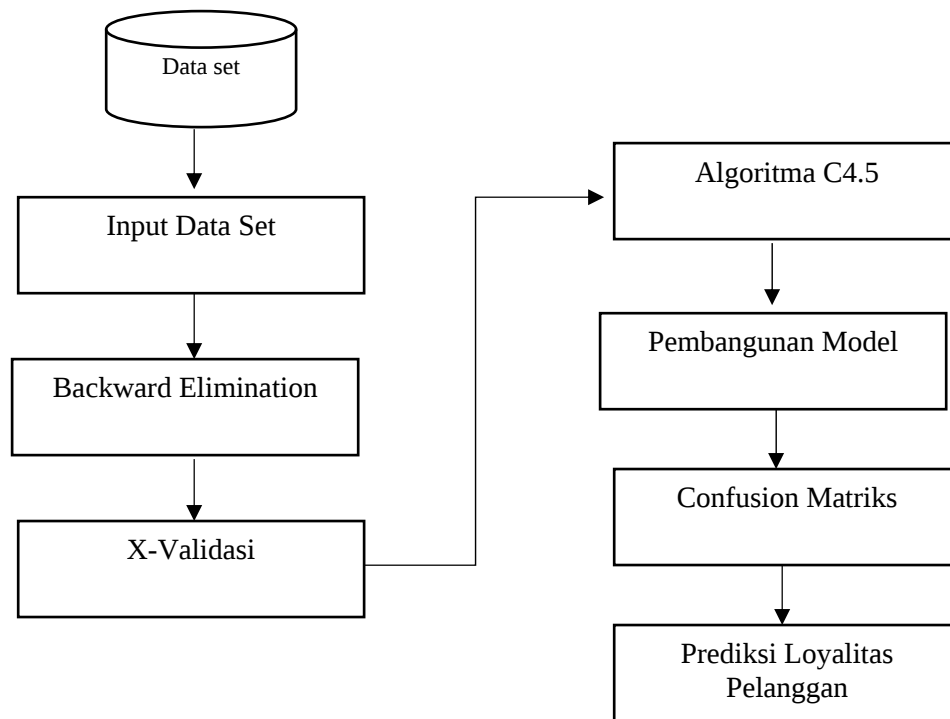
Dataset yang digunakan pada penelitian ini adalah data set perusahaan yang didapat dari salah satu perusahaan layanan TV kabel di Bantaeng yaitu PT. Cipta visual prima dataset akan melalui tahap *Preprocessing*. Dalam tahap ini dilakukan pembersihan data terhadap beberapa atribut yang ada.

Variabel start berlangganan dihilangkan dimana nilai atribut tersebut bulan dimana pelanggan mulai berlangganan akan dirubah menjadi lama berlangganan. Variabel registrasi langganan dihilangkan dimana nilai untuk variabel registrasi langganan paralel dan non paralel yang mewakili nilai dari variabel jenis langganan sedangkan untuk tunggakan memiliki nilai jumlah tunggakan dalam hitungan bulan yang tidak konsisten dalam penentuan status langganan, sehingga untuk pengolahan awal didapatkan 6 variabel independen dengan 1 variabel dependen seperti contoh dataset dibawah ini :

- a. Jenis Kelamin
- b. Penghasilan Per bulan
- c. Jenis Langganan
- d. Biaya Registrasi
- e. Biaya Tagihan
- f. Lama Berlangganan
- g. Status berlangganan (Class).

3.3. Eksperimen

Hasil pengolahan awal yang ada selanjutnya diuji menggunakan algoritma C4.5 berbasis *backward elimination*, proses uji dataset dilakukan dengan memanfaatkan *tools Rapid Miner*. Model yang diusulkan menggambarkan alur metode yang diusulkan serta menjelaskan cara kerja model yang diusulkan. Model akan dibentuk dari data yang sudah diolah, dan hasil pengelolaan model akan diukur dengan model yang ada saat ini.



Gambar 1. Eksperimen

4. HASIL DAN PEMBAHASAN

4.1. Preprocessing

Data yang digunakan pada penelitian ini dikumpulkan berdasarkan data pelanggan PT. Cipta Visual Prima Kabupaten Bantang, perusahaan tersebut bergerak dibidang jasa layanan TV kabel. Data yang didapatkan adalah berupa data mentah yang didalamnya terdapat 1053 record.

Dalam tahap ini dilakukan pembersihan data terhadap beberapa atribut yang ada, dimana variabel start berlangganan dihilangkan dimana nilai atribut tersebut bulan dimana pelanggan mulai berlangganan, dan dirubah menjadi lama berlangganan. Variabel registrasi langganan dihilangkan dimana nilai untuk variabel registrasi langganan paralel dan non paralel yang mewakili nilai dari variabel jenis langganan sedangkan untuk tunggakan memiliki nilai jumlah tunggakan dalam hitungan bulan yang tidak konsisten dalam penentuan status langganan, sehingga untuk pengolahan awal didapatkan 6 variabel independen dengan 1 variabel dependen.

- a. Jenis Kelamin
- b. Penghasilan Per bulan
- c. Jenis Langganan
- d. Biaya Registrasi
- e. Biaya Tagihan
- f. Lama Berlangganan
- g. Status berlangganan (Class).

4.2. Eksperimen Menggunakan Algoritma C4.5

Eksperimen diawali dengan proses perhitungan nilai *entropy* total data set, selanjutnya menghitung nilai *information gain*, untuk mendapatkan nilai *gain ratio* untuk masing – masing atribut perlu dilakukan perhitungan *split info*, dimana nilai *information gain* akan dibagi dengan nilai *split info* masing masing atribut. Nilai *gain ratio* tertinggi akan menjadi akar simpul dari pohon keputusan.

Perhitungan nilai *entropy* total data set menggunakan persamaan 2.1 dengan penjelasan dibawah ini :

Aktif = 927 Record

Tidak Aktif = 126 Record

Entropy Aktif = $(-927/1053) \cdot \text{LOG}(927/1053;2) = 0,161863349$

Entropy Tidak Aktif = $(-126/1053) \cdot \text{LOG}(126/1053;2) = 0,366513993$

Entropy Total Aktif dan tidak aktif

= $0,161863349 + 0,366513993 = 0,528377342$

Perhitungan nilai *entropy* masing – masing atribut dilakukan menggunakan persamaan 2.1

Tabel 2. Perhitungan Kasus Atribut Penghasilan

Penghasilan	Status_Langganan	Jumlah
>2000000	AKTIF (+)	503
>2000000	TIDAK AKTIF (-)	20
<2000000	AKTIF (+)	356
<2000000	TIDAK AKTIF (-)	40
<1000000	AKTIF (+)	68
<1000000	TIDAK AKTIF (-)	66

$>2000000 = (-503/523) \cdot \text{LOG}(503/523;2) + (-20/523) \cdot \text{LOG}(20/523;2) = 0,2341679$

$<2000000 = (-356/396) \cdot \text{LOG}(356/396;2) + (-40/396) \cdot \text{LOG}(40/396;2) = 0,472189385$

$<1000000 = (-68/134) \cdot \text{LOG}(68/134;2) + (-66/134) \cdot \text{LOG}(66/134;2) = 0,99983930$

Setelah mendapatkan hasil *entropy* dari masing – masing nilai atribut penghasilan langkah selanjutnya adalah menghitung nilai *gain* atribut penghasilan menggunakan persamaan 2.2 sebagai berikut :

Gain = $(0,528377342) - ((523/1053) \cdot 0,2341679 + (396/1053) \cdot 0,472189385 + (134/1053) \cdot 0,9998393) = 0,107261222$

Perhitungan *split info* menggunakan persamaan 2.3

Split info = $(-523/1053) \cdot \text{LOG}(523/1053;2) + (-396/1053) \cdot \text{LOG}(396/1053;2) + (-134/1053) \cdot \text{LOG}(134/1053;2) = 0,308610156$

Menghitung *gain ratio* untuk menentukan atribut sebagai akar dari pohon keputusan menggunakan persamaan 2.4

Gain Ratio (Total, Penghasilan) = $0,107261222 / 0,308610156 = 0,347562191$

Perhitungan nilai *entropy* atribut biaya tagihan menggunakan persamaan 2.1

Tabel 3. Perhitungan Kasus Atribut Biaya Tagihan

Biaya Tagihan	Status_Langganan	Jumlah
Rp. 35.000	AKTIF (+)	148
Rp. 35.000	TIDAK AKTIF (-)	10
Rp. 25.000	AKTIF (+)	779
Rp. 25.000	TIDAK AKTIF (-)	116

$$Rp.35.000=(-148/158)*\text{LOG}(148/158;2)+(-10/158)*\text{LOG}(10/158;2)=0,340373286$$

$$Rp.25.000=(-779/895)*\text{LOG}(779/895;2)+(-116/895)*\text{LOG}(116/895;2)=0,556364721$$

Setelah mendapatkan hasil *entropy* dari masing – masing nilai atribut biaya tagihan langkah selanjutnya adalah menghitung nilai gain atribut biaya tagihan menggunakan persamaan 2.2 sebagai berikut :

$$\text{Gain (Total, Biaya Tagihan)}=0,528377342-((158/1053)*0,340373286+(895/1053) *0,556364721) = 0,004421592$$

Perhitungan nilai *split info* menggunakan persamaan 2.3

$$\text{Split info (total, Biaya Tagihan)} =(-158/1053)*\text{LOG}(158/1053;2)+(-895/1052)*\text{LOG} (895/1053;2) = 0,610148618.$$

Menghitung *gain ratio* menggunakan persamaan 2.4

$$\text{Gain Ratio (Total, Biaya Tagihan)} = 0,004421592/0,610148618 = 0,007246746$$

Perhitungan nilai *entropy* atribut lama berlangganan menggunakan persamaan 2.1, untuk nilai atribut berlangganan peneliti membuat 3 range untuk masing – masing nilai, nilai lama berlangganan 100 bulan keatas masuk range >100 Bulan, 50 bulan masuk range <100 bulan.

Tabel 4. Perhitungan Kasus Atribut Lama Berlangganan

Lama Berlangganan	Status_Langganan	Jumlah
< 100 Bulan	AKTIF (+)	757
< 100 Bulan	TIDAK AKTIF (-)	104
< 50 Bulan	AKTIF (+)	95
< 50 Bulan	TIDAK AKTIF (-)	15
> 100 Bulan	AKTIF (+)	75
> 100 Bulan	TIDAK AKTIF (-)	7

$$<100\text{Bulan}=(-757/861)*\text{LOG}(757/861;2)+(-104/861)*\text{LOG}(104/861;2)=0,531626809$$

$$< 50 \text{ Bulan }=(-95/110)*\text{LOG}(95/110;2)+(-15/110)*\text{LOG}(15/110;2) =0,574635698$$

$$> 100 \text{ Bulan }=(-75/82)*\text{LOG}(75/82;2)+(-7/82)*\text{LOG}(7/82;2) = 0,420809489$$

Setelah mendapatkan hasil *entropy* dari masing – masing nilai atribut lama berlangganan langkah selanjutnya adalah menghitung nilai gain atribut lama berlangganan menggunakan persamaan 2.2 sebagai berikut :

$$\text{Gain}= 0,528377342-((110/1053)*0,574635698+(861/1053)*0,531626809+ (82/1053)*(0,420809489)) = 0,000887325$$

Perhitungan nilai *split info* menggunakan persamaan 2.3

$$\text{Split info} = (-110/1053)*\text{LOG}(110/1053,2)+(-861/1053)*\text{LOG}(861/1053;2)+(-82/1053)*\text{LOG}(82/1053,2)= 0,426294421$$

Menghitung *gain ratio* menggunakan persamaan 2.4

$$\text{Gain Ratio (Total, lama berlangganan)} = 0,000887325/0,426294421 = 0,002081485$$

Perhitungan nilai *entropy* atribut jenis kelamin menggunakan persamaan 2.1.

Tabel 5. Perhitungan Kasus Atribut Jenis Kelamin

Jenis Kelamin	Status_Langganan	Jumlah
Laki - laki	AKTIF (+)	732
Laki - laki	TIDAK AKTIF (-)	101
Perempuan	AKTIF (+)	195
Perempuan	TIDAK AKTIF (-)	25

Laki – Laki $(-732/833) * \text{LOG}(732/833;2) + (-101/833) * \text{LOG}(101/833;2) = 0,532939022$

Perempuan $= (-195/220) * \text{LOG}(195/220;2) + (-25/220) * \text{LOG}(25/220;2) = 0,510787823$

Setelah mendapatkan hasil *entropy* dari masing – masing nilai atribut jenis kelamin, langkah selanjutnya adalah menghitung nilai gain atribut jenis kelamin menggunakan persamaan 2.2 sebagai berikut :

Gain (Total, Jenis Kelamin) = $0,528377342 - ((833/1053) * 0,532939022) + (220/1053) * (0,510787823) = 0,213500909$

Perhitungan nilai *split info* menggunakan persamaan 2.3

Split info (total, Jenis Kelamin) = $(-833/1053) * \text{LOG}(833/1053;2) + (-220/1053) * \text{LOG}(220/1053;2) = 0,739426488$

Menghitung *gain ratio* menggunakan persamaan 2.4

Gain Ratio (Total, Jenis Kelamin) = $0,213500909 / 0,739426488 = 0,288738518$

Perhitungan nilai *entropy* atribut jenis langganan menggunakan persamaan 2.1

Tabel 6. Perhitungan Kasus Atribut Jenis Langganan

Jenis Langganan	Status_Langganan	Jumlah
Paralel	AKTIF (+)	148
Paralel	TIDAK AKTIF (-)	10
Non Paralel	AKTIF (+)	779
Non Paralel	TIDAK AKTIF (-)	116

Paralel $= (-148/158) * \text{LOG}(148/158;2) + (-10/158) * (\text{LOG}(10/158;2)) = 0,340373286$

Non Paralel $= (-779/895) * \text{LOG}(779/895;2) + (-116/895) * \text{LOG}(116 / 895;2) = 0,556364721$

Setelah mendapatkan hasil *entropy* dari masing – masing nilai atribut jenis kelamin, langkah selanjutnya adalah menghitung nilai gain atribut jenis kelamin menggunakan persamaan 2.2 sebagai berikut :

Gain (Total, Jenis Langganan) = $0,528377342 - ((158/1053) * 0,340373286 + (895/1053) * 0,556364721) = 0,004421592$.

Perhitungan nilai *split info* menggunakan persamaan 2.3

Split info (total, Jenis Langganan) = $(-158/1053) * \text{LOG}(158/1053;2) + (-895/1052) * \text{LOG}(895/1053;2) = 0,610148618$.

Menghitung *gain ratio* menggunakan persamaan 2.4

Gain Ratio (Total, Jenis Langganan) = $0,004421592 / 0,610148618 = 0,007246746$

Perhitungan nilai *entropy* atribut biaya registrasi menggunakan persamaan 2.1.

Tabel 7. Perhitungan Kasus Atribut Biaya Registrasi

Biaya Registrasi	Status_Langgan	Jumlah
IDR 500,000	AKTIF (+)	148
IDR 500,000	TIDAK AKTIF (-)	10
IDR 300,000	AKTIF (+)	779
IDR 300,000	TIDAK AKTIF (-)	116

$$500.000 = (-148/158) * \text{LOG}(148/158;2) + (-10/158) * \text{LOG}(10/158;2) = 0,340373286$$

$$300.000 = (-779/895) * \text{LOG}(779/895;2) + (-116/895) * \text{LOG}(116/895;2) = 0,556364721$$

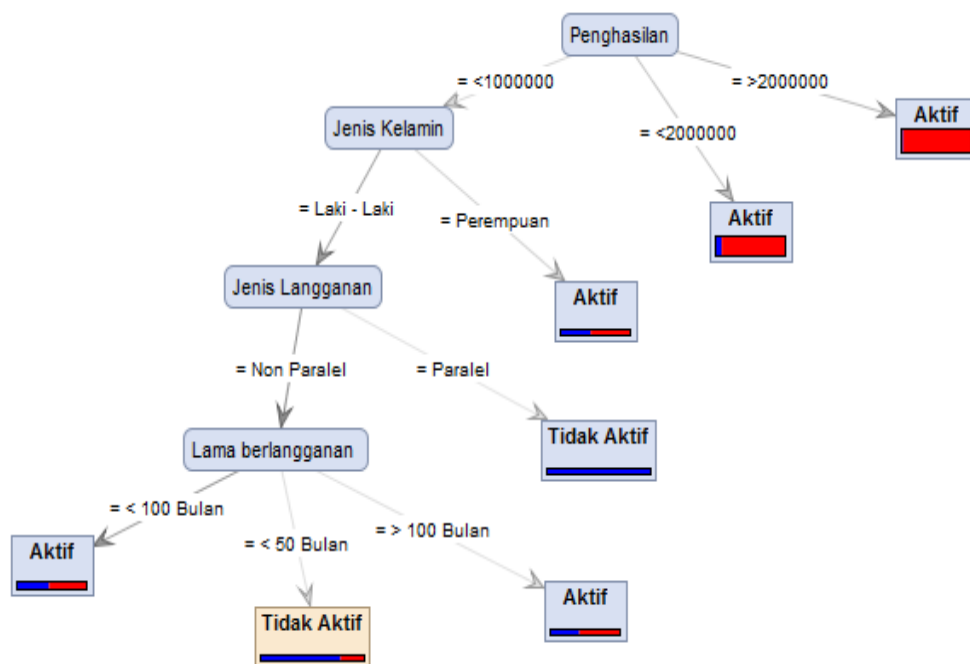
$$\text{Gain (Total, Biaya Registrasi)} = 0,528377342 - ((158/1053) * 0,340373286 + (895/1053) * 0,556364721) = 0,004421592$$

$$\text{Split info (total, Biaya Registrasi)} = (-158/1053) * \text{LOG}(158/1053;2) + (-895/1052) * \text{LOG}(895/1053;2) = 0,610148618$$

$$\text{Gain Ratio (Total, Biaya Registrasi)} = 0,004421592 / 0,610148618 = 0,007246746$$

Untuk nilai *gain ratio* yang paling tinggi akan menjadi node akar dari pohon. Dari hasil yang didapatkan nilai *gain ratio* tertinggi dimiliki oleh atribut penghasilan yaitu 0.07923, sehingga ditetapkan penghasilan menjadi akar dari pohon keputusan sesuai dengan tabel 4.8 dibawah ini :

Berdasarkan perhitungan *gain ratio* tertinggi, pohon keputusan yang terbentuk untuk node 1 seperti gambar di bawah ini.



Gambar 2. Pohon Keputusan Algoritma C4.5

5. KESIMPULAN DAN SARAN

5.1. Kesimpulan

Berdasarkan hasil dan pembahasan, maka dapat disimpulkan bahwa :

- a. Dalam penelitian ini dilakukan pengujian model dengan menggunakan algoritma *C4.5* berbasis *Backward Elimination* dengan menggunakan data pelanggan TV kabel . Model yang dihasilkan diuji untuk mendapatkan nilai *accuracy*, *precision* dan *Recall*. selain itu juga peneliti mengkomparasi dengan algoritma lain yaitu K-NN berbasis *Backward Elimination* yang mana hasilnya lebih bagus menggunakan algoritma *C4.5* berbasis *Backward Elimination*. Maka dapat disimpulkan pengujian model loyalitas pelanggan dengan menggunakan algoritma *C4.5* berbasis *Backward Elimination* lebih baik dari pada K-NN berbasis *Backward Elimination* dan *C4.5* sendiri.
- b. Dengan demikian dari hasil pengujian model diatas dapat disimpulkan bahwa *C4.5* berbasis *Backward Elimination* memberikan pemecahan untuk permasalahan prediksi loyalitas pelanggan lebih akurat.

5.2. Saran

Berdasarkan hasil penelitian dan pembahasan, maka beberapa saran yang dapat diberikan adalah sebagai berikut :

- a. Penggunaan teknik lain dalam memprediksi loyalitas pelanggan sangat dibutuhkan, sehingga dapat menghasilkan pola prediksi loyalitas pelanggan dengan tingkat akurasi yang lebih baik.
- b. Hasil prediksi loyalitas dapat menjadi kontribusi untuk perusahaan namun ke depan untuk memprediksi loyalitas pelanggan dapat menggunakan data kuisioner.

DAFTAR PUSTAKA

- [1] A. Keropyan and A. M Gil Lafuente, "Costumer loyalty programs to sustain consumer fidelity in mobile telecommunication market" *Expert System with applications*. vol. 39. no. 12, pp, 11269-11275, 2012
- [2] M.A.H. Farquad, Vadlamani Ravi, S. Bapi Raju, "Churn prediction using comprehensible *Support Vector Machine* : An analytical CRM application", *Applied Soft Computing*, 19, 31 – 40, 2014
- [3] Jonathan Burez, Dirk Van den Poel, "CRM at a pay – TV company : using analytical models to reduce customer attrition by targeted marketing for subscription service", *Expert System With Application*. Vol. 32, 277-288, 2007
- [4] Yance Sonatha, "Customer Churn Prediction Using CRISP – DM (Case Study : Customer of TelkomFlexi Bandung)", *Jurnal Elektron*, ISSN :2085-6989, Vol 5 No. 1, Edisi Desember 2013
- [5] Eko Prasetyo, Rr Ani Dijah Rahajoe, Arif Arizal, "Perbandingan K-Support Vector Nearest Neighbor Terhadap Decision Tree Dan Naive Bayes", *artitel*, UPN Jatim Repository, 2013.
- [6] Matthias Reif, Faisal Shahait, "Efficient Feature Size Reduction Via Predictive Forward Selection". *Pattern Recognition* ,47, 1664–1673. 2014.
- [7] L. Ladha, T.Deepa, "Feature Selection Methods And Algorithms", *International Journal on Computer Science and Engineering (IJCSE)*, ISSN : 0975-3397 Vol. 3 No. 5 May 2011.
- [8] Hasbi Yasin, "Pemilihan Variabel pada Model Geographically Weighted Regression" *Media Statistika*, Vol. 4, No. 2 : 63-72, Desember 2011.
- [9] Ali Narin, Yalcin Isler & Mahmut Ozer, "Investigating the Performance Improvement of HRV Indices in CHF Using Feature Selection Methods Based on Backward Elimination and Statistical Significance", *Computer in Biology and Medicine*, 2014.
- [10] X. Wu, V. Kumar, J. Ross Quinlan, J. Ghosh, Q. Yang, H. Motoda, G. J. Mc Lachlan, A. Ng, B. Liu, P. S. Yu, D. H. Zhou, M. Steinbach, D. J. Hand, and D. Steinberg, "Top 10 Algorithms in data mining", Vol. 14, No. 1, pp. 1-37, 2009.

- [11] Chih-Fong Tsai, Mao-Yuan Chen. "Variable selection by association rules for customer churn prediction of multimedia on demand", *Expert Systems with Applications*, 37, 2006–2015, 2010
- [12] GuangliNie, Wei Rowe, LinglingZhang, YingjieTian, Yong Shi. "Credit card churn forecasting by logistic regression and decision tree", *Expert Systems with Applications* 38,15273–15285, 2011.
- [13] Y. Richter, Elad Yom – Tov, N. Slonim , "Predicting customer churn in mobile networks through analisys of social groups", 'Proceedings Of The SIAM International Conference on Data Mining, pp. 732-741, 2010
- [14] S.M.S Hosseini, A. Maleki and M. R Gholamian, "Cluster analisys using data mining approach to develop CRM methodology to assess the customer loyalty", *Expert Systems with Applications*, Vol 37, No. 7 pp. 5259-5264, 2010.
- [15] Larose, D. T. "Discovering Knowledge In Data". Canada: Wiley- Interscience. 2005.
- [16] Susanto, S., &Suryadi, D. "Pengantar Data Mining menggaliPengetahuandariBongkahan Data". Yogyakarta: C.V ANDI OFFSET. 2010
- [17] Kusrini, & Luthfi, E. T. "Algoritma Data Mining". Yogyakarta: Andi Publishing. 2009.
- [18] Nurma Jayanti, Sulistyo Puspitodjati, Tety Elida. "TeknikKlasifikasiPohonKeputusanuntukMempredikasiKebangkrutan Bank BerdasarkanRasioKeuangan Bank". *Proceeding Seminar Ilmiah Nasional Komputer dan Sistem Intelegen (KOMMIT)*.Depok. 2008.
- [19] Raden Sofian Bahri, Irfan Maliki, "Perbandingan Algoritma Template Matching dan Feature Extraction Pada Optical Character Recognition", *Jurnal Komputer dan Informatika (KOMPUTA)*, Edisi. 1, Vol. 1, Maret 2012.
- [20] Novelysa. S, Partano. S, Pengarapen. B, " Analisa Metode Backward dan Metode Forward untuk Menentukan Persamaan Regresi Linier Berganda," *Saintia matematika. Vol. 2, No. 4*, 2014
- [21] K. Y. Chana, C. K. Kwong, T. S. Dillona, Y. C. Tsimb," Reducing Overfitting in Manufacturing Process Modeling Using a Backward Elimination Based Genetic Programming," *Applied Soft Computing*, 2011
- [22] Daniel T.Larose, "Data Mining: Methods and Models", *Published simultaneously*, 125, 2006
- [23] Ian H. Witten &Eibe Frank, Data Mining : Practical Machine *Learning Tools and Techniques, Second Edition. University of Waikato: Elsevier*, xxiii, 2005
- [24] Kohavi R, "A Study of Cross Validation and Bootstrap For Accuracy Estimation and Model Selection", *Lecture Note In Computer Science*, 114 – 124, 1995.
- [25] M.Bramer. "Principles of Data Mining". London : Springer - Verlag. 2007.