



SNESTIK

Seminar Nasional Teknik Elektro, Sistem Informasi,
dan Teknik Informatika

<https://ejurnal.itats.ac.id/snestik> dan <https://snestik.itats.ac.id>



Informasi Pelaksanaan :

SNESTIK I - Surabaya, 26 Juni 2021

Ruang Seminar Gedung A, Kampus Institut Teknologi Adhi Tama Surabaya

Informasi Artikel:

DOI : 10.31284/p.snestik.2021.1878

Prosiding ISSN 2775-5126

Fakultas Teknik Elektro dan Teknologi Informasi-Institut Teknologi Adhi Tama Surabaya
Gedung A-ITATS, Jl. Arief Rachman Hakim 100 Surabaya 60117 Telp. (031) 5945043
Email : snestik@itats.ac.id

Penerapan Metode C.45 untuk Mendukung Strategi Promosi pada Paket Data Seluler

Anggi Yhurinda Perdana Putri¹, Ruli Utami², dan Fanida Almas Nadhifah³

Institut Teknologi Adhi Tama Surabaya^{1,2,3}

e-mail: anggi@itats.ac.id

ABSTRACT

The need for information is so high that many providers give telecommunication services and play profitable business from selling data packets. To promote data packet to the customers, a telecommunication provider will consider some aspects related to the uses of provider services and the loyalty of customers upon the provider. Thus, every customer will get different promotions based on different usages of pulse, data packet, and network. Those differences due to credit using, data package and network in every customer. Basically, technology can be used for managing data so as to be more efficient such as classifying data of customers by algorithm C4.5. It can categorize customers into two groups namely "getting the promotion" and "not getting promo". The test carried out to the algorithm C4.5 taking 100 training data and 35 testing data obtained the accuracy 88.57%.

Keywords: Classification; Promotion Strategy; Cellular Data Package; C4.5 Algorithm

ABSTRAK

Kebutuhan informasi yang sangat tinggi serta banyaknya provider yang menyediakan jasa telekomunikasi menjadikan bisnis penyedia paket data termasuk dalam bisnis yang sangat menguntungkan. Dalam memberikan promosi paket data kepada pelanggan, sebuah provider telekomunikasi akan mempertimbangkan beberapa aspek yang berkaitan dengan penggunaan produk layanan provider maupun loyalitas pelanggan terhadap provider. Setiap pelanggan akan mendapatkan promosi yang berbeda antara pelanggan satu dengan pelanggan yang lain. Perbedaan tersebut dikarenakan penggunaan pulsa, paket data, dan jaringan tiap pelanggan berbeda-beda. Dalam pemanfaatan teknologi untuk pengelolaan data agar proses kerja dapat lebih efisien, dilakukan dengan mengklasifikasi data pelanggan dengan menggunakan algoritma C4.5 yaitu dengan mengklasifikasikan menjadi dua kategori pelanggan yaitu "dapat promo" dan "tidak dapat promo". Hasil

pengujian yang dihasilkan menggunakan algoritma C4.5 dengan menggunakan data training data training sebanyak 500 data dan data testing 35 menghasilkan nilai akurasi sebesar 88,57%.

Kata kunci: Klasifikasi; Strategi Promosi; Paket Data Seluler; Algoritma C4.5.

PENDAHULUAN

Perkembangan teknologi informasi yang saat ini berkembang sangat pesat berdampak bagi perusahaan telekomunikasi dalam menjaga loyalitas pelangganya. Perusahaan telekomunikasi berlomba dengan memberikan penawaran yang menarik bagi pelangganya agar setia menggunakan layanannya. Salah satu cara yang dilakukan adalah dengan memberikan promo bagi para pelangganya. Namun dalam pemberian promo yang telah dilakukan masih kurang efektif karena hanya berdasarkan satu aspek tanpa mempertimbangkan aspek yang lain. Berdasarkan dari latar belakang yang telah dijelaskan sebelumnya, maka perlu adanya sebuah sistem klasifikasi yang dapat membantu menentukan pemberian promo dengan mempertimbangkan semua aspek agar pemebrian lebih efektif. Metode kalsifikasi yang akan digunakan yaitu algoritma C4.5. Algoritma C4.5 dipilih karena dapat menghasilkan pohon keputusan yang mudah untuk di implementasikan sehingga penyelesaian permasalahan dapat dilakukan dengan cara yang sederhana dan tepat sasaran.

METODE

Algoritma C4.5 merupakan algoritma yang digunakan untuk membangun pohon keputusan (Utami & Atmojo, 2019). Pohon keputusan berguna untuk mengekspolari data, menemukan hubungan tersembunyi antara sejumlah calon variabel input dengan sebuah variabel target (Nurmuslimah, Putri & Mulya, 2020). Langkah-Langkah untuk membangun pohon keputusan secara umum adalah sebagai berikut (Kalsum, 2009):

1. Pilih atribut sebagai akar (root) – yaitu berdasarkan nilai gain tertinggi dari setiap atribut
2. Membuat cabang untuk setiap nilai – jumlah cabang sesuai dengan jumlah kelompok data pada atribut terpilih sebelumnya.
3. Bagi kasus dalam cabang
4. Mengulangi proses pada setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Pada perhitungan algoritma C4.5, atribut dengan nilai *gain* tertinggi akan menjadi *root node*. Sehingga untuk menentukan atribut mana yang akan dijadikan sebagai *root*, maka terlebih dahulu menghitung nilai *gain* dan *entropy* pada masing-masing atribut dengan rumus (1) dan (2) (Wendi & Hamidah, 2014):

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

$$\text{Entropy}(S) = \sum_{i=1}^n - p_i * \log_2 p_i \dots \dots \dots (2)$$

Keterangan: S= Himpunan kasus; A= Atribut/Fitur; n= Jumlah partisi atribut; |Si|= Jumlah kasus partisi ke-I; |S|= Jumlah kasus dalam S; pi= Proporsi dari Si terhadap S

HASIL DAN PEMBAHASAN

Data Training

Data training yang digunakan = 500 data dengan 5 atribut yaitu nama provider, pemakaian pulsa, pemakaian kuota, jaringan, lama pemakaian dan atribut class untuk hasil klasifikasi. Sampel data training yaitu 25 data dari 500 data ditunjukkan pada tabel 1.

Tabel 1. Sampel Data Training

ID PEL	PROVIDER	PULSA	KUOTA	JARINGAN	LAMA PAKAI	DAPAT PROMO ?
1	C	RENDAH	SEDANG	3G	LAMA	TIDAK
2	C	SEDANG	TINGGI	4G	CUKUP	YA
3	B	TINGGI	SEDANG	3G	LAMA	YA
4	B	TINGGI	SEDANG	4G	LAMA	YA
5	B	SEDANG	SEDANG	3G	BARU	TIDAK
6	C	TINGGI	RENDAH	3G	LAMA	TIDAK
7	A	RENDAH	RENDAH	4G	BARU	TIDAK
8	A	RENDAH	TINGGI	4G	CUKUP	YA
9	C	RENDAH	TINGGI	4G	LAMA	YA
10	A	TINGGI	TINGGI	3G	BARU	YA
11	B	RENDAH	SEDANG	3G	BARU	TIDAK
12	B	SEDANG	RENDAH	3G	CUKUP	TIDAK
13	A	TINGGI	TINGGI	4G	CUKUP	YA
14	A	TINGGI	RENDAH	3G	BARU	TIDAK
15	A	RENDAH	SEDANG	4G	LAMA	TIDAK
16	B	SEDANG	RENDAH	3G	BARU	TIDAK
17	A	TINGGI	RENDAH	3G	CUKUP	YA
18	A	TINGGI	RENDAH	4G	LAMA	YA
19	C	TINGGI	SEDANG	3G	LAMA	YA
20	A	RENDAH	TINGGI	4G	CUKUP	YA
21	C	SEDANG	TINGGI	4G	LAMA	YA
22	B	TINGGI	RENDAH	4G	LAMA	YA
23	A	SEDANG	TINGGI	4G	LAMA	YA
24	A	RENDAH	SEDANG	4G	BARU	TIDAK
25	C	RENDAH	TINGGI	3G	CUKUP	YA

Dari tabel 1 terdapat dua *Class* yaitu YA dan TIDAK untuk menentukan apakah pelanggan berhak mendapat promo atau tidak. Kemudian tahap selanjutnya yaitu menghitung *entropy total* dan *entropy* tiap atribut.

Menghitung Jumlah kasus

Jumlah Total kasus (S) = 45; Jumlah kasus “YA” (S1) = 28

Jumlah kasus “TIDAK” (S2) = 17

Menghitung Entropy Total

$Entropy(Total) = (-2845 \cdot \log_2(2845)) + (-1745 \cdot \log_2(1725))$

$Entropy(Total) = 0,982474087$

Menghitung Entropy Provider C

Diketahui: S = 12; S1 = 7; S2 = 5

$Entropy Provider C = (-712 \cdot \log_2(712)) + (-512 \cdot \log_2(512))$

$Entropy Provider C = 0,979868757$

Kemudian hitung entropy pada seluruh kelompok data dengan rumus yang sama. Setelah seluruh entropy ditemukan selanjutnya hitung nilai Gain pada masing-masing atribut.

Menghitung Gain Provider

$Gain Provider = Entropy(Total) - \sum |Provider_i| \cdot S_i \cdot \log_2(S_i)$

$Gain Provider = 0,9825 - ((1245 \cdot 0,9799) + (1945 \cdot 0,9819) + (1445 \cdot 0,9852))$

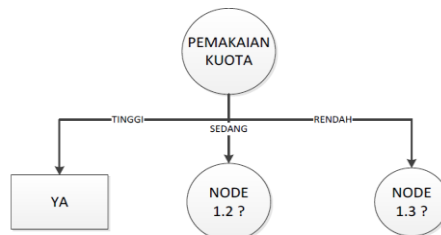
$Gain Provider = 0,0001$

Berdasarkan perhitungan diatas nilai *Gain* terbesarnya yaitu *atribut* Pemakaian *kuota* dengan nilai *gain* = 0,4096 (*root node*). Hasil perhitungan *entropy* ditunjukkan pada tabel 2

Tabel 2. Hasil Perhitungan Gain dan Entropy

NODE	ATRIBUT	VALUE	JUMLAH DATA	YA	TIDAK	ENTROPY	GAIN
1.	TOTAL		45	26	19	0,982474087	
	Provider	C	12	7	5	0,979868757	0,000063
		A	19	11	8	0,981940787	
		B	14	8	6	0,985228136	
	Pemakaian Pulsa	Tinggi	20	16	4	0,721928095	0,122628
		Sedang	12	5	7	0,979868757	
		Rendah	13	5	8	0,961236605	
	Pemakaian Kuota	Tinggi	15	15	0	0	0,409568
		Sedang	15	8	7	0,996791632	
		Rendah	15	3	2	0,721928095	
	Jaringan	4G	22	17	5	0,773226674	0,110905
		3G	23	9	14	0,965636133	
	Lama Pemakaian	Lama	19	11	8	0,981940787	0,077017
		Cukup	11	9	2	0,684038436	
		Baru	15	6	6	0,970950595	

Berdasarkan tabel 2 diketahui bahwa nilai gain tertinggi adalah atribut: Pemakaian Kuota”, maka atribut tersebut akan menjadi root node. Pada atribut pemakaian kuota terdapat 3 kelompok data yaitu rendah, sedang, dan tinggi pada pemakaian kuota tinggi seluruh data masuk pada kelas YA. Sedangkan untuk kelompok sedang dan rendah belum diketahui kelasnya. Sehingga untuk mengetahui kelasnya perlu dilakukan perhitungan ulang. Dari kondisi tersebut dihasilkan pohon keputusan sementara seperti gambar 1 berikut.

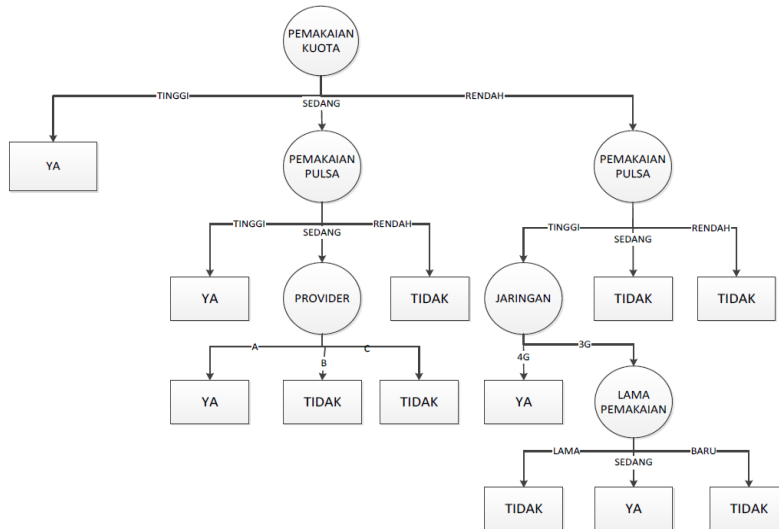


Gambar 1. Pohon Keputusan Sementara

Langkah selanjutnya adalah menghitung kembali nilai entropy serta nilai gain untuk mengetahui atribut apa yang ada pada node 1.2 dan 1.3. Proses perhitungan gain dan entropy harus dilakukan hingga seluruh data sudah diperoleh hasil keputusan menerima promo atau tidak. Hasil akhir dari proses perhitungan gain dan entropy ditunjukkan pada gambar 2.

Berdasarkan pohon keputusan pada gambar 3.10 dapat menghasilkan kriteria kelas “YA” apabila memenuhi salah satu syarat sebagai berikut:

- Pemakaian kuota = Tinggi
- Pemakaian kuota = Sedang; Pemakaian pulsa = Tinggi
- Pemakaian kuota = Sedang; Pemakaian pulsa = Sedang; Provider = A
- Pemakaian kuota = Rendah; Pemakaian pulsa = Tinggi; Jaringan = 4G
- Pemakaian kuota = Rendah; Pemakaian pulsa = Tinggi; Jaringan = 3G; Lama pemakaian = Sedang



Gambar 2. Hasil Akhir Pohon Keputusan

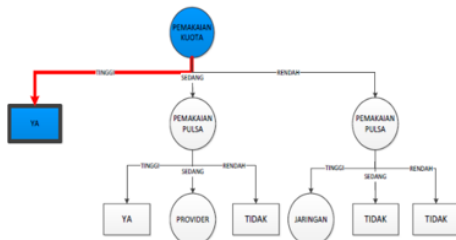
Data Testing

Setelah dilakukan training data, dilanjutkan pada proses testing data menggunakan data baru berjumlah 35 data yang belum diketahui kelasnya. Tabel menunjukkan sampel data testing yaitu 3 data dari 35 data testing.

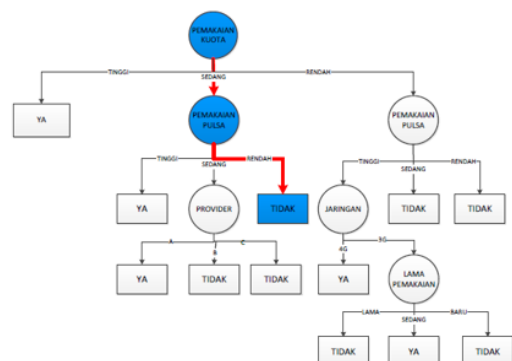
Tabel 3. Sampel Data Testing

ID PEL	PROVIDER	PULSA	KUOTA	JARINGAN	LAMA PAKAI	DAPAT PROMO?
1	A	SEDANG	TINGGI	3G	LAMA	?
2	C	RENDAH	SEDANG	3G	LAMA	?
3	C	TINGGI	RENDAH	3G	LAMA	?

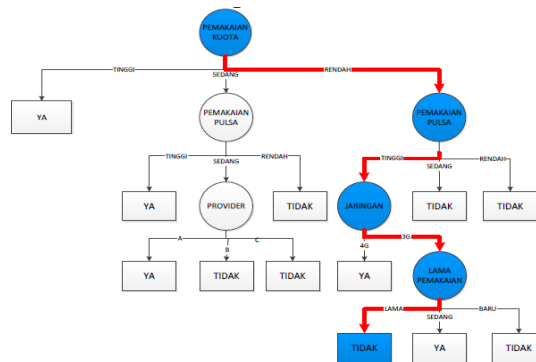
Dari data testing pada tabel 3, menghasilkan pohon keputusan yang ditunjukkan pada gambar 3 (a-c)



(a)



(b)



(c)

Gambar 3. a) Pemakaian kuota = Tinggi maka masuk pada kelas “YA”, b) Pemakaian kuota = Sedang, Pemakaian pulsa = Rendah maka masuk pada kelas “TIDAK”, c) Pemakaian kuota = Rendah, Pemakaian pulsa = Tinggi, Jaringan 3G, Lama pemakaian = Lama maka masuk pada kelas “TIDAK”, c)

KESIMPULAN

Berdasarkan hasil klasifikasi data pelanggan yang berhak mendapatkan promo paket data menggunakan algoritma C4.5. Dari data training sebanyak 500 data dan data testing 35 data menghasilkan 19 pelanggan yang masuk kelas “YA” (mendapatkan promo) dan 16 masuk kelas “TIDAK” (tidak mendapatkan promo)

DAFTAR PUSTAKA

- [1] R. Utami & S. Atmojo. (2019). Implementasi algoritma Apriori untuk Penentuan Weekend Product Promotion pada Minimarket “XYZ”. Prosiding Seminar Nasional Sains dan Teknologi Terapan 1 (1), 315-320
- [2] S. Nurmuslimah, A. Y. P. Putri, D. A. Mulya. (2020). Aplikasi Fuzzy Mamdani untuk Meminimalkan Biaya Produksi Petis. Jurnal Teknologi dan Manajemen 1 (2), 87-94
- [3] Kalsum, Umi. (2009). Penggunaan Pohon Keputusan (Decision Tree) Untuk Pengambilan Keputusan Dalam Penerimaan Pegawai.
- [4] Wendi & hamidah, Ida. (2014). Analisis Dan Penerapan Algoritma C4.5 Dalam Data Mining Untuk Memprediksi Masa Studi Mahasiswa Berdasarkan Data Nilai Akademik. Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNAST) ISSN: 1979-911X, Yogyakarta