

Penerapan Optimasi Algoritma C45 dengan Naïve Bayes pada Pemilihan *Internet Service Provider*

I Made Darma Susila¹, Ida Bagus Suradarma²

Program Studi Sistem Komputer, Sekolah Tinggi Manajemen Informatika dan Komputer STIKOM Bali
Jalan Raya Puputan No 86 Renon - Denpasar, Tlp. (0361) 244445/Fax. (0361) 264773
e-mail: ¹darma_s@stikom-bali.ac.id, ²ibsuradarma@stikom-bali.ac.id

Abstrak

Internet Service Provider adalah sebuah badan usaha yang bergerak di bidang jasa pengelola internet. Sebagai sebuah badan usaha penyedia jasa internet yang bersifat komersial, ISP memiliki penilaian tersendiri dari calon customer. Penilaian yang biasanya digunakan sebagai tolak ukur pemilihan ISP oleh calon customer antara lain adalah penilaian service hotline, nilai Customer Interconnect Rate (CIR), Harga, Customer Satisfactions, SLA, dan Perangkat Last Mile. Permasalahan yang biasanya muncul pada calon customer adalah penentuan prioritas penilaian pemilihan terhadap beberapa parameter penilaian ISP. Hal yang dapat dilakukan adalah dengan menggunakan teknik klasifikasi untuk memprioritaskan penilaian dalam pemilihan ISP. Pada penelitian yang dilakukan, pemilihan algoritma klasifikasi C45 digunakan sebagai algoritma yang membentuk prioritas penilaian dalam pemilihan ISP dalam bentuk pohon keputusan. Hal ini didasarkan atas keunggulan algoritma C45 yang mampu memangkas struktur hierarki pohon keputusan terhadap parameter yang dimiliki dalam proses klasifikasi sehingga memudahkan dalam implementasi pengambilan keputusan. Untuk menangani adanya pembentukan akar yang memiliki nilai ganda pada perhitungan gain ratio di algoritma C45, pada penelitian yang dilakukan disempurnakan dengan konsep probabilitas yaitu Naïve Bayes. Hasil dari penelitian yang dilakukan adalah adanya model optimasi dari pohon keputusan yang dihasilkan oleh algoritma C45 dengan optimasi algoritma Naïve Bayes, dimana Naïve Bayes mampu memilih akar keputusan saat nilai dari perhitungan gain ratio pada C45 menunjukkan hasil nilai yang sama. Pada pengujian, dari data uji yang disiapkan, mampu ditelusuri keputusan dalam bentuk klasifikasi yang berdasarkan pada pembentukan pohon keputusan yang telah dibentuk oleh C45 dengan optimasi Naïve Bayes pada studi kasus pemilihan ISP.

Kata kunci: Algoritma C45, Naïve Bayes, *Internet Service Provider*.

Abstract

Internet Service Provider is a business engaged in the services of Internet management. As an Internet service provider business that is commercial, ISP has its own assessment of customer candidate. Assessments commonly used as a benchmark for ISP selection by prospective customers include service hotline assessments, Customer Interconnect Rate (CIR), Price, Customer Satisfactions, SLA and Last Mile Devices. The problem that arises in the prospective customer is the determination of the priority of the election assessment on the ISP assessment parameters. The solution for this case is using the classification techniques to prioritize assessments in the selection of ISPs. In this research, the selection of C45 classification algorithm is used as an algorithm that forms the priority of assessment in the selection of ISPs in the form of decision trees. This is based on the superiority of the C45 algorithm that is able to prune the hierarchical structure of the decision tree against the parameters held in the classification process so as to facilitate the implementation of decision making. To deal with the existence of root formation that has a double value in the calculation of gain ratio in the C45 algorithm, the research conducted is refined with the concept of probability is Naïve Bayes. The result of this research is an optimization model of the decision tree produced by C45 algorithm with Naïve Bayes algorithm optimization, where Naïve Bayes is able to choose the root of the decision when the value of the calculation of gain ratio at C45 shows the same value. In the test, from the prepared test data, traceable decision in the form of classification based on the formation of decision trees that have been formed by C45 with optimization Naïve Bayes on ISP selection case studies.

Keywords: C45 Algorithm, Naïve Bayes, *Internet Service Provider*.

1. Pendahuluan

Pemilihan jasa *Internet Service Provider* atau sering disebut dengan ISP pada sebuah perusahaan merupakan hal yang penting. Semakin baik kualitas ISP yang digunakan oleh perusahaan akan berdampak pada kualitas kinerja karyawannya. Beberapa kriteria dalam penilaian ISP yang umumnya digunakan sebagai parameter pemilihan oleh *customer* antara lain adalah yaitu Jaminan CIR (*Customers Interconnect Rate*), Jaminan SLA (*Service Level Agreement*), Dukungan Teknis dan *Hotline*, Perangkat *Last Miles*, Perbandingan Harga dan Kualitas Layanan, *Customer Satisfactions*, *Features*, IP Addressing dan Legalitas Pemerintah [1]. Permasalahan utama pada pemilihan ISP oleh calon *customer* adalah prioritas terhadap parameter penilaian. Dalam hal ini yang dimaksud adalah mana parameter penilaian yang menjadi urutan pertama pemilihan ISP. Salah satu contoh penentuan pemilihan ISP yang dikenalkan adalah dengan teknik pembobotan atau *weighted product* sebagai alternatif penghitungan terhadap penilaian kelayakan ISP [2]. Teknik *weighted product* yang digunakan berfokus pada pembobotan pada masing-masing parameter dengan hasil penilaian bobot terbesar.

Pada penelitian yang dilakukan, pemilihan penilaian kelayakan ISP dengan parameter yang sudah ditentukan dilakukan dengan teknik klasifikasi dimana salah satu algoritma yang dipilih adalah algoritma yang mengadopsi teori pengambilan keputusan dengan konsep pohon keputusan yaitu C45. Keunggulannya adalah, dengan adanya pohon keputusan, membangun struktur pemilihan akan lebih mudah dan secara khusus dalam klasifikasi C45 memiliki kemampuan untuk memangkas struktur hierarki pohon keputusan terhadap parameter yang sehingga memudahkan dalam implementasi pengambilan keputusan. Keunggulan lainnya pada algoritma C45 antara lain adalah dapat menangani atribut yang bersifat kontinyu dan diskrit, C45 dapat menangani *training* data dengan *missing value*, hasil pohon keputusan C45 akan dipangkas setelah dibentuk dan pemilihan atribut yang dilakukan dengan menggunakan *Gain ratio* dan menjadikan sebuah akar pengambilan keputusan dengan nilai parameter sebagai cabang keputusan [3]. Dan untuk penyempurnaan permasalahan pada C45 yaitu permasalahan nilai perhitungan yang sama pada parameter yang berbeda dari hasil perhitungan nilai *Gain Ratio*, dilakukan dengan konsep probabilitas yaitu Naïve Bayes. Penggunaan Naïve Bayes dilakukan untuk memastikan nilai dari setiap hasil *gain ratio* pada C45 yang sama untuk diambil keputusan yang optimal pada struktur pohon keputusan.

Data parameter yang digunakan dalam penelitian ini sebagai penilaian kelayakan pemilihan ISP adalah *Hotline* ISP, nilai *Customer Interconnect Rate* (CIR), Harga, *Customer Satisfactions*, SLA, Perangkat *Last Mile* dan memiliki kelas kelayakan untuk diterima dan tidak diterima. Metode C45 yang diimplementasikan dengan optimalisasi Naïve Bayes dan menghasilkan sebuah pohon keputusan dapat digunakan sebagai pedoman serta membantu dalam melakukan pertimbangan mengenai layak atau tidak ISP ini untuk dipilih oleh calon *customer* baru.

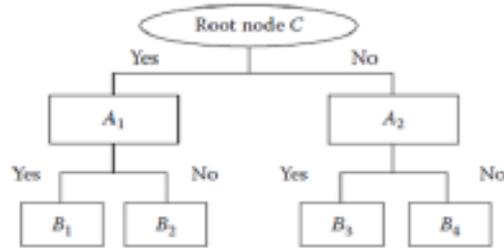
2. Metodologi Penelitian

2.1. Decision Tree

Algoritma *decision tree* menggunakan pendekatan *divide-and-conquer*. Umumnya digunakan untuk penyelesaian masalah dengan teknik klasifikasi. Model kerja algoritma *decision tree* adalah bekerja dari atas ke bawah, untuk mencari pada setiap tahap atribut dan membaginya ke dalam bagian terbaik *class* yang terbentuk. Langkah akhirnya adalah memproses secara rekursif sub masalah yang dihasilkan dari pembagian yang dihasilkan. Strategi ini menghasilkan sebuah *decision tree* yang dapat diubah menjadi satu kumpulan aturan klasifikasi [4].

Decision tree merupakan salah satu metode klasifikasi yang menggunakan konsep struktur pohon (*tree*) di mana setiap *node* merepresentasikan atribut, cabangnya merepresentasikan nilai dari atribut, dan daun merepresentasikan kelas. *Node* yang paling atas dari *decision tree* disebut sebagai *root*. Pada *decision tree* terdapat 3 jenis *node*, yaitu:

- a. *Root Node*, merupakan *node* paling atas, pada *node* ini tidak ada *input* dan bisa tidak mempunyai *output* atau mempunyai *output* lebih dari satu.
- b. *Internal Node*, merupakan *node* percabangan, pada *node* ini hanya terdapat satu *input* dan mempunyai *output* minimal dua.
- c. *Leaf node* atau *terminal node*, merupakan *node* akhir, pada *node* ini hanya terdapat satu *input* dan tidak mempunyai *output*.



Gambar 1. Contoh struktur pohon keputusan.

Gambar 1 merupakan ilustrasi *decision tree* tergantung pada aturan *if-then*, tetapi tidak membutuhkan parameter dan metrik. Struktur sederhana dan dapat ditafsirkan. *Decision tree* memungkinkan untuk memecahkan masalah atribut *multi-type*. *Decision tree* juga dapat mengelola nilai-nilai yang hilang atau *data noise* [5].

2.2. Algoritma C45

Algoritma C45 adalah salah satu algoritma klasifikasi data dengan teknik keputusan yang terkenal dan disukai karena memiliki kelebihan – kelebihan. Kelebihan ini misalnya: dapat mengolah data numerik (kontinyu) dan diskret, dapat menangani nilai atribut yang hilang, menghasilkan aturan – aturan yang mudah diinterpretasikan dan tercepat diantara algoritma – algoritma yang menggunakan memori utama di komputer. Algoritma C45 sendiri merupakan algoritma yang dikembangkan dari algoritma ID3 dimana algoritma ID3 sendiri ditemukan oleh J. Ross Quinlan sejak tahun 1986. Algoritma ID3 memiliki kepanjangan *Iterative Dichotomizer 3* dan dikategorikan sebagai algoritma learning. Algoritma ID3 dikatakan algoritma *learning* karena pada dasarnya pada kecerdasan buatan, terdapat tiga teknik dasar yaitu *searching reasoning* dan *planning* yang memerlukan aturan baku dalam penyelesaian masalah. Dalam ID3 ada istilah *entropy* yaitu ukuran kemurnian suatu atribut sedangkan *information gain* merupakan pengurangan *entropy* yang disebabkan oleh partisi berdasarkan suatu atribut. Untuk melakukan perhitungan *Entropy* menggunakan rumus sebagai berikut:

$$Entropy(S) = \sum_i^c -p_i \log_2 p_i \tag{1}$$

Keterangan dari rumus tersebut adalah :
 C = Jumlah nilai yang ada pada atribut target.
 p_i = Jumlah sampel untuk kelas i.

Kemudian untuk menghitung *information gain*, yang digunakan untuk mengukur efektivitas suatu atribut dalam pengklasifikasian data adalah:

$$Gain(S,A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v) \tag{2}$$

Keterangan dari rumus tersebut adalah:
 A = Atribut.
 V = Menyatakan suatu nilai yang mungkin untuk nilai A.
 Values_(A) = Himpunan nilai-nilai yang mungkin untuk nilai A.
 |S_v| = Jumlah sampel untuk nilai v.
 |S| = Jumlah seluruh sampel data.
 Entropy = Entropy untuk sampel-sampel yang memiliki nilai v.

Algoritma C45 sering juga dikatakan sebagai algoritma pohon keputusan yang populer pada kelompok algoritma pohon keputusan. Secara mendasar algoritma C45 memiliki kesamaan dengan ID3 dari pembentukan model pohon keputusan.

Perbedaan utama C45 dari ID3 adalah:

- a. C45 dapat menangani atribut kontinyu dan diskrit.
- b. C45 dapat menangani *training data* dengan *missing value*.
- c. Hasil pohon keputusan C45 akan dipangkas setelah dibentuk.
- d. Pemilihan atribut yang dilakukan dengan menggunakan *Gain ratio*.

Gain ratio yang dimaksud dalam C45 adalah untuk mengatasi bias yang terjadi pada ID3. Ini semacam bentuk normalisasi untuk mendapatkan informasi menggunakan ‘*split* informasi’, yang didefinisikan sebagai berikut:

$$SplitInfoA(D) = \sum_{j=0}^v \frac{|D_j|}{|D|} \times \log_2 \left(\frac{|D_j|}{|D|} \right) \quad (3)$$

Dimana:

D = ruang (data) *sample* yang digunakan untuk *training*.

D_j = jumlah *sample* untuk atribut i .

Nilai ini merupakan informasi yang potensial yang dihasilkan dengan memisahkan set data pelatihan, D , menjadi partisi v , sesuai dengan hasil v dari tes pada atribut A . Untuk mencari nilai *gain ratio* didefinisikan seperti berikut.

$$GainRatio(A) = \frac{Gain(A)}{SplitInfo(A)} \quad (4)$$

Pohon keputusan yang dibangun adalah dengan cara membagi data secara rekursif hingga tiap bagian terdiri dari data yang berasal dari kelas yang sama. Bentuk pemecahan (*split*) yang digunakan untuk membagi data tergantung dari jenis atribut yang digunakan dalam *split*. *Split* untuk atribut numerik yaitu mengurutkan contoh berdasarkan atribut kontinyu A , kemudian membentuk minimum permulaan (*threshold*) M dari contoh-contoh yang ada dari kelas mayoritas pada setiap partisi yang bersebelahan, lalu menggabungkan partisi-partisi yang bersebelahan tersebut dengan kelas mayoritas yang sama. *Split* untuk atribut diskret A mempunyai bentuk *value* (A) $\in X$ dimana $X \subset \text{domain}(A)$. Jika suatu set data mempunyai beberapa pengamatan dengan *missing value* yaitu *record* dengan beberapa nilai variabel tidak ada. Jika jumlah pengamatan terbatas maka atribut dengan *missing value* dapat diganti dengan nilai rata-rata dari variabel yang bersangkutan.

2.3. Naïve Bayes

Metode Bayes merupakan pendekatan statistik untuk melakukan inferensi induksi pada persoalan klasifikasi. Metode ini menggunakan probabilitas bersyarat sebagai dasarnya. Metode Klasifikasi Naïve Bayes didasarkan pada teorema Bayes, dengan asumsi bahwa efek dari nilai atribut pada kelas tertentu tidak tergantung pada nilai dari atribut lainnya. asumsi ini sering dikatakan sebagai “model fitur independen” [6].

Probabilitas untuk klasifikasi model kondisional adalah sebagai berikut:

$$P(C|F_1, \dots, F_n) \quad (5)$$

Kondisional di atas merupakan variabel kelas dependen C dengan sejumlah kecil hasil atau kelas, tergantung pada beberapa variabel fitur F_1 sampai F_n . Sehingga penulisan teori Bayes adalah:

$$P(C|F_1, \dots, F_n) = \frac{p(C)p(F_1, \dots, F_n|C)}{p(F_1, \dots, F_n)} \quad (6)$$

Persamaan di atas dapat ditulis sebagai:

$$posterior = \frac{prior \times likelihood}{evidence} \quad (7)$$

Asumsi kemandirian bersyarat yang "naif" memegang peranan. Menganggap bahwa setiap fitur F_i adalah secara kondisi independen terhadap setiap fitur lainnya F_j untuk $j \neq i$. Ini berarti bahwa:

$$P(F_i | C, F_j) = p(F_i | C) \quad (8)$$

Untuk $i \neq j$, sehingga *joint* model dapat dinyatakan sebagai:

$$P(F_i | C, F_j) = p(F_i | C) \quad (9)$$

Ini berarti bahwa di bawah asumsi independen di atas, distribusi bersyarat dari variabel kelas C dapat dinyatakan seperti ini:

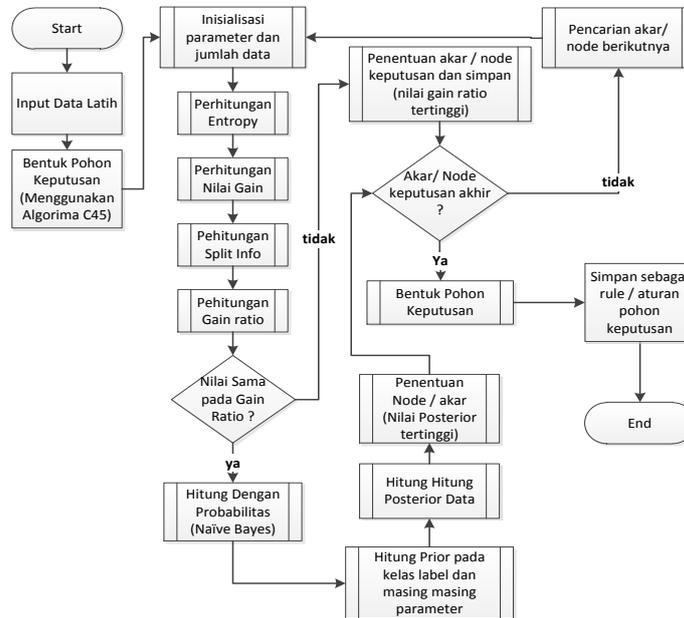
$$P(C|F_1, \dots, F_n) = \frac{1}{Z} p(C) \prod_{i=1}^n p(F_i|C) \tag{10}$$

Dimana Z (bukti) adalah faktor skala tergantung hanya pada F_1, \dots, F_n , yaitu, sebuah konstanta jika nilai dari variabel fitur diketahui.

Model dari bentuk ini jauh lebih mudah dikelola, karena mereka memecah menjadi *class prior* $p(C)$ dan distribusi probabilitas independen $p(F_i|C)$. Jika ada k kelas dan jika model untuk masing-masing $p(F_i|C = c)$ dapat dinyatakan dalam bentuk parameter, maka model Naïve Bayes yang sesuai memiliki $(k - 1) + n r k$ parameter. Dalam praktiknya, sering $k = 2$ (klasifikasi biner) dan $r = 1$ (variabel Bernoulli sebagai fitur) yang umum, sehingga jumlah parameter model Naïve Bayes adalah $2n + 1$, dimana n adalah jumlah fitur biner yang digunakan untuk klasifikasi dan prediksi.

3. Pembahasan dan Hasil

Pada penelitian yang dilakukan, proses pembentukan pohon keputusan pemilihan ISP ditunjukkan pada Gambar 2.



Gambar 2. Alur pembentukan pohon keputusan pemilihan ISP.

Pada Gambar 2 dapat dijelaskan bahwa proses pembentukan pohon keputusan pemilihan ISP diawali dengan meng-*input*-kan data latih sebagai klasifikasi. Data latih yang dimaksud adalah data pemilihan kriteria. Pada penelitian yang dilakukan, data *training* yang didapatkan menggunakan 6 parameter yaitu *Hotline* (dengan nilai parameter ada dan tidak), *CIR* (dengan nilai parameter 1:1, 1:2 dan 1:4), *Harga* (dengan nilai parameter murah, standar, dan mahal), *Customer Satisfaction* (dengan nilai parameter baik, biasa, cukup baik), *SLA* (dengan nilai parameter 99.5 dan 99.9) dan *Perangkat Last Mile* (dengan nilai parameter modem dan *router*). Berdasarkan parameter dan nilai dari masing-masing parameter dilakukan kombinasi data sehingga didapatkanlah total data *training* sebanyak 216 data. Dari 216 data kemudian menggunakan kepakaran yang didapat dari pakar yang mampu memberikan label kelas pada data *training*, dengan penilaian kelayakan pemilihan ISP yang ideal dalam sebuah perusahaan. Pada penelitian yang dilakukan data kriteria pemilihan ISP ditunjukkan pada Tabel 1.

Tabel 1. Penilaian kriteria pemilihan ISP pada calon *customer*.

No	Hotline	CIR	Harga	Customer Satisfactions	SLA	Perangkat Last Mile	Status ISP Menurut Customer
1	Ada	1:2	Standar	Baik	99.5	Router	Ya
2	Ada	1:2	Standar	Baik	99.5	Modem	Ya
3	Ada	1:2	Standar	Baik	99.9	Router	Ya

4	Ada	1:2	Standar	Baik	99.9	Modem	Ya
5	Ada	1:2	Standar	Cukup Baik	99.5	Router	Tidak
6	Ada	1:2	Standar	Cukup Baik	99.5	Modem	Tidak
7	Ada	1:2	Standar	Cukup Baik	99.9	Router	Ya
8	Ada	1:2	Standar	Cukup Baik	99.9	Modem	Ya
9	Ada	1:2	Standar	Biasa	99.5	Router	Tidak
10	Ada	1:2	Standar	Biasa	99.5	Modem	Tidak
11	Ada	1:2	Standar	Biasa	99.9	Router	Ya
12	Ada	1:2	Standar	Biasa	99.9	Modem	Ya
13	Ada	1:2	Mahal	Baik	99.5	Router	Tidak
13	Ada	1:2	Mahal	Baik	99.5	Modem	Tidak
14	Ada	1:2	Mahal	Baik	99.9	Router	Ya
15	Ada	1:2	Mahal	Baik	99.9	Modem	Ya
16	Ada	1:2	Mahal	Cukup Baik	99.5	Router	Tidak
17	Ada	1:2	Mahal	Cukup Baik	99.5	Modem	Tidak
18	Ada	1:2	Mahal	Cukup Baik	99.9	Router	Tidak
19	Ada	1:2	Mahal	Cukup Baik	99.9	Modem	Tidak
20	Ada	1:2	Mahal	Biasa	99.5	Router	Tidak
21	Ada	1:2	Mahal	Biasa	99.5	Modem	Tidak
...
216	Ada	1:2	Mahal	Biasa	99.9	Modem	Tidak

Setelah memiliki data kualifikasi ISP, berikutnya dilakukan inisialisasi parameter dan jumlah data. Untuk memudahkan perhitungan pada penelitian menginisialisasi kelas label status ISP menurut *customer*, dimana untuk kelas label Ya menjadi Golongan A dan Tidak menjadi Golongan B. Didapatkan hasil inisialisasi pada Table 2.

Tabel 2. Inisialisasi parameter dan jumlah data.

Atribut	Jumlah Kasus	Golongan	
		A	B
Total	216	64	152
<i>Hot line</i>	Ada	108	44
	Tidak	108	0
CIR	1:1	72	40
	1:2	72	50
	1:4	72	62
Harga	Murah	72	40
	Standar	71	50
	Mahal	73	62
<i>Customer Satisfactions</i>	Baik	74	45
	Biasa	70	57
	Cukup Baik	72	50
SLA	99.9	108	72
	99.5	108	80
Perangkat <i>Last Mile</i>	Router	108	76
	Modem	108	76

Setelah melakukan inisialisasi dilakukan perhitungan C45 untuk mencari akar atau *node* pertama dari pohon keputusan. Diawali dengan menghitung entropi pada algoritma C45. Contoh penghitungan entropi pada parameter CIR dengan nilai 1:1, entropi awal adalah sebagai berikut:

$$E(Total) = \left(\left(\frac{-32}{72} \right) \log_2 \frac{32}{72} \right) + \left(\left(\frac{-40}{72} \right) \log_2 \frac{40}{72} \right) = 0.991076$$

Berikutnya adalah menghitung nilai *gain* dan didapatkan pada nilai *gain* pada parameter CIR sebagai berikut:

$$Gain(Total, CIR) = 0.876716289 - \left(\left(\frac{72}{216} \right) 0.991076 \right) + \left(\left(\frac{72}{216} \right) 0.887976 \right) + \left(\left(\frac{72}{216} \right) 0.581321 \right) = 0.056592$$

Setelah mendapatkan nilai *Gain* dilakukan penghitungan *Split Info*.

$$SplitInfo(CIR) = \left(- \left(\frac{72}{216} \right) \log_2 \frac{72}{216} \right) + \left(- \left(\frac{72}{216} \right) \log_2 \frac{72}{216} \right) + \left(- \left(\frac{72}{216} \right) \log_2 \frac{72}{216} \right) = 1.584963$$

Setelah mendapatkan *Split* info dilakukan perhitungan terakhir yaitu *gain ratio*, dan didapatkan nilai pada parameter CIR sebagai berikut:

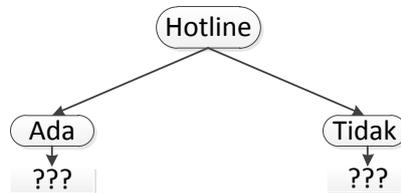
$$GainRatio(CIR) = 0.056592 / 1.584963 = 0.035705$$

Hasil perhitungan pertama di dapatkan *node* awal adalah parameter *hotline* ditunjukkan pada Tabel 3. Parameter *hotline* dipilih karena memiliki nilai *gain ratio* tertinggi dan parameter lainnya tidak ada yang sama dengan nilai *gain ratio* parameter *hotline*.

Tabel 3. Perhitungan akar keputusan pertama.

Atribut	Jumlah Kasus	Entropi		Gain	Split Info	Gain Ratio
		A	B			
Total	216	0.876716289				
Hot line	Ada	108	0.975119065	0.389157	1.000000	0.389157
	Tidak	108	0			
CIR	1:1	72	0.99107606	0.056592	1.584963	0.035705
	1:2	72	0.88797632			
	1:4	72	0.581321499			
Harga	Murah	72	0.99107606	0.051710	1.584870	0.032627
	Standar	71	0.876064368			
	Mahal	73	0.611551695			
Customer Satisfactions	Baik	74	0.966009606	0.025381	1.584591	0.016018
	Biasa	70	0.692419486			
	Cukup Baik	72	0.88797632			
SLA	99.9	108	0.918295834	0.004755	1.000000	0.004755
	99.5	108	0.825626526			
Perangkat Last Mile	Router	108	0.876716289	0.000000	1.000000	0.000000
	Modem	108	0.876716289			

Pembentukan *node* awal ditunjukkan pada Gambar 3.



Gambar 3. Pembentukan *node* pertama pohon keputusan.

Setelah pembentukan akar pertama, maka penghitungan dilakukan perulangan dengan menghitung dari data yang memiliki nilai parameter *hotline* ada dan nilai parameter *hotline* tidak ada. Pada Iterasi kedua pada penentuan Akar dari parameter Harga ditunjukkan pada Tabel 4, memiliki kesamaan nilai *gain ratio* yang sama.

Tabel 4. Penghitungan iterasi 2 hasil sama pada *gain ratio*.

Atribut	Jumlah Kasus	Golongan		Entropi Total	Nilai Gain	Split Info	Gain Ratio
		A	B				
Total	36	32	4	0.503258			
Harga	Murah	12	0	0.000000	0.197160	1.584963	0.124394
	Standar	12	0	0.000000			
	Mahal	12	4	0.918296			
Customer Satisfactions	Baik	12	0	0.000000	0.197160	1.584963	0.124394
	Biasa	12	4	0.918296			
	Cukup Baik	12	0	0.000000			
SLA	99.9	18	2	0.503258	0.000000	1.000000	0.000000
	99.5	18	2	0.503258			
Perangkat Last Mile	Router	18	2	0.503258	0.000000	1.000000	0.000000
	Modem	18	2	0.503258			

Pada saat terjadi nilai sama pada perhitungan *gain ratio* maka digunakan algoritma Naïve Bayes untuk memilih *node* keputusan berikutnya. Pada penelitian yang dilakukan, data perhitungan yang digunakan untuk memilih *node* harga atau *Customer Satisfaction*, digunakan kemungkinan muncul pada data latih yang ditunjukkan pada Tabel 5.

Tabel 5. Kemungkinan pilihan node pada data latih.

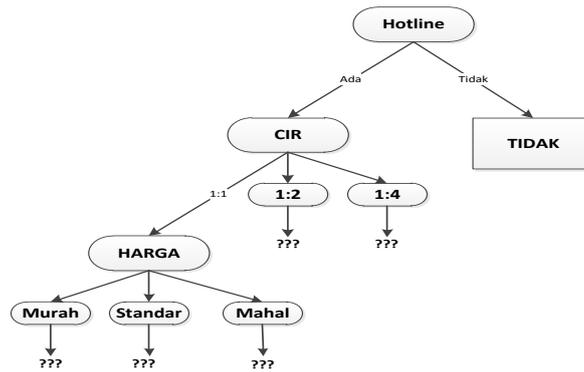
Pilihan Pertama	Pilihan Kedua	Hasil Pilihan Node
Murah	Baik	Harga
Murah	Biasa	Harga
Murah	Cukup Baik	Harga
Standar	Baik	Customer Satisfactions
Standar	Biasa	Harga
Standar	Cukup Baik	Harga
Mahal	Baik	Customer Satisfactions
Mahal	Biasa	Customer Satisfactions
Mahal	Cukup Baik	Customer Satisfactions
Baik	Murah	Harga
Baik	Standar	Customer Satisfactions
Baik	Mahal	Customer Satisfactions
Cukup Baik	Murah	Harga
Cukup Baik	Standar	Harga
Cukup Baik	Mahal	Customer Satisfactions
Biasa	Murah	Harga
Biasa	Standar	Harga
Biasa	Mahal	Harga

Berdasarkan data kemungkinan pemilihan *node* pada data pada Tabel 5 maka dilakukan perhitungan Naïve Bayes yang menghitung nilai *Prior* dan *Posterior* dan ditunjukkan pada Tabel 6.

Tabel 6. Hasil perhitungan *prior* dan *posterior*.

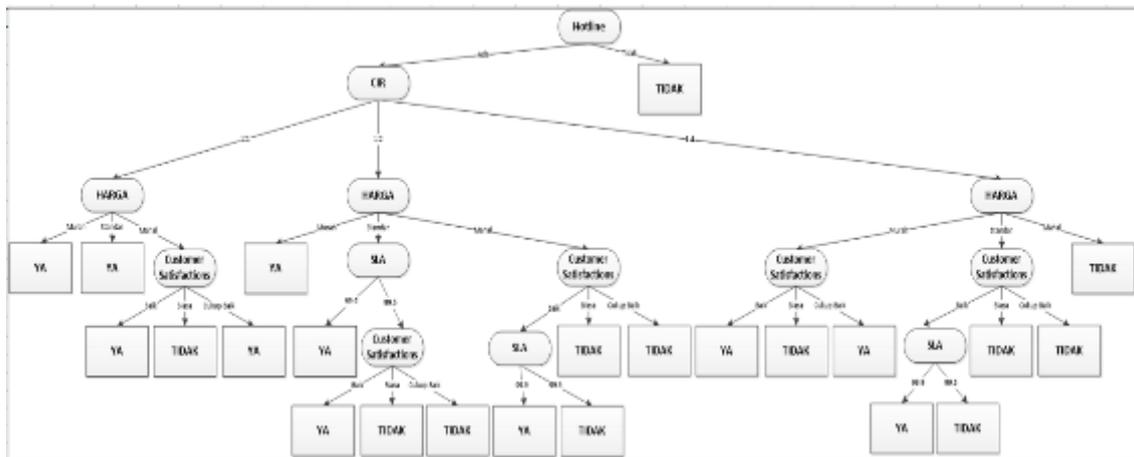
Prior Pilihan Pertama		Prior Pilihan Kedua		Posterior		Klasifikasi Bayes (Pilihan Node)
Harga	Cust. Satisf	Harga	Cust. Satisf	Harga	Cust. Satisf	
0.2727	0.0000	0.0909	0.2857	0.0152	0.0000	Harga
0.2727	0.0000	0.1818	0.1429	0.0303	0.0000	Harga
0.2727	0.0000	0.1818	0.1429	0.0303	0.0000	Harga
0.1818	0.2500	0.0909	0.2857	0.0101	0.0278	Cust. Satisf
0.1818	0.2500	0.1818	0.1429	0.0202	0.0139	Harga
0.1818	0.2500	0.1818	0.1429	0.0202	0.0139	Harga
0.0000	0.7500	0.0909	0.2857	0.0000	0.0833	Cust.Satisf
0.0000	0.7500	0.1818	0.1429	0.0000	0.0417	Cust.Satisf
0.0000	0.7500	0.1818	0.1429	0.0000	0.0417	Cust.Satisf
0.0909	0.5000	0.2727	0.0000	0.0152	0.0000	Harga
0.0909	0.5000	0.1818	0.1429	0.0101	0.0278	Cust.Satisf
0.0909	0.5000	0.0909	0.2857	0.0051	0.0556	Cust.Satisf
0.1818	0.2500	0.2727	0.0000	0.0303	0.0000	Harga
0.1818	0.2500	0.1818	0.1429	0.0202	0.0139	Harga
0.1818	0.2500	0.0909	0.2857	0.0101	0.0278	Cust.Satisf
0.2727	0.0000	0.2727	0.0000	0.0455	0.0000	Harga
0.2727	0.0000	0.1818	0.1429	0.0303	0.0000	Harga
0.2727	0.0000	0.0909	0.2857	0.0152	0.0000	Harga

Dari hasil perhitungan yang didapat, dalam penghitungan *posterior*, dapat dilihat bahwa dominan *node* terpilih adalah *node* harga. Maka untuk melanjutkan iterasi pada C45 akan di pilih *node* harga pada akar parameter CIR nilai 1:1 yang ditunjukkan pada Gambar 4.



Gambar 4. Pembentukan *node* hasil optimasi Naïve Bayes.

Perhitungan algoritma C4.5 dilakukan berulang hingga mendapatkan kedalaman pohon keputusan paling dalam. Hasil yang didapat pada penelitian menunjukkan bentuk Pohon Keputusan sebagai berikut:



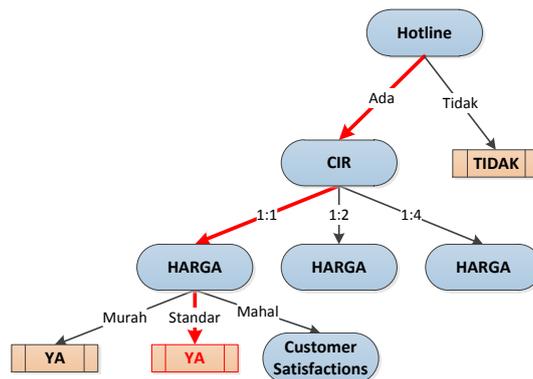
Gambar 5. Hasil akhir pembentukan pohon keputusan.

Setelah mendapatkan struktur pohon keputusan yang dihasilkan dari algoritma, maka dilakukan pengujian dengan data uji. Pada penelitian ini, data uji diambil secara acak sebanyak 10 buah data uji yang ditunjukkan pada Tabel 7.

Tabel 7. Penilaian kriteria pemilihan ISP pada calon *customer*.

No.	Nama ISP	Hotline	CIR	Harga	Customer Satisfactions	SLA	Perangkat Last Mile	Status ISP
1	def	Ada	1:4	Murah	Cukup Baik	99.9	Router	???
2	abc	Ada	1:4	Mahal	Biasa	99.9	Router	???
3	pqr	Tidak	1:2	Mahal	Baik	99.5	Modem	???
4	ghi	Ada	1:4	Murah	Biasa	99.9	Modem	???
5	xyz	Ada	1:1	Standar	Baik	99.5	Modem	???
6	stu	Tidak	1:1	Mahal	Biasa	99.5	Router	???
7	Adc	Tidak	1:1	Murah	Biasa	99.5	Router	???
8	jkl	Tidak	1:1	Mahal	Biasa	99.9	Router	???
9	gno	Tidak	1:2	Standar	Baik	99.9	Router	???
10	mno	Ada	1:4	Standar	Cukup Baik	99.9	Modem	???

Dari data uji pada Tabel 7, dilakukan penelusuran hasil keputusan pemilihan ISP sesuai pada struktur pohon keputusan yang dihasilkan oleh algoritma C45 dengan optimasi Naïve Bayes (Gambar 5). Salah satu contoh proses penelusuran yang dilakukan, mengambil data uji pada no. 5 pada Tabel 7. Proses penelusuran ditunjukkan pada Gambar 6.



Gambar 6. Hasil penelusuran pohon keputusan pada data uji.

Dari Gambar 6, dapat dijelaskan bahwa ISP xyz yang akan dicek pertama adalah parameter ketersediaan layanan *Hotline* dengan nilai “ada”, kemudian CIR dengan nilai 1:1. Pada pohon keputusan CIR dengan nilai 1:1 akan dilihat parameter harga. Data uji memiliki nilai harga standar, maka menghasilkan keputusan status ISP menurut calon *customer* adalah iya diterima untuk digunakan. Pada Gambar 6 juga dapat dijelaskan keuntungan Algoritma C45 bahwa untuk menghasilkan keputusan diterima, tidak diperlukan penelusuran ke semua parameter yang ada, cukup hanya terhadap tiga parameter dari tujuh parameter penilaian ISP yang ada. Dan Hal ini membuktikan bahwa penelusuran dengan C45 lebih cepat dan efisien. Hasil dari penelusuran pohon keputusan pada data uji (Tabel 7), ditunjukkan pada Tabel 8.

Tabel 8. Hasil penelusuran data uji.

No.	Nama ISP	Hotline	CIR	Harga	Customer Satisfactions	SLA	Perangkat Last Mile	Status ISP
1	def	Ada	1:4	Murah	Cukup Baik	99.9	Router	Ya
2	abc	Ada	1:4	Mahal	Biasa	99.9	Router	Tidak
3	pqr	Tidak	1:2	Mahal	Baik	99.5	Modem	Tidak
4	ghi	Ada	1:4	Murah	Biasa	99.9	Modem	Tidak
5	xyz	Ada	1:1	Standar	Baik	99.5	Modem	Ya
6	stu	Tidak	1:1	Mahal	Biasa	99.5	Router	Tidak
7	Adc	Tidak	1:1	Murah	Biasa	99.5	Router	Tidak
8	jkl	Tidak	1:1	Mahal	Biasa	99.9	Router	Tidak
9	gno	Tidak	1:2	Standar	Baik	99.9	Router	Tidak
10	mno	Ada	1:4	Standar	Cukup Baik	99.9	Modem	Tidak

Berdasarkan hasil yang ditunjukkan pada Tabel 8, terlihat bahwa keputusan terpendek adalah saat penelusuran parameter *Hotline* dengan nilai dari parameter adalah tidak yang menghasilkan hasil klasifikasi tidak. Dan penelusuran akan dilanjutkan apabila penelusuran pada parameter *hotline* dengan nilai parameter ada. Penelusuran pada data uji akan terus ditelusuri hingga menemui hasil akhir atau berupa keputusan sesuai dengan struktur pohon keputusan pada Gambar 5. Hal ini menyimpulkan bahwa pohon keputusan mampu untuk digunakan sebagai model pohon keputusan untuk pemilihan ISP.

4. Kesimpulan

Berdasarkan pada pembahasan dan hasil yang telah dipaparkan, dapat disimpulkan bahwa algoritma C45 dapat membentuk sebuah struktur pohon keputusan khususnya pada studi kasus pemilihan prioritas dalam penilaian ISP. Permasalahan pada algoritma C45 yang memungkinkan menghasilkan nilai *gain ratio* ganda, dapat dioptimalkan dengan algoritma Naïve Bayes. Dari hasil penelusuran pohon keputusan yang dilakukan terhadap contoh data uji, menghasilkan bahwa pohon keputusan yang dibentuk dapat mengefisienkan penelusuran, karena tidak memerlukan penelusuran ke semua parameter dalam penilaian ISP.

Daftar Pustaka

[1] Stiawan, D. (2015). Konsep Dasar Internet Tips Memilih ISP. Retrieved from <http://www.academia.edu>: https://www.academia.edu/11250391/Konsep_Dasar_Internet_Tips_Memilih_ISP

- [2] Marwa, S. (2015). Sistem Pendukung Keputusan Pemilihan Layanan Internet Service Provider. Indonesian Journal on Networking and Security, Vol. 4, No. 4.
- [3] Quinlan, J. (1993). C 45 : Progrmas for Machine Learning. San Mateo: Morgan Kaumann.
- [4] Witten, I. F. (2011). Data Mining Practical Machine Learning Tools and Techniques(3rd ed). USA: Elsevier.
- [5] Dua, S. X. (2011). Data Mining and Machine Learning in Cybersecurity. USA: Taylor and Francis Group.
- [6] Larose, D. T. (2005). Discovering Knowledge in Data. New Jersey: John Willey & Sons Inc.
- [7] Frankish K., Ramsey WM., "The Cambridge Handbook of Artificial Intelligence", Cambridge University : United Kingdom, 2014.
- [8] J. Wu and Z. Cai, "A naive Bayes probability estimation model based on self-adaptive differential evolution," Journal of Intelligent Information Systems, vol. 42, no. 3, pp. 671–694, 2014.